FORECASTING THE PEAK-PERIOD STATION-TO-STATION ORIGIN-DESTINATION MATRIX IN URBAN RAIL TRANSIT SYSTEM: CASE STUDY OF CHONGQING, CHINA

Yan Cheng, Corresponding Author
Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University
4800 Cao’an Road, Jiading District, Shanghai, P.R. China, Post Code 201804
Tel: (86)021-69583757; Email: patty_1234@126.com

Xiafei Ye
Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University
4800 Cao’an Road, Jiading District, Shanghai, P.R. China, Post Code 201804
Tel: (86)021-69589875 Fax: (86)021-69583712; Email: yxf@tongji.edu.cn

Lifeng Zhou
Chongqing Rail Transit (Group) Co., Ltd.
West Jinkai Avenue, Yubei District, Chongqing, P.R. China, Post Code 401120
Email: 70106925@qq.com

Word Count: 5,030 words text + 9 tables/figures x 250 words (each) = 7,280 words

Number of References 13
Number of Figures 5
Number of Tables 4

Submission Date: 7/31/2017
ABSTRACT
The maximum one-direction section passenger flow within peak hour is an important indicator for planning and design of urban rail transit. To determine it, it is necessary to forecast passengers’ departure time and route choice during peak period. As the basis of this process, the peak-period station-to-station origin-destination (OD) matrix reflects the passengers’ travel needs. This paper tests traditional gravity models in forecasting the peak-period station-to-station OD matrix in urban rail transit with a real-world case study of Chongqing, China. To solve its over-estimation when deterrence between two stations is too little, the gravity-model-based Peak Period Coefficient (PPC) model is introduced. Comparing results show that with the same dataset, the PPC model is superior to the gravity model. Its standard deviation is only 12.90 passengers, reduced by 56.02%.

Keywords: Urban rail transit, Station-to-station ridership, Gravity model, Peak period coefficient, Deterrence function
INTRODUCTION

With its rapid growth, urban rail transit is now playing a more important role in supporting and promoting city development. By the end of 2016, there are 133 operating lines of 30 cities in mainland of China, reaching the total length of 4152.8 km. The annual patronage reaches 16.1 billion passengers, increasing 16.6% than the year of 2015. There are 5636.5 km of lines under construction in 48 cities.

The planning and design of urban rail transit need foresight because of its high construction cost and big difficulty to modify after construction. As a key indicator in planning and design phase, the maximum one-direction section passenger flow (MSPF) within peak hour directly influences the determination of vehicle selection and train formation. It is accumulated by the passengers who depart from different stations before this section and pass through it during peak hour, whose departure time should be before or during peak hour. Conventionally, this indicator is forecasted through multiplying all-day passenger flow of this section by a peak hour coefficient, which depends on the line attribute. However, this method hasn’t taken the forming mechanism of MSPF within peak hour into account, so it cannot reflect the complex process.

In order to overcome this shortcoming, it is necessary to extend the time range from peak hour to peak period, and forecast passengers’ departure time and route choice during this time range. For this dynamic passenger assignment process, time-varying peak-period station-to-station origin-destination (OD) matrices are critical inputs, because they represent the travel needs of passengers at different moments. However, most of previous studies on passenger assignment model for urban rail transit assumed that time-varying peak-period station-to-station OD matrices were given (1-3). This becomes invalid at the planning and design stage, because the actual operational data hasn’t generated. To forecast time-varying peak-period station-to-station OD matrices, the first step is forecasting the peak-period station-to-station OD matrix, which is of fundamental importance. Errors in it are transferred onto time-varying peak-period station-to-station OD matrices, leading to errors in MSPF within peak hour ultimately.

There is little research on the forecasting of the peak-period station-to-station OD matrix at present. Multiplicative models are adopted by (4) and (5). The former model chose 32 factors, which can be categorized into 4 areas: built-environment, travel impedance, intermodal connection and other variables. The model was calibrated with the data of Seoul Metro in Korea. Meanwhile, the latter one selected 22 variables which can be divided into 4 areas: land use, intermodal connection, station context and travel impedance, estimated with data of Metro system in Nanjing, China. Although this kind of models performs well according to regression results, they are totally different from traditional trip distribution models. The theoretical foundation is relatively weak, which makes production and attraction constraints cannot be guaranteed. In addition, it is difficult to acquire some data (such as several intermodal connection variables) at the planning and design stage.
Traditional trip distribution models concentrate on trips between traffic analyses zones (TAZs) of all modes, whose time range is a whole day. One of the most important models is the gravity model \((6,7)\), which is widely used because it is simple in concept and has been well documented. Nevertheless, it has several drawbacks such as it may overestimate when the distance between origin and destination is too short. It still needs to be verified whether the gravity model is suitable for the forecasting of the peak-period station-to-station OD matrix in urban rail transit. Apart from traditional trip distribution models, disaggregate destination choice models based on utility-maximization theory have been proposed by some scholars in recent years to forecast OD matrix of all modes \((8-11)\). This kind of models has a good performance; however, the data requirement is too high for forecasting OD matrix in urban rail transit. It cannot be clear who will choose urban rail transit at the planning and design stage, let alone the investigation of their information such as socio-economic characteristics and travel characteristics. So these models are not taken into consideration. In summary, the forecasting of urban rail transit peak-period station-to-station OD matrix still needs more in-depth study.

A long-term all-day station-to-station OD matrix with relatively high accuracy can be forecasted by planners through traditional four-step procedure, which is not time-dependent. So it is assumed that the all-day station-to-station OD matrix is given. Based on this assumption, the paper tests traditional gravity models in forecasting the peak-period station-to-station OD matrix in urban rail transit with a real-world case study of Chongqing, China, and analyzes their merits and demerits. To solve the drawbacks, an improved forecasting model called Peak Period Coefficient (PPC) model is proposed. According to this model, the peak-period station-to-station OD matrix is obtained by multiplying the given all-day station-to-station OD matrix by forecasting PPC, whose expression is deduced from the gravity model. Models with different configurations are estimated with the same dataset, and their performance is compared with the ones of gravity models.

The rest of this paper is organized as follows: Section 2 describes the data of Chongqing rail transit. Section 3 presents the performance of gravity models and the proposal of the improved model. Section 4 compares the results of these models. Section 5 provides conclusions.

DATA AND SCOPE

Case Study Object
Chongqing is located in the southwest part of China, covering an area of 82400 km², 650 km² of which is the built area of the main city. The registered population of Chongqing is 33.752 million while the permanent resident is 29.914 million, 818.98 million of which belongs to the main city.

By the end of 2016, the urban rail transit network of Chongqing consists of 4 rail transit lines. Line 1 and Line 6 are metro systems, while Line 2 and Line 3 are
monorail systems. All lines are located in the main city, the total operating mileage of which reaches 202 km, with 120 stations included (Figure 1). After removing the stations with zero ridership and merging interchange stations, 110 stations are left. Due to the significant changes in land use around Guangdianyuan station and a high proportion of undeveloped land around Jiuquhe station, 108 stations are left on the list of study objects eventually.

![FIGURE 1 Chongqing rail transit stations.](image)

**Data Source**

To compare the performance of different models in forecasting the peak-period station-to-station OD matrix, the same dataset is used in this study. It mainly includes 3 parts.
1. Prerequisites: The all-day station-to-station OD matrix of the study objects;
2. Dependent variable: The peak-period station-to-station OD matrix of the study objects;
3. Independent variables: The ridership of study objects in different periods, the time shortest path between any two stations and its interchange times.

According to the automated fare collection (AFC) data of August 26, 2015 provided by Chongqing Rail Transit (Group) Co., Ltd., the station-to-station OD matrices within all-day and peak period as well as the ridership of study objects are counted. It is worth noting that in order to ensure the consistency of data, one trip is classified into different periods according to its boarding time, no matter when it ends. Statistic shows that there are 8669 pairs of OD stations with ridership larger than 0 during the peak period.

The time shortest path is calculated by Dijkstra algorithm, according to the network information about Chongqing Rail Transit, including the information about lines, stations, travel time between stations, stop time, interchange time and so on.

Temporal Scope: Peak Period
Urban rail transit system has the advantages of fast speed, punctuality and large capacity. Owing to these, it undertakes a lot of long- and moderate-distance trips. Their aims are mainly commuting and going to school. This type of passengers has rigid travel demands. Their specified arrival time (SAT) of destination is fixed and relatively close, so that the peak hour of urban rail transit system is usually within the morning peak period. This situation is also applicable for Chongqing Rail Transit (Figure 2). All 4 lines reach MSPF during all day between 8 a.m. and 9 a.m.

![FIGURE 2 MSPF of lines in Chongqing rail transit on one weekday.](image)
The same trip may be divided into different periods according to various criteria of the classification. Here, one trip is divided into different periods according to its boarding time. Because MSPF within peak hour is accumulated by the passengers departing from different stations before or during the peak hour, so the starting time of peak period should be earlier than that of peak hour. The paper provides that peak period is from 6 a.m. to 9 a.m.

**METHODOLOGY**

**Deterrence Function**

Deterrence function is one of the most important influence factors of station-to-station OD matrix; its mathematical form and components directly affect the final estimation results. Currently, there are 3 commonly used mathematical forms, which is power function, exponential function and the combination of both. These formulations are written as equation (1)-(3) when applied in urban rail transit system.

Form 1: Power function

\[ f(c_{ij}) = f(d_{ij}) = d_{ij}^{-\gamma} \]  

Form 2: Exponential function

\[ f(c_{ij}) = f(d_{ij}) = \exp(-\eta \cdot d_{ij}) \]  

Form 3: The combination of power function and exponential function

\[ f(c_{ij}) = f(d_{ij}) = \exp(-\eta \cdot d_{ij}) \cdot d_{ij}^{-\gamma} \]  

In these formulations, \( c_{ij} \) = the generalized travel cost between station i and station j; \( d_{ij} \) = the travel time between station i and station j; and \( \gamma, \eta \) = parameters to be estimated.

One essential difference between urban rail transit and road traffic as well as bus transit is that its passengers can only change between various lines through interchange stations. This difference brings about the result that when measuring the generalized travel cost between two stations, interchange times is a key factor except the travel time. Since when the exponent of a power function is negative, the base mustn’t be 0, and the interchange times between two stations are non-negative integers, so it is not appropriate to introduce interchange times into a deterrence function in the form of power function. This paper introduces it in the form of exponential function.

Form 4: Power function

\[ f(c_{ij}) = f(d_{ij}, n_{ij}) = d_{ij}^{-\gamma} \cdot \exp(-\tau \cdot n_{ij}) \]  

Form 5: Exponential function

\[ f(c_{ij}) = f(d_{ij}, n_{ij}) = \exp(-\eta \cdot d_{ij} - \tau \cdot n_{ij}) \]
Form 6: The combination of power function and exponential function

\[ f(c_i) = f(d_i, n_i) = \exp(-\eta \cdot d_i - \tau \cdot n_i) \cdot d_i^{-\gamma} \]  

(6)

In these formulations, \( n_i \) = the interchange times between station i and station j and \( \tau = \) a parameter to be estimated.

With the expansion of urban rail transit network, there may be several routes between origin and destination stations for passengers to choose. To ensure the data consistency of the study, the travel time and interchange times are the corresponding values of the time shortest path between them. Models with these deterrence functions will be estimated subsequently. Their performance would be compared to support the analysis of the influence of deterrence function on the peak-period station-to-station OD matrix.

**Gravity Model**

**Modeling**

The gravity model is similar to Newton’s law of gravity. According to this model, in urban rail transit system, the ridership between an origin station and a destination station depends directly on the total boardings of the origin station, the total alightings of the destination station, and depends inversely on the deterrence between two stations. The peak-period ridership between two stations can be expressed as an unconstrained gravity model (7), given in equation (7)

\[ t_{ij}^p = k \left[ O_i^p \right]^\alpha \left[ D_j^p \right]^\beta \cdot f(c_{ij}^p) \]  

(7)

where \( t_{ij}^p \) = the ridership between station i and station j during peak period; \( O_i^p \) = the boardings of station i during peak period; \( D_j^p \) = the sum of passengers who depart from other stations during peak period and alight on station j; \( f(c_{ij}^p) \) = the deterrence function of station i and j; and \( k, \alpha, \beta = \) parameters to be estimated.

On this basis, if the total boardings of origin stations are known, a production-constrained gravity model can be developed, given in equation (8)

\[ t_{ij}^p = \left[ O_i^p \left( D_j^p \right)^\rho \cdot f(c_{ij}^p) \right] \left[ \sum_j \left( D_j^p \right)^\rho \cdot f(c_{ij}^p) \right]^{-1} \]  

(8)

where \( \rho = \) a parameter to be estimated; and the meanings of other symbols are as same as the ones in the unconstrained gravity model.

**Estimation**

For an unconstrained gravity model, since it is non-linear, the authors take logarithms of equation (7) as follows

\[ \ln t_{ij}^p = \ln k + \alpha \ln O_i^p + \beta \ln D_j^p + \ln f(c_{ij}^p) \]  

(9)

where \( \ln f(c_{ij}^p) \) varies with the deterrence function, and it is always a linear combination of \( \ln d_i^p, d_j^p \) and \( n_i^p \). According to equation (9), the parameters can be
estimated by Ordinary Least Squares (OLS).

For a production-constrained gravity model, (12) proposed the estimation method. The linear equations to be estimated are summarized in equation (10)-(13)

\[
\ln \left( \frac{r_i^p / r_i^c}{r_j^p / r_j^c} \right) = \rho \ln \left( D_i^p / D_j^c \right) + \ln \left( \overline{f(c_i^p)} / \overline{f(c_j^p)} \right)
\]

\[
r_i^p = \left( \prod_{j=1}^{J} c_{ij}^p \right)^{1/J}
\]

\[
\overline{D}^c = \left( \prod_{j=1}^{J} D_j^c \right)^{1/J}
\]

\[
\overline{f(c_i^p)} = \left( \prod_{j=1}^{J} \overline{f(c_j^p)} \right)^{1/J}
\]

where \( \ln \left( \overline{f(c_i^p)} / \overline{f(c_j^p)} \right) \) varies with the deterrence function, and it is always a linear combination of \( \ln \left( d_i^p / d_j^p \right) \), \( d_i^p - \frac{1}{J} \sum_{j=1}^{J} d_j^p \) and \( n_i^p - \frac{1}{J} \sum_{j=1}^{J} n_j^p \). \( \overline{d}_i^p \) is calculated by equation (14).

\[
\overline{d}_i^p = \left( \prod_{j=1}^{J} d_j^p \right)^{1/J}
\]

### TABLE 1 Bivariate Correlation Analysis of Gravity Models

<table>
<thead>
<tr>
<th>Unconstrained Gravity Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>Asymptotic significance of K-S test</td>
</tr>
<tr>
<td>Correlation coefficient</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Production-constrained gravity models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>Asymptotic significance of K-S test</td>
</tr>
<tr>
<td>Correlation coefficient</td>
</tr>
</tbody>
</table>

Note: ** Correlation is significant at the 0.01 level (2-tailed).

The result of Kolmogorov–Smirnov test shows the asymptotic significances of variables’ distributions mentioned above are all less than 0.050, which means none of them follows a normal distribution. Bivariate correlations among variables in different kinds of gravity models are calculated through Spearman correlation analysis. It has proven that no matter in which kind of gravity models, the independent variables are
significantly correlated with the dependent variable, and the value is relatively high. The signs of correlation coefficients are in accordance with the inference of models. The values of correlation coefficients among variables in the production-constrained gravity model are comparatively higher (Table 1).

There are 12 unconstrained gravity models and 6 production-constrained gravity models to be estimated. The main differences among unconstrained gravity models are whether \( k \) is equal to one and deterrence functions. For production-constrained gravity models, because there is no constant to be estimated, so the main difference is their deterrence functions.

The forecast peak-period station-to-station OD matrices are iterated by Fratar Method (13) to ensure the satisfaction of production and attraction constraints, which is represented by equation (15)-(16).

\[
\sum_i t_{ij}^p = O_i^p \tag{15}
\]

\[
\sum_j t_{ij}^p = D_j^p \tag{16}
\]

As a measurement of model performance, the standard deviation, which is also called root-mean-square deviation (RMSD), is taken as the comparison standard. In this study, it is calculated by equation (17)

\[
\sigma = \left[ \left( \epsilon_1^2 + \epsilon_2^2 + \cdots + \epsilon_i^2 + \cdots + \epsilon_n^2 \right) / n \right]^{1/2} \tag{17}
\]

where \( \epsilon_i \) = the difference between predicted ridership and observed ridership between the \( l \)th pair of OD stations; and \( n \) = the number of station pairs whose peak-period ridership is larger than 0.

Note: The slash-filled model fails t-test.

FIGURE 3 The standard deviation of gravity models.
Figure 3 indicates that:

1. Some models’ variables fail t-test, because there are two variables related to travel time included in deterrence functions. When the influence of travel time is mainly reflected by one of them, the other one cannot pass the significant test;
2. The models with deterrence functions including interchange times have better performance. The obvious decrease of standard deviation proves that it is necessary to consider the influence of interchange times on deterrence in urban rail transit system;
3. The optimal model in the framework is the unconstrained gravity model with a deterrence function in Form 5 and k=1, the standard deviation of which is 29.33 passengers.

**Analysis**

Although gravity model is widely used, the following drawbacks still exist: (1) it hasn’t taken human behavior into consideration basically; (2) it may overestimate when the distance between origin and destination is too short; (3) the trips within one TAZ is hard to be forecasted; (4) trip distance in one TAZ is not a fixed value; (5) deterrence between TAZs varies significantly according to traffic modes.

In urban rail transit system, passengers move between stations instead of TAZs. Thanks to this change, drawback (3)-(5) of gravity model can be avoided. For (3), there is no passenger travelling from a station to its own in one trip; for (4), when origin and destination stations are determined, trip distance is fixed because of the infrastructure characteristics of urban rail transit; and for (5), the deterrence between stations is not influenced by traffic modes because there is only one mode. But drawback (2) is further aggravated with travel modes reducing from all modes to urban mass transit alone. When the deterrence is too little, the passengers travelling between two stations are more likely to choose bus transit and non-motorized transportation due to travel cost. This phenomenon makes the ridership between OD stations decline. But traditional gravity models are strictly in accordance with the trend that with the decrease of deterrence, the ridership between two stations increases, which doesn’t conform to the fact.

**Peak Period Coefficient (PPC) Model**

**Modeling**

The Peak Period Coefficient (PPC) model is proposed to solve the over-estimation caused by the gravity model. According to this model, the peak-period station-to-station OD matrix is obtained by multiplying the given all-day station-to-station OD matrix by forecasting PPC.

The PPC of station-to-station ridership is the proportion of station-to-station ridership during peak-period to the one within all day. The trip purposes of passengers who travel during various periods are quite different. For passengers travelling during
the morning peak period, their main purpose is commuting and going to school, so the ridership between residential-land-based origin stations and destination stations whose surrounding land is mainly related to jobs and schools is tend to be high. When the time range is extended to all day, the composition of passengers is more complex, with three parts included. The first one is the morning-peak-period passengers working around the station, whose trip purpose is going home. Because of the difference and the diversification of land use around urban rail transit stations, it is not recommended to multiply the all-day station-to-station OD matrix by an identical PPC to obtain the peak-period station-to-station OD matrix. This simple conversion hasn’t considered the properties of origin and destination stations.

Based on the hypothesis that the station-to-station ridership follows an unconstrained gravity model, the PPC is calculated as follows

\[ P_y = t^i_D t^j_O = \left[ k_1 \left( O^i_r \right)^{\alpha} \cdot \left( D^j_r \right)^{\beta} \cdot f(c^i_j) \right] + \left[ k_2 \left( O^j_r \right)^{\alpha} \cdot \left( D^i_r \right)^{\beta} \cdot f(c^j_i) \right] \]  

(18)

where \( P_y \) is the PPC of ridership between station \( i \) and station \( j \); \( t^i_D, t^j_O \) is the ridership between station \( i \) and station \( j \) during peak period and all day, respectively; \( O^i_r, O^j_r \) is the boardings of station \( i \) during peak period and all day, respectively; \( D^i_r = \) the sum of passengers who depart from other stations during peak period and alight on station \( j \), \( D^j_r = \) the alightings of station \( j \) during all day; and \( f(c^i_j), f(c^j_i) = \) the deterrence function of station \( i \) and station \( j \) during peak period and all day, respectively.

Equation (19) can be obtained by dividing the numerator and the dominator of \( P_y \) by \( \left( O^i_r \right)^{\alpha} \cdot \left( D^j_r \right)^{\beta} \) simultaneously, which is represented by

\[ P_y = \left( k_1 / k_2 \right) \left[ f(c^i_j) / f(c^j_i) \right] \left[ \left( O^i_r / O^j_r \right)^{\alpha} \cdot \left( D^j_r / D^i_r \right)^{\beta} \right] \left[ \left( O^i_r \right)^{\alpha - \alpha} \cdot \left( D^j_r \right)^{\beta - \beta} \right] \]  

(19)

After substituting \( P_\alpha = O^i_r / O^j_r \) and \( P_\beta = D^j_r / D^i_r \) into equation (20), it can be written as follows

\[ P_y = k \left( k_1 / k_2 \right) \left[ f(c^i_j) / f(c^j_i) \right] \left[ \left( P_\alpha \right)^{\alpha} \cdot \left( P_\beta \right)^{\beta} \right] \left[ \left( O^i_r \right)^{\alpha - \alpha} \cdot \left( D^j_r \right)^{\beta - \beta} \right] \]  

(20)

where \( P_\alpha = \) the proportion of peak-period boardings of station \( i \); \( P_\beta = \) the proportion of peak-period alightings of station \( j \); and \( k, \alpha_i, \beta_i, \alpha', \beta' = \) parameters to be estimated, where \( k = k_1 / k_2; \alpha' = \alpha_1 - \alpha \), and \( \beta' = \beta_2 - \beta \).

Equation (20) shows that \( P_\alpha, P_\beta, O^i_r, D^j_r \) and deterrence between two stations are the main influence factors of PPC. The former two factors need to be forecasted according to land use characteristics around stations, while the other four factors can be calculated from given data.

Based on equation (20), the PPC model is written in equation (21)
The model can be further simplified with the following assumptions.

**Assumption (1):** The station-to-station OD matrices during different time periods are only affected by the boardings of origin station, the alightings of destination station and deterrence between two stations.

In an unconstrained gravity model, the parameter $k$ is regarded as a comprehensive correction coefficient caused by other influence factors except the three influence factors mentioned in Assumption (1). If Assumption (1) holds then $k_1 = k_2$, that is to say, $k' = 1$. Equation (21) can be simplified as shown in equation (22).

\[
t_{ij}^o = k' \left[ f(c_{ij}^o) / f(c_{ij}^d) \right] \left[ (P_{ij})^\alpha (P_{ij})^\beta \right] \left[ (O_{ij})^\alpha (O_{ij})^\beta \right] . t_{ij}^d
\]

**Assumption (2):** The deterrence between origin and destination stations during different time periods keep consistent.

Usually, the deterrence function is a function of the travel time between origin and destination stations as well as interchange times. The characteristic of urban rail transit system leads to the result that once the lines and stations construction is complete, these two independent variables have little changes during different time periods. If Assumption (2) holds then $f(c_{ij}^o) = f(c_{ij}^d)$, simplifying the model to equation (23).

\[
t_{ij}^o = k' \left[ (P_{ij})^\beta \right] \left[ (O_{ij})^\beta \right] . t_{ij}^d
\]

Whether these two assumptions are valid or not in the model still needs more analysis.

**Estimation**

Since the all-day station-to-station OD matrix is given, the forecasting part is the PPC. The logarithms of PPC model is shown in equation (24)

\[
\ln P_{ij} = \ln k^o + \alpha_1 \ln P_{ij} + \beta_1 \ln P_{ij} - \alpha' \ln O_{ij}^o - \beta' \ln D_{ij}^d + \ln f(c_{ij}^o) - \ln f(c_{ij}^d)
\]

The parameters can be estimated by Ordinary Least Squares (OLS) since the equation is linear. It should be noted that to prevent the inaccuracy of forecast $P_{ij}$ and $P_{ij}$ influencing the performance of the PPC model, here the actual values of both variables are used for estimation.

Table 2 shows that the signs of correlation coefficients are consistent with the inference of PPC model. All variables are significantly correlated. The coefficients of $\ln P_{ij}$ and $\ln P_{ij}$ are higher than the ones of $\ln O_{ij}^o$ and $\ln D_{ij}^d$, while the ones of $d_{ij}^o$, $\ln d_{ij}^o$ and $n_{ij}^o$ are a little lower. Compared with the traditional gravity model, the correlation coefficients of $d_{ij}^o$, $\ln d_{ij}^o$ and $n_{ij}^o$ in the PPC model decline markedly,
proving that the influence of these variables on the PPC model gets weaken.

**TABLE 2 Bivariate Correlation Analysis of PPC Models**

<table>
<thead>
<tr>
<th>Variables</th>
<th>$\ln P_i$</th>
<th>$\ln P_o$</th>
<th>$\ln O^d_i$</th>
<th>$\ln D^d_i$</th>
<th>$\ln d^p_i$</th>
<th>$d^p_i$</th>
<th>$n^p_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptotic significance of K-S test</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>1.000</td>
<td>0.498**</td>
<td>0.516**</td>
<td>-0.241**</td>
<td>-0.162**</td>
<td>-0.044*</td>
<td>-0.044*</td>
</tr>
</tbody>
</table>

Note: ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

\[d^p_i = d^d_i, n^p_i = n^d_i.\]

**FIGURE 4 The standard deviation of PPC models.**

The combination of two assumptions and different deterrence functions forms 14 PPC models to be estimated. Similarly, the forecast peak-period station-to-station OD matrices are iterated by Fratar Method (13) to ensure the satisfaction of production and attraction constraints.

Excluding the models with variables failing t-test (Figure 4), the optimal model in the PPC model framework is the one with a deterrence function in Form 3, whose $k\neq1$. Its standard deviation is 12.90 passengers. In the model, there is a difference between the deterrence during peak period and all day, which means the perceived travel time during peak period is longer than the one during other periods even though the actual travel time is totally the same. This is caused by the crowding discomfort in trains during the peak period.

For Assumption (1), the estimation results show that the standard deviation differences between models whose $k=1$ and $k\neq1$ are less than 1 passenger, except the
models with deterrence function in the form of a combination of power function and exponential function. So when the difference between these two kinds of models is acceptable, it is assumed to be valid. For Assumption (2), whether it is valid is supported by the proofs from two aspects. On one hand, $n_y^p$ and $d_y^p$ fail t-tests in some models (Table 3). The failure of $n_y^p$ is owing to that it doesn’t change through all day, so its impact on the PPC is negligible. Meanwhile, both $d_y^p$ and $\ln d_y^p$ are proxies for travel time. When only one of them is used to represent the deterrence, the modulus of regression coefficient is little or the variable fails the t-test; when both of them are in the function, the signs of their regression coefficients are opposite. So the influence of travel time is balanced. On the other hand, it is easy to find from Figure 4 that there is no big difference among models with various deterrence functions and the one without a deterrence function. The standard deviation of the model without a deterrence function is 13.30 passengers, which is 0.40 higher than the one of the optimal PPC model, but still lower than most of other models. So Assumption (2) is regarded to be valid when the difference of standard deviation is tolerable or the network is not very crowded during peak period.

TABLE 3 The Parameters of PPC Models

<table>
<thead>
<tr>
<th>PPC Models</th>
<th>$\ln k^p$</th>
<th>$\ln P_o$</th>
<th>$\ln P_d$</th>
<th>$\ln O_j^d$</th>
<th>$\ln D_j^d$</th>
<th>$\ln d_y^p$</th>
<th>$d_y^p$</th>
<th>$n_y^p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.952</td>
<td>0.818</td>
<td>0.853</td>
<td>-0.124</td>
<td>-0.089</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>3.088</td>
<td>0.819</td>
<td>0.849</td>
<td>-0.126</td>
<td>-0.091</td>
<td>-0.031</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>3.140</td>
<td>0.820</td>
<td>0.841</td>
<td>-0.130</td>
<td>-0.096</td>
<td>-0.003</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>2.812</td>
<td>0.817</td>
<td>0.836</td>
<td>-0.132</td>
<td>-0.098</td>
<td>-0.009</td>
<td>0.170</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>3.086</td>
<td>0.819</td>
<td>0.850</td>
<td>-0.126</td>
<td>-0.091</td>
<td>-0.030</td>
<td>-0.002</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>3.143</td>
<td>0.818</td>
<td>0.840</td>
<td>-0.131</td>
<td>-0.096</td>
<td>-0.003</td>
<td>0.011</td>
<td>-0.008</td>
</tr>
<tr>
<td>7</td>
<td>2.803</td>
<td>0.818</td>
<td>0.836</td>
<td>-0.132</td>
<td>-0.097</td>
<td>-0.009</td>
<td>0.173</td>
<td>-0.008</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>0.790</td>
<td>0.778</td>
<td>0.044</td>
<td>0.049</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>0.789</td>
<td>0.797</td>
<td>0.029</td>
<td>0.039</td>
<td>0.078</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>0.789</td>
<td>0.780</td>
<td>0.044</td>
<td>0.049</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>0.790</td>
<td>0.782</td>
<td>-0.004</td>
<td>0.010</td>
<td>-0.014</td>
<td>0.391</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>0.795</td>
<td>0.800</td>
<td>0.030</td>
<td>0.039</td>
<td>0.090</td>
<td>-0.038</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>0.789</td>
<td>0.780</td>
<td>0.044</td>
<td>0.049</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>-</td>
<td>0.796</td>
<td>0.786</td>
<td>-0.004</td>
<td>0.009</td>
<td>-0.014</td>
<td>0.406</td>
<td>-0.042</td>
</tr>
</tbody>
</table>

Note: The variable with the coefficient in underlined and bold style fails t-test.

RESULT AND DISCUSSION

The optimal models in two frameworks are given in equation (25)-(26)

The gravity model

$$t_y^p = \left( O_j^d \right)^{0.210} \left( D_j^d \right)^{0.314} \exp \left( -0.033 d_y^p - 0.730 n_y^p \right)$$

(25)

The PPC model

$$t_y^p = 16.64(P_\alpha)^{0.817}(P_\alpha)^{0.836}/(O_j^d)^{0.132}(D_j^d)^{0.098} \left[ (d_y^p)^{0.170} \exp \left( -0.009 d_y^p \right) \right] t_j^d$$

(26)
The gravity model

The PPC model

FIGURE 5 The forecast deviation of the optimal model.

TABLE 4 Statistics of Model Deviations

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Gravity model</th>
<th>PPC model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>29.33</td>
<td>12.90</td>
</tr>
<tr>
<td>Variance</td>
<td>860.26</td>
<td>166.51</td>
</tr>
<tr>
<td>Minimum</td>
<td>-684.00</td>
<td>-237.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>384.00</td>
<td>199.00</td>
</tr>
<tr>
<td>Percentile 25</td>
<td>-2.00</td>
<td>-1.00</td>
</tr>
<tr>
<td>Percentile 50</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Percentile 75</td>
<td>5.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Average deviation of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>little-deterrence OD stations</td>
<td>29.95</td>
<td>-0.54</td>
</tr>
</tbody>
</table>
Figure 5 shows the deviation between actual value and the results forecasted by the optimal gravity model and PPC model and iterated by Fratar method. In Figure 5 (a), the deviation near the 45° diagonal is relatively high, because the deterrence of these OD stations is little. But this problem doesn’t exist in Figure 5 (b). The origin-destination stations with travel time between them less than 5 min and no interchange are defined as little-deterrence OD stations. Table 4 shows that the average deviation of little-deterrence OD stations of Model I is as much as 29.95 passengers, while the one of Model II is only 0.54 passengers. It is proved that the over-estimation is weakened effectively in the PPC model.

From the figure and table, it also can be seen that the PPC model is obviously superior to the gravity model in terms of overall performance or case result. The standard deviation of the PPC model is 16.43 passengers less than the one of the gravity model, reduced by 56.02%.

**CONCLUSIONS**

With the case study of Chongqing, China, this study proposes a gravity-model-based Peak Period Coefficient (PPC) model. According to this model, the peak-period station-to-station OD matrix in urban rail transit is forecasted by multiplying the given all-day station-to-station OD matrix by the forecasting PPC. Results show that:

1. When adopting traditional gravity models in forecasting the peak-period station-to-station OD matrix in urban rail transit, several drawbacks of these models can be avoided. But the drawback of over-estimation is further aggravated with travel modes reducing from all modes to urban mass transit alone.

2. The PPC model can effectively weaken the over-estimation when the deterrence between OD stations is too little. Its performance is much better than the traditional gravity model, with standard deviation reduced by 56.02%.

3. In the PPC model, interchange times isn’t the main influence factor any longer. And there is no big difference in the performance among models with various deterrence functions and the one without a deterrence function. This result means there may be some difference among the deterrence during various periods, but the influence of the difference on the PPC is not obvious.

When applying the PPC model in planning and design phase, the proportion of peak-period boardings of the origin station and the proportion of peak-period alightings of the destination station can be forecasted based on the socio-economic and land use characteristics around stations.
REFERENCES


