Sigma J Eng Nat Sci, Vol. 42, No. 2, pp. 555−565, April, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00043

Research Article

A new modelling approach for air transportation: A case study for total number of air passengers per month

Reşit ÇELİK[1], Hasan Aykut KARABOĞA[2,*], İbrahim DEMİR[3], Erdal GÜL[4]

[1]*Department of Statistics, Yıldız Technical University, Istanbul, 34220, Türkiye*
[2]*Department of Educational Sciences, Amasya University, Amasya, 05100, Türkiye*
[3]*Turkish Statistical Institute, Ankara, 06420, Türkiye*
[4]*Department of Mathematics, Yıldız Technical University, Istanbul, 34220,Türkiye*

ARTICLE INFO

Article history
Received: 23 December 2022
Revised: 21 April 2023
Accepted: 04 May 2023

Keywords:
Modelling Air Passengers;
Studentized Residuals;
Correcting Heteroscedasticity;
Studentized RCEV Test;
Weighted Regression;
Forecasting

ABSTRACT

Modeling the number of air passengers correctly is essential for management policy in the global world. Based on seasonality (depending on the season of the year), data about the number of air passengers are heteroscedastic. Heteroscedasticity violates “Homoscedasticity” which is one of the central assumptions of linear regression analysis. In this study, a new weighting approach called “Weighting Absolute Centered External Variable” (WCEV) is applied to the Turkish total monthly air passenger’s data to obtain correct statistical inference and forecasting. Besides scatter plot months vs. studentized residuals, the homoscedasticity assumption is checked with the studentized RCEV test as well. Consequently, the WCEV method is shown superior performance against multiple linear regressions and exponential weighted moving average (EWMA) methods. The study also provides insights into the seasonal patterns of air passenger demand in Turkey, with passenger mobility increasing in the last quarter of each year and the lowest demand in January and February. This information can be used to optimize airport and airplane maintenance schedules and increase capacity during peak months.

Cite this article as: Çelik R, Karaboğa HA, Demir I, Gül E. A new modelling approach for air transportation: A case study for total number of air passengers per month. Sigma J Eng Nat Sci 2024;42(2):555−565.

INTRODUCTION

Air transportation has been a driving power of economic and social development [1–3]. Furthermore, the growth of the aviation industry enhances the trade development in a country, in turn; it provides a great deal of contributions to its wealth [4]. However, unqualified employees, inadequate ground operations, and political and economic conflicts have negative effects on civil aviation. In addition, poor service quality [5], trade barriers [6], ticket prices [7], competition with high-speed trains [8], market shocks and terrorist attacks [4] or pandemics -like Covid-19- create undesired impacts on air transportation industry. The

***Corresponding author.**
*E-mail address: h.aykut.karaboga@amasya.edu.tr

This paper was recommended for publication in revised form by Regional Editor-in-Chief Ahmet Selim Dalkılıç

Published by Yıldız Technical University Press, İstanbul, Türkiye

COVID-19 pandemic has had a significant impact on air passenger demand, causing a sharp decline in the number of air traveling demands. Despite the pandemic, air passenger demand is expected to grow steadily in the long term due to factors such as rising disposable incomes, the growth of international trade, and the increasing popularity of air travel among the middle class. In conclusion, air passenger demand is a critical aspect of the aviation industry, driving growth and development but also presenting significant challenges that must be addressed. As the world emerges from the pandemic and the aviation industry adapts to the new normal, it will be essential to balance growth with sustainability and address the economic and social impacts of air travel. Thus, accurate demand forecasting models could minimize the mentioned social, economic, and political effects.

Demand forecasts reduce stakeholders' risks and create accurate transportation systems [9]. Demand forecasting is critical in airport operations to manage immediate operational reflections [10,11]. The consequence of operating inaccurate forecast models falls into several factors such as capacity optimization faults, increased operation costs, airport accumulation, and customer dissatisfaction. In this regard, managers, policy makers, and airport planners should be able to use a simple and efficient forecasting model to take necessary precautions. Long-term and short-term strategic transportation plans [12] -with accurate forecasting models- play a primary role in the air transport market, where volatility and risk are high.

Some discussions about demand forecasting can be seen in Hakim & Merkert [9], Wang & Song [13] and Wang & Gao [14]. Moreover, for macro (overall sales projections for strategic planning) or micro level (day to day or hour to hour) planning approaches and analysing methods we recommend to the readers detailed review [15].

In this study, we create a robust forecasting model for the demand of air the number of passengers, using a weighted regression model that has never been cited in the air transportation studies.

To sum up, our major research questions are:

- Can we provide a simple and efficient model for policy makers and airport planners?
- Is our model providing linear regression assumptions?
- Does a Weighted Regression model show a satisfying forecasting performance for the demand for the total number of air passengers per month?

The parts of the study are as follows: In the section 2, brief literature on air passenger demand forecasting was introduced. In the section 3, methodological approach used in the study was clarified. The data were presented data in the section 3. As for the section 4, analysis, results, and forecast results were presented. The final section is devoted to the evaluation of the research results, conclusions, and future research suggestions.

LITERATURE REVIEW

The studies about the air passengers' demand of generally focus on developed countries and especially due to their aviation history on the United States and the United Kingdom. Asia and in Africa focused studies provide insight on developing industries as well [16]. Thus, the developments in this field can be easily examined.

In the literature review, demand modeling approaches are generally categorized under three different headings: time series based, artificial intelligence-based, and linear regression-based approaches.

Regarding time series-based approach; Pitfield [17] modelled air passengers of United Kingdom with autoregressive integrated moving average (ARIMA). Bermúdez et al. [12] forecasted United Kingdom monthly air passenger data with additive Holt–Winters and maximum likelihood methods. Since the usage of the Holt–Winters forecasts inappropriate with sessional data, they transformed stochastic part of the data and the model provided homoscedastic, uncorrelated, and multivariate normal distributed residuals. Gelhausen et al. [18], forecasted German air passenger demand with co-integration theory-based model and compared with the classical four-step forecasting model. Additionally, they supported model success with the Brexit effect on traffic volume for German airports. Karlaftis [19] used the dynamic TOBIT model to estimate the monthly number of passengers of Corfu airport. In this study, Karlaftis underlined the accurate modelling for airports with high seasonal volatility.

Considering the hybrid artificial intelligence approach, Ming et al. [20] built a Hybrid ARIMA-SVMs model to predict multistep-ahead air transportation demand. However, the overtraining of the Support Vector Machine (SVM) got worsened the predictive performance of the hybrid ARIMA-SVM model. Carson et al. [21] compared aggregate and disaggregate air travel demand forecasts with quasi-aggregating individual markets (quasi-AIM) approach and forecasted monthly US civil aviation data. But data showed heterogeneity so, their disaggregate model outperformed to other models. According to Xiao et al. [22] decomposed data into three parts: seasonal, trend and irregular components. The model is composed of three parts too: generalized regression neural network (GRNN), singular spectrum analysis (SSA), and radial basis function networks (RNFNs). Their ensemble model overachieved in forecasting HKIA air passengers. Xiong et al. proposed an integrated approach using the MI-SVR machine learning model to predict air passenger traffic at Shanghai Pudong International Airport. The model incorporates key influencing factors based on mutual information, resulting in improved prediction accuracy compared to conventional methods [23]. Tang et al. developed a model to forecast short-term daily passenger traffic at a major airport terminal in China during the COVID-19 pandemic, which reduces model error by 27.7% compared to a baseline model and provides implications for airport operations [24]. In their study, Jin et al. (2020) evaluate various forecasting techniques, including ARIMA, ARMA, and machine learning, and propose that a hybrid approach using variational mode decomposition (VMD), ARMA, and kernel extreme learning machine (KELM) provides the most accurate predictions for future passenger traffic at three major Chinese airports [25].

Regarding linear regression-based forecasting, Abed et al. [26] used stepwise regression method in order to modeling Saudi Arabia's international air passengers demand. They tried to prevent multicollinearity, and they examined the variable relationships with 4 different models. It was revealed that total expenditures and the population were the most important variables that affect international flight demand in the country. Aderamo [27] analysed effecting factors of Nigerian domestic air transport demand with multiple linear regression method. He underlined that governmental support to improve air transportation systems in developing countries. Baikgaki [28] obtained a linear regression model for Republic of South Africa's domestic air passengers demand. They used stepwise regression to solve the multicollinearity problem. Sivrikaya and Tunç [29] forecasted domestic air passenger demand of Turkey for city pairs with a semi-logarithmic regression model. The adjusted R squared score of the model was 86.4%. The authors emphasized the potential of air travel demand forecasting for a new airport. Hakim and Merkert investigated the factors influencing air transport demand in South Asian nations between 1973 and 2015 using fixed-effects models and a three-step error correction model [9]. Naghawi et al. developed an econometric air passenger demand model in Jordan with 6 demand determinants using stepwise regression and multiple linear regression analysis with annual data from 2006 to 2017 [30].

In the literature review, we realized that air passenger data are irregular, high volatile and seasonal [31]. As stated Karlaftis [19], correct and unbiased forecast models become more significant in touristic regions like Antalya, İzmir or İstanbul. Air traffic or demand forecasting with daily and weekly data is more troublesome compared to monthly data as well [17]. However, the demand of forecasting the seasonal, auto-correlated, and non-stationary data is troublesome for artificial intelligence-based or econometric-based approaches. But adding seasonal dummies to model improve forecasts [32], and simple models give better results compared to complex models and external factors [33,34].

In the literature review, we have not found any similar application with Weighting Absolute Centered External Variable Method (WCEV) [35] forecast the demand of air the number of passengers, except WCEV method was successfully applied International airline passengers-(monthly total from jan 1948 to dec 1960) in [35] (data is available at https://www.kaggle.com/datasets/andreazzini/international-airline-passengers). The study does not contain forecasting. As a different approach to modeling air passenger demand, this study contains a forecasting model in which the months are independent variables and the total number of air passengers per month is dependent variable. Additionally, we eliminated the external factors and seasonality with a trend variable. Normality assumption is checked with the histogram of Studentized residual and homoscedasticity assumption is checked with studentized residuals plots [36] and studentized RCEV test [37].

MATERIALS AND METHODS

Data Description

The contributions supplied by the state to the aviation industry have provided strong growth in the aviation industry in recent years in Turkey. Turkish air transportation industry has still been developing greatly [38]. The development of the last fifteen years is a result of the liberalization of the markets. In this context, geographical position, growth potential and touristic attraction provide momentum to Turkey with its technological support, such that purchasing tickets, booking and check-in can be easily carried out from the websites of companies or other e-commerce websites. For example, liberalization in Turkish Civil Aviation has been a driving force in industry growth with its annual average 10% since 2003 [29].

The General Directorate of Turkish State Airports Administration (TSAA) has been serving the civil aviation of Turkey since 1933. After the year 2003, when domestic flights opened to competition of private airways, there has been an enormous increase in the total numbers of passengers, flights, number of destinations, and operating companies. The total number of destinations in 2014 was 53 from 7 centers with 6 airway companies for domestic flights and 237 from 108 countries for international flights. TSAA forecasted the total number of passengers 238 million for 2020 and 247 million for 2021, including direct transit passengers. Therefore, we should create accurate and simple models in order to manage air transportation potential. We obtained the data from the monthly reports of TSAA.

Methods

Regression analysis is well-known method in the statistics. Normality, homoscedasticity, uncorrelated residuals, and independent variables are the basic assumptions in classic linear regression analysis [39,40]. Homoscedasticity is one of the classical linear regression analysis assumptions. The homoscedasticity assumption is a central assumption of linear regression theory [41]. The distribution theory and statistical inference based on confidence intervals and tests of hypotheses are valid and meaningful only if the standard linear regression assumptions are satisfied [35]. Therefore, checking the homoscedasticity assumption is vital in linear regression analysis.

In this study we checked normality of residuals and outliers with the histogram of the studentized residuals;

homoscedasticity is checked with studentized RCEV test and scatter plots of studentized residuals vs months (as stated by Bischof et al. [42], if model specified correctly, studentized residuals are homoscedastic). The model sufficiency is evaluated with the F test, adjusted R^2, and standard errors. For extended information and application codes we encourage readers to Çelik's studentized RCEV test development study [37] and WCEV studies [35,36]. Some simulation of applications of the WCEV and Studentized RCEV test can be seen in Çelik [36,37] respectively.

Weighting Absolute Centered External Variable (WCEV) Method

WCEV algorithm [35,36] that corrects heteroscedasticity is given below. For the regression model given in equation 1,

$$Y = X\beta + \varepsilon \tag{1}$$

Ordinary least squares (OLS) parameter estimates, fit values, and residuals are obtained by $\hat{\beta} = (X'X)^{-1}(X'Y)$, $\hat{Y} = X\hat{\beta}$ and $\hat{\varepsilon}_i = y_i - \hat{y}_i$ respectively. Data is sorted by fitted values ascending order (In time series data is sorted by months ascending order). The external variable d is demonstrated as $d = \{1,2,\dots n\}$ in sorted data, where d denotes the external variable and n denotes the sample size. The regressing external variable on absolute residuals $\hat{\varepsilon}_i$ or is given below

$$\hat{s}_i = |\hat{\varepsilon}_i| = \alpha_0 + \alpha_1 d_i + \alpha_2 d_i^2 \tag{2}$$

Where $\hat{s}_i$ denotes standard deviation of estimates and α_1, α_2 are parameters to be estimated. As stated in Carroll and Ruppert [43] squared residuals can be thought of as estimates of the variance and absolute residuals can be thought of estimating relative standard deviations. Squared residuals can also be used in this step. The first-order derivative of this function with respect to the external variable d is

$$\frac{\partial s}{\partial d} = \alpha_1 + 2\alpha_2 d \tag{3}$$

The point that makes the derivative function equal to zero is the widest vertically part of the BDR surface. This point is called the nodal point of absolute residuals. Nodal point is obtained by setting the derivative equal to zero and solving for d: $(\alpha_1 + 2\alpha_2 d = 0 \rightarrow d = -\alpha_1/2\alpha_2)$

$$\widehat{m} = -0.5\frac{\widehat{\alpha}_1}{\widehat{\alpha}_2} \tag{4}$$

where $\widehat{m}$ is the nodal point estimation (dotted line in Figure 1-a and 1-b). The transformed centred external variable with respect to nodal point is given in equation 5.

$$o_i = (d_i - \widehat{m}) \tag{5}$$

where o_i denotes centred external variable. The weight estimations are obtained by taking absolute values of the centred external variable as in equation 6.

$$\widehat{w}_i = 1/|o_i|^{\gamma}\ ,\ |o_i| \neq 0\ ,\ \gamma \in [-2,2] \tag{6}$$

where $\widehat{w}_i$ denotes weight estimation, $|o_i|$ denotes the absolute value of o_i and γ denotes a scalar that maximizes log likelihood functions is given in equation 7 along a grid of values in a range of interest under the normality assumption.

$$L = \frac{n}{2}\{-2\ln(2\pi) - n\ln MSE_w + \sum \ln w_i - (n-k)\} \tag{7}$$

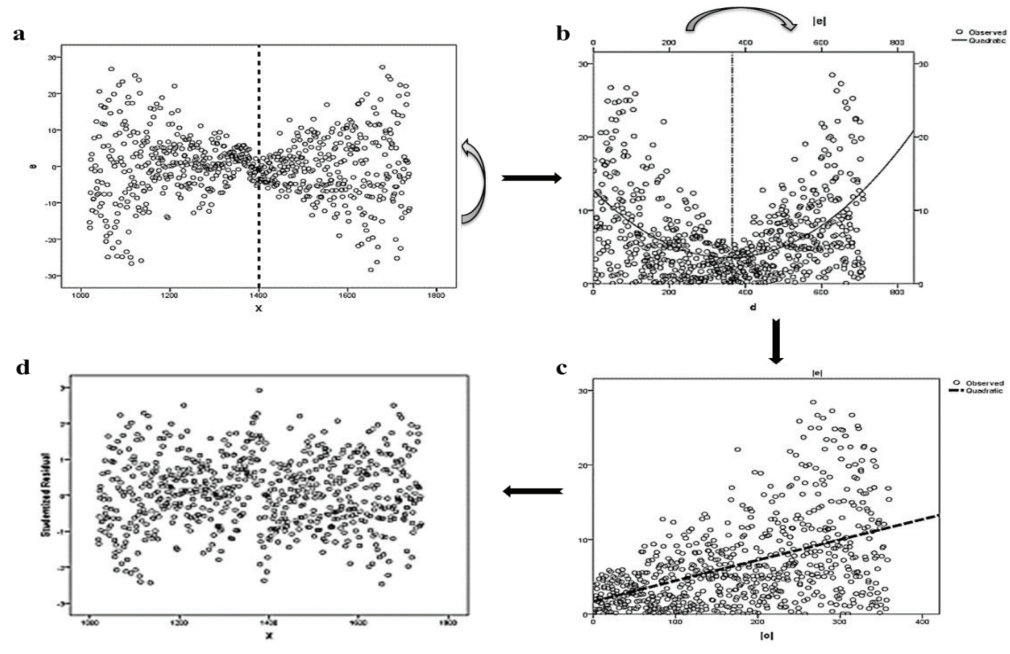

Figure 1. Graphically WCEV.

Where, MSE_w denotes the mean square error as in equation 9, n denotes sample size, and k is the number of parameters to be estimated. Finally weighted least squares are performed as in equation 8**.**

$$\hat{\beta}_w = \left(X'\widehat{W}X\right)^{-1}X'\widehat{W}Y \tag{8}$$

Where, $\hat{Y} = X\hat{\beta}_w$.

$$MSE_w = \frac{\sum(Y-\hat{Y})^2}{n-k} \tag{9}$$

$$S_{\hat{\beta}} = \sqrt{MSE_w diag\left(X'\widehat{W}X\right)^{-1}} \tag{10}$$

As Stated by Son and Kim [44], WCEV is superior to other robust estimation methods because WCEV provides the less auto-correlated residuals, smaller CV and RMSE than other traditional estimation methods. Since WCEV does not violate regression assumptions and is applicable to data with various heteroscedasticity types, WCEV can be practically used in correcting heteroscedasticity problems. In particular, WCEV provides more robust estimations when the variance of residual is a function of explanatory variable whose grid values are not smooth. Figure 1 shows how the WCEV method works.

RCEV Heteroscedasticity Test Based on the Studentized Residuals (Studentized RCEV) Test

The studentized residuals have a mean value of 0 and a variance of 1 [45]. For the regression model in equation 1 OLS parameter estimates, fit values, raw residuals and are obtained by matrix form follows as:

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{11}$$

$$\hat{Y} = X\hat{\beta} = \overbrace{X(X'X)^{-1}X'}^{H}Y \tag{12}$$

$$\hat{\varepsilon} = Y - \hat{Y} = Y - HY = (I-H)Y \tag{13}$$

The vector of the studentized residuals r is obtained by the follow equation:

$$r_i = \frac{\hat{\varepsilon}_i}{RMSE\sqrt{1-h_{ii}}} \tag{14}$$

where h_{ii}'s are diagonal elements of H and $RMSE = \sqrt{\frac{1}{n-k}\hat{\varepsilon}'\hat{\varepsilon}}$ where k denotes the number of parameters to be estimates and n denotes the number of observations in the model. After obtaining studentized residuals, then RCEV studentized test is performed for auxiliary regression:

$$|r_i| = \sigma_0 + \gamma|o_i| \tag{15}$$

where a denotes intercept, γ denotes the slope parameter (Figure 1-c), and $|o_i|$ denotes absolute transformed centered external variable as in equation 5.

Now the interest is whether slope of Equation 15 $\gamma = 0$ or not. If $\gamma = 0$, $\hat{\sigma}_i = \hat{\sigma}_0 = 1 \rightarrow \sigma_i^2 = \hat{\sigma}_0^2 = 1$ residuals are homoscedastic, otherwise residuals are heteroscedastic.

Hypotheses:

$H_0: \gamma = 0 \rightarrow \sigma_i = \sigma_0 \rightarrow \sigma_i^2 = \sigma_0^2 = 1$ (i = 1,.., n) $\rightarrow$ Homoscedasticity

$H_a: \gamma \neq 0 \rightarrow \sigma_i = \sigma_0 + \hat{\gamma}|o_i| \rightarrow \sigma_i^2 \neq 1$ (i = 1,.., n) $\rightarrow$ Heteroscedasticity

Significance of γ is tested by the t-test by:

$$t = \frac{\hat{\gamma}}{s_{\hat{\gamma}}} \sim t_{(n-2)} \tag{16}$$

In this test, where $\hat{\gamma} = \frac{\sum|r|_i|o_i| - n|r|\,|o|}{\sum|o_i|^2 - n\overline{|o|}^2}$; $s_{\hat{\gamma}} = \frac{RMSE}{\sqrt{(|o_i|-\overline{|o|})^2}}$ and $RMSE = \sqrt{\frac{\sum(|r_i|-|\hat{r}_i|)^2}{n-2}}$.

For $|t| \leq t_{(1-\alpha/2;(n-2))}$ or $(Pr > |t|) > \alpha/2$ homoscedasticity assumption is valid and $\alpha/2$ homoscedasticity assumption is valid. The flow diagram of the application is given below:

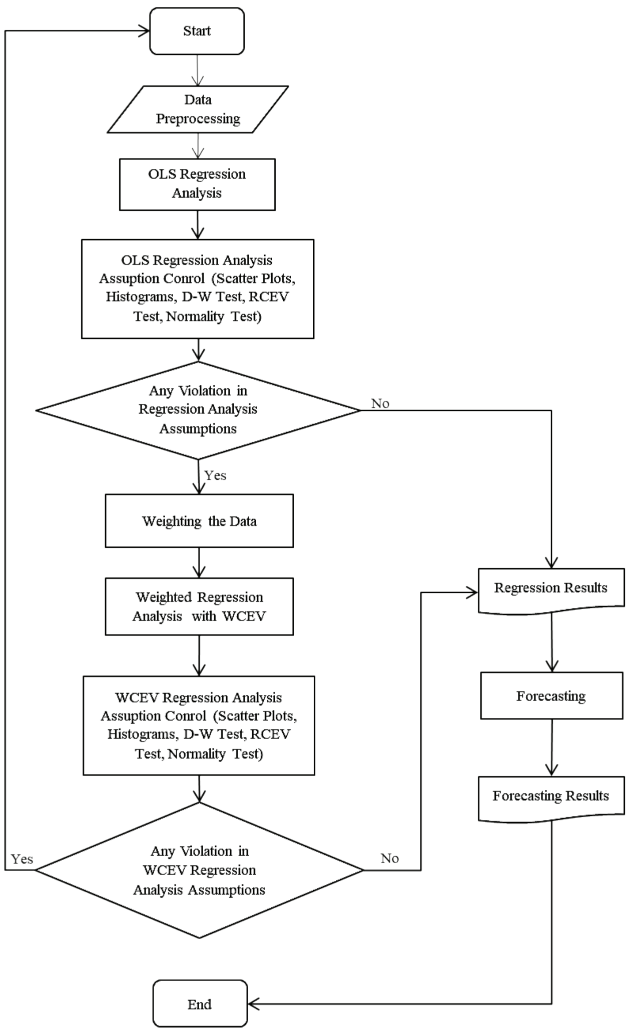

Figure 2. Flow diagram of application.

Our study contains a forecasting model in which the months are independent variables with a trend variable and the total number of air passengers per month is the dependent variable. We also used a linear regression model to forecast monthly total passengers from February 2017 to December 2018. Specifically, the approach recently proposed by Çelik [35] to solve heteroscedasticity problem in linear regression models and to obtain a robust model.

RESULTS AND DISCUSSION

In this study, at the first step, we built an Ordinary Least Squares (OLS) regression model. Since regression assumptions are not provided, the WCEV weighting method is applied. According to the model used in this study, independent variables refer to months (**X**) and dependent variable (**Y**) refers to total monthly air passengers. The equation is given in (17).

$$E(Y_i) = \beta_0 + \sum_{j=1}^{11} \beta_j X_{ji} + \tau Trend_i \qquad i = 1, \dots, n \quad (17)$$

Figure 3 represents the variation in the total monthly number air of passengers of Turkey from January 2007 to December 2018. The total numbers of passengers increased year by year. When we examined the data, we realized that the number of air passengers demanding flight seemed to be accelerated in the months of October, November, and December. Parameter estimations and statistics of OLS and WCEV are given in Table 1. In our model, we avoided a dummy variable trap by selecting December as the reference month. For this reason, β coefficient of the November is not significant with respect to the reference month. The coefficient of the other variables and trends are significant.

The adjusted R^2 of the model was 92.01%. The significance of the dummy variables was evaluated according to the reference month. The variations in the accumulated number of passengers per-month are significant, except for November, which that means the mobility of number of air passengers are the same in November and December in Turkey.

From the histogram of Figure 4 (left side), it can be concluded that, OLS provides the normality of the residuals. According to the studentized RCEV test of the OLS model in Table 1, and scatter plot of the residuals against month (Figure3 right side) homoscedasticity assumption is violated. Scatter plots of studentized residuals vs months indicates butterfly distributed residuals pattern, which is a particular type of heteroscedasticity.

Graphical analysis of the residuals demonstrates that the normality assumption is provided (Figure 4). According to the studentized RCEV test, it was found that the homoscedasticity assumption was ensured and the butterfly distributed residuals problem is eliminated. Also, the distribution of residuals indicates that the weighted model's residual distribution is closer to the normal distribution compared to OLS residuals. Monthly and yearly increases demonstrate that the weighted model can be more efficient and so, forecasting results can be more accurate.

Forecasting Results

Demand forecasting is a major topic for strategic planning and policy making [15]. A lot of paper modeled air transport demand based on demographic factors, socio-economic and competition factors [14]. Differently, our study focuses on the data of passengers who are demanding flights monthly. Thus, other variables are shown with the trend and error variables in the model. Additionally, all air passenger data analyzed instead of a single airport data. In this way, it is aimed to formulate the general lines of aviation policies more easily.

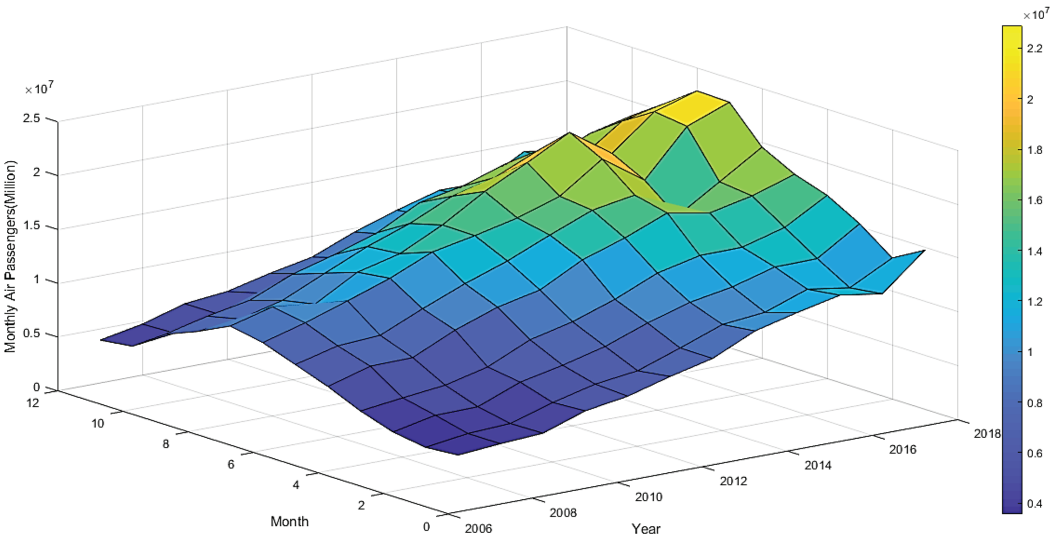

Figure 3. Total monthly passengers according to months of years

Table 1. Regression analyses results of OLS and WCEW

	The OLS Regression Model		**WCEV Regression Model**	
Variables	**β Coefficient**	**S.E.**	**β Coefficient**	**S.E.**
(Constant)	86.328***	4.548	85.279***	6.074
January	-107.064***	5.636	-107.227***	7.868
February	-101.276***	5.635	-101.578***	7.460
March	-94.611***	5.634	-95.090***	7.049
April	-85.788***	5.634	-86.464***	6.641
May	-73.193***	5.634	-73.286***	6.242
June	-62.400***	5.783	-61.448***	5.985
July	-47.881***	5.634	-48.524***	6.198
August	-34.319***	5.634	-34.816***	6.597
September	-25.828***	5.634	-25.68***	7.006
October	-15.232**	5.635	-15.142**	7.417
November	-4.911 (p=0.385)	5.636	-5.035 (p=0.521)	7.826
trend	6.136***	0.399	6.176***	0.244
Studentized RCEV Test Results	$\hat{\sigma}_0 = 0.1622524, \quad s_{\hat{\sigma}_0} = 0.100709$		$\hat{\sigma}_0 = 0.7122956, \quad s_{\hat{\sigma}_0} = 0.106023$	
	$\hat{\gamma}=0.0203116, \quad s_{\hat{\gamma}}=0.0029368$		$\hat{\gamma} = 0.003267, \quad s_{\hat{\gamma}} = 0.00030539$	
	$\left\lvert t_{\hat{\gamma}} \right\rvert=6.916, \quad t_{116;0.975}=1.980626$		$\left\lvert t_{\hat{\gamma}} \right\rvert = 1.06083, \quad t_{116;0.975} = 1.980626$	
	P value < 0.000		**P value**= 0.144	
	Decision: Heteroscedastic residuals		**Decision:** Homoscedastic residuals	

Analysis of Variance Results: F=122.93<0.000. Significance: *p<0.10, **p<0.05, ***p<0.001.

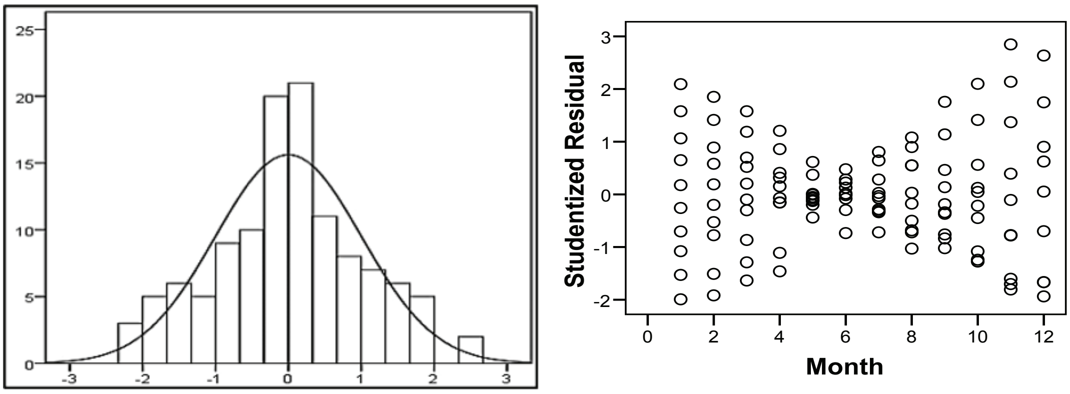

Figure 4. Histogram plot and studentized residuals and scatter plot of studentized residual vs. months of OLS.

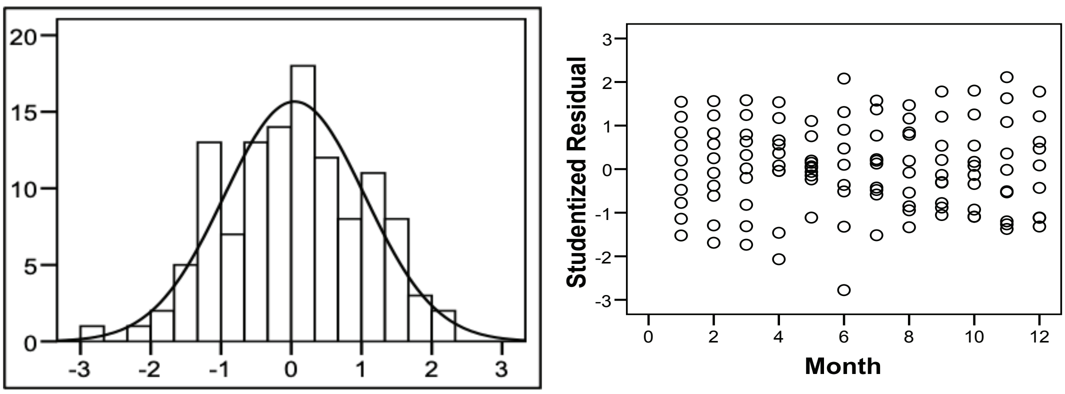

Figure 5. Histogram plot and studentized residuals and scatter plot of studentized residual vs months of the WCEV.

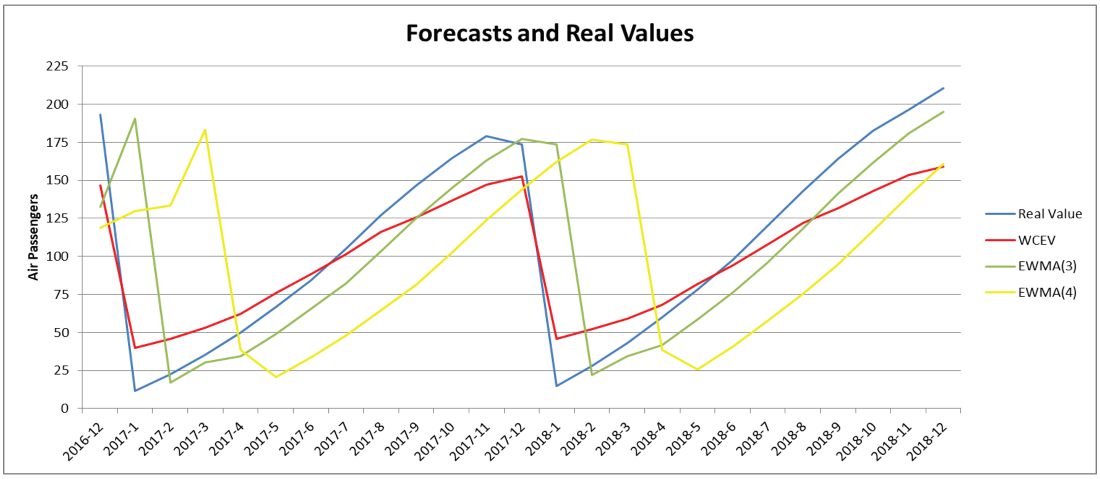

Figure 6. Forecast results of the EWMA models for comparison with WCEV.

Table 2. Comparison of forecast results

Date	Real Value	WCEV	EWMA(3)	EWMA(4)
2016-12	193.045	146.535 [143.418; 149.651]	132.627[125.131; 140.123]	129.896[122.400; 137.392]
2017-1	11.495	39.803 [36.686; 42.920]	190.247[182.751; 197.743]	133.432[125.937; 140.928]
2017-2	22.528	46.007 [42.890; 49.124]	17.074[9.578; 24.569]	183.173[175.677; 190.669]
2017-3	35.355	52.989 [49.872; 56.106]	30.336[22.840; 37.832]	38.623[31.127; 46.119]
2017-4	50.123	62.107 [58.990; 65.224]	34.262[26.766; 41.758]	20.951[13.455; 28.447]
2017-5	66.866	75.838 [72.721; 78.955]	48.859[41.363; 56.355]	33.521[26.025; 41.017]
2017-6	83.963	88.177 [85.60; 91.294]	65.423[57.927; 72.919]	48.010[40.514; 55.506]
2017-7	105.257	101.628 [98.511; 104.745]	82.414[74.918; 89.910]	64.461[56.965; 71.957]
2017-8	126.968	115.9 [112.783; 119.017]	103.499[96.003; 110.995]	81.528[74.032; 89.024]
2017-9	146.832	125.532 [122.415; 128.649]	124.998[117.502; 132.494]	102.198[94.702; 109.694]
2017-10	164.577	136.566 [133.449; 139.683]	144.928[137.433; 152.424]	123.836[116.340; 131.332]
2017-11	178.851	147.229 [144.112; 150.346]	162.855[155.359; 170.350]	143.965[136.469; 151.461]
2017-12	173.744	152.715 [149.598; 155.832]	177.385[169.889; 184.881]	162.007[154.511; 169.503]
2018-1	14.749	45.983 [42.866; 49.100]	173.324[165.828; 180.820]	176.696[169.200; 184.192]
2018-2	27.818	52.187 [49.070; 55.304]	22.264[14.768; 29.760]	173.658[166.162; 181.154]
2018-3	43.121	59.169 [56.052; 62.286]	34.501[27.005; 41.997]	38.525[31.029; 46.021]
2018-4	60.057	68.287 [65.170; 71.404]	41.822[34.326; 49.318]	25.951[18.455; 33.447]
2018-5	78.056	82.018 [78.901; 85.135]	58.581[51.085; 66.077]	40.930[33.434; 48.426]
2018-6	97.623	94.357 [91.240; 97.474]	76.456[68.961; 83.952]	57.628[50.132; 65.124]
2018-7	120.281	107.808 [104.691; 110.925]	95.903[88.407; 103.399]	75.477[67.981; 82.973]
2018-8	143.066	122.08 [118.963; 125.197]	118.347[110.851; 125.843]	94.826[87.330; 102.322]
2018-9	163.834	131.712 [128.595; 134.824]	140.985[133.489; 148.481]	117.024[109.528; 124.520]
2018-10	182.543	142.746 [139.629; 145.863]	161.839[154.343; 169.335]	139.780[132.284; 147.275]
2018-11	196.481	153.409 [150.292; 156.526]	180.735[173.239; 188.231]	160.837[153.341; 168.333]
2018-12	210.498	158.895 [155.778; 162.012]	194.986[187.490; 202.482]	179.828[172.332; 187.323]
Correlation Coeff.		0.996***	0.636***	0.001(ns)
T-test for difference		1.498 (df: 24, sig: 0.147)	0.313 (df: 23, sig: 0.757)	0.509 (df: 23, sig: 0.615)
MAD		21.797	30.601	72.785
RMSE		25.718	52.091	82.033

Significance: * p<0.10 ** p<0.05 *** p<0.001. MAD: Mean Absolute Deviation; RMSE: Root Mean Squared Error.

In this section, the WCEV and EWMA methods in the weighted estimation methods are compared. The EWMA method is a control procedure, and it is highly effective in detecting systemic changes in continuous variable processes. The method makes a continuously updated prediction by assigning a certain weight to the current observation values and previous observation values [46]. The method reduces the impact of previously observed data while placing more emphasis on the most recent observations. This makes it faster and more accurate in detecting systemic changes in the process. The EWMA method is particularly useful in many areas of industrial applications, such as quality control and process control. Compared to other control chart methods where the previous observation values are associated only with the latest value, the EWMA method is more precise and faster. Therefore, the EWMA method is a valuable tool for monitoring and controlling continuous variable processes in industrial applications [47].

Figure 6 displays the compatibility between the forecasted values and the real values. The forecasted values are obtained using 3 and 4 lag EWMA models with the WCEV method. As evident from the graph, the forecasting results obtained using the WCEV method is closer to the real values. Further detailed comparison results can be found in Table 2. As Stated by Son and Kim [44], WCEV is superior to other robust forecasting methods because WCEV provides the less auto-correlated residuals, smaller CV and RMSE than other traditional estimation methods.

When the values in the table are examined, it can be observed that the WCEV method provides forecasts with a higher confidence level. The RMSE and MAD values also indicate a lower deviation compared to the EWMA methods. Mean for real values and forecasts are 103.635 and 97.945; standard deviations are 62.286 and 97.945 respectively. The correlation with real and forecast values is very high (0.996). The mean difference of real and forecasted values is not significant as well. Also, the mean absolute deviation and root mean squared error values are smaller than EWMA (3) and EWMA (4). Although, data is highly volatile, successful forecast results are obtained.

CONCLUSION

Given all applications of new techniques for modeling research in the air transportation industry, the smallest improvements can provide a great competitive advantage among players. In this industry, where even the smallest costs are calculated, any contribution is welcomed to reach more accurate demand forecasts. Furthermore, in areas with high investment costs, especially in the aviation sector, modeling errors may cause great economic losses. However, we observed that the simplest, most accurate, and useful models are preferable.

Recently, hybrid and weighted methods have been commonly preferred. However, hybrid methods generally complicate the modeling of the process and make it difficult to interpret the coefficients of the variables in the obtained models. Instead, it would be beneficial to use more practical, understandable, and simple models. The WCEV method is one of these simple methods, which is a weighted regression model.

This study provides a detailed description of the steps followed to develop a weighted econometric model of air travel demand in Turkey. The model analyses the number of air passengers as a function of months try forecasting the number of air passengers by establishing a statistical relationship between the air passengers and months.

To compare the models, whether the model obtained satisfies the statistical criteria were evaluated with the RCEV test and graphically. The model obtained by the WCEV method is excellent in terms of goodness-of-fit criteria; there is no heteroscedasticity and multicollinearity problem. Weighting according to months in the model ensured that the model produces more accurate forecast results than the EWMA methods and all regression assumptions provided. It has been understood that the results are more robust and reveal narrower confidence intervals.

One of the prominent findings of the study is that the passenger mobility increases at the last quarter of each year for Turkey. Furthermore, Turkey's total monthly number of passengers is the lowest in January and February. When this information is taken into account, the necessity of weighting is revealed. As a result, the WCEV method produced better results compared to the classical OLS model.

The second finding of the study is that January or February can be the best time period for the heavy maintenance of airports and airplanes with the lowest air passenger demand in Turkey. Therefore, it is possible to attract quality-sensitive or price-sensitive passengers with developing competitiveness based on technological, operational, engineering, and services capabilities that are based on flight security will be based on this productive time assessment.

In conclusion, the capacity for runway, number of planes, and total service personal can be increased in peak months. Finally, the forecasting results show that the growth of the sector will continue due to the growth in the country. We predict a significant increase in the number of passengers thanks to the construction of new airports, cheap ticket prices, and low-cost carriers. In further studies, we are planning to analyse effect of the increase in air passenger demand and unqualified human resources on the firms' operations and performance. Because we believe all these conditions should be taken into consideration to eliminate the economic loss.

DATA AVAILABILITY

All datasets are available at the General Directorate of the Turkish State Airports Administration (TSAA) statistics website.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Green RK. Airports and economic development. Real Estate Economics 2007;35:91−112. [CrossRef]
[2] Bilotkach V. Are airports engines of economic development? A dynamic panel data approach. Urban Studies 2015;52:1577−1593. [CrossRef]
[3] Brida JG, Rodriguez-Brindis MA, Lanzilotta B, Rodriguez-Collazo S. Testing linearity in the long-run relationship between economic growth and passenger air transport in Mexico. Int J Transport Econ 2016;43.
[4] Chi J, Baek J. Dynamic relationship between air transport demand and economic growth in the United States: A new look. Transport Policy 2013;29:257−260. [CrossRef]
[5] Abrahams M. A service quality model of air travel demand: an empirical study. Transport Res A Gener 1983;17:385−393. [CrossRef]
[6] Kulendran N, Wilson K. Is there a relationship between international trade and international travel? Appl Econ 2000;32:1001−1009. [CrossRef]
[7] Gallet CA, Doucouliagos H. The income elasticity of air travel: A meta-analysis. Ann Tourism Res 2014;49:141−155. [CrossRef]
[8] Abrate G, Viglia G, García JS, Forgas-Coll S. Price Competition within and between Airlines and High-Speed Trains: The Case of the Milan-Rome Route. Tourism Econ 2016;22:311−323. [CrossRef]
[9] Hakim MM, Merkert R. Econometric evidence on the determinants of air transport in South Asian countries. Transport Policy 2019;83:120−126. [CrossRef]
[10] Kim S, Shin DH. Forecasting short-term air passenger demand using big data from search engine queries. Automat Construct 2016;70:98−108. [CrossRef]
[11] Kurdel P, Sedláčková AN, Novák A. Analysis of using time series method for prediction of number of passengers at the airport. J KONBiN 2020;50:203−216. [CrossRef]
[12] Bermúdez JD, Segura JV, Vercher E. Holt-Winters forecasting: an alternative formulation applied to UK air passenger data. J Appl Stat 2007;34:1075−1090. [CrossRef]
[13] Wang M, Song H. Air travel demand studies: a review. J China Tourism Res 2010;6:29−49. [CrossRef]
[14] Wang S, Gao Y. A literature review and citation analyses of air travel demand studies published between 2010 and 2020. J Air Transport Manag 2021;97:102135. [CrossRef]
[15] Banerjee N, Morton A, Akartunalı K. Passenger demand forecasting in scheduled transportation. Eur J Oper Res 2019;285:797−810. [CrossRef]
[16] Wang M, Song H. Air travel demand studies: a review. J China Tourism Resh 2010;6:29−49. [CrossRef]
[17] Pitfield D. Predicting air-transport demand. Environ Plan A 1993;25:459−466. [CrossRef]
[18] Gelhausen MC, Berster P, Wilken D. A new direct demand model of long-term forecasting air passengers and air transport movements at German airports. J Air Transport Manag 2018;71:140−152. [CrossRef]
[19] Karlaftis MG. Demand forecasting in regional airports: dynamic Tobit models with GARCH errors. Sitraer 2008;7:100−111.
[20] Ming W, Bao Y, Hu Z, Xiong T. Multistep-ahead air passengers traffic prediction with hybrid ARIMA-SVMs models. Sci World J 2014;2014. [CrossRef]
[21] Carson RT, Cenesizoglu T, Parker R. Forecasting (aggregate) demand for US commercial air travel. Int J Forecast 2011;27:923−941. [CrossRef]
[22] Xiao Y, Liu JJ, Xiao J, Hu Y, Bu H, Wang S. Application of multiscale analysis-based intelligent ensemble modeling on airport traffic forecast. Transport Lett 2015;7:73−79. [CrossRef]
[23] Xiong H, Fan C, Chen H, Yang Y, Antwi CO, Fan X. A novel approach to air passenger index prediction: Based on mutual information principle and support vector regression blended model. SAGE Open 2022;12:215824402110711. [CrossRef]
[24] Tang H, Yu J, Lin B, Geng Y, Wang Z, Chen X, et al. Airport terminal passenger forecast under the impact of COVID-19 outbreaks: A case study from China. J Build Eng 2023;65:105740. [CrossRef]
[25] Jin F, Li Y, Sun S, Li H. Forecasting air passenger demand with a new hybrid ensemble approach. J Air Transport Manag 2020;83. [CrossRef]
[26] Abed SY, Ba-Fail AO, Jasimuddin SM. An econometric analysis of international air travel demand in Saudi Arabia. J Air Transport Manag 2001;7:143−148. [CrossRef]
[27] Aderamo AJ. Demand for air transport in Nigeria. J Econ 2010;1:23−31. [CrossRef]

[28] Baikgaki OA. Determinants of domestic air passenger demand in the republic of South Africa. Doctorial Thesis. 2014. [CrossRef]
[29] Sivrikaya O, Tunç E. Demand forecasting for domestic air transportation in Turkey. Open Transport J 2013;7. [CrossRef]
[30] Naghawi H, Alobeidyeen A, Abdel-Jaber M. Econometric modeling for international passenger air travel demand in Jordan. Jordan J Civil Eng 2019;13.
[31] Xiao Y, Liu Y, Liu JJ, Xiao J, Hu Y. Oscillations extracting for the management of passenger flows in the airport of Hong Kong. Transport A Transport Sci 2016;12:65–79. [CrossRef]
[32] Franses PH. Seasonality, non-stationarity and the forecasting of monthly time series. Int J Forecast 1991;7:199–208. [CrossRef]
[33] Karlaftis MG, Zografos KG, Papastavrou JD, Charnes JM. Methodological framework for air-travel demand forecasting. J Transport Eng 1996;122:96–104. [CrossRef]
[34] Karlaftis MG, Vlahogianni EI. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. Transport Res C Emerg Technol 2011;19:387–399. [CrossRef]Çelik R. Correcting double outward box distributed residuals by WCEV. Commun Stat Theory Methods 2017;46:9566–9590. [CrossRef]
[35] Çelik R. Correcting double outward box distributed residuals by WCEV. Commun Stat Theory Methods 2017;46:9566–9590. [CrossRef]
[36] Çelik R. Stabilizing heteroscedasticity for butterfly-distributed residuals by the weighting absolute centered external variable. J Appl Stat 2015;42:705–721. [CrossRef]
[37] Çelik R. RCEV heteroscedasticity test based on the studentized residuals. Commun Stat Theory Methods 2019;48:3258–3268. [CrossRef]
[38] Deveci M, Demirel NÇ, Ahmetoğlu E. Airline new route selection based on interval type-2 fuzzy MCDM: A case study of new route between Turkey-North American region destinations. J Air Transport Manag 2017;59:83–99. [CrossRef]
[39] Draper NR, Smith H. Applied regression analysis. vol. 326. Hoboken: John Wiley & Sons; 1998. [CrossRef]
[40] Gujarati DN. Basic econometrics. India: Tata McGraw-Hill Education; 2009.
[41] Carapeto M, Holt W. Testing for heteroscedasticity in regression models. J Appl Stat 2003;30:13–20. [CrossRef]
[42] Bischoff W, Heck B, Howind J, Teusch A. A procedure for estimating the variance function of linear models and for checking the appropriateness of estimated variances: a case study of GPS carrier-phase observations. J Geodesy 2006;79:694–704. [CrossRef]
[43] Carroll RJ, Ruppet D. Transformation and Weighting in Regression. New York: Chapman and Hall; 1988. [CrossRef]
[44] Son N, Kim M. A study on robust regression estimators in heteroscedastic error models. J Korean Data Inform Sci Soc 2017;28:1191–1204. [CrossRef]
[45] Ott RL, Longnecker M. An Introduction to Statistical Methods and Data Analysis. 6th ed. Australia; United States: Brooks/Cole; 2008.
[46] Montgomery DC. Introduction to statistical quality control. 6th ed. Hoboken, N.J: Wiley; 2009.
[47] Sukparungsee S, Areepong Y, Taboran R. Exponentially weighted moving average-Moving average charts for monitoring the process mean. PLoS One 2020;15:e0228208. [CrossRef]