IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 30, 2019, accepted September 29, 2019, date of publication October 7, 2019, date of current version October 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2945963

Investigation of the Relationships and Effects of Urban Transformation Parameters for Risky Structures: A Rapid Assessment Model

SEREL AKYOL[1,2] **AND EYYUP GULBANDILAR**[3]
[1]Computer Engineering Department, Faculty of Engineering, Eskisehir Technical University, 26470 Eskisehir, Turkey
[2]Sivrihisar Vocational School Computer Programming Program, Eskisehir Osmangazi University, 26600 Sivrihisar/Eskisehir, Turkey
[3]Computer Engineering Department, Faculty of Engineering, Eskisehir Osmangazi University, 26040 Eskisehir, Turkey

Corresponding author: Eyyup Gulbandilar (egulbandilar@ogu.edu.tr)

ABSTRACT Most of the works on the literature on urban transformation focus on the outcomes of transformation in legal, psychosocial, socioeconomic, and geographical aspects; and employ rapid screening models to assess the areas and multiple structures subject to urban transformation. Aiming the contribute to the literature, this study investigates the causal relationships between the parameters used in risk assessment of the individual masonry structures undergoing transformation. The causal relationship, which expresses the cause-effect relationship between two variables, shows that the independent variable has a direct or indirect effect on the dependent variable. The results of statistical analysis and theory should be considered concurrently in building a causal relationship model. The structural assessment reports for risky structures of 183 individual masonry buildings were examined and the relationships between the dependent and independent variables were assessed using path analysis. In order to establish a rapid assessment technique for risk assessment, the variables were chosen by binary logistic regression analysis due to the discrete nature of the dependent variable, and the final model was built accordingly. According to the model analysed by binary logistic regression, direct and indirect effects between the variables were determined using path analysis. While path analysis is applied to continuous data and evaluates linear regression results, an evaluation was performed based on logistic regression with discrete data results in this study. According to the path model analysis, the city where the building was located had the largest direct effect (path coefficient). It was concluded that the model, built with 6 effective variables selected among 25 independent variables generating the risk result, was acceptable in terms of engineering, and the proposed rapid assessment model could be used for risk assessment because of its high correct classification rate.

INDEX TERMS Causality, logistic regression, path analysis, urban transformation.

I. INTRODUCTION

Cities are settlements that have multifaceted and dynamic structures [1]. Since the day they are founded, cities are worn out due to natural disasters, increasing population, unplanned construction, and many other problems occurring in this process. This situation has raised the topic of urban transformation in the process of reconstructing our cities in light of the principles of modern urbanism and planning [2].

As in many other countries, Turkey has performed considerable research and development work on urban transformation over the recent years [3]–[7]. Most of this work, which makes comparative analyses of the urban transformation projects carried out in Turkey and the world [8] or evaluates the transformation from a local perspective [9], tends to focus on the objectives of urban transformation and the results of the implemented projects. Much of the literature on urban transformation focuses on the outcomes of transformation in legal, psychosocial, socioeconomic and geographical aspects [2], [10], [11].

The high number of parameters involved in the risk assessment of risky structures and areas subject to urban transformation poses challenges in establishing a decision mechanism. Therefore, there has been a need for rapid screening methods that allow for performing efficient assessment in a short time. For this purpose, various regulations,

The associate editor coordinating the review of this manuscript and approving it for publication was Lu Liu.

studies and assessments, known as rapid screening methods in the literature, have been performed in many areas subject to urban transformation. Early research on rapid screening methods presented a method referred to as the SST Format, which was developed using data from the Tokachi-Oki earthquake in 1968 [12]. Having been introduced in the literature on earthquake engineering by Federal Emergency Management Agency (FEMA) 154 and FEMA 155 reports (1988), rapid visual screening methods are still in use today [13]. Tezcan et al. proposed the P5 Method for rapid assessment of existing buildings to prevent the loss of life during an earthquake [14]. This method was then improved as the P24 Method and it was finally updated as the P25 Method [15]. These methods provide data on whether a structure has potential to collapse based on its performance score.

The Turkish Ministry of Environment and Urbanization has been authorized to administer the procedures and principles regarding rehabilitation, evacuation and renewal of areas at risk of disasters and lands and parcels with risky buildings outside these areas in order to provide proper, healthy and safe living environments that meet technical and aesthetic norms and standards. Having been enacted with this purpose, the Turkish Law on Transformation of Areas under Disaster Risk (No.6306) provides practical guidelines and methods to assess the risk levels of the areas with individual or multiple structures [16]. "The Methods for Regional Seismic Risk Assessment of Buildings", a section on regulations authorized by this law, is a rapid screening technique that is based on an information form to be applied in the field for seismic risk assessment of reinforced concrete and masonry buildings. In order to obtain accurate data while filling in building data collection forms, a team of experts should carry out the field investigation. With these forms, firstly, the identity information (i.e., address and coordinate information) of the examined buildings and then the technical information are collected, and the performance scores of the buildings are calculated in light of the technical data obtained. The risk priority among the regions can be determined with the distribution of the calculated scores by ordering the calculated performance scores from highest to lowest. The lower the performance score obtained from the structure, the higher the risk of the building. The current rapid screening methods serve the purpose of risky area prioritization. The methods to be used in defining the regional risk situation can be implemented in areas with a statistically significant number of buildings as required by scientific techniques. These methods cannot be used for risk assessment in an individual building, which poses a limitation [17].

The existing rapid screening forms are applied to multiple structures in areas declared as risky areas and consist of a large number of parameters. Experts should collect this data only after the required observations and measurements have been performed. In determining risky structures to be assessed within the scope of transformation, the excess of the number of parameters complicates the assessment process, while increasing the cost and giving damage to the structures. We have reviewed the literature on urban transformation and concluded that the majority of relevant studies focus on the outcomes of transformation in legal, psychosocial, socioeconomic and geographical aspects and employ rapid screening models to assess the areas and multiple structures subject to urban transformation, but there are currently no research results available on the risk assessment of individual structures.

The data obtained from the relevant Provincial Directorates for Environment and Urbanization are used in this study in order to optimize the decision process. For this purpose, by selecting effective parameters for risk assessment of individual masonry structures, we built a comprehensive path model to sort out the pathways underlying the effects between them. The proposed rapid assessment model is intended to contribute to the decision process and to minimize the time and costs of risk assessment procedure and the damage occurred in the structure in this process. The risk assessment model, which typically lasts weeks due to fieldwork, data collection from a building and simulation, is optimized to provide information about the risk assessment of the building (Risky/Safe). Therefore, an initial assessment is carried out to determine whether the actual risk assessment can be performed with advanced analysis techniques. Developed particularly for individual masonry buildings, this model is significant for assessing the current risk situation of a building in a short time without causing permanent damage to the building or paying additional costs. This rapid assessment model could contribute to the decision-making process since the existing rapid assessment models are used to identify "risky areas" based on multiple structures and the risk assessment of individual structures is carried out only by for-profit companies.

II. MATERIALS AND METHODS

Path analyses and diagrams, which contribute to the decision-making process and have been a popular research topic in many areas of science such as statistics, econometrics, epidemiology, genetics and other related disciplines, are graphical models used to encode assumptions about the data generation process [18]–[20]. Interactions and causality relationships of the variables in the model become complicated as the number of variables increases. Causality refers to the cause-effect relationships between variables. The causal relationship is that events and phenomena are interdependent or each case can be explained based on a cause [21]. A cause-effect relationship is established between two variables and a causal model is formed as a result of linking these reciprocal relations with each other by means of impact pathways [22].

The term "exogenous variable" is used to refer to the dependent variable in the model and the term "endogenous variable" is used for the independent variable for causal analysis. An exogenous variable is a variable whose variability is determined by causes outside the causal model. In contrast, an endogenous variable is a variable whose variability is accounted for by the exogenous and other endogenous variables in the causal model [23]. For each endogenous variable y, there is a function $y = f(x, u)$ representing the causal relationship from exogenous variables u and other endogenous variables x to y. When assuming that no hidden confounders exist, exogenous variables can be omitted from the analysis.

The function $y = f(x)$, which shows a causal relationship, indicates that the endogenous variable x has a direct effect on the exogenous variable y [24]. The causality shown in the function can alternatively be expressed as $x \rightarrow y$. In this expression, which is preferred particularly in the representation of graph structure, causality is characterized by a unidirectional relation (from one cause to one consequence), represented by a directional arrow. The arrow begins from the variable that induces the change and points towards the variable that shows the effect, and its direction indicates the direction of the effect [22].

Paul and Anderson built a causal model using multivariate species data [25]. Whenever there is an independent variable in a cause-effect relationship, a dependent variable either appears or changes. In cases where there is no cause-effect relationship, although it may seem like there is a relationship between the variables, in fact, this relationship is due to the effects of other variables on these variables [26]. A variable being affected by another variable does not necessarily indicate a causal (i.e., cause-effect) relationship [27]. A causal relationship involves hypothesis as well as knowledge and facts [28]. Relationships are defined by evaluating the results obtained from knowledge and theories and statistical analysis together. These interactions resulting from causal or non-causal relationships should be analysed cautiously. Path analysis, which is one of the methods allowing for causal modelling analysis, is used to determine whether the data is consistent with the model built rather than creating a causality structure. This method is highly effective for examining complex models and comparing different models [29].

There are a number of reporting and analysis parameters that should be applied in determination of risky structures to be assessed within the scope of urban transformation. According to these parameters, institutions or organizations licensed by the Ministry of Environment and Urbanization assess the results and prepare risk assessment reports. Institutions and organizations authorized in accordance with the Law no. 6306 for determining risky structures make risk decisions by carrying out essential examinations and assessments based on the parameters under the main headings of structure overview and data collection from the structure.

Turkish cities with masonry structures that have undergone a risk assessment process in different geographical regions and high-risk seismic zones were included in this study. Seismic zones in Turkey are classified into five zones based on tectonic maps: 1st, 2nd, 3rd, 4th and 5th degree seismic zones [30]. Due to the geographical location of Turkey, 42% of the land in the country is in the 1st degree, 24% is in the 2nd degree and 18% is in the 3rd degree seismic zones [31]. While the 1st degree seismic zones are areas with the highest seismic risk, those in the 5th degree zone are areas where seismic activity is minimal or not felt at all.

The data analysed in this study were obtained from Kutahya, Afyon and Eskisehir Provincial Directorates of the Ministry of Environment and Urbanization, and Istanbul-Zeytinburnu Municipality by gaining necessary permissions under legal conditions. We analysed *the Risk Assessment Forms of Vulnerable Buildings in accordance with the Code on Detection of Risky Buildings* for a total of 183 masonry buildings that underwent a risk assessment process between 2014 and 2017 and that were selected by random sampling from cities in three different geographical regions. Out of 25 independent variables showing technical data of the structures, 12 variables have continuous data and 13 had discrete data. The *Risk* dependent variable, which showed whether a structure was risky or not, consisted of discrete data and it took either "Risky" or "Safe" values. Significant parameters that would be included in the model to be built for rapid assessment were determined by a logistic regression analysis. Logistic regression is a type of analysis suitable for situations where the dependent variable (i.e., the predicted variable) is not a continuous or quantitative variable, in other words, for situations where it is categorical or classified [32]. The primary objective of this analysis is to build an appropriate model that has maximum accuracy for predicting the value of a categorical dependent variable for a dataset of independent variables [33]. All the parameters analysed in this study were the ones used in *the Code on Detection of Risky Buildings* in accordance with the Law no. 6306 for detecting risky masonry structures, and they were used to build a model for rapid assessment. The logistic regression model built was tested using path analysis, which is commonly used to test causal models with a theoretical basis.

A. LOGISTIC REGRESSION

Modelling in cause-effect based studies varies depending on the data structure of predicted and predictor variables. A model is the formation of information or thoughts related to an event based on certain rules. Like other model configuration techniques used in statistics, logistic regression analysis aims to build an acceptable model which could define the correlation between dependent and independent variables with the highest accuracy and the least variables [33]. Logistic regression is one of the most used machine learning method is for classification and clustering [34].

Logistic regression analysis, which is used to examine the cause-effect relationship between dependent and independent variables [35] and to make a classification, is preferred because it allows the dependent variable to take categorical and discrete values [36], [37]. The independent variables used in analysis are not restricted to be continuous or categorical. When the dependent variable has two categories, such as 0 and 1, analysis is carried out using "binary (dichotomous) logistic regression" [38].

Different methods such as forward selection, backward elimination and stepwise selection are used in the selection of the variables in the model to be formed by logistic regression analysis [39].

In forward selection method, there is only a constant term in the first step without variables. In the next step, an independent variable that provides the most significant contribution to the model according to a statistical decision rule that controls the significance of variables enters the model. Analysis is repeated according to the chi-square values in each step and independent variables are selected. The analysis continues until there is no significant variable to be added to the model. In this method, a variable that is entered in the model is never removed from the model even if it becomes insignificant for the model [40].

As the first step in the backward elimination method, all of the variables are included in the model. In contrast to the forward selection method, the least significant variables are removed from the model in the following steps, starting with the variable that provides the least contribution to the model. The main purpose of this method is to determine the least significant independent variable and the model that accounts for the dependent variable most. As a result of the analysis, it is possible that there are no independent variables in the model or that all of them are included in the model without removing any independent variables from the model [40]. Backward elimination method is recommended more than forward selection method.

In the stepwise selection method, an analysis is performed by entering all variables into the model or by deciding which variables will be entered into the model in which order according to a mathematical criterion. The use of this method allows a large number of variables to be examined quickly and efficiently, and it ensures that the variables are compatible with regression equations. The stepwise selection method incorporates the combined the applications of forward selection and backward elimination methods [38].

Logistic regression analysis can be performed as ''enter'' and ''stepwise'' in IBM SPSS statistical software [41]. Enter method, where all variables are entered into the model and analysed, is useful in situations where it is desired to observe the effects of the variables together. In stepwise backward elimination or forward selection methods, variable selection can be performed according to conditional, Wald and Likelihood Ratio (LR) statistics. The main difference between these methods is the statistical rules considered in the selection of independent variables to be included and excluded from the model. LR determines the variable to be included in or excluded from the model at each step of the analysis according to the probability of a likelihood-ratio statistic value. The conditional method makes variable selection with less sensitive statistics than LR. Variable selection is based on a probability of a likelihood-ratio statistic based on conditional parameter estimates. Variable selection according to the Wald method is made based on the significance of Wald statistical coefficients. The Wald statistics, also used in the causality analysis, is the square of the ratio of a non-standardized logistic coefficient to its standard error. The Wald statistics corresponds to the significance test of the β coefficient in logistic regression [36].

In logistic regression analysis, the ''model chi-square'' test is used to test the fitness of a model. The model to be analysed is built by calculating the natural logarithm of the ratio of the probability of occurrence of an event and the probability of non-occurrence to each other. The significance test for each independent variable in the model is tested using the Wald statistics. *Odds*, a basic term in logistic regression analysis, is the ratio of the probability of occurrence of an event to the probability of non-occurrence based on available data. *Odds ratio (OR)* value, on the other hand, is the ratio of the odds values of the two variables to each other and summarizes the relationship between the two variables. *Logit*is obtained by calculating the natural logarithm of the Odds ratio taken from the analysis. The purpose of using logarithmic distribution is to normalize distribution. Logit, which linearizes logistic regression, takes the natural logarithm of Odds ratio because of its asymmetrical structure and makes it symmetrical, and it corresponds to β coefficient in linear regression analysis [42].

In logistic regression, where the effects of the predictor variables on the dependent variable are obtained as probability and the risk factors are determined as probability, the probability of occurrence of the examined event is expressed as shown in (1) [43].

$$P = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}} \tag{1}$$

β values in the (1) show the regression coefficients of the independent variables, X values show the independent variables, and k shows the number of independent variables [36].

B. PATH ANALYSIS

Path coefficient was defined in 1921 by Dr. Sewall Wright as part of the standard deviation resulting from the independent variable observed in the dependent variable when the independent variables other than the variable whose effect is determined are constant [23]. Direct and indirect effects of variables can be determined by using the path analysis technique [44]. Path analysis, which allows for analysis of more complex models, can be used where there is a chain of dependent and independent variables affecting each other.

Path analysis is similar to causal analysis as it includes causal and non-causal relationships other than multiple regression [45]. Path analysis can prove a model that highlights the causal relationships between variables, but it cannot prove causality. Any causal model to be built should be based on a theory, expert knowledge, and research-based logic, followed by path analysis [29].

III. RESULTS

A total of 25 independent variables used as predictors for risk assessment were analysed by binary logistic regression

TABLE 1. Model summary of variable selection methods.

Method	-2 Log Likelihood	Cox & Snell R^2	Nagelkerke R^2
Backward Elimination (Conditional)	24.646	0.368	0.822
Backward Elimination (Wald)	24.044	0.370	0.827

TABLE 2. Coefficient estimates of model variables.

Variables	β	Wald	Sig.	Exp. (β)
X_1		9.561	0.00	
X_1 (1)	-5.100	5.412	0.254	0.006
X_1 (2)	-6.895	9.542	0.00	0.001
X_1 (3)	-15.923	0.000	0.756	0.000
X_2	2.041	3.107	0.00	7.698
X_3	0.070	4.279	0.00	1.073
X_4 (1)	-21.291	0.000	0.195	0.000
X_5	0.042	0.487	0.07	1.043
X_6	0.236	0.610	0.006	1.266

for the predicted Risk dependent variable. All statistical data analyses were performed using SPSS version 22.0.

In logistic regression analysis, which method to choose for variable selection depends on the structure of the data set and the researchers' decision. Stepwise methods give better results in model building studies. In forward selection method, a variable that is significant at some entry level and entered in the model is never removed from the model even if the significance level of the variable decreases in the following steps. For this reason, the analyses are based on the backward elimination Wald and conditional methods in order to build a model that could account for the dependent variables most with the least number of significant independent variables.

In the backward elimination-conditional method, removal testing for the variables excluded from the model is based on the probability of the likelihood-ratio statistic based on conditional parameter estimates. When the probability for stepwise entry value is set at 0.05 and the removal value is set at 0.10, there were four variables in the model: *Region, Number of Storeys, Shear Force X* and *Seismic Zone*. The method parameters are determined according to the values accepted in the literature. Models with different variables can be created by changing these values.

In the backward elimination-Wald method, the variables removed from the model are determined according to the Wald statistics values used in the causality analysis. In the final model obtained in this analysis, the variable *Building Height* is also included in addition to *Region, Number of Storeys, Shear Force X* and *Seismic Zone* variables.

The model summary values of the models built by two different methods are presented in Table 1. The correct class rate for the two different models with 4 and 5 variables is the same (97.3%). This does not mean that the added variable does not help improve the model. Evaluating the model with this ratio alone is not the right approach.

The value 0, which is the minimum possible -2 Log likelihood value, shows that the likelihood value is 1. It takes this value under the most ideal conditions or in certain circumstances. The -2 Log likelihood values of the models built in the analyses are compared and interpreted. A decrease in this value shows that the model fit is improving.

If the value of other parameters expressing the model fit (i.e. Cox and Snell R^2 and Nagelkerke R^2 values) is 1, this indicates a perfect model fit. The larger the values, the better the model fit. Nagelkerke R^2 value is an improved version of Cox and Snell R^2 coefficient. It is recommended to use Nagelkerke R^2 results, which show how much the independent variables account for the dependent variable as a percentage [46].

In the backward elimination method, the analysis steps should be examined in order to eliminate the deficiencies caused by the fact that a variable removed from the model cannot be added into the model again. Accordingly, it is seen that the correct class value decreases in the step where the variable *Age* is removed from the model in both methods.

The variable *Age* is a parameter that directly increases the risk level in determining risky structures [17]. In addition to its direct affect, compliance with the current earthquake regulations is also evaluated according to the construction year.

The variable *Building Height*, which is not included in the first model, is an effective parameter in creating a building load-bearing system model, determining the risk of a structure, calculating the column and carrying out an assessment based on critical storey, which are important for risk assessment of buildings [17].

Since causality is a statistical analysis and knowledge-theory-based approach, the variables Age and Building Height are included in the final model. Table 2 shows the model fit, significance values and model coefficients generated by parameters selected based on the literature. The following are six independent variables and one dependent variable in the model built:

$X_1 =$ *Region*
$X_2 =$ *Number of Storeys*
$X_3 =$ *Shear Force X*
$X_4 =$ *Seismic Zone*
$X_5 =$ *Age*
$X_6 =$ *Building Height*
$Y =$ *Risk*

The variable X_1-*Region* in the model is a categorical variable showing the city where the building to be assessed is located (i.e., Eskişehir, Kütahya, Afyon, Istanbul). In the analysis carried out for four different cities, an assessment is made according to the category taken as the reference. The variable X_2-*Number of Storeys*, provided among information on the existing structural system in the risk assessment forms for masonry buildings, is one of the important parameters affecting the earthquake behaviour of masonry structures. The variable X_3-*Shear Force X* shows the X value of *"The Contribution of the Walls with Inadequate Strength to Shear Force Resistance at the Critical Storey (%)"*, which is obtained based on the existing situation performance analysis results. The variable X_4-*Seismic Zone* shows the seismic zone (i.e., 1st, 2nd, 3rd, 4th or 5th degree seismic zones) where the structure to be assessed is located according to its geographical region. The seismic zone variable was not found to be statistically significant in the finally model because the masonry structures assessed in this study were in the 1st and 2nd degree seismic zones. However, the impact of the seismic zone on the risk status of a structure is critical [47]. The variable X_5-*Age* variable is produced using the detected date and Building Construction Year information in the risk assessment reports. The variable X_6-*Building Height* is a value that is entered according to the approximate building dimensions and is an effective parameter in risk detection.

The variables included in the model were not only analysed statistically but they were also included in the model based on their practical impacts as reported in the literature.

According to the Cox and Snell R^2 value presented in Table 3, *Risk* explained 37.20% of the variance in the predicted variable when predictive variables were analysed. Nagelkerke R^2 value showed that 83% of the predicted variable was explained by the predictor variables.

TABLE 3. Model summary.

Step	-2 Log Likelihood	Cox & Snell R^2	Nagelkerke R^2
1	23.542[a]	0.372	0.830

New parameters to be included in the model could increase the predicting power of the model. According to the results of logistic regression analysis without performing variable elimination process, the Chi-square value in the test of model significance was found to be 0.000 with 2 degrees of freedom and the significance level was found to be 1.000. In the analyses performed with all of the available parameters, -2 Log Likelihood value is 0, Cox and Snell R^2 is 0,447 and Nagelkerke R^2 is 1, and the independent variables predict the dependent variable accurately.

The Hosmer and Lemeshow test is defined as the model of fit test using Chi-square values. Whether there is a significant difference between the estimated values and the observed values is examined. Significance levels are controlled. It is concluded that the model estimates for Sig>0.05 do not differ from the observations and that the predictive ability of the model is similar to the actual situation. In the Hosmer and Lemeshow model of fit test in Table 4, the Chi-square value was 0.433 with 8 degrees of freedom and the significance level was found to be 1.000. According to this value, when the predictive variables were analysed, the result of the Hosmer and Lemeshow test was not significant (p>0.05), indicating that the model-data fit was adequate [43]. Therefore, there was no significant difference between the observed and model-predicted values [33].

TABLE 4. Hosmer and Lemeshow test.

Step	Chi-square	df	Sig.
1	0.433	8	1.000

While the correct class rate of the model created without variable elimination is near 100%, that of the final model is 97.3%. The increase in the Chi-square value indicating the difference between the estimated values and the observed values is interpreted in association with the model accuracy.

Path analysis, which is a causal analysis technique, has been used to analyse the relationships between the independent variables and the structures in the model according to the dependent *"Risk"* variable. The analyses were conducted using LISREL 8.8 Software. The effects on *Risk* were analysed using binary logistic regression and a model was built. Hypotheses for the relationships between the variables in the model were formulated based on Annex-A of the Regulations on the Implementation of the Law No. 6306 on Transformation of Areas under Disaster Risk [16]. The geographic region where the structure is located determines the seismic zone degree [48]. The variable *Region*, which shows the city where the structure is located, has a direct effect on the variable *Seismic Zone*. It also has a direct on the *Risk* dependent variable as well as its indirect effect via the *Seismic Zone*. *Number of storeys*, which is one of the most important parameters affecting the earthquake behaviour of masonry structures, is related to seismic zone. The maximum number of storeys allowed in masonry buildings is 2 storeys in first degree seismic zones, 3 storeys in second and third degree seismic zones and 4 storeys in fourth degree seismic zones [49]. There is a direct relationship between the seismic zone where the structure is located and the *Number of Storeys*. This situation is stated in the relevant regulations. The *Number of Storeys* has also an indirect effect on the *Risk* result via the earthquake effect.

The model in question is composed of six independent and one dependent variables correlated with each other. After defining the relationships between the variables, the maximum likelihood estimation method was used in the path analysis for the parameters [50]. Fig. 1 shows the path diagram of the model.

As can be seen in Fig. 2, according to the path coefficient analysis, the *Risk (Y)* dependent variable had the highest direct effect on the *Region* (X_1) variable among the independent variables ($\beta = -0.51$). The negative value of the path coefficient indicated that the correlation was a negative one. The path coefficient analysis also showed that the dependent variable had the lowest direct effect on the *Building Height* (X_6) variable ($\beta = 0.01$). The effect sizes of the variables X_1 and X_2, which indirectly affected the dependent variable Y through X_4, were -0.20 and 0.17, respectively. As discussed theoretically above, X_1 affects Y variable both directly and, at the same time, indirectly via X_4. Likewise,

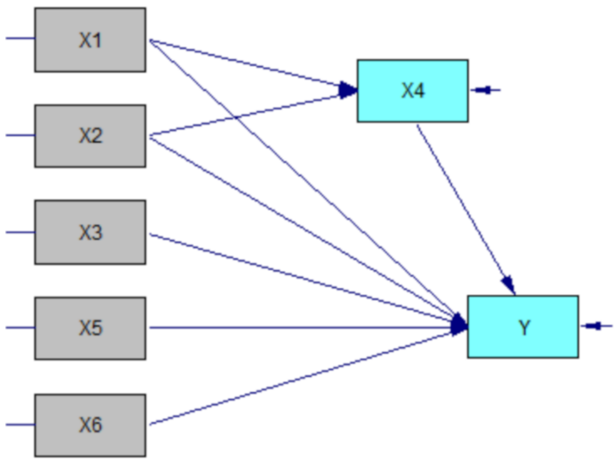

FIGURE 1. Path diagram of the model.

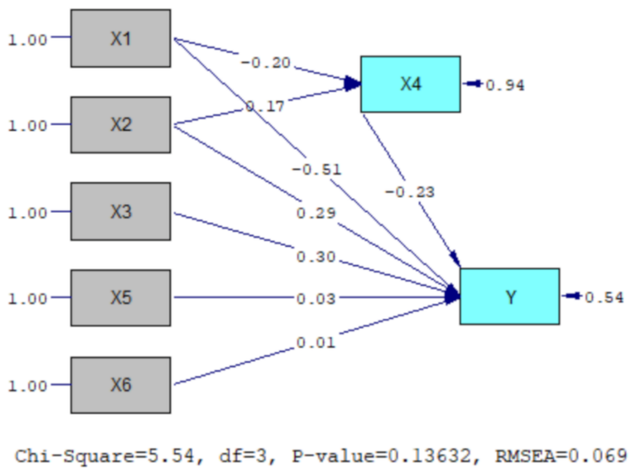

FIGURE 2. Path diagram with standardized path coefficients for the model built.

the variable X_2 affects the Y variable both directly and, at the same time, indirectly via X_4.

Analysis results suggest that there were latent variables with consequent effects that could not be included in the measurement. According to Kline, path coefficients with absolute values less than 0.10 indicate a small effect size, values around 0.30 indicate a medium effect size; and values greater than 0.50 indicate a large effect size [42]. In light of this, we found that the variable X_1 had a large negative effect but the variables X_2 and X_3 had medium positive effect on predicting the result directly.

The most important advantage of path analysis is that it allows for measuring the direct and indirect effects of one variable on the other. In this way, the sizes of the direct and indirect effects can be compared and the total effect can be obtained (Table 5) [51].

In this study, a model was built based on the variables analysed within the scope of the related laws, regulations and literature, and the RMSEA, x^2/sd, CFI, GFI, AGFI, NFI and NNFI fit indices of the model were examined according to the criterion values in Table 6 [42]. The RMSEA refers to Root Mean Square Error of Approximation and is a measure of approximate fit in the population [52]. Since chi-square statistic is rapidly affected by sample size, x^2/sd value, which is obtained by x^2 value by degree of freedom, was used instead of it [53]. The GFI (Goodness of Fit Index) indicates the extent to which the model reproduces the sample covariance matrix. The GFI tends to decline in cases where the degree of freedom is higher compared to the sample size. The AGFI (Adjusted Goodness of Fit Index), on the other hand, is an index used to overcome the shortcomings of GFI in cases with a large sample size [52]. The AGFI index is used in cases where excess sample volume increases the GFI value and prevents accurate results. However, it is not recommended to use it with a low sample volume.

The model in our study, which consists of 183 samples, is not suitable to be assessed by the AGFI index of fit [53]. The CFI (Comparative Fit Index) value assumes that there is no correlation between the variables and indicates the difference of the model built from the null model [52]. The

TABLE 5. Effect values of the variables in the model.

$X_1 \rightarrow Y$	Effect Values
Direct Effect	-0.51
Indirect Effect	0.088
Total Effect	-0.422
$X_2 \rightarrow Y$	
Direct Effect	0.286
Indirect Effect	0.051
Total Effect	0.337
$X_3 \rightarrow Y$	
Direct Effect	0.297
Indirect Effect	0.111
Total Effect	0.408
$X_4 \rightarrow Y$	
Direct Effect	-0.228
Indirect Effect	0.132
Total Effect	-0.096
$X_5 \rightarrow Y$	
Direct Effect	0.032
Indirect Effect	0.166
Total Effect	0.198
$X_6 \rightarrow Y$	
Direct Effect	0.009
Indirect Effect	0.194
Total Effect	0.203
$X_1 \rightarrow X_4$	
Direct Effect	-0.020
$X_2 \rightarrow X_4$	
Direct Effect	0.17

TABLE 6. Criterion Values for Model Fit Indices.

Fit Indices	Criterion Values	
	Perfect Fit	Good Fit
RMSEA	≤ 0.05	≤ 0.08
x^2/sd	≤ 2.00	≤ 3.00
GFI	$\geq$0.95	$\geq$0.90
AGFI	$\geq$0.95	$\geq$0.90
CFI	$\geq$0.95	$\geq$0.90
NFI	$\geq$0.95	$\geq$0.90
NNFI	$\geq$0.95	$\geq$0.90

NFI (Normed Fit Index) is based on rescaling x^2 in a range between 0 and 1. This fit index is calculated using the null model [54]. Finally, the NNFI (Non-normed Fit Index) is used to compare alternative models or to measure alternative models with the null model. In cases where the sample volume is not too large, the NFI value does not approach 1 even if the existing model is accurate. The NNFI provides a solution in such cases [55]. However, where small volume samples are used, the value of the NNFI may show poor fit despite other indexes of fit indicating good fit [42], [46], [53].

According to the criterion values for model fit indices, the RMSEA, AGFI and NNFI values of the model indicated a good fit while the other values indicated a perfect fit (RMSEA = 0.069; x^2/sd = 5.54/3; GFI = 0.99; CFI = 0.99; NFI = 0.97; NNFI = 0.90).

IV. DISCUSSION AND CONCLUSION

In this study, the parameters that were assessed based on the decisions made by the authorized institutions and organizations in the process of urban transformation about the risk status of masonry buildings undergoing a risk assessment process were chosen, and a comprehensive rapid assessment model was developed. The path analysis method was used to assess the direct and indirect effects of various parameters on the assessment process by using the risk assessment data for risky structures obtained from different regions.

Having been built based on data from three different geographical regions, this rapid assessment model is intended to strongly contribute to the decision process and to minimize the time and costs of risk assessment and the damage created in the structure in this process. The proposed rapid assessment model includes a case study and path model process based on a complex and multi-parameter conceptual framework, and it can be applied to different regions and structures. Urban transformation is in development stage in Turkey and hence only a part of the aimed transformation has been realized throughout the country. In this regard this work is of critical importance especially for countries near an active seismic belt in assessment of risks in structures.

The model built in the study was developed as an overall initial assessment method that can be applied to conduct the risk assessment of a masonry structure in any geographical area. In addition, the model was developed in accordance with the Annex-A rapid evaluation and Annex-2 comprehensive assessment guidelines suggested by the current legislation in Turkey and in light of the relevant literature. In conclusion, risk analysis is an expensive, destructive and time-consuming process requiring expert evaluation. The risk assessment process typically takes weeks. This process is completed in one week under optimum conditions with fieldwork that should be performed by an expert team, data collection from the structure as a result of destructive inspection and computer simulation. The current studies identify risky area with rapid assessment methods. However, a structure in a risky area may be ''safe'' or a structure in an area without risk assessment may be ''risky''. The model developed in this work can be conveniently used to gain insights about the risk status of structures. By determining the risk status of individual masonry structures in particular, the model aims to provide a quick assessment about whether to make a risk assessment or not with advanced analysis techniques. Further analyses are advised to be performed on the structures determined risky. Although path analysis statistically tests causal models, this is essentially a causality that the researcher creates. The obtained ''causal'' result is confirmed in terms of model fit, but it does not necessarily mean that such causality exists in absolute terms [56].

Most existing machine learning algorithms, such as logistic regression, can only be applied to vector-based data [57]. Applying traditional machine learning methods to a graph requires initial processing of the original graphical structure data [58]. In the continuation of this study, we seek to examine the causal graphs and parameters among causal inference techniques, to create a rapid assessment model and to design specialized software compatible with the model.

ACKNOWLEDGMENT

The authors would like to thank our gratitude to Kutahya, Afyon and Eskisehir Provincial Directorates of the Ministry of Environment and Urbanization and to Mr. M. Zafer Alsac, Deputy Mayor of Zeytinburnu, for their support within legal terms. They would also like to thank Prof. Dr. Yasar Hoscan, Asst. Prof. Dr. Hakan Ozbasaran and Asst. Prof. Dr. Murat Dogan for sharing their pearls of wisdom with us during the course of this study.

REFERENCES

[1] A. Karadag and G. Mirioglu, ''Bayrakli kentsel dönüsüm projesi üzerine cografi degerlendirmeler,'' *Türk Cografya Dergisi*, no. 57, pp. 21–32, 2012. [Online]. Available: http://dergipark.org.tr/tcd/issue/21224/227776

[2] M. R. Montgomery, ''The urban transformation of the developing world,'' *Science*, vol. 319, no. 5864, pp. 761–764, Feb. 2008. doi: 10.1126/science.1153012.

[3] O. Anil, M. Sahmaran, and M. K. Kockar, ''Field applications for risk assessment methods of urban transformation law no. 6306: A case study of Beyoglu,'' in *Proc. Int. Conf. Earthquake Eng. Seismol.*, Eskisehir, Turkey, 2017, pp. 28–29.

[4] F. N. Genc, ''Türkiye'de kentsel dönüşüm: mevzuat ve uygulamaların genel görünümü,'' (in Turkish), *Yönetim Ekonomi*, vol. 15, no. 1, pp. 115–130, 2008. [Online]. Available: https://dergipark.org.tr/tr/download/article-file/145982

[5] H. Karaman, ''Earthquake loss assessment study for Zeytinburnu district,'' in *Proc. Sci. Tech. Congr. Geomatics Turkey*, Ankara, Turkey, May 2009, pp. 1–8. [Online]. Available: https://www.hkmo.org.tr/resimler/ekler/04cb95ba2bea9fd_ek.pdf

[6] S. Polat and N. Dostoglu, ''On the concept of urban regeneration: Samples of Kükürtlü and Mudanya in Bursa,'' *Uludağ Univ. J. Fac. Eng.*, vol. 12, no. 1, pp. 61–76, 2007.

[7] H. I. Solak and A. Alaybeyoglu, ''A fuzzy logic based system design for determination of risk areas in urban regeneration,'' *Selcuk Univ. J. Eng. Sci. Technol.*, vol. 5, no. 4, pp. 402–413, Dec. 2017. doi: 10.15317/Scitech.2017.100.

[8] A. Şişman and D. Kibaroğlu, ''The urban renovation projects in the world and Turkey,'' (in Turkish), in *Proc. UCTEA Chamber Survey Cadastre Eng. 12, Sci. Tech. Congr. Geomatics Turkey*, Ankara, Turkey, May 2009, pp. 1–9. [Online]. Available: http://kisi.deu.edu.tr/erkin.baser/Kentsel%20D%C3%B6n%C3%BC%C5%9F%C3%BCm.pdf

[9] G. C. S. Lin, ''Reproducing spaces of Chinese urbanisation: New city-based and land-centred urban transformation,'' *Urban Stud.*, vol. 44, no. 9, pp. 1827–1855, Aug. 2007. doi: 10.1080/00420980701426673.

[10] K. McCormick, S. Anderberg, L. Coenen, and L. Neij, "Advancing sustainable urban transformation," *J. Cleaner Prod.*, vol. 50, pp. 1–11, Jul. 2013. doi: 10.1016/j.jclepro.2013.01.003.

[11] O. Dundar, "Models of urban transformation: Informal housing in Ankara," *Cities*, vol. 18, no. 6, pp. 391–401, Dec. 2001. doi: 10.1016/S0264-2751(01)00031-2.

[12] T. Shiga, A. Shibata, and T. Takahashi, "Earthquake damage and wall index of reinforced concrete buildings," in *Proc. Tohuku District Symp.*, 1968, pp. 29–32.

[13] *Rapid Visual Screening of Buildings for Potential Seismic Hazards: Supporting Documentation*, 3rd ed., Federal Emergency Manage. Agency, Washington, DC, USA, 2015.

[14] S. S. Tezcan, Gursoy, E. Kaya, and I. E. Bal, "Project to prevent loss of life in eartquakes," (in Turkish) in *Proc. Kocaeli Emergency Conf. (Kocaeli)*. İstanbul, Turkey: İstanbul Technical Univ., Jan. 2003, pp. 146–153.

[15] I. E. Bal, S. S. Tezcan, and F. G. Gulay, "P25 rapid screening method to determine the collapse vulnerability of RC buildings," in *Proc. 6th Nat. Conf. Earthquake Eng.*, Istanbul, Turkey, Oct. 2007, pp. 661–674.

[16] *The Law of Transformation of Areas under Disaster Risk*, Turkey Legal Gazette, Ankara, Turkey, 2012, pp. 11579–11590.

[17] *Implementation Regulation of Law 6306*, Turkey Legal Gazette, Ankara, Turkey, 2012.

[18] L. A. Cox, Jr., and P. F. Ricci, "Causation in risk assessment and management: Models, inference, biases, and a microbial risk–benefit case study," *Environ. Int.*, vol. 31, no. 3, pp. 377–397, Apr. 2005. doi: 10.1016/j.envint.2004.08.010.

[19] J. D. Ramsey and D. Malinsky, "Comparing the performance of graphical structure learning algorithms with TETRAD," 2016, *arXiv:1607.08110*. [Online]. Available: https://arxiv.org/abs/1607.08110

[20] S. Souravlas, A. Sifaleras, and S. Katsavounis, "A parallel algorithm for community detection in social networks, based on path analysis and threaded binary trees," *IEEE Access*, vol. 7, pp. 20499–20519, 2019. doi: 10.1109/ACCESS.2019.2897783.

[21] A. Perdicoúlis, M. Hanusch, H. D. Kasperidus, and U. Weiland, "The handling of causality in SEA guidance," *Environ. Impact Assessment Rev.*, vol. 27, no. 2, pp. 176–187, Mar. 2007. doi: 10.1016/j.envint.2004.08.010.

[22] G. Voegeli, W. Hediger, and F. Romerio, "Sustainability assessment of hydropower: Using causal diagram to seize the importance of impact pathways," *Environ. Impact Assessment Rev.*, vol. 77, pp. 69–84, Jul. 2019. doi: 10.1016/j.eiar.2019.03.005.

[23] E. J. Pedhazur, *Multiple Regression in Behavioral Research: Explanation and Prediction*, 3rd ed. Boston, MA, USA: Thomson Learning, 1997.

[24] C. Glymour, K. Zhang, and P. Spirtes, "Review of causal discovery methods based on graphical models," *Frontiers Genet.*, vol. 10, pp. 1–15, Jun. 2019. doi: 10.3389/fgene.2019.00524.

[25] W. L. Paul and M. J. Anderson, "Causal modeling with multivariate species data," *J. Exp. Mar. Biol. Ecol.*, vol. 448, pp. 72–84, Oct. 2013. doi: 10.1016/j.jembe.2013.05.028.

[26] V. Sumbuloglu, R. Alpar, and P. Ozdemir, "Investigation of relationships between variables," *Turkey Int. Med. J.*, vol. 5, no. 6, pp. 416–419, 1998.

[27] A. Gosain and M. Bhugra, "A comprehensive survey of association rules on quantitative data in data mining," in *Proc. IEEE Conf. Inf. Commun. Technol. (ICT)*, Apr. 2013, pp. 1003–1008. doi: 10.1109/CICT.2013.6558244.

[28] E. Goldvarg and P. N. Johnson-Laird, "Naive causality: A mental model theory of causal meaning and reasoning," *Cogn. Sci.*, vol. 25, no. 4, pp. 565–610, Jul. 2001. doi: 10.1016/s0364-0213(01)00046-5.

[29] D. L. Streiner, "Finding our way: An introduction to path analysis," *Can. J. Psychiatry*, vol. 50, no. 2, pp. 115–122, Feb. 2005. doi: 10.1177/070674370505000207.

[30] K. Kayabali and M. Akin, "Seismic hazard map of Turkey using the deterministic approach," *Eng. Geol.*, vol. 69, nos. 1–2, pp. 127–137, Apr. 2003. doi: 10.1016/S0013-7952(02)00272-7.

[31] M. E. Sonmez, "Cografi bilgi sistemleri (CBS) tabanli deprem hasar riski analizi: Zeytinburnu (Istanbul) örnegi," *Turkish Geogr. Soc.*, vol. 56, pp. 11–22, 2011. [Online]. Available: https://dergipark.org.tr/tr/download/article-file/198442

[32] C. A. Mertler and R. V. Reinhart, *Advanced and Multivariate Statistical Methods: Practical Application and Interpretation*, 3rd ed. New York, NY, USA: Routledge, 2017.

[33] O. Cokluk, G. Sekercioglu, and S. Buyukozturk, *SPSS and LISREL Applications of Multivariate Statistics for the Social Sciences*, (in Turkish), 2012.

[34] Z.-Y. Yang, Y. Liang, H. Zhang, H. Chai, B. Zhang, and C. Peng, "Robust sparse logistic regression with the L_q(0<q<1) regularization for feature selection using gene expression data," *IEEE Access*, vol. 6, pp. 68586–68595, Dec. 2018. doi: 10.1109/ACCESS.2018.2880198.

[35] A. Agresti, *An Introduction to Categorical Data Analysis*, 3rd ed. Maitland, FL, USA: Wiley, 2019.

[36] F. E. Harrell, "Binary logistic regression," in *Regression Modeling Strategies* (Springer Series in Statistics), 2nd ed. Cham, Switzerland: Springer, 2015, ch. 10, pp. 219–274. doi: 10.1007/978-3-319-19425-7_10.

[37] D. Lei, M. Du, H. Chen, Z. Li, and Y. Wu, "Distributed parallel sparse multinomial logistic regression," *IEEE Access*, vol. 7, pp. 55496–55508, 2019. doi: 10.1109/ACCESS.2019.2913280.

[38] G. E. Bonney, "Logistic regression for dependent binary observations," *Biometric*, vol. 43, no. 4, pp. 951–973, Dec. 1987. doi: 10.2307/2531548.

[39] D. G. Kleinbaum, L. L. Kupper, K. E. Müller, and A. Nizam, *Applied Regression Analysis and Other Multivariable Methods*, 3rd ed. Belmont, CA, USA: Duxbury Press, 1997.

[40] Z. Bursac, C. H. Gauss, D. K. Williams, and D. W. Hosmer, "Purposeful selection of variables in logistic regression," *Source Code Biol. Med.*, vol. 3, no. 17, pp. 1–17, 2008. doi: 10.1186/1751-0473-3-17.

[41] *IBM SPSS Regression 20*, IBM Corp., Chicago, IL, USA, 2011.

[42] R. B. Kline, *Principles and Practice of Structural Equation Modeling*, 3rd ed. New York, NY, USA: Guilford Press, 2011.

[43] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, 2nd ed. Hoboken, NJ, USA: Wiley, 2000.

[44] Q. Wang, G. Zhang, C. Sun, and N. Wu, "High efficient load paths analysis with U* index generated by deep learning," *Comput. Methods Appl. Mech. Eng.*, vol. 344, pp. 499–511, Feb. 2019. doi: 10.1016/j.cma.2018.10.012.

[45] E. Oktay, M. M. Akinci, and A. Karaaslan, "Research on the interaction of statistics with the courses in business administration curriculum with using path analysis," *Atatürk Univ. J. Graduate School Social Sci.*, vol. 16, no. 1, pp. 513–527, 2012. [Online]. Available: http://dergipark.org.tr/ataunisosbil/issue/2829/38369

[46] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, 3rd ed. New York, NY, USA: Harper Collins, 1996.

[47] T. Rashed and J. Weeks, "Assessing vulnerability to earthquake hazards through spatial multicriteria analysis of urban areas," *Int. J. Geogr. Inf. Sci.*, vol. 17, no. 6, pp. 547–576, Sep. 2003. doi: 10.1080/1365881031000114071.

[48] S. Durak, "Masonary buildings comman used in Aegean region and safety of these buildings," (in Turkish), 2007.

[49] *Specification for Structures to Be Built in Disaster Areas*, Turkey Legal Gazette, Ankara, Turkey, 2007.

[50] Y. Peng, T. Feng, and H. Timmermans, "A path analysis of outdoor comfort in urban public spaces," *Build. Environ.*, vol. 148, pp. 459–467, Jan. 2019. doi: 10.1016/j.buildenv.2018.11.023.

[51] H. B. Asher, *Causal Modeling* (Quantitative Applications in the Social Sciences). Thousand Oaks, CA, USA: SAGE, 1983.

[52] B. H. Munro, *Statistical Methods for Health Care Research*, 5th ed. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2005.

[53] D. Hooper, J. Coughlan, and M. R. Mullen, "Structural equation modelling: Guidelines for determining model fit," *Electron. J. Bus. Res. Methods*, vol. 6, no. 1, pp. 53–60, 2008.

[54] M. Dogan, "Influence of sample size, estimation method and normality on fit indices in confirmatory factor analysis," M.S. thesis, Dept. Statist., Osmangazi Univ., Eskisehir, Turkey, 2013.

[55] R. E. Schumacker and R. G. Lomax, *A Beginner's Guide to Structural Equation Modeling*, 3rd ed. New York, NY, USA: Routledge, 2010.

[56] N. Sumer, "Yapisal esitlik modelleri: Temel kavramlar ve örnek uygulamalar," *Türk Psikoloji Yazilari*, vol. 3, no. 6, pp. 49–74, 2000. [Online]. Available: https://www.researchgate.net/profile/Nebi_Suemer/publication/281981476_Yapidotlesssal_esitlik_modelleri_Temel_kavramlar_ve_ornek_uygulamalar/links/5678718208aebcdda0ebd8df/Yapidotlesssal-esitlik-modelleri-Temel-kavramlar-ve-oernek-uygulamalar.pdf

[57] W. Liu, H. Liu, D. Tao, Y. Wang, and K. Lu, "Manifold regularized kernel logistic regression for Web image annotation," *Neurocomputing*, vol. 172, pp. 3–8, Jan. 2016. doi: 10.1016/j.neucom.2014.06.096.

[58] T. Ma, W. Shao, Y. Hao, and J. Cao, "Graph classification based on graph set reconstruction and graph kernel feature reduction," *Neurocomputing*, vol. 296, pp. 33–45, Jun. 2018. doi: 10.1016/j.neucom.2018.03.029.