Development of a model to forecast the external migration rate in development projects reflecting city characteristics

Ki-Bum Kim¹, Joon Park², Jee-Won Seo¹, Young-Jun Yu³, In-Hwan Hyun⁴, Ja-Yong Koo⁷

¹Department of Environmental Engineering, University of Seoul, Seoul 02504, Republic of Korea
²International School of Urban Sciences, University of Seoul, Seoul 02504, Republic of Korea
³Business Development Division, K-water, Daejeon 34350, Republic of Korea
⁴Department of Civil and Environment Engineering, Dankook University, Yongin 31116, Republic of Korea

ABSTRACT

In planning public service systems such as waterworks, the design population is a very important factor. Owing to the limitations of the indirect method, two new models, which take into consideration urban characteristics, were developed to accurately predict external migration rate (EMR), which is an essential component in estimating reliably the design population. The root mean square error (RMSE) between the model values and observed values were 10.12 and 15.58 for the metropolitan cities and counties respectively and were lower compared to RMSE values of 27.31 and 28.79 obtained by the indirect method. Thus, the developed models provide a more accurate estimate of EMR than the indirect method. In addition, the major influencing factors for external migration in counties were development type, ageing index, number of businesses. On the other hand, the major influencing migration factors for cities were project scale, distance to city center, manufacturing size, population growth rate and residential environment. Future medium and long-term studies would be done to identify emerging trends to appropriately inform policy making.

KEYWORDS: City characteristics, Development project, External migration rate, Quantification theory type I, Urban characteristics, Water supply

1. Introduction

Water demand forecasting in the mid- and long-term begins with predicting the design population. For this reason, the design population is the most important factor in planning public service systems, such as waterworks. However, the design population is characterized by uncertainty and variability, due to the nature of long-term planning; thus, it does not necessarily serve as a precise guide and may simply be utilized to determine a vision for the region. Given that proposing a design population unilaterally can create a psychological obstacle that immobilizes future development potential, there is one trend of opinion that its importance is low [1].

However, when the size of the design population is incorrectly estimated, all water demand plans are either over- or underestimated [2]. Insufficient water demand planning creates an unstable water supply; as water is a basic human right, this cause an imbalance in water welfare. Meanwhile, excessive water demand planning reduces the operational efficiency of limited water resources and, further, causes waste at the national level. To ensure the effectiveness of the National Waterworks Master Plan and Waterworks Basic Plan in Korea, an accurate population forecast is required and this means setting a realistic and feasible target. At present, predictions of the design population have too much elasticity. A more accurate and reasonable estimation method is needed.

Korean Ministry of Land, Infrastructure and Transport (MOLIT) and Korean Ministry of Environment (MOE) [3, 4] have presented a methodology for estimating population for various water supply plans, including the Korean National Planning and Local Government Waterworks Basic Plans. Both studies claim that the cohort-component method using the survival model should be employed to predict the design population, while a social increment should be considered as an auxiliary means. The social increment takes into account external migration (EM) for development projects.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Fig. 1 shows the change in population with the implementation of a development project is City A. The external migration rate (EMR) is defined as the ratio between the population entering the development project area and the cities outside the development project area. Regarding the migration mobility diagram shown in Fig. 1, the EMR is defined as Eq. (1) [3, 4]. The terms SM_{oc} and SM_{co} in Fig. 1 are definitions of social migration that occurs irrespective of development projects. SM_{c} in the figure also, represents social migration that occurs irrespective of development projects, however, such migrations do not change the population of City A.

\[
EMR = \frac{EM}{EM + IM} \times 100\%
\]  

(1)

MOLIT and MOE [3] have presented the EMR for cities and provinces in Korea, while MOLIT and MOE [4] have demonstrated the value of subdividing development project types, based on the results of previous studies. However, the values in previous studies have indirectly estimated the EMR, due to difficulties in data acquisition. The EMR derived using the indirect method proposed by MOLIT and MOE [3, 4], which is commonly used for designing populations can be calculated using Eq. (2). This method is problematic for the following reasons.

\[
EMR = \frac{CPGR \times CP - NPGR \times CP}{DP} \times 100\%
\]  

(2)

\(\text{if } CPGR \times CP - NPGR \times CP > 0\)

First, the EMR has been derived based on the EMR of 66 residential development projects examined in 2006. It does not reflect the trends of the latest development projects. Second, it has been derived using the indirect method. Indirect method has some shortcomings. It does not take into consideration of the actual EM. The indirect method is based on the assumption that the amount by which the growth rate of city and county population (CPGR) exceeds the average growth rate of the national population (NPGR) represents the increase in the social population. In addition, the increase in social population is assumed to be the result of EM. However, the social population can increase regardless of development projects. This causes a difference in population estimates, given that social population growth occurs in areas without any development projects. In addition, when estimating the EMR, the value derived by dividing the social population growth by the design population of the development project (DP) is used. For this reason, some errors may occur when calculating the EMR. Third, since the EMR is reported for provinces, the different conditions affecting external inflows to different cities in the same province are not taken into account.

To address the above problems and apply an accurate EMR, researchers have been asked to calculate an EMR that takes into consideration the actual transfer population in development project areas, based on the latest development projects trends, and applies the EMR best suited to each city. Accordingly, this study has developed a model that takes into consideration the urban characteristics that affect the overall EMR. The main objective of this study is to provide basic data, based on reasonable population forecast results for predicting water demand.

2. Literature Review

Balogun et al. [5] examined the impact of population changes on water supply and proposed a practical solution to reduce the gap between water supply and demand. As changes in population may cause great changes to public services such as the water supply, many studies have attempted to estimate the accurate population.

The present study examines prior research related to external migration mobility. As the EMR comprises social population movement, it is necessary to estimate it considering the factors that influence population movements; that is, urban characteristics. The previous studies that have identified urban characteristics that affect population movement are summarized below.

Molloy et al. [6] have argued that job changing has an effect on domestic migration, with gender, age, education, race/ethnicity, nativity, the presence of children in the household, the number of working adults, income, employment status, homeownership, and the percentage of married households made an internal migration in the United States. McCatty [7] has developed three models (migration and the dual sector model of economic development, the family/household migration model, and the Todaro model of migration), concluding that job finding, family changing, health, and education affected migration. Clark and Withers [8] have suggested that patterns of migration change due to changes in family composition. Miheretu [9] has claimed that the distance between two areas, density of population, education, and employment affect population movement.

Previous studies have indicated that social population movements are affected by the characteristics of people behaviors. However, previous studies involving Korea, which is a developing country, have indicated that social population movements are more affected by city characteristics than by people's
behavior. The previous studies that have identified urban characteristics that affecting population movement are summarized below.

In researching factors that influence population movement, Yoo [10] has found that when wage levels and housing supply rates increase, EM also increases. Kim and Jang [11] have observed that the expected income ratio and educational opportunity ratio can create population growth, triggering significant EM. Lee [12] has analyzed the correlation between population movement and influence factors, discovering that the number of pharmacies, the gas coverage rate, and the housing area all affect population movement. Moreover, Lee [13] has discovered that income opportunities, job opportunities, the tax burden, comfort, housing construction, capital area readjustment zone separation, population size, and migration distance as population movement factors in Gyeonggi Province. Lee and Park [14] have noted that population movement into a residential development project area occurs because people want to purchase a house or improve workplace proximity, education, or their residential environment. In particular, those aged 20-29 usually move for the purpose of purchasing a house, those aged 30-39 predominantly move in order to be closer to their workplace, those aged 40-49 relocate for the purpose of education, and those aged 50-59 migrate in order to improve their residential environment. Migration has thus been shown to be associated with the life cycle. Furthermore, the shorter the migration distance, the higher the rate of relocating for individuals aged 20-29 and 40-49, whereas those aged 30-39, there is a higher rate of migration for longer distances.

Kim and Jang [15] have found that in many cases of population movement, migration occurs to areas that have social and economic characteristics that are similar to their existing residential environments. Kim and Choi [16] calculated the EMR of housing development projects in capital and metropolitan cities by analyzing the external inflow rate for 11 residential land development project districts in cities across the country during the 1990s. Kwon [17] has found that age, gender, educational level, manufacturing ratio, labor rate, and rural/urban status are factors that affect population migration to metropolitan areas, while Choi and Kim [18] have found that land development areas, design population, local downtown distances, population density, and industrial structures affect population movement. Lee and Lee [19] have analyzed the characteristics of population movement into new cities, discovering that, in the case of migration within the capital city, educational level, housing scales, occupancy types, and household type had the biggest impact.

Jeon [20] has analyzed the EMR of 594 new apartment buildings in Gyeonggi-do and subsequently comparing the results with the EMR applied when establishing the city basic plan. Areas involving large-scale housing development projects were found to have high EMR, in comparison to other areas. It is necessary to consider the in-migration rate when making such calculations, given the strong correlation between the type of development project, the size of development project, and the EMR. Hur [21] has analyzed the EMR in relation to local government, development project size, development project type, and distance from Seoul by targeting 594 apartment buildings constructed in Gyeonggi Province since 2005. Based on the analysis results, the factors were divided into project district factors and regional factors, and while project type and project enforcer influenced the EMR. As a complement to Hur’s [21] study, Lee [22] has found the EMR to be higher as the distance from Seoul decreases and employment conditions become more favorable.

From this comprehensive analysis of prior research, the common factors influencing the EMR are found to be development project size and type, distance from capital city, and factors related to economic activity. However, these results reflect the fact that most previous studies have used data drawn solely from Gyeonggi Province. In this study, additional city characteristics are identified and applied to the model to predict the EMR.

3. Research Methods and Data

3.1. Investigation of Development Projects

To create a prediction model of EMR according to urban characteristics reflecting the latest trends in development projects, the present study has used a sample of 113 development projects completed by the end of 2014. These projects were selected from among development plans listed in the land use regulation information system of the MOLIT. They have been categorized by development project type and area, as shown in Table 1.

The first limitation of this research is in here. Due to the time and effort required to obtain population data associated with migration to development project areas, a relatively small number of samples was used for model development. The analysis results regarding factors influencing EMR can therefore change, depending on the level of data scale. Future studies need to consider data in the medium to long term in order to analyze the influence factors in more detail.

Fig. 2 shows the location of the development projects. The project name in Fig. 2 is expressed as “NNN_XX_DDDM”, where “NNN” represents the development project number, “XX” is an abbreviation that stands for the area listed in Table 1, “DDD” indicates the type of development project; and “M” represents the number of development projects.

3.2. Investigation of External Migration Rates

Since the indirect method used previously is inaccurate, data relating to the population moving into a development project area must be used to calculate the EMR of the development project accurately. The population data used to derive the EMR of the 113 development projects sampled in this study, have been drawn from the resident-registration data of populations that moved into development project areas during the year that followed the completion date of each development project. These
Table 1. Investigated Development Project by Project Type and Area

| Area      | Abbreviation in Fig. 2 | Development project type | Total |
|-----------|-------------------------|--------------------------|-------|
|           |                         | RSD | CID | RED | RCD | REI |       |
| Seoul     | SO                      | 1   | 2   | 1   | 5   | 1   | 10    |
| Busan     | BS                      | 3   | 1   | -   | -   | 1   | 5     |
| Daegu     | DG                      | 3   | -   | 1   | 3   | -   | 7     |
| Incheon   | IC                      | 3   | 3   | 2   | 4   | 3   | 15    |
| Gwangju   | GJ                      | 3   | -   | -   | -   | -   | 3     |
| Daejeon   | DJ                      | -   | 1   | -   | 1   | 2   | 4     |
| Ulsan     | US                      | 1   | -   | -   | -   | -   | 1     |
| Gyeonggi  | GG                      | 19  | 5   | 2   | 4   | -   | 30    |
| Gangwon   | GW                      | 3   | 1   | -   | 2   | -   | 6     |
| Chungbuk  | CB                      | 5   | -   | -   | -   | -   | 5     |
| Chungnam  | CN                      | 1   | -   | -   | -   | -   | 1     |
| Jeonbuk   | JB                      | 3   | -   | -   | -   | -   | 3     |
| Jeonnam   | JN                      | 2   | 1   | -   | -   | -   | 3     |
| Gyeongbuk | GB                      | 5   | 3   | -   | -   | 2   | 10    |
| Gyeongnam | GN                      | 4   | 3   | -   | 3   | -   | 10    |
| Jeju      | -                       | -   | -   | -   | -   | -   | -     |
| Total     |                         | 56  | 20  | 6   | 22  | 9   | 113   |

Fig. 2. Location of investigated development project.
data have been used in cooperation with the MOLIT and Korean Ministry of Public Administration and Security (MOPAS); they cover, 60 of the total 229 cities, counties, and districts in which the 113 development projects were implemented.

Using these data, the present study has analyzed both the populations that migrated to development project areas within cities and counties and the populations that migrated to development project areas outside cities and counties. To calculate the EMR, among the population that migrated into the development project areas, the population that migrated from outside cities was used to calculate the ratio of the total population that moved into development project area.

3.3. Investigation of Factors Affecting the External Migration Rate

As mentioned previously, one of the problems related to the EMR presented by the MOLIT and MOE [4] was the application of the same EMR throughout the province, regardless of different city characteristics within the province. To address this problem, this study has considered the urban characteristics of each city, county, and district; that is, those urban characteristics that affect the EMR.

As a result of a comprehensive review of previous studies, the factors affecting the EMR can be largely classified into three types: factors related to development projects (such as development project size), factors related to population characteristics in historical development projects, and economic and industrial factors of the development project area, including the employment rate and the number of businesses. In this study, 11 sub-factors derived from the three main factors affecting the EMR have been analyzed for the 113 development projects sampled, as shown in Table 2. The effect of these factors on the EMR has also been examined.

The second limitation of this research is in here. The analyses of various influential factors also need to be considered. For example, the direction of government policy can have a significant impact on population migration and water supply planning. If a government-level development project is undertaken, as described in the example above, population migration rates may be much larger than expected.

To discover the relationship between EMR and its influencing factors, the standard correlation coefficient was calculated using Eq. (3). When a sample is small, the standard correlation coefficient is used as an index to determine the correlation in the significance probability of the t distribution. If the correlation coefficient between the factors is larger than the standard correlation coefficient, there is a correlation between the two factors.

$$R_0 = \frac{t_{\alpha}}{\sqrt{n - 2 + t_{\alpha}^2}}$$ (3)

Here, $R_0 = \text{standard correlation coefficient}$, $n = \text{number of sample (113 development projects)}$, $t_{\alpha} = \text{t value of t distribution at significance } \alpha$ ($t$ value for 113 samples at 95% significance level = 1.661)

Table 2. Factors Influencing External Migration Rate

| Category                | Research contents                                                                 |
|-------------------------|----------------------------------------------------------------------------------|
| Development project     | 5 development project types (residential development project, city development project, redevelopment project, reconstruction project, residential environment improvement project) |
| Development project type ($P_t$) | Number of projected households (number of households)                              |
| Development project scale ($P_s$) | Year of completion of development project (year)                                   |
| Completion year ($YC$)  | Distance from development project area to metropolitan city hall or shortest distance among calculated distances to provincial government office (shortest distance by vehicle) (km) |
| Distance to city center ($D$) | Average population growth rate for the three years prior to completion (%)         |
| Population              | Average aging index for three years prior to completion (rate of population under 14 y of age compared to population over 65) (%) |
| Population growth rate ($PGR$) | Daytime active population ratio compared to the most recent resident population based on year of completion of development project (%) |
| Demographic structure ($DGS$) | Average percentage of employed persons in the population aged 15 y or older for three years prior to completion (%) |
| Daytime population index ($DPI$) | Number of industry for businesses compared to average area for three years prior to completion (number of businesses/km²) |
| Employment rate ($R_e$)  | Rate of the number of manufacturing businesses compared to total number of businesses for three years prior to completion (%) |
| Economy and industry    | Gross regional product per person ($GRP$) (million won/person)                     |

410
3.4. Development of Model to Forecast the External Migration Rate

The prediction method for the EMR used quantification I analysis, which comprises quantitative numerical analysis of qualitative variables. This method was used because, as shown in Table 2, the qualitative variables of the factors influencing the EMR, such as the development project type, were included. Quantification I analysis can typically be understood as a dummy regression analysis, and multiple regression analysis is an analytical method used to develop a model that can describe or predict the value of a dependent variable using more than two variables. The following Eq. (4) indicates the general formula that was derived through the quantification theory type I analysis.

\[ Y = A_0 + A_1 X_1 + A_2 X_2 + \cdots + A_P X_P + \epsilon \]  (4)

Here, \( Y \): Dependent variable  
\( X_1, X_2, \ldots, X_P \): Independent variable  
\( A_0, A_1, \ldots, A_P \): Regression coefficient  
\( \epsilon \): error

Each independent variable was selected as an influencing factor in the EMR, and the quantification I analysis was carried out using SPSS 23.0. The step selection method was used for the process of developing the regression model. In this method, the statistical significance of each independent variable is verified as the variables are added, step by step. If the F value for the input value is greater than the entry value, the variable is included in the regression model. If a new variable is included, when the F value for the value that is input by recalculating the F value is greater than the rejection value, the variable contained in the regression model is removed. When there is no independent variable for which the F value is greater than the entry value and no variable that needs to be removed, the phase of the regression model ends. In this case, as the sample size was small, 0.15 and 0.25 were applied as the significance probabilities of the F value including the independent variable, for the significance probability of the F value’s entry and the significance probability of the F value’s removal.

Regarding the derived final quantification I model, the significance of the model itself was examined by conducting a significance test (variance analysis). A significance test was also carried out on the regression coefficient included in the final model. Lastly, the fitness of the estimation model was derived through a collinearity test of the independent variable included in the final model.

4. Research Results

4.1. Result of Investigation of External Migration Rate in Development Project

Using the examined resident registration computer processing information data, with the cooperation of the MOLIT, the MOPAS, and 60 cities, counties, and districts, the actual EMR was calculated for 113 development projects. The development project with the highest EMR was found to be the RSD in Guyeong, Ulju-gu, Ulsan, with an EMR of 86.74%. The project with the lowest EMR was Chuncheon Hyoil RCD with a rate of 12.52%.

Overall, the development projects were found to have higher EMRs than those presented by the MOLIT and MOE [4]. The EMR of metropolitan cities was higher than that of cities and counties. This is because the EMR determined by the MOLIT and MOE [4] was derived by dividing an increase in social population, estimated using a specific point in time by the design population. Some deviation occurred because the method of estimating the social population increment was based on the national average population growth rate. The population data
surveyed in this study consisted of resident registration information data. As this dataset was based on the actual lot number, it is a more accurate way of determining the EMR of development projects.

The mean value of the estimated EMR for the 113 development projects was 47.25%. The EMRs by development project type (Pt) are shown in the left figure in Fig. 3. The average EMR of CID was the highest, at 49.92%, the average EMR of RCD was the lowest, at 42.42%. The right figure in Fig. 3 illustrates the box plot for the EMR by Pt. There is a significant deviation within the Pt and for this reason, it is difficult to clearly identify the characteristics of the EMR using Pt alone.

Fig. 4 shows the calculated average EMR, in which 113 development projects are classified by area (metropolitan cities/Gyeonggi-do/other cities and counties) and results of the box plot diagram.

The average EMR of metropolitan cities was 55.87%. The EMR of Gyeonggi-do was 44.74%, and that of other areas was 39.01%. The average EMR of metropolitan cities was found to be 10% higher than the average EMR of other areas. This was consistent with the results of previous studies, which found that shorter distances to the city center were associated with higher EMR.

The results of Jeon’s [20] study found that the EMR of Gyeonggi-do was 39.9%. Jeon [20] conducted a census-type survey, with a different sample for Gyeonggi-do than the one used in this study. Accordingly, it is difficult to directly compare results. However, for the EMR of Gyeonggi-do, the resulting value does not show any large deviation from the previous result; it can thus be concluded that the reliability of the present survey is high.

Meanwhile, as shown in the right figure in Fig. 4, when some outliers were removed, the EMR showed deviations between metropolitan cities and other areas. When that was taken into account, it was judged that the prediction model of EMR should be divided into metropolitan cities and other areas. The prediction model of EMR developed in this study also used these divisions.

4.2. Result of Analysis of Factors Influencing External Migration Rate

Table 3 displays the results of an analysis of the correlation between factors that influence the EMR, as shown in Table 2 and the derived EMR. As the influencing factors vary by period, data based on development project completion date were analyzed.

As the D was shorter, the EMR was found to be higher. The higher in-migration population rate of cities and counties in areas that are close to the city center, which tends to be economically active areas, is judged to be the natural result of economic activity focused in city centers in Korea.

Although the increasing population growth rate (PGR) has been associated historically with an overall high EMR, the range was found to be insignificant. However, when this finding examined in detail, the trend was found to be based on a PGR of 1.0%, as shown in Fig. 5. When a development project is being constructed, population outflows will be larger than population inflows in comparison to metropolitan cities until the development project is completed. When a development project is completed, the population that has temporarily migrated to other cities and counties will relocate to areas surrounding the development project. In other words, the phenomenon of a high EMR in areas that have a PGR of less than 1.0% compared to areas with a rate exceeding 1.0% is probably due to this pattern.

However, when several factors are considered simultaneously, a phenomenon occurs whereby the correlation between certain factors may change or reverse. Accordingly, it is necessary to estimate influence factors by comprehensively considering of various factors.

As is evident from the results of previous studies, one cause of population migration to other areas involves economic activity. Fig. 6 depicts the relationship between the daytime population index (DPI) and the EMR. Areas with an index of more than 100 can be categorized as business areas, while those with an index of less than 100 can be categorized as residential areas.
The correlation between business/residential areas and the EMR has also been analyzed.

In business areas, an increasing DPI was associated with an increasing EMR. This phenomenon implies that the rate of population migration to business areas for economic activity is high. This result is consistent with the results of previous studies, which have found that economic activity is one of the factors responsible for population migration. Conversely, for residential areas, a lower DPI was associated with an increasing EMR. This result implies that people pursuing a more comfortable living environment regardless of economic activity migrate when there is a newly developed residential complex, regardless of the distance from their current residence.

The results shown in Fig. 7, Fig. 8, and Fig. 9 show the EMR according to the number of industry for businesses (Nib), the manufacturing ratio (Rm), and the gross regional product per capita (GRP). Each influencing factor demonstrates a tendency that is directly proportional to the EMR.

For Nib, the EMR characteristics can be divided based on 250 businesses/km², while they can be divided based on 10% in the case of Rm. For GRP, they can be classified based on

---

**Table 3. Results of Analysis on Influencing Factors of External Migration Rate**

| Category                  | Correlation coefficient with external migration rate | Standard correlation coefficient (113 samples, confidence level of 95%) | Results of correlation review |
|---------------------------|------------------------------------------------------|------------------------------------------------------------------------|------------------------------|
| Development project factor| \( P_t \)                                           | 0.1111                                                               | 0.1557 + (Low correlation)   |
|                           | \( P_s \)                                           | -0.0089                                                             | - (Low correlation)          |
|                           | \( YC \)                                             | -0.2203                                                             | 0.1557 + (Correlated)        |
| Population factor         | \( PGR \)                                            | 0.1747                                                              | 0.1557 + (Correlated)        |
|                           | \( DGS \)                                            | 0.1236                                                              | 0.1557 + (Low correlation)   |
| Economy and industry factor| \( DPI \)                                         | 0.3187                                                              | 0.1557 + (High correlation)  |
|                           | \( R_e \)                                            | 0.0831                                                              | 0.1557 + (Low correlation)   |
|                           | \( N_{ib} \)                                         | 0.2640                                                              | 0.1557 + (High correlation)  |
|                           | \( R_m \)                                            | 0.1899                                                              | 0.1557 + (Correlated)        |
|                           | \( GRP \)                                            | 0.3721                                                              | 0.1557 + (High correlation)  |

---

**Fig. 5.** Box plot of external migration rate by population growth rate.

**Fig. 6.** External migration rate by daytime population index.

**Fig. 7.** Box plot of external migration rate by number of industry for businesses.
4.3. Results of the Model Developed To Predict the External Migration Rate

As shown in the EMR investigation results, there was a clear difference between metropolitan cities and other cities and counties. Taking this finding into consideration, the prediction model was divided into one model (model YA) for metropolitan cities and another model (model YB) for other cities and counties. The development results are as follows. The model derived as the result of quantification I analysis for influencing factors of EMR Y_A corresponds to Eq. (5), and Table 4 shows the description of regression coefficient in Eq. (5).

\[
Y_A = 30.959 + 16.600 \times RCD + 12.768 \times CID + \\
10.438 \times RED + 6.598 \times RCD + \\
7.608 \times PGR + 12.994 \times DGS + (-8.420) \times \\
N_b + 13.835 \times GRP
\]  

(5)

Here, Y_A: External migration rate of metropolitan cities (%)

As a result of the development of model YA, the hit ratio of the predicted value for the investigated value was 71.11% with a confidence level of 95%. This shows that the influencing factor of the EMR used as an explanatory variable expresses the EMR well.

For model Y_A, the factors with a statistically significant effect on the EMR were found to be P_t, PGR, DGS, N_b, and GRP. Regarding the EMR by P_t, RSD had the highest rate, followed
This tendency is similar to that found by Jeon [20]. The EMR was higher when the development project was completed in an area with a recent PGR of less than 1.0%. This tendency is due to the outflow of population resulting from the development project. Areas featuring a DGS exceeding 70.0% were associated with an increasing EMR. This reflects the trend in Korea in which the DGS is rising, along with the higher migration rate associated with the completion year of development projects (YC) in metropolitan cities with a higher EMR. The EMR of areas with high GRP was also found to be high. A variance analysis carried out to determine the adequacy of the model found that the significance probability of the F value was less than 0.05 at 0.000, this demonstrates that the applicable model had statistically significant results.

The regression coefficient significance test results for model $Y_A$ are shown in Table 5. Of the explanatory variables comprising the model, the dummy variables indicating the $P$, (RED, RCD) and the PGR variables had low explanatory power. All other variables were statistically significant. The statistical sig-

### Table 5. Model $Y_A$ Regression Coefficient’s Significance Test Results

| Variable | Standardized coefficient | t-value | Significance probability | Variance inflation factor |
|----------|--------------------------|---------|--------------------------|--------------------------|
| Constant | -                        | 4.227   | 0.000                    | -                        |
| $RSD^{**}$ | 0.515                    | 2.818   | 0.008                    | 2.614                    |
| $CID^*$ | 0.310                    | 2.020   | 0.051                    | 1.844                    |
| $RED$ | 0.199                    | 1.408   | 0.168                    | 1.565                    |
| $RCD$ | 0.201                    | 1.160   | 0.250                    | 2.335                    |
| $PGR$ | 0.219                    | 1.609   | 0.116                    | 1.452                    |
| $DGS^{**}$ | 0.427                    | 3.238   | 0.003                    | 1.358                    |
| $N_b^*$ | -0.274                   | -1.841  | 0.074                    | 1.728                    |
| $GRP^{**}$ | 0.437                    | 3.189   | 0.003                    | 1.470                    |

* Significant at confidence level of 90%      ** Significant at confidence level of 95%

### Table 6. Description of Model $Y_B$ Regression Coefficient

| Development project type | $RSD$ | $CID$ | $RED$ | $RCD$ |
|--------------------------|-------|-------|-------|-------|
| $RSD$ | 1     | 0     | 0     | 0     |
| $CID$ | 0     | 1     | 0     | 0     |
| $RED$ | 0     | 0     | 1     | 0     |
| $RCD$ | 0     | 0     | 0     | 1     |
| $REI$ | 0     | 0     | 0     | 0     |

| Development project scale (number of households) | More than 3,000 households | Less than 3,000 households |
|--------------------------------------------------|----------------------------|---------------------------|
| $P$ | 1 | 0 |
| Distance to city center (km) | Less than 15 km | More than 15 km |
| $D$ | 1 | 0 |
| Population growth rate (%) | Less than 1.0% | More than 1.0% |
| $PGR$ | 1 | 0 |
| Daytime population index (%) | More than 100% | Less than 100% |
| $DPI$ | 1 | 0 |
| Manufacturing ratio (%) | More than 10.0% | Less than 10.0% |
| $R_m$ | 1 | 0 |
| Gross regional product per capita (1 million KRW/person) | More than 20 million KRW/person | Less than 20 million KRW/person |
| $GRP$ | 1 | 0 |
significance was low due to the small number of variables included in each dummy group. However, these variables had a significance probability of less than 0.3 and explained the dependent variable sufficiently with a confidence level of 70%. Thus, they were not excluded from the variable. The PGR could explain the dependent variable at the 90% level and was thus not excluded from the variable.

Model YB, for cities and counties, is equal to Eq. (6), and Table 6 shows the description of regression coefficient in Eq. (6).

\[
\begin{align*}
Y_B &= 28.252 + 0.772 \times RSD + 10.540 \times CID + \\
&\quad (-6.892) \times RED + (-15.577) \times RCD \\
&\quad 6.824 \times P_s + 12.170 \times D + (-5.909) \times PGR \\
&\quad (-8.439) \times DPI + 12.340 \times R_m + 6.695 \times GRP \\
\end{align*}
\]

Here, \( Y_B \): external migration rate of cities and counties (%)

As a result of the development of model YB, the hit ratio of the predicted value for the investigated value was 75.00% at a confidence level of 95%. Accordingly, the influencing factors in the EMR that were used as explanatory variables expressed the EMR well.

In the case of model YB, factors that had a statistically significant effect on the EMR differed from model YA. The EMR increased as the development project scale \( (P_s) \) increased, the \( D \) decreased, and the \( R_m \) and \( GRP \) increased. The EMR was high for areas with a \( PGR \) of less than 1.0% and for areas classified as residential areas with less than 100% of the DPI. This pattern indicates that factors related to economic activity have a significant effect on the EMR. Meanwhile, the association of a higher EMR with larger \( P_s \) implies that the population that migrates to development projects in cities and counties does not migrate simply to be close to economic activity. Instead, they move after having comprehensively considered factors related to economic activity as well as factors related to the living residential environment. When predicting EMR in future, these trends need to be considered.

The results of a variance analysis confirming the model’s adequacy found that the significance probability for the F value was less than 0.05 at 0.002, demonstrating that the applicable model had statistically significant results.

The significance test results of the regression coefficient for Model YB are shown in Table 7. Of the explanatory variables that comprised the model, dummy variables indicating the \( P_s \) \((RSD, RED, \text{and } RCD) \) and \( PGR \), as well as the \( GRP \), variables had low explanatory power. All other variables were statistically significant. The reason for this low statistical significance is that the number of variables included in each dummy group was small. If additional samples are considered in the future, a more sophisticated model will result.

The developed models YA and YB are statistically significant, given the results of the variance analysis and coefficient significance test, with a hit ratio over 70%. Furthermore, the variables included in the model are also considered logically appropriate. Comprehensively, the models are considered valid.

### 4.4. Comparison to the Investigated External Migration Rate

The investigated EMR of the 113 development projects, the estimated values resulting from the newly developed model in this study, and the estimated values achieved using the previous indirect method are shown in Fig. 10 and Fig. 11. The model developed in this study is closer to the investigated value than the value derived from the previous indirect method. In case of model YA, the root mean square error (RMSE) between the predicted value and the investigated value was 10.12. On the other hand, the RMSE between the value derived from the previous indirect method and the investigated value was 27.31. This pattern indicates that the newly developed model has been shown to be more accurate and useful than the indirect method.

| Variable | Standardized coefficient | t-value | Significance probability | Variance inflation factor |
|----------|--------------------------|---------|--------------------------|--------------------------|
| Constant | -                        | 2.501   | 0.015                    | -                        |
| RSD      | 0.024                    | 1.071   | 0.244                    | 9.026                    |
| CID*     | 0.260                    | 1.952   | 0.063                    | 6.712                    |
| RED      | -0.073                   | -1.470  | 0.150                    | 2.175                    |
| RCD      | -0.331                   | -1.314  | 0.194                    | 5.712                    |
| \( P_s \) | 0.203                    | 1.566   | 0.123                    | 1.505                    |
| \( D^{**} \) | 0.381                   | 2.681   | 0.010                    | 1.818                    |
| PGR      | -0.179                   | -1.454  | 0.151                    | 1.360                    |
| DPI*     | -0.255                   | -1.766  | 0.083                    | 1.879                    |
| \( R_m^{**} \) | 0.353                   | 2.952   | 0.005                    | 1.285                    |
| GRP      | 0.210                    | 1.517   | 0.135                    | 1.719                    |

* Significant at confidence level of 90%  ** Significant at confidence level of 95%
4.5. Results of the Application of the Model to Predict External Migration Rate

The EMR presented by the MOLIT and MOE [4] is used for the water service maintenance basic plan, which is currently derived from water supply demand forecasts. However, the presented value, as described in the introduction, contains inaccuracies.

First, the presented value was established using the value of the EMR presented in the MOLIT and MOE [4]. Table 8 shows the results of using the value presented by the MOLIT and MOE [4] to derive the EMR of future development plans in comparison to the results obtained using the EMR derived from models Y_A and Y_B, which were developed in this study for five examples of development projects completed since 2009. These samples were not included when the model was developed.
In the case of metropolitan cities, compared to the actual EMR, the EMR applied by the MOLIT and MOE [4] was significantly low and, as a result, the error between the two values was significant. When the prediction model developed in this study was applied, the error was reduced to less than 3.0%.

Furthermore, although City B in Province G and City P in Province G applied the same EMR because they fell within the same province, the resulting errors were significant, at 20.16% and 6.40%, respectively. When the prediction model developed in this study was applied, different EMR values were used to reflect the individual characteristics of each city, the error was significantly reduced.

The KC City development project shown in Table 8 is a large-scale development project with a planned population of 80,000. Following the government announcement of the new city development policy, this is an area in which the population growth rate was already high before the project. The actual increase in population has been 61,610, as a result of EM. When employing the indirect method, the EM expected for the development project was 19,520, an underestimation totaling less than 35% of the actual EM. However, when the model developed in this study is applied, the EM is estimated at 59,512, a figure very similar to the actual EM. The baseline population of the development project area is approximately 300,000. The predicted population error is about 42,000, or 14%, which can cause problems, even when considering the 10% margin in water supply planning.

5. Conclusions

This study has shown that the indirect method incorrectly estimates EMR, which is a key component in estimating design population for infrastructural projects. Underestimating the design population can lead to insufficient water demand plans, which can cause unbalanced water welfare issues. Meanwhile, excessive water demand planning reduces the operational efficiency of limited water resources and causes waste at the national level.

To overcome this challenge, two new models were developed using quantification analysis. The models were developed for both counties and cities, after identifying the major factors influencing EM.

The major findings of this study were:

1) A high EMR was associated with the following factors: a shorter distance to the city center, a business area with a higher population growth rate and weekly population index, and areas with favorable conditions for economic activity, including a high number of businesses, a high manufacturing business ratio, and a high gross regional product per capita.

2) Two prediction models, one using quantification analysis for metropolitan cities (YA) and for cities and counties (YB) were developed; both models were shown to be statistically significant. These models can be used to predict EMR based on data drawn from target cities and counties.

3) RMSE between the model values and the observed values were 10.12 and 15.58 for metropolitan cities and counties respectively; they were lower than the RMSE values of 27.31 and 28.79 obtained using the indirect method. Thus, the newly developed models provide a more accurate estimate of EMR than the indirect method.

However, as described previously, this research has the following limitations. First, a relatively small number of samples were used for model development. Second, 11 influential factors were considered in this research; however, various influential factors need to be considered additionally.

Consequently, future medium and long-term studies should aim to cover more counties and cities in order to comprehensively identify important trends and significant factors that affect EMR. Such results would be useful for both governments and the private sector when planning public infrastructural projects.

Acknowledgments

This study was funded by the Ministry of Land Infrastructure and Transport (MOLIT) and K-water.

Nomenclature

EMR  The ratio of the population entering the development
References

1. Kent TJ. The urban general plan. Chicage: Planners Press; 1990.
2. Kwon HB. A study on the application of external migration rate in the population forecast – Focused on the Pyeongtaek city [master thesis] Gyeonggi: Gachon Univ.; 2012.
3. MOLIT and MOE. Guidelines for predict water supply. Daejeon: MOLIT and MOE; 2008.
4. MOLIT and MOE. Guidelines for predict water supply. Daejeon: MOLIT and MOE; 2014.
5. Balogun II, Sojobi AO, Galkaye E. Public water supply in Lagos State, Nigeria: Review of importance and challenges, status and concerns and pragmatic solutions. Cogent Eng. 2017:4:1329776.
6. Molloy R, Smith CL, Wozniak A. Job changing and the decline in long-distance migration in the United States. Demography 2017:54:631-653.
7. McCatty M. The process of rural-urban migration in developing countries. An honours essay submitted in fulfilment of the degree of Bachelor of Arts to Department of Economics, Carleton University, Ottawa, Ontario; 2004.
8. Clark W, Withers S. Family migration and mobility sequences in the United States: Spatial mobility in the context of the life course. Demogr. Res. 2008;17:591-622.
9. Mihenetu BA. Causes and consequences of rural-urban migration: The case of Woldiya town [doctoral dissertation]. North Ethiopia; 2011.
10. Yoo KM. An empirical analysis on the determinants of population movement: In case of Korea (1966-1985). The Korean Econ. Assoc. Kyongjehak Yongu 1991;39:157-209.
11. Kim ST, Jang JH. Empirical analysis on the economic determinants of regional migration in Korea during 1970-1991 period. The Korea Int. Economic Assoc. Kukje Kyungje Yongu 1997;3:175-197.
12. Lee EW. Quality of life and interregional migration in Korea. J. Korean Natl. Econ. 1997;3:175-197.
13. Lee W. A study on the determinants on migration in Kyonggi-Do. J. Korea Plan. Assoc. 2000;35:67-76.
14. Lee W, Park EK. A study on the migration to large-scale residential development areas. Gyeonggi: Kyonggi Research Institute; 2001.
15. Kim GS, Jang W. The study on characteristics of population movement within Busan. J. Korea Plan. Assoc. 2001;36:39-55.
16. Kim HB, Choi JS. A study on population projection for infrastructure sizing in housing development sites. Korea Planning Association – 2001 autumn conference proceeding 1; 2001. p. 483-501.
17. Kwon SC. Migration to the capital region in Korea: Assessing the relative importance of place characteristics and migrant selectivity. J. Korean Assoc. Region. Geogr. 2005;11:571-584.
18. Choi DS, Kim TG. The estimation of the infra-regional ratio of moving-in population into residential land development area. Seoul Stud. 2009;10:105-119.
19. Lee W, Lee JE. A study on in-migration characteristic in the second generation of newtown, the capital area. Gyeonggi: Kyonggi Research Institute; 2007.
20. Jeon HS. A study on population immigration effect of different types of development project in Kyonggi-do. J. Korea Plan. Assoc. 2009;44:113-123.
21. Hur JW. An empirical analysis on the determinants of the external migration rate in the housing development sites in Gyeonggi province. J. Korean Region. Dev. Assoc. 2010;22:47-61.
22. Lee MH. A study on the determinants of the external migration rate to the residential development project areas – In and around the metropolitan residential development project areas [master thesis]. Seoul: Chungang Univ.; 2011.