Intercity train dwell time estimation by using robust regression method: a study case on Surabaya-Yogyakarta line

P Dewi\textsuperscript{1} and H Widyastuti\textsuperscript{2*}
\textsuperscript{1} The Indonesian Railways Polytechnic, Madiun 63129, East Java, Indonesia
\textsuperscript{2} Department of Civil Engineering, Sepuluh Nopember Institute of Technology, Surabaya 60111, East Java, Indonesia

*Corresponding author's e-mail: hera@ce.its.ac.id

Abstract. Dwell time is important an issue to optimise not only for train operation efficiency but also punctuality. This paper proposes a model based on a field survey of intercity trains along the Surabaya–Yogyakarta line by using a robust regression analysis of 138 observations from 13 trainsets operating on the line. Assuming no differences in door or platform width, a model has been formulated to include the number of people who are alighting and boarding, the type of platform and the number of train doors on one side. There were three robust estimators applied to the model. With a 2.037\% mean absolute percentage error, the robust regression using the M estimator formed the dwell time model for this study. The results show that, between the dummy variable for platform type using a binary number (0 or 1), dummy variables for high-level platforms and mid-level platforms have opposing tendencies. High-level platform use reduces dwell time by 2.014 sec based on continuous variables; in contrast, dwell time increases by 9.353 sec based on continuous variables by using mid-level platforms. Furthermore, people who are alighting need less time than those boarding. Dwell time will increase by 1.222 sec per alighting passenger and 1.314 per boarding passenger, while it will decrease by 2.727 sec per door on a one-sided trainset. Moreover, dwell time estimates with the model have reduced trainset travel time.

1. Introduction
Dwell time is an essential issue in train scheduling. Indonesian railway scheduling is charged to the Ministry of Transportation as the infrastructure owner and to PT Kereta Api Indonesia (Persero) as the intercity train operator. The Surabaya–Yogyakarta line is serviced by 13 trainsets; however, there are only two trainsets that service this line’s main business. The rest trainset has further terminal station far off Yogyakarta and Surabaya. However, passenger demand on this line is high. PT Kereta Api Indonesia transported 2,478,064 people in 2017, 2,894,819 people in 2018 and 2,093,638 people from January to September 2019 on the Surabaya–Yogyakarta line with their 13 trainsets. Dwell time and stop stations for each trainset vary by classes (economy, executive and mixed class). The longest total dwell time along the line is 1 hour and 8 minutes with 13 stops, and the shortest is 15 minutes with four stops based on the 2019 Indonesian Train Schedule.

The Surabaya–Yogyakarta line has upgraded to a double track recently. It serves 45 stations in three provinces. For passenger purposes, there are only 19 stations where loading and/or unloading activities for the intercity trainset occur. In order to make dwell times for passenger flows in stations run effectively and fairly without dwell time classifications, mathematical models were established to estimate the effective dwell time for each serviced station.
2. Literature review
Railway scheduling determines short stops in stations for loading and/or unloading passengers, train operation crossing and overtaking. There are no standard provisions for assigning dwell times in stations. Many researchers have studied dwell time using varied methods and approaches. A dwell time model including the number of alighting, boarding and through passengers using multiple linear regression was established in the Massachusetts Bay Transportation Authority (MTBA) Red Line. The model resulted in a linear liaison between the number of alighting and boarding passengers, but there was nonlinear relation in vehicle congestion. [1]

Another study was conducted in Mass Rapid Transit Swiss that developed a multiple linear regression model including the number of alighting and boarding passengers at critical doors, delays, morning peak hours, train type and rainy days. The study indicated that passenger boarding (1.2 sec/person) takes more time than alighting (0.9 sec/person) [2]. A study applied to London Underground Ltd (LUL) proposed a linear method to estimate stop times by including critical door factors, the number of boarding passengers, the number of alighting passengers, the number of train doors, the number of train seats, the number of through passengers and the door widths. However, it resulted in a nonlinear liaison for dwell time and number of alighting and boarding passengers [3].

A multi-agent simulation was established in Japan and found agent (passenger) decisions to alight and board were influenced by the environment and other agents. The environment consists of the platform and the train. The platform attributes were its length and width, stair locations, elevators, walls, pillars and where the train stops. The train attributes were the number of cars; car length and width; door locations; number of doors; door width; and seat location, width and length. For the agents, calculations of walking speed and direction towards the nearest door based on the environment were completed [4].

Multiple linear regression was chosen to model Istanbul Turkey Tramways’ dwell time on weekends and weekdays. There were some variables that have important effects on the dwell time that were considered in the models, which included the number of alighting and boarding passengers, platform occupancy, wagon occupancy, peak hours and off-peak hours on weekdays and weekends and time spending for operation by the driver. High platform occupancy had a greater influence on tramway dwell time than high wagon occupancy on weekdays, but this was the opposite on weekends. Among the continuous variables, in both the weekday and weekend models, the time spent by the driver had the highest effect on dwell time. The number of passengers who were alighting in the weekday model and the number of passengers who were boarding in the weekend model had the lowest impact on dwell time [5]. A linear regression also resulted in a statistically significant liaison in a Taiwan Railway Administration (TRA) study. Through a field survey with digital applications for surveyors on each platform, the study resulted in average time of 1.64 sec/person for boarding and 1.38 sec/person for alighting [6].

Regression is one of the most commonly used statistical techniques. Out of many possible regression techniques, the ordinary least squares (OLS) method has been generally adopted because of its ease of computation. However, the OLS estimation of regression weights in a multiple regression is affected by outliers, non-normality, multicollinearity and missing data [7]. Outliers are observations that appear inconsistent with the rest of the data. Often, such influential points remain hidden to the user, because they do not always appear in least squares residual plots. To remedy this problem, new statistical techniques have been developed that are not so easily affected by outliers. These are robust methods, such as the Huber M Estimation, MM Estimation and S Estimation. [8]

In order to find the best regression model, mean absolute percentage error (MAPE) can be used as a benchmark. The best MAPE model is equivalent to a weighted mean absolute error (MAE) regression. Its universal consistency in empirical risk minimisation remains possible using MAPE instead of MAE [9].

3. Data collection
The data used in this paper were obtained from on-board observations of the Surabaya–Yogyakarta line on 13 trainsets for each station stop. The dwell time in this paper is the time between when the trainset fully stopped on the station platform and when the on-board operator (usually the on-board janitor) closed the door. The surveyor would stand in the middle of the trainset (in front of the dining car) to
observe passenger flows and measured dwell time using a stopwatch. The platform types were observed directly for each stop and written in an observation form (as below). High-level platforms (1000 mm from the head rail) were defined as having no step differences between the platform height and the train floor; mid-level platforms (430 mm from the head rail) were defined as having a small step between the platform and train floor, which less requires stairs to accommodate passenger alighting and boarding; and the low-level platforms (180 mm from the head rail) were described as having a huge step between the platform and the train floor, which require stairs to ease passenger flow. The number of passengers who were alighting and boarding was informed by on-board officers and was updated during the observation day by the passenger section of Daerah Operasi (DAOP) 8 Surabaya, a regional office of PT Kereta Api Indonesia who manage train operations for certain line on the Surabaya–Yogyakarta route. The observation days were conducted from 3 February to 13 February 2020. The observations were gathered from Surabaya to Yogyakarta on 13 trainsets (one run), comprising 138 station observations using the observation form in figure 1 below.

Figure 1. Observation form example

To complete the observation form in figure 1, before the observation started, the surveyor had to write the trainset identity, one-sided doors number and observation date and strikethrough one route choice (trainset origin–destination). The following columns were written during the observation period with the help of an on-board officer for passenger number data. To determine the platform type, the surveyor had to assume platform types based on the vertical distance between the car floor and the platform, where some stations used portable stairs to help passengers up onto mid-level or low-level platforms. The time column was filled in based on the duration of passenger activities on platform, which the surveyor stopped counting when no more alighting and boarding activities took place. The arrival column was filled in with the time the trainset spent fully stopped on the platform, and the departure column was filled in with the time the trainset started to run to depart the station.

There are 19 stations where passenger loading and/or unloading occurs. The types of platform and serviced stations are presented in table 1.

Table 1. Station and platform types

| No | Station     | Platform Type |
|----|-------------|--------------|
| 1  | Yogyakarta  | High-Level   |
| 2  | Lempuyangan | High-Level   |
| 3  | Klaten      | Mid-Level    |
| 4  | Purwosari   | Mid-Level    |
| 5  | Solo Balapan| High-Level   |
| 6  | Sragen      | Mid-Level    |
| 7  | Walikukun   | Low-Level    |
| 8  | Ngawi       | Mid-Level    |
| 9  | Magetan     | Mid-Level    |
| 10 | Madiun      | High-Level   |
| 11 | Caruban     | Mid-Level    |
| 12 | Nganjuk     | Mid-Level    |
| 13 | Kertosono   | Mid-Level    |
Based on Table 1, there are five stations that use high-level platforms, 11 stations that use mid-level platforms and only three stations that use low-level platforms for passengers to board and alight. Furthermore, one predicting variable for dwell time was the number of doors on one side of the trainset, which were opened during passenger boarding and alighting. The number of doors varied depending on the trainset; the lowest number was 12 on Sritanjung, and the highest was 18 on a number of trainsets. Passenger flow in each station was divided into two steps. The first step was passenger alighting, and the next step was passenger boarding, during the transition period in which passengers who intended to board had to wait until there were no more alighting passengers on the platform area. Then, there was mostly no friction between alighting and boarding passengers on the platform area.

4. Data analysis
The modelling was approached by a linear liaison assumption between the independent variables, including number of doors, number of alighting passengers, number of boarding passengers and type of platform for each stop station. For the platform type, a dummy variable was generated. The dummy variable using a binary number (1 or 0) for each platform type quantified the originally qualitative platform data. Although there were three types of platform, the dummy variable included only two. In this case, mid-level and high-level platforms were chosen because of their frequency in the observation data. OLS was the first method conducted followed by the classic assumption test and outlier identification. The violation of the classic assumption and the presence of outliers transformed the OLS regression into a robust regression with M, S and MM estimators. To choose the best model, MAPE was calculated.

4.1. Ordinary least square model
An OLS is introduced to build the first estimation model. The independent variables are fitted by running a regression analysis and collinearity test. Furthermore, the classic assumptions for OLS were conducted, which consisted of a linearity test, a normality test, a heteroscedasticity test and a multicollinearity test. The OLS model is sensitive to outliers, a small number of which could affect its results. To investigate outliers, Cook’s distance, DIFIT and DrBETA were applied. The general OLS model is:

\[ Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \epsilon_i \]  

1. Linearity test
   The linearity test is conducted to confirm that the mean of the data group is placed in a straight line.

2. Normality test
   The normality tests are supplementary to the graphical assessment of normality. One of main tests for the assessment of normality is the Shapiro–Wilk test, which compares the scores in the sample to a normally distributed set of scores with the same mean and standard deviation. The null hypothesis is that the sample distribution is normal.

3. Heteroscedasticity test
   Heteroscedasticity is the unequal variance of the data along the regression line (i.e., when the spread of the data points is not constant). Heteroscedasticity has serious consequences for the OLS estimator. Although the OLS estimator remains unbiased, the estimated standard error (SE) is wrong.
Because of this, confidence intervals and hypotheses tests cannot be relied on. In addition, the OLS estimator is no longer Best Linear Unbiased Estimation (BLUE).

4. Multicollinearity test
Multicollinearity occurs when the independent variables in a regression model are correlated. This correlation is a problem because independent variables should be independent. If the degree of correlation between the variables is high enough, it can cause problems when fitting the model and interpreting the results. Variance inflation factors (VIFs) measure inflation in the variances of the parameter estimates due to collinearities that exist among the predictors. It is a measure of how much the variance of the estimated regression coefficient $\beta_k$ is inflated by the existence of a correlation among the predictor variables in the model. A VIF of 1 means that there is no correlation among the $k^{th}$ predictor and the remaining predictor variables, and the variance of $\beta_k$ is hence not inflated. The general rule is that VIFs exceeding 4 warrant further investigation, while VIFs exceeding 10 are signs of serious multicollinearity requiring correction.

The VIF is calculated as:

$$VIF = \frac{1}{1 - R_k^2} = \text{Tolerance}$$

(2)

4.2. Outliers
A common problem in regression analysis is the presence of outliers. As defined by Barnett and Lewis (1994), outliers are observations that appear inconsistent with the rest of data. Outliers may occur as a result of unusual, but explainable, events, such as faulty measurements, the incorrect recording of data, the failure of a measurement instrument and so on.

1. Leverage point
A data point has high leverage if it has extreme predictor $X$ values. With a single predictor, an extreme $X$ value is simply one that is particularly high or low. With multiple predictors, extreme $X$ values may be particularly high or low for one or more predictors or unusual combinations of predictor values.

$$h_{ii} = X_i^T(X^TX)^{-1}X_i; i: 1, 2, \ldots, n$$

(3)

$$\bar{2h_{ii}} = \frac{2}{n^2} \sum_{i=1}^{n} h_{ii} = \frac{2k}{n} = 2(p-1)$$

(4)

An observation data-i would be defined as a leverage on $X$, if $h_{ii} > 2\bar{h_{ii}}$.

2. Cook’s distance
Cook’s distance is the scaled change in fitted values, which is useful for identifying outliers in the $X$ values (observations for predictor variables). Cook’s distance shows the influence of each observation on the fitted response values. An observation with a Cook’s distance greater than three times the mean Cook’s distance might be an outlier.

$$(\text{Cook’s Distance})_i = \left[\frac{R_{\text{standard}}^2}{2}\right] \times \left[\frac{h_{ii}}{1 - h_{ii}}\right]$$

(5)

3. Difference in fits (DfFITS)
The DfFITS is a scaled measure of the change in the predicted value for the $i^{th}$ observation and is calculated by deleting the $i^{th}$ observation with the formula given as:

$$(\text{DfFITS})_i = \frac{\hat{y}_i - \hat{y}_{i-1}}{s_{i-1} - \sqrt{h_{ii}}}$$

(6)
Data-i would be defined as an outlier if:
\[ |DfFITSi| > 1 \quad \text{for } \leq 30 \] (7)
\[ |DfFITSi| > 2 \left( \frac{p}{n} \right)^{0.5} \quad \text{for } n > 30 \] (8)

As the data for this research have important value that could not be deleted at any cost, a robust regression method that provides an alternative to the least squares regression and that works with less restrictive assumptions was applied. Specifically, it provides much better regression coefficient estimates when outliers were present in the data. Outliers violate the assumption of normally distributed residuals in least squares regression. They tend to distort the least squares coefficients by having more influence than they deserve.

4.3. Robust regression model

Robust regression is a regression method that is used when the distribution of residuals is not normal or there are outliers that affect the model. This method is an important tool for analysing the data affected by outliers so that the resulting models are stout against outliers. The robust estimation method is a form of regression analysis designed to circumvent the limitations of traditional parametric and non-parametric methods. Robust estimation methods are designed to not be overly affected by outliers. Under these conditions, a robust regression that is resistant to the influence of outliers is the best method [10]. Therefore, we introduce a comparison of robust regression methods to determine the best method.

1. M estimation

The most common general method of robust regression is the M estimation method, introduced by Huber (1964). The M estimation method is a generalisation to a maximum likelihood estimation in the context of location models. This is nearly as efficient as an OLS regression. Rather than minimising the sum of squared errors as the objective, the M estimation method principle is minimising the residual function. The M estimation objective function is:
\[ \sum_{i=1}^{n} \rho(e_i) = \sum_{i=1}^{n} \rho(y_i - x_i^Tb) \] (9)

With the Huber weight function:
\[ p(e) = \begin{cases} \frac{e^2}{2} ; & |e| \leq k \\ k|e| - \frac{k^2}{2} ; & |e| > k \end{cases} \] (10)

The general function for the M estimator is:
\[ \hat{\beta}_M = (x'wx)^{-1} (x'wY) \] (11)

2. S estimation

The regression estimators associated with M scales are the S estimators proposed by Yohai (1987). The S estimation method is based on the residual scale of the M estimation method. S estimators are a generalisation of Least Median Square (LMS) and Least Trimmed Squares (LTS). They have the same asymptotic properties corresponding to M estimators and handle 50% of the outliers appearing in the data. The weakness of the M estimation method is the lack of consideration of the data distribution and that it is not a function of the overall data because it only uses the median as the weighted value. S estimators refer to the fact that this estimator essentially is based on the minimisation of a (robust) scale M estimator. This method uses the residual standard deviation to overcome the weaknesses of median [11]. The S estimator is defined by:
\[ \hat{\beta}_S = \min \hat{s}(e(\beta)) \] (12)
With \( \hat{s}(e) \) being obtained from the M estimation residual scalar \( s(e) \), which is the solution from
\[
\frac{1}{n} \sum_i \rho \left( \frac{e_i(\beta)}{\hat{s}(e_i)} \right) = \delta.
\]

3. Modified M estimation
The modified M (MM) estimation method is a special type of M estimation method developed by Yohai (1987). MM estimation is a combination of the high breakdown value estimation method and the efficient estimation method. Yohai’s MM estimator was the first estimation with a high breakdown point and high efficiency under normal error. The MM estimator regression is defined as:
\[
\hat{\beta}_{MM} = \min \beta \sum_{i=1}^{n} \rho \left( \frac{e_i}{\hat{\sigma}} \right) = \min \beta \sum_{i=1}^{n} \rho \left( \frac{y_i - \sum_{j=0}^{k} x_{ij} \beta_j}{\hat{\sigma}} \right)
\]
\[\text{(13)}\]

4.4. Model determination
In order to choose the best model, Mean Absolute Percentage Error (MAPE) was applied to get the best estimation model with the lowest MAPE value:
\[
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \times 100\%
\]
\[\text{(14)}\]

5. Result and discussion
5.1. Data description
The on-board observation for 13 trainset stop stations resulted in 138 data point to further analyse. The descriptive statistics of the observation results show that the number of doors is between 12 and 18 unit per trainset, excluding the wagon, dining and generator cars. The number of boarding and alighting passengers varied for each stop station; there were no passengers alighting or boarding in some cases. The summary of the observation results is shown in Table 2 below.

| Variable     | Obs | Mean   | Std.Dev. | Min | Max |
|--------------|-----|--------|----------|-----|-----|
| Door         | 138 | 15.899 | 2.069    | 12  | 18  |
| Boarding     | 138 | 41.13  | 70.37    | 0   | 386 |
| Alighting    | 138 | 32.833 | 62.037   | 0   | 493 |
| Dwell time   | 138 | 132.797| 133.431  | 7   | 665 |

Based on table 2, the standard deviation of the continuous variables was high; therefore, outliers and leverages were assumed to be present. The highest number of boarding passengers was 386, while there was no boarding activity for some stop stations.

| Table 3. Dummy variable summarize |
|----------------------------------|
| Dummy Variables                | Freq. | Percent | Cum.  |
| Low-Level Platform             | 3     | 2.17    | 2.17  |
| Mid-Level Platform             | 87    | 63.04   | 65.22 |
| High-Level Platform            | 48    | 34.78   | 100.00|
| Total                          | 138   | 100.00  |       |

As shown in table 3, varied platforms were used to service passengers at each stop station. Mid-level platform use dominated with more than 50% of cases, while low-level platforms were rarely used. For
the statistical analysis, the type of platform was defined as a dummy variable of 0 if the serviced station used another type platform and 1 if the station used the specific platform type.

**Table 4. Correlation matrix among variables**

| Variables                  | (1) | (2) | (3) | (4) | (5) | (6) |
|----------------------------|-----|-----|-----|-----|-----|-----|
| (1) Time                   | 1.000 |     |     |     |     |     |
| (2) Door Number            | -0.089 | 1.000 |     |     |     |     |
| (3) Boarding               | 0.801* | 0.027 | 1.000 |     |     |     |
| (4) Alighting              | 0.619* | -0.106 | 0.184* | 1.000 |     |     |
| (5) Mid-Level Platform     | -0.583* | -0.067 | -0.497* | -0.484* | 1.000 |     |
| (6) High-level Platform    | 0.623* | 0.065 | 0.528* | 0.514* | -0.954* | 1.000 |

* shows significance at the 0.05 level

**Figure 2. Matrix of variables Scatter-plot**

Based on table 4 and figure 2, among the six independent variables, four had statistically significant connections to dwell time. However, the multiple linear regression analysis result was not satisfactory (see table 7). The studentized residual of the multiple linear regression model deviated from the linear assumption, while its adjusted R2 was 0.87. The normality test for the studentized residual with the Shapiro–Wilk test resulted in 0.00000, which forced the rejection of the null hypothesis (that the residual is normally distributed). The result was supported by a kernel density estimation, as seen in figure 3.
Figure 3. Kernel density estimate

In Figure 3, the kernel density estimation based on observation data has a similar shape and pattern to the normal density curve; however, the peak point of the observation data exceeds the normal density. This indicates that the observations contain extreme data, which potentially influence the regression model’s accuracy.

In terms of multicollinearity, the VIF values for continuous variables were below 5, while the dummy variable VIFs were greater than 10. For this case, the multicollinearity test can be ignored for the dummy variable because it was sourced from three types of platform that interact with each other. In addition, the heteroscedasticity test using the Breusch–Pagan/Cook–Weisberg test did not have a satisfactory result at valued 0.0000, which forced the rejection of H0 (constant variance).

5.2. Outlier identification

In order to identify outliers and leverage points, the Cook’s distance and DfFITS methods indicated the same data points as outliers with the highest exceeding values from the limit. The limitation for Cook’s distance is $4/n$, and the boundary value of DfFITS for detecting outliers is $2\sqrt{p/n}$, where $p$ is number of parameters and $n$ is number of data points. The top ten data points that most violated the value based on Cook’s distance and DfFITS are listed in Table 5 and Table 6 below.

| ID | Door | Boarding | Alight-g | Mid_Le-m | High_L-m | Time | d |
|----|------|----------|----------|----------|----------|------|---|
| 7. | s7   | 18       | 0        | 269      | 0        | 1    | 373 | .8485313 |
| 15. | s15  | 18       | 269      | 0        | 0        | 1    | 290 | .1038137 |
| 67. | s67  | 12       | 0        | 493      | 0        | 1    | 379 | .4376512 |
| 81. | s81  | 12       | 129      | 143      | 0        | 1    | 572 | .1788606 |
| 82. | s82  | 18       | 130      | 68       | 0        | 1    | 532 | .1442872 |
| 183. | s103 | 16       | 386      | 0        | 0        | 1    | 633 | .1074935 |
| 111. | s111 | 16       | 171      | 126      | 0        | 1    | 295 | .8481431 |
| 121. | s121 | 18       | 331      | 0        | 0        | 1    | 311 | .4398887 |
| 130. | s130 | 14       | 91       | 133      | 0        | 1    | 434 | .6480041 |
| 138. | s138 | 14       | 236      | 125      | 0        | 1    | 665 | .1849594 |
Comparing both table 5 and table 6, outlier points were detected in s7, s15, s67, s81, s82, s103, s111, s112, s130 and s138. The outlier data contain important data that represent real situations in the field where there were no passengers boarding but hundreds of passengers alighting. In addition, to uncover more influencing data, the leverage points and normalised squared residual were calculated.

Figure 4. Outlier and leverage identification

The lines on figure 4 show the average values of leverage and the (normalised) residuals squared. Points above the horizontal line have higher-than-average leverage, and points to the right of the vertical line have larger-than-average residuals. As the chart above shows, there are many data point that have unbalanced values of leverage and normalised residual squared. For instance, s82 has larger-than-average residuals while s81 has larger-than-average residuals and leverage. The data points exceeding the average greatly affect the slope of the regression line. Based on table 5, table 6 and figure 4, the observation data have many outliers and influencing data. Next, robust regression methods were conducted to get the best estimation model.

5.3. Robust regression result
Given the deviation from the OLS assumptions, robust regressions were developed using M estimation with Huber’s weight, S estimation and MM estimation. The comparison between them and OLS is presented in table 7.

Table 6. The result of identifying outlier by estimate DfFITS

| ID | Door | Boarding | Alight | Mid_Le | High_Le | Time | dfit |
|----|------|----------|--------|--------|---------|------|------|
| 7  | s7   | 18       | 0      | 269    | 0       | 1    | 373  | .5411851 |
| 15 | s15  | 18       | 269    | 0      | 0       | 1    | 290  | -.8038641 |
| 67 | s67  | 12       | 0      | 493    | 0       | 1    | 379  | -5.778908 |
| 81 | s81  | 12       | 129    | 143    | 0       | 1    | 572  | 1.105113  |
| 82 | s82  | 18       | 130    | 68     | 0       | 1    | 532  | 1.043153  |
| 103| s103 | 16       | 386    | 0      | 0       | 1    | 633  | .8078269  |
| 111| s111 | 16       | 171    | 126    | 0       | 1    | 295  | -.4992299 |
| 121| s121 | 18       | 331    | 0      | 0       | 1    | 311  | -1.717546 |
| 130| s130 | 14       | 91     | 133    | 0       | 1    | 434  | .5010158  |
| 138| s138 | 14       | 236    | 125    | 0       | 1    | 665  | 1.105176  |
Table 7. Comparison model

| VARIABLES           | OLS DwellingTime | M-Estimator DwellingTime | S Estimator DwellingTime | MM Estimator DwellingTime |
|---------------------|------------------|--------------------------|--------------------------|----------------------------|
| Door Number         | -3.685*          | -2.747**                 | -2.727***                | -2.122**                   |
|                     | (2.021)          | (1.218)                  | (1.020)                  | (0.971)                    |
| Boarding            | 1.349***         | 1.314***                 | 1.567***                 | 1.514***                   |
|                     | (0.0694)         | (0.144)                  | (0.0251)                 | (0.204)                    |
| Alighting           | 1.031***         | 1.222***                 | 1.194***                 | 1.267***                   |
|                     | (0.0790)         | (0.131)                  | (0.0484)                 | (0.0812)                   |
| Mid-Level Platform  | 11.08            | 9.353**                  | 1.779                    | 6.699                      |
|                     | (28.30)          | (4.638)                  | (5.014)                  | (5.492)                    |
| High-Level Platform | 12.07            | -2.014                   | -7.912                   | -7.912                     |
|                     | (30.06)          | (7.833)                  | (5.829)                  | (8.510)                    |
| Constant            | 90.88**          | 76.06***                 | 74.48***                 | 65.05***                   |
|                     | (42.65)          | (20.15)                  | (16.77)                  | (15.76)                    |
| Observations        | 138              | 138                      | 138                      | 138                        |
| R-squared           | 0.874            |                          |                          |                            |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As Table 7 shows, the four models show that mid-level platforms increase dwell time while high-level platform reduce it, except the high-level platform on OLS model. Among the robust estimators, the M estimator had four independent variables with a less than 0.05% probability error. The comparison of R-square values between the OLS and robust models is inaccurate because R-square in robust regressions is different from R-square in OLS.

5.4. Model determination
The determination between the models in Table 7 is based on MAPE calculation result as below:

Table 8. MAPE comparison between model

| No | Method                              | MAPE  |
|----|-------------------------------------|-------|
| 1  | Ordinary Least Square               | 3.368%|
| 2  | Robust Regression S Estimation       | 3.507%|
| 3  | Robust Regression M Estimation       | 2.037%|
| 4  | Robust Regression MM Estimation      | 2.352%|

Based on Table 8 above, the M estimation had the lowest MAPE value, indicating the lowest error of all the models. Here, a robust regression model using M estimation for dwell time was taken. The model is written as:

\[
\text{Dwell Time} = -2.727 \text{ Door number} + 1.314 \text{ Boarding} + 1.222 \text{ Alighting} \\
+ 9.353 \text{ MidLevel Platform} - 2.014 \text{ HighLevel Platform} + 76.06
\]

Based on the model, alighting passengers need less time than boarding, with 1.222 sec added to dwell time per alighting passenger and 1.314 sec per boarding passenger. By using high-level platforms on the stop stations, dwell time will be reduced by 2.014 sec based on continuous variables, and using mid-level platforms will increase dwell time by 9.353 sec based on continuous variables.
6. Applied model
The model application reduced dwell time for each station per trainset. For instance, the dwell time model applied to the Sancaka and Argo Wilis data resulted in more than 40% in time savings compared to the actual dwell time based on passenger volume on the day of observation.

From table 9 above, the total actual dwell time for Argo Wilis was 29 minutes, while the model resulted in 18 minutes of total dwell time, whereas in the actual operation, total dwell time for Sancaka was 35 minutes. Furthermore, with dwell time modelling, total travel time per trainset will be reduced.

Table 9. Argo Wilis and Sancaka saving time from dwell time

| Data | Trainset         | STATION     | Platform | Boarding (person) | Alighting (person) | Door (unit) | Dwelltime in Field | Dwelltime With Model in second | Roundup Dwelltime With Model in Minute | Saving Time | Saving per Trainset |
|------|------------------|-------------|----------|-------------------|-------------------|-------------|--------------------|-------------------------------|----------------------------------------|-------------|---------------------|
| s1   | ARG0 WILIS/1/SGU-GMR | YOGYAKARTA  | High     | 0                 | 117               | 16          | 00:01              | 00:02                         | 00:11                                   |             |                     |
| s2   | ARG0 WILIS/1/SGU-GMR | SOLO BALAPAN| High     | 20               | 26                | 16          | 00:07              | 88,146                        | 00:02                                   | 00:05       |                     |
| s3   | ARG0 WILIS/1/SGU-GMR | MADUN       | High     | 79               | 62                | 16          | 00:05              | 209,664                       | 00:04                                   | 00:01       | 00:19               |
| s4   | ARG0 WILIS/1/SGU-GMR | KERTOSONO   | Mid      | 7                | 0                 | 16          | 00:02              | 50,659                        | 00:01                                   | 00:01       |                     |
| s5   | ARG0 WILIS/1/SGU-GMR | JOMBANG     | Mid      | 7                | 8                 | 16          | 00:02              | 60,435                        | 00:01                                   | 00:04       |                     |
| s6   | SANCACA/179/SGU-YK | KLATEN      | Mid      | 4                | 12                | 18          | 00:02              | 55,847                        | 00:01                                   | 00:01       |                     |
| s7   | SANCACA/179/SGU-YK | SOLO BALAPAN| High     | 19               | 74                | 18          | 00:05              | 139,994                       | 00:03                                   | 00:02       |                     |
| s8   | SANCACA/179/SGU-YK | MADUN       | High     | 42               | 17                | 18          | 00:16              | 100,562                       | 00:02                                   | 00:14       |                     |
| s9   | SANCACA/179/SGU-YK | NGANJUK     | Mid      | 7                | 4                 | 18          | 00:02              | 50,053                        | 00:01                                   | 00:01       | 00:23               |
| s10  | SANCACA/179/SGU-YK | KERTOSONO   | Mid      | 8                | 2                 | 14          | 00:03              | 48,923                        | 00:01                                   | 00:02       |                     |
| s11  | SANCACA/179/SGU-YK | JOMBANG     | Mid      | 15               | 18                | 18          | 00:03              | 77,673                        | 00:02                                   | 00:01       |                     |
| s12  | SANCACA/179/SGU-YK | MOJOKERTO   | Mid      | 39               | 7                 | 18          | 00:04              | 95,767                        | 00:02                                   | 00:02       |                     |
| s13  | SANCACA/179/SGU-YK | MOJOKERTO   | Mid      | 39               | 7                 | 18          | 00:04              | 95,767                        | 00:02                                   | 00:02       |                     |

7. Conclusions
The dwell time study on the Surabaya–Yogyakarta trainset estimation using statistical approaches has resulted in a model with robust regression methods due to the violation of the OLS assumptions of heteroscedasticity and normality and the presence of outliers. In the comparison of OLS and the three robust regression estimators, the robust regression with the Huber M estimator had best results with minimal error. According to the model, alighting passengers need less time than boarding passengers. The model’s applications to passenger volume, number of doors and platform type caused the trainset travel time to run effectively.

Acknowledgements
We would like to acknowledge the fruitful discussion and help during our observations from Asad Triadi and team from DAOP 8 Surabaya PT Kereta Api Indonesia. The author would also thank to the vice president of passenger marketing at PT Kereta Api Indonesia for providing passenger data on the Surabaya–Yogyakarta line.

References
[1] Puong A 2000 Dwell Time Model and Analysis for MBTA Red Line Massachusetts Institute of Technology Research Memo 1-10
[2] Gysin K 2018 An Investigation of the Influences on Train Dwell Time (Zurich: Swiss Federal Institute of Technology, ETH Zurich) 601-700
[3] Harris N G 2006 Train Boarding and Alighting Rates at High Passenger Loads J. of Adv. Trans. 40 249–63
[4] Yamamura A, Koresawa M, Inagi T, Tomii N 2013 Proc. of World Conf. on Transportation Research (WCTR) (Rio de Janeiro: World Conference on Transport Research Society) 13
[5] Amirnazmiafshar E 2019 Tramway Dwell Time Estimation and Analysis by using Multiple Linear Regression Method European J. Eng. Sci. Tech. 2 36-48
[6] Jong J-C and Chang E-F 2011 Proc. of World Cong. on Railway Research (WCRR) (Lilie: The Rail Safety and Standards Board(UIC)) 9
[7] Ho K and Naugher J R 2000 Outlier Lies: An Illustrative Example of Identifying Outliers and Applying Robust Models Mult. Linear Regres. Viewpoints 26 2–6
[8] Berk R A 1990 A Primer on Robust Regression Modern Methods of Data Analysis Eds J Fox and J S Long (Newbury Park: Sage Publications) chapter 7 292–323
[9] Myttenaere A D, Golden B, Grand B L, Rossi F 2016 J. Adv. in artific. neural net., mach. learning and comp. intel.: Selected papers from the 23rd Eur. Symp. on Artific. Neural Networks (ESANN 2015) 192 38-48

[10] Draper N R and Smith H 1998 Applied Regression Analysis (Canada: John Wiley and Sons, Inc) Third Edition 567-584

[11] Susanti Y and Pratiwi H 2014 M-estimation, S-estimation, and MM-estimation in robust regression Int. J. Pure Applied Math. 91 349-360