مجلة البحوث المالية
المجلد (23) – العدد الأول – يناير 2022

The improvement of Forecasting accuracy Egyptian tourism demand using combining statistical and Judgmental forecasts

BY

Azza Elshahat Mohammed

Faculty Of Commerce Port Said University

Supervised by

Prof. Bahgat Thabet
Professor Emeritus
Department of Statistics,
Mathematics & Insurance
Faculty of Commerce
Port Said University

Dr. Samah Kamal
Lecturer in Statistics
Department of Statistics,
Mathematics & Insurance
Faculty of Commerce
Port Said University

رابط المجلة: https://jsst.journals.ekb.eg/

Abstract

This study aims to examine the forecasting accuracy of a combined statistical forecasts and expert judgments method by using annually tourism arrival data in Egypt in period from 1993-2017, using an econometric model, the autoregressive distributed lag model - error correction model (ARDL-ECM), statistical forecasts adjusted by the Delphi experts.

Over the forecasting period of 2018 -2022, the combined forecasts outperform the baseline forecasts produced by (ARDL- ECM) models, demonstrating the value of implementing this integration procedure , and determine if the adjusted forecasts are unbiased. Various error measures such as (APE), (MAPE), and (RMSPE) are used to determine forecasting effectiveness, and statistical tests were performed to evaluate forecast accuracy using the Delphi method and a range of expert judgment adjustments to integrate statistical forecasts and expert judgments.

Several forecasting models are compared as part of the study to evaluate the performance of the combined method by examining the statistical and judgmentally adjusted forecasts with regression analysis we were able to determine whether or not they were biased. Based on the hypothesis tests, it was concluded that the Delphi panel adjustments increased forecast accuracy. However, for some sample markets, the group-adjusted forecasts were biased. There are several advantages to integrating expert judgments into statistical forecasts; however, combined forecasting does not always produce satisfactory results, especially when historical information is unavailable.

Keywords: Tourism forecasts, Statistical Forecasts, Judgmental Forecasts, Integrating ,ARDL- ECM, Delphi method, accuracy, bias.

Introduction

Methods for integrating judgment and statistics are investigated, quantitative studies of forecasting rare in comparison with the prevalence of quantitative forecasting methods each quantitative and judgmental forecasting methods have strengths and weaknesses, A lack of contextual information in the forecasting and combination process has led to poor accuracy due to the integration of statistical forecasts.

In this study, we apply an econometric model for estimation of Egypt tourism demand data using the autoregressive distributed lag error correction model (ARDL-ECM) .In Egypt, annual forecasts of visitor arrivals were developed from three source markets (Germany, Saudi, and the UK) .The Delphi technique was used to incorporate experts' domain knowledge into the statistical forecasts.In both quantitative and qualitative analyses, the accuracy, bias, and efficiency of statistical, judgmental, and expert forecasts were analyzed.

A test of Hypotheses for the study was conducted by comparing accuracy across three different Delphi rounds, source markets, and regression analyses, in the three markets studied, statistical forecasts adjusted by Delphi experts improved forecast accuracy on average. Compared to the original statistical model and the simple average of individual experts' responses, the consensus group forecast in the final round of the Delphi survey provided materially more accurate forecasts.

It was discovered that some of the group forecasts were biased and inefficient ,but overall they achieved satisfactory accuracy, The two types of forecast need to be integrated in order to make better tourism demand forecasts, because everything we plan is dependent on demand forecasts, Forecast accuracy has a direct impact on procurement, production, and delivery efficiency, as well as increasing the effectiveness of the sales forecasting process, As explained by the experts interviewed, this study found numerous reasons for an improvement in accuracy.

Quantitative methods are more accurate than qualitative approaches when enough historical data is available You should choose quantitative methods that take into account how much the change is

(large or small), what type of data you have (cross-sectional or time-series), and your prior knowledge about the future relationships.

(1) Problem of research:

Forecasting represents one of the main goals in any country for policy making and planning ,In an attempt to increase forecast accuracy ,the two methods of quantitative and judgmental method are combined, Hence the topic of this study and its interest ,which is the combination of quantitative and judgmental methods to predict tourism demand in Egypt.

a. The necessity and need for integration

Incorporating judgmental forecasts into statistical forecasts of tourism demand would improve their accuracy. Using this integration, policymakers in the Egyptian tourism industry are expected to be able to make better and more reliable decisions, as there will be reduce the risk of forecasting failures.

Statistical methods have a number of significant advantages over judgments.

- **When using heuristics, speed and efficiency usually prevail over accuracy.**
- **In forecasting, judgment plays a significant role, Bias may be present when forecasters are evaluating or forecasting outcomes.**
- **In addition to suffering from biases inherent in judgmental forecasting methods, formal quantitative methods struggle when past data is scarce, and also suffer from major problems when handling special events or significant changes in the environment, such as the introduction of new government policies.**

In addition, there is evidence that the forecast accuracy increases when special events are taken into account when the statistical forecast is adjusted, Statistical methods allow managers and forecasters to access a large amount of data and a consistent way of handling these data.

b. The need for tourism demand forecasting

To increase the accuracy of forecasting through a combination of quantitative and qualitative methods. Tourism having accurate demand forecasts allows planning to start as early as possible to the next, Many organizations public and private sectors use tourism demand forecasts to improve the efficiency of their decision-making

process ,Tourism demand forecasts can lead to greater business efficiency, profit growth, and a stronger economy.

c. Tourism demand forecasting characteristics

d. A specific set of challenges affects tourism demand forecasters and practitioners that do not apply to those in other industries such as (External interventions, Complexity of tourism behavior, Measures of tourism demand, Methods and models of tourism demand forecasting).

e. Forecasting tourism demand with methods and models

A forecasting method simply involves organizing data in order to predict future events while a forecasting model is "an expression of a forecasting method".There are two categories of forecasting methods in tourism studies, qualitative and quantitative. These methods (also known as judgmental methods, as the study has done), help project developments into the future. Expert judgments or opinions are used rather than mathematical rules to organize past information on the forecast variable of interest .

Quantitative methods use mathematical rules to quantify past information about a phenomenon by exploiting the underlying patterns and relationships, The subtypes of these methods consist of time series (or explanatory, noncausal) methods and econometric (or causal) methods.

In a Delphi process, expert judgments are sought in response to a given forecasting problem from a group of trained and knowledgeable people.

(2) Importance of the research:

The study is significant for the following reasons:

a. In this research, two methods were combined as an attempt to improve the accuracy of forecasting the demand for Egyption tourism and fewer forecast errors.

b. The research indicates that combining forecasts is more effective than selecting a single forecasting model to fit all practitioners. The mechanical integration of judgmental and statistical forecasts provides better results than judgmental adjustment alone.

c. Research using Forecasting based on historical and current data that allows reasonable and measurable targets .

(3) Objective of the research

Research 's main objective is to improve our ability to forecast Egypt's tourist demand more accurately,and this was done through the following sub objectives :

1. **To test the effectiveness of combining judgmental and statistical Egyption tourism demand forecasting.**
2. **To examine the accuracy and bias of expert estimates of tourism demand based on group judgmental adjustments.**
3. **To develop a research framework for the integration of statistical forecasts and judgmental forecasts based on Egypt tourism demand data.**
4. **To provide recommendations, suggestions to decision makers in Egypt, both public and private, on how to utilize judgmental forecasting models and approaches in the tourism industry.**

(4) Literature review

1. <u>Paul Goodwin (1999)</u>

<u>Judgmental forecasts of time series affected by special events: does providing a statistical forecast improve accuracy?</u>

There are times when patterns of marketing or sales are influenced by exogenous factors, such as sales promotions. A statistical forecast follows the principle of extrapolating regular patterns from series; however, judgmental forecasts take into account events such as external influences, which can happen too seldom for statistical estimation.

Researchers analyzed how judgmental forecasters used statistical time series methods when encountering sporadic special events. In several conditions, varying the complexity, the degree of noise, the salience of the cue, the predictive value of the statistic and availability of the statistical forecast led to better accuracy of the statistical forecast in some instances, but judgmental forecasters did not use personal judgment optimally. When a statistical forecast was highly reliable, it was adjusted; if it was not, it can be adjusted.

2. <u>Philip Hans Franses (2010)</u>

<u>Do experts' adjustments on model-based SKU-level forecasts improve forecast quality?</u>

This study aims to evaluate test (i) if expert forecasts are distinct from model forecasts and (ii) whether expert forecasts are more accurate than model forecasts, a statistical methodology is developed to compare expert forecasts to model forecasts. When experts prepare their forecasts, they frequently consult the model forecasts. This study examines whether expert forecasts differ from model forecasts systematically and significantly, and whether this would also lead to better forecast accuracy.

It uses autoregressive dynamics in his statistical methodology to compare model forecasts with expert predictions. The authors analyzed 35 countries' experts who adjusted SKU-level forecasts for pharmaceutical items in seven different categories. They found that expert forecasts were better at best than model-based forecasts, but were usually lower than model-based forecasts.

Despite the fact that forecasts can differ significantly from model forecasts, the study points out that forecast gains are not large. Positively, whenever they do better, this is mostly due to their adjustment. Furthermore, it examines whether experts may exert too much influence over forecasts, and finds overwhelming evidence that this hypothesis is correct.

3. Vera Shanshan Lin (2013)

Improving Forecasting Accuracy by Combining Statistical and Judgmental Forecasts in Tourism

The objective of this study is to determine whether a combined method is accurate in forecasting tourism arrivals in Hong Kong. In a Web-based Tourism Demand Forecasting System (TDFS), a Delphi method is used to integrate statistical forecasts and expert judgments. The forecasting accuracy is determined by calculating the absolute percentage error (APE), mean absolute percentage error (MAPE) and root mean square percentage error (RMSPE).

In addition, the forecast performance of the combined method is compared to other forecasting models, such as the Nave model, exponential smoothing, and the Box-Jenkins time-series model. Over the forecasting period 2008Q1-2011Q4, the combined forecasts showed a significant improvement over the basic forecasts derived from VARs. This suggests the value of applying this integration procedure.

Statistical forecasts can benefit from incorporating expert judgments.

4. <u>Vera Shanshan Lin et,All (2014)</u>

<u>Accuracy and bias of experts adjusted forecasts</u>

This study examines if expert-based econometric forecasts of tourism demand improved by group-based judgmental adjustments, and whether the adjustments. In addition to the Delphi method, several statistical tests were carried out to determine expert judgmental adjustments used to see how accurate the forecasts were. In order to determine whether forecasts were unbiased,

we used regression analysis. Although the adjusted Delphi panel forecasts were found to be more accurate in general, independent bias was found in many individual markets when the group-adjusted forecasts were evaluated. An in-depth interview with the Delphi panellists shed more light on the bias that accompanied the Delphi surveys.

5. <u>Teerada Khamphinit,et all,.(2015)</u>

<u>Combining Qualitative and Time Series Forecasting to Increase the Forecasting Accuracy for Instant Noodle Sales in Thailand.</u>

The purpose of this case study about maximizing the accuracy of sales forecasts for instant noodles in Thailand. According to the research, instant noodles' sales patterns were adjusted to improve sales forecasts based on past events , A selected instant noodle brand was studied for 48 months through sales data. Second, A test set comprised of the first 36 months of data, and a verified set comprised of the rest based on the most recent twelve months.

Graphed the data to analyze Checking whether any factors such as promotional offers, capacity expansions, or natural disasters distorted the data and the pattern of the data. After obtaining the forecast data, the best time-series method for calculating the 36-month actual and adjusted data was determined. To forecast actual and adjusted sales using time series forecasting methods, the Crystal Ball software was used. Further, actual sales data was used to verify the results.

The (MAPE) of the forecast is a measure of forecast accuracy, accuracy affects, and accuracy influences the efficiency of procurement, production, and delivery processes. Processes. an

improvement in accuracy of 46.14 %, 22.53%, and 56.42%, respectively. The mean average percentage error in sales forecasting was 6.07%-11.62% after the adjustment, This method is therefore significantly more effective than the existing approach.

6. <u>Vera Shanshan Lin (2018)</u>

<u>Judgmental adjustments in tourism forecasting practice: How good are they?</u>

This study Through examining judgemental forecasting procedures, analysing the practices of tourism professionals, and evaluating their judgment based on Hong Kong visitor arrivals forecasts from 2011Q2 to 2015, this study emphasizes the accuracy of judgmental forecasting procedures. Tourism which collected and combined experts' forecasts. Statistical tests and error measures are used in this study to evaluate forecasting performance and explore the determinants of judgmental adjustment behavior.

According to the results, forecast accuracy is positively for data with more uncertainty, forecast accuracy is positively correlated with degree of variation, and expert adjustments are particularly helpful in terms of improving forecast accuracy.

(5) Hypotheses of research

A number of research hypotheses were developed with the purpose of achieving the objectives;

1- An improvement in the accuracy of judgmental forecasts based on statistical forecasts

- **The null hypothesis H_0: Delphi-based judgmental adjustments to statistical forecasts improve the accuracy of tourism forecasts.**
- **The alternative hypothesis H_1:Delphi-based judgmental adjustments to statistical forecasts unimproved the accuracy of tourism forecasts.**
- **Accept the Null hypothesis , reject the alternative hypothesis**

2- Naive forecasts do not perform as well as judgmentally adjusted forecasts.

- **The null hypothesis H_0: Delphi On average, judgmentally adjusted forecasts are more accurate than Naïve forecasts.**
- **The alternative hypothesis H_1 : Delphi On average, judgmentally adjusted forecasts are less accurate than Naïve forecasts.**

- **Accept the Null hypothesis , reject the alternative hypothesis**

3- Forecasts adjusted for judgment are biased.

- **The null hypothesis H_0: Delphi-based judgmentally adjusted forecasts of tourism demand are biased.**
- **The alternative hypothesis H_1:Delphi-based judgmentally adjusted forecasts of tourism demand are unbiased.**
- **Reject the Null hypothesis , accept the alternative hypothesis**

4- Forecasts based on judgemental adjustments are inefficient.

- **The null hypothesis H_0: Judgmentally adjusted forecasts are inefficient.**
- **The alternative hypothesis H_1 : Judgmentally adjusted forecasts are efficient.**
- **Accept the Null hypothesis , reject the alternative hypothesis**

5 - Forecast accuracy improves via the Delphi approach.

- **The null hypothesis H_0: Final Delphi forecasts are more accurate than the average of the statistical group).**
- **The alternative hypothesis H_1 : Final Delphi forecasts are less accurate than the average of the statistical group).**
- **Accept the Null hypothesis , reject the alternative hypothesis**

(6) Limits of study

Visitors from a country or region of origin to a destination is the variable used in the study to measure international tourism demand, The demand model drew on data from a range annual data from 1993 to 2017 were used to forecast demand from 2018 to 2022 using the models estimated in this study.

(7) Egyption tourism demand

Model elements relevant to tourism demand forecasting are proposed, along with criteria for selecting the right model, The forecast is accurate and reliable to assist decision-makers in making more effective and efficient decisions as well as choosing the most suitable method (casual or non-casual) to forecast tourism demand.

In spite of its rapid growth, the tourism industry has been affected by several changes and events, such as human-caused conflicts, natural disasters, and economic crises UNWTO, 2011).

Example 2011 was a year marked by turbulent economic conditions, political uncertainty in the Middle East and North Africa, and a terrible earthquake in Japan. After the global financial/economic crisis of 2008-2009, global tourism continues to recover from the blow of the 2008-2009 global financial/economic crisis and it is estimated that international visitors will exceed 1 billion by 2012, reach close to 1.4 billion by 2020, and be close to 1.8 billion by 2030. There will be an average increase of 43 million international visitors annually between 2010 and 2030, During the period 1995-2010, the number of tourists increased by about 3.9 percent a year (UNWTO, 2011).

Tourism demand has increased significantly due to a significantly increased accommodation capacity, improved tourism facilities, and better facilities to service the tourism sector, Approximately 2.5 million jobs were directly and indirectly associated with the tourism industry. (Egypt SIS, 2012: 4).

- **On 27 September, 2015, tourists approx 2.35 million visited Egypt during the period between January and June 2016, In contrast, 4.8 million tourists visited during the same period last year. The decline, totaling 51.2 percent, There has been a decrease in Russian tourists by 54.9 percent and tourists from the UK by 14.9 percent.**
- **In 2015, the majority of visitors to Egypt came from Eastern Europe, accounting for 37.7%, with Russian tourists making up 67.9% of that number, and 35.1 percent from Eastern Europe, with Germany accounting for 31.2 percent, All 224 passengers and crew aboard a charter flight operated by Russian Metrojet flights which crashed 23 minutes after taking off from Sharm El-Sheikh in Egypt and was en route to St. Petersburg, lost all its passengers, An investigation into the crash has forced Moscow to suspend all flights to Egypt. Sharm el-Sheikh was closed to all flights from and to the UK.**

(8) Study Plan

This will be done through:

Chapter 1 Introduction

Chapter 2 Literature Review

Chapter 3 Statistical Analysis Autoregressive Distributed Lag – Error Correctiom Model (ARDL- ECM)

Chapter 4 Delphi forecasting method
Chapter 5 the applied study
Chapter 6 Conclusion

(9) METHODOLOGY

This method is divided into three stages ,The properties of the variables were examined to determine if they were unit roots, and the bounds test was used to examine their cointegration . Next, we estimate demand with a generalized model

A number of tests were performed to ensure that ARDL-ECMs for every relevant market were correctly identified. Second, we calculated the forecasts of the explanatory variables . On this basis, baseline forecasts of tourist arrivals from each source market were based on the complete ARDL-ECMs. Lastly, Delphi adjustments using various scenarios were carried out to determine different levels the Delphi panel reviewed the baseline forecasts and adjusted them based on the Delphi panel's judgement .

To evaluate forecasting performance and identify characteristics related to judgmental adjustments, a number of error measures and statistical tests are applied in this study. Practitioners and researchers are both better forecasters than academicians.

9.1 Econometric Analysis of Tourism Demand

- **Variables and functionl form**

The purpose of study is to model the demand for Egypt tourism from travelers of a selected origin country/region

This can be expressed as: $VA_{it} = A\, Y_{it}^{B1}\, P_{it}^{B2}\, P_{st}^{B3}\, e_{it}$ **Equ 1**

Where : VA_{it} indicates the number of tourists arriving from a country or region i^th^ to Egypt at time t;Y_{it} is an index of the real GDP from i th origin country/region at time t 2010=100;P_{it} Own-price variable based on the exchange-rate-adjusted consumer price index 2010=100 (CPI) a definition is $P_{it}=(CPI_t^{EG} / Ex^{EG})/(CPI_t^i / EX_t^i)$ at time t where i CPI_t^{EG} and CPI_t^i are the CPI_s for Egypt and i^{th} origin country/region at time t, respectively,P_{ist} is the substitute price variable computed as a weighted index of CPI based on the share of international tourists arriving in each substituent market at time t,that

is, $P_{ist}=\sum_{j=1}^{3}$ (CPI_{jt}/EX_{jt}) W_{jt}^{i} $(j=1,2,3)$,**Representing Jordan, Algeria ,Tunisia respectively ;**W^{i}_{jt} **is calculated as** TVA_{jt}^{i}/($\sum_{j=1}^{3} TVA_{jt}^{i}$)**Representing the number of international visitor arrivals for the** j_{jt}^{th} **country/region at time t, and** TVA_{jt}^{i} **is the visitor arrivals of a substitution country j from origin country/region i at time t , (**e_{it}**) a residual term refers to those factors that have been left out Three dummy variables (D1(D97), D2(2001), and D3(2011)) A dummy variable was included as a way to measure impact on visitors' arrivals The dummy variables assume a value of 1 in the year where they have an effect, and 0 otherwise.**

1. Data sources

We applied data from 1993 to 2017 to estimate the demand models to produce Annul forecasts for 2018-2022; the data of the dependent variable measured by visitor arrivals were collected from visitor arrivals Statistics (Ministry of Tourism), Ministry of interior: (Passports immigration and nationality dministration .Tourism and antiquities police ,Central Bank of Egypt), the selected three origins generated more than 26.78% of the inbound market share in Egypt during 2017 :Germany (14.86%), Saudi Arabia (8.07%), and the UK (3.85%).

CPIs (2010=100) and exchange rates , three competitive destinations of Egypt including Jordan, Algeria ,Tunisia were selected to calculate the substitute prices,the income variable, Y, measured by the real GDP index (2010=100) CPIs (2010=100) and exchange rates , International Financial Statistics (IFS) of the International Monetary Fund (IMF, 2015) as well as official websites of the statistical bureaus or departments of each country and region were used. three competitive destinations of Egypt including Jordan, Algeria ,Tunisia were selected to calculate the substitute prices.

In order to incorporate domain knowledge into statistical forecasts using the Delphi method, a statistical model, the auto autoregressive distributed lag model error correction model (ARDL-ECM), was used. Panel members included government officials, hospitality industry

representatives, and academics from various universities.

Statistical forecasts were integrated with human judgment using both quantitative and qualitative analysis, Our analysis of statistical and judgmental forecasts was based on Accuracy, bias, and efficiency are three dimensions we used the quantitative analysis, In order to test the research hypotheses, employing statistical tests,regression analyses ,examining the values of the error measures.

Equation (1),can be written in logarithm form:

$$LnVA_{it}=\beta_0+\beta_1 lnY_{it}+\beta_2 lnP_{it}+\beta_3 lnP_{st}+dummies+\varepsilon_{it} \qquad \text{Equ 2}$$

Where: β_0 **= lnA ,** ε_{it} **= ln**e_{it}**,** ε_{it}**=ln**e_{it} **(**ε_{it}**~N(0.** σ^2**) and** β_1 **,** β_2 **and** β_3 **are elasticities for income, own prices, and substitute prices, respectively. A positive impact of income level and substitute price on tourism demand is expected for** β_1 **and** $\beta 3 > 0$ **while** $\beta 2 < 0$ **should have a negative effect (due to the own price of tourism being negative).**

In order to evaluate the model, we estimate a conditional ARDL - ECM using Equation (3):

$$\Delta\ln VA_{it}=\alpha_0+\pi_1\ln VA_{i,t-1}+\pi_2\ln Y_{i,t-1}+\pi_3\ln P_{i,t-1}+\pi_4\ln P_{is,t-1}+$$
$$\sum_{j=1}^{p1}\psi_{Qj}\Delta\ln VA_{i,t-j}+\sum_{j=0}^{p2}\psi_{Yj}\Delta\ln Y_{i,t-j}+\sum_{j=0}^{p3}\psi_{Pi,t-j}\Delta\ln P_{i,t-j}+$$
$$\sum_{j=0}^{p4}\psi_{Pis.j}\Delta\ln P_{is,t-j}+\delta_1 D_1+\delta_2 D_2+\delta_3 D_3+\sum_{d-1}^{D}\theta_d Dummies+u_{it} \qquad \text{Equ 3}$$

We have taken into account the time path followed by tourists when making their decisions. For annual data, p = 1. The Akaike Information Criterion (AIC) normally determines the length of lags for a model, according to Song, Witt, and Li (2009). ε_{it} **is the random error whose normally and variance are assumed to be zero, that is,** $\varepsilon_{it} \sim N(0, \sigma^2)$**,**$\Delta$ **is the first difference operator (i.e** $\Delta X_1=X_t-X_{t-1}$ **),**

The above equation describes the short-term dynamic interactions between the tourist arrival and its determinants, the π **coefficients the long-run relationship between the demand and its determinants, when X is zero no long-run relationship exists, F-test is used to determine whether there's a long-run relationship and whether there's no long-run relationship** π **is non zero.**

It is used to test the extent of the variables' inactivity, and after conducting the (ADF) testand P.P test through the two hypotheses test

1- Null hypothesis p-value > 0.05 (the data has a unit root and is non-stationary)
2- Alternative hypothesis p-value <= 0.05, the data does not have a unit root and is stationary).
3- Decision: if the calculated **τ** is > Critical Value , p-value > 0.05 then we not reject the null hypothesis and accept the alternative hypothesis, the data does not have a unit root and is stationary,And vice versa, if the calculated **τ** is less than Critical Value, p-value < 0.05 then we : Reject the null hypothesis the data does not have a unit root and is stationary.

In the ADF the null hypothesis of a unit root was not rejected for all variables test the calculated **τ** is greater than the tabular and p-value more than 0.05, (the null hypothesis is that the data are non-stationary) .

It was only rejected in these cases: VA Visitor arrival in the model in Saudi, own price in Germany are less negative than the table value so we take first deference .

The null hypothesis of a unit root was not rejected for all variables in the PP test P-value more than 0.05 it was only rejected in this case: GDP in UK and substitute price in the UK, we take First deference, First differences rendered all series stationary, with the ADF statistics and PP in all cases being less than the critical values at either the 1% or 5% significance level.

Accordingly, the ADF and P-P test shows that all variables were stationary after the first difference, ADF and PP test conclude that the variables used are integrated order of I (0),I (1), data analysis steps can be performed.

2. Testing for long-run relationships

A long-run relationship between demand and its determinants is delineated through the π coefficients in Equation (4). The null hypothesis states that a long-run relationship doesn't exist if π is zero. For examining long-run relationships, the F-test is used for testing the alternative hypothesis that at least one π is non-zero.

$$\ln VA_{it} = \lambda_0 + \lambda_1 \ln Y_{it} + \lambda_2 \ln P_{it} + \lambda_3 \ln P_{st} + v_{it} \qquad \text{Equ 4}$$

$$\lambda_0 = -\frac{\alpha_0}{\pi_1},\ \lambda_1 = -\frac{\pi_2}{\pi_1},\ \lambda_2 = -\frac{\pi_3}{\pi_1},\ \text{and}\ \lambda_3 = -\frac{\pi_4}{\pi_1}$$

Test for Co-integration (Pesaran et al., 2001)

Bounds test (partial F-statistic)

- **Null hypothesis :H_0: $\pi_1 = \pi_2 = \pi_3 = \pi_4$**

No cointegration among the variables in Equation (4)

- **The alternative hypothesis H1:$\pi_1 \neq \pi_2 \neq \pi_3 \neq \pi_4$ At least one π is non-zero**

If H_0 is not rejected, proceed no further ; Otherwise, the second step is required

3. Lag Length Criteria

When choosing ARDL bounds, make sure that the lag is appropriate to make sure the model is reliable. Inappropriate lag lengths will cause incorrect estimates. The AIC (Akaike information criteria) has been used to identify appropriate lag lengths for three models: Germany, Saudi and the UK.

4. Tourism Demand Estimation Results

Table (1)Tourism Demand Estimation Results

	Germany	Saudi	UK
ln VA (-1)	-0.091799	0.291906	1.041072
ln VA (-2)	-0.318364	-	-
ln GDP	0.471878	1.256251	-5.57748
ln GDP (-1)	30.96377	-	-32.014
ln PI	0.483003	0.540859	-4.83227
ln PI (-1)	-0.751668	-1.309847	0.677508
ln PS	-0.792562	-0.088916	0.448041
Ln PS(-1)	-	0.238607	-1.36809
D2001	0.318554	-0.247685	-0.43261
D2001(-1)	-	-0.421485	-
D20011	0.111477	0.119012	0.005755
D97	-0.06032	-0.201837	-0.50444
D97(-1)	-132644	-0.13149	-
C	-29.5701	3.145458	31.59169
Adj$-R^2$	0.85	0.74	0.78

5. ARDL Bounds Tests

The result of bounds test for ARDL

Table (2) Results of ARDL bound tests

Test	Germany	Saudi	UK
F statistic	5.31***	3.29**	2.46*
I(0)	2.88	2.27	1.99
I(1)	3.99	3.28	2.94
Lag	2	1	2
Fixed regressors	Rest,constant	Rest,constant	Rest,constant

Note: *, ** and * indicate significance at the 1%, 5%, and 10% level, respectively**
Source: Researcher's preparation using EViews 10

The demand model for Germany and Saudi tourists the relevant F statistic were ((5.31),(3.29)) at the 10% level, those that exceeded the upper critical bound were listed below, As a result, the null hypothesis of no co-integration will be rejected and implying long-run cointegration relationships amongst the variables. The demand model for UK tourists, the relevant F statistic was (2.46) In the case where F Located on the upper and lower bounds, the result could not be definitive so the null hypothesis of no cointegration reject.

- **The null hypothesis (H0: $\pi_1 = \pi_2 = \pi_3 = 0$) the variables are not cointegrated.**
- **The alternative hypothesis (H1: $\pi_1 \neq \pi_2 \neq \pi_3 \neq 0$)**

6. Tourism demand elasticities:

As shown in table (3) , the long-run tourism demand model for three major source markets in Egypt was constructed by normalizing visitor arrivals, as presented in the following Table:

Table (3) Demand elasticities

Demand elasticities	Germany	Saudi	UK
Income	22.2***	1.77***	-2.01
Own price	-0.19	-1.08**	1.16
Cross price	-0.56***	0.21	-3.43

Notes: *, **, * 10%, 5% and 1% level of significance, respectively.**

Increase in income 1% will lead to a (22.2%), (1.77%), and increase in visitor arrivals from Germany and Saudi.

The point estimates of income elasticity for Germany and the Saudi were greater than one, however, suggesting that these countries are not

income elastic when it comes to tourism to Egypt. There was a decline in demand for Egypt tourism after Egypt's own-price elasticity was negative, support for assuming that higher prices would lead to a decline in itself in demand.

The t test results show that the income elasticity's from three of the above mentioned countries are elastic as they are significantly greater than 1 at 5% significance level.

The calculated t_0 is greater than the critical value For (Germany and Saudi) then, reject Null hypothesis and accepted the alternative hypothesis that is suggest the income elasticity is elastic.

Since $|t|>$ critical value, a level of 5% rejects the Null hypothesis of significance and concloud income elasticity of demand is not equal to unity,most point estimates of the income elasticity's are positive only uk as an exception (-2.01) but this value is statistically insignificant, Estimates of own-price elasticity coefficient (UK) is negative indicating that an increase in the price of tourism products/ services in Egypt will lead to a decline in the demand for Egypt tourism .

As long as the associated confidence intervals cover the value of zero, these price elasticity does not differ statistically from zero, there was a decline in demand for Egypt tourism after Egypt's own-price elasticity was negative, support for assuming that higher prices would lead to a decline in itself in demand, the estimated cross –price elasticities are positive in the Saudi models which mean that increas in the costs of tourism will lead to an increase in the demand for Egypt

7. Model fitting

Table (4) Model fitting

Test statistic	Germany	Saudi	UK
R^2	0.91	0.86	0.93
Adjusted R^2	0.85	0.74	0.78
F statistic	13.53***	6.99***	6.35*
AIC	-0.43	-0.73	0.54

Note: ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively

Source: Preparing Researcher's by using EViews 10

Table above shows that:

- **The high values of the all models adjusted R^2 :Germany (0.85), Saudi (0.74) and UK (0.78), had a high goodness of fit and also means that about 85% (Germany,74% (Saudi) and 78% (UK), of the changes in visitor arrivals from the main markets during the period 1993−2017 .**
- **The F-statistic for all models are high & significant at the 1% significance level that is the estimated model is significant.**

8. Estimation of Short-Run

Effects of short run dynamic coefficients on the long run relationships that are derived from the ECM equation. This impact is persistent through the long run.The error correction variable (ECT) is significant at the level of 1 percent with a value of Germany (-1.41) with probability low 1 percent which indicates that the ECM specification used is Good ,The expected negative sign of ECM is not statistically significant .The sign of the error correction term indicates that the model is fit.

9. Residual Diagnostics results

A series of diagnostic tests, including the Breusch–Godfrey Lagrange multiplier for serial correlation, the Breusch–Pagan test for heteroscedasticity, the Jarque–Bera for normality to ensure that the final models were accurate, three final models were evaluated , If the models pass all the tests, they are finalized Forecasting will be done

Only the model the Germany passed all diagnostic tests, according to diagnostic statistics (residual diagnostic tests). All two other models failed a few tests but passed most of them.

1. **J-B test: Model (Germany) has P values, 0.80 (was greater than 0.05) so accept the null hypothesis the J-B normality test, where as models Saudi ,UK have P values 0.017, 0.000 (were smaller than 0.05) so reject the null hypothesis the J-B normality test.**
2. **the LM tests:In all models (Germany, Saudi Arabia, and the UK) which satisfy the assumption of independent error terms , there was no evidence of autocorrelation with the explanatory variable.**
3. **Heteroscedasticity test**

The Breusch-Pagan-Godfrey test: The test indicates no serial correlation problem since the p-value is greater than 0.05. All models

Germany, Saudi and the UK were free of the heteroscedasticity problem according to the test.

<u>White's heteroscedasticity test :</u>The White tests (without White cross terms) suggested that all three Heteroscedasticity were not a problem so accept the null hypothesis.

<u>The ARCH tests</u> :the homoscedasticity of the errors and the independence of them from regressors were observed in three different models.

<u>Stability Diagnostics</u>

The RESET test:the null hypothesis is not reject which p>0.05 for (Germany, Saudi ,UK)0.93, 0.90, 0.10 .

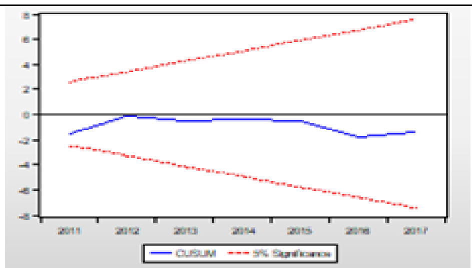
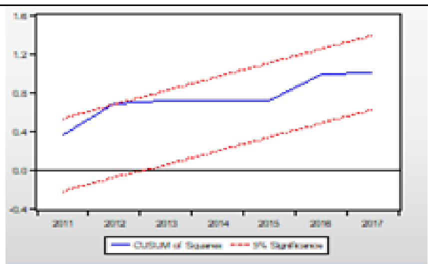
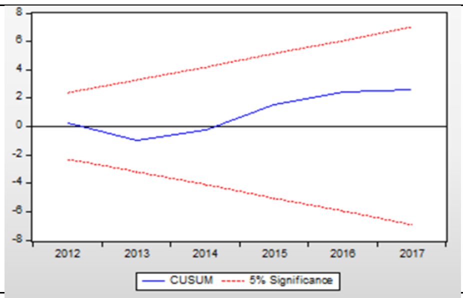
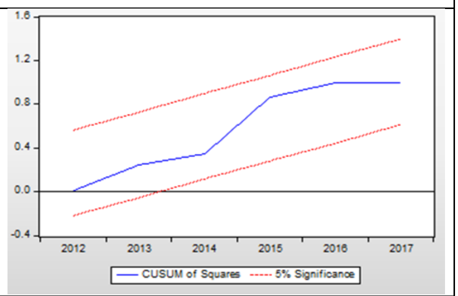
10. <u>Stability Tests:</u>

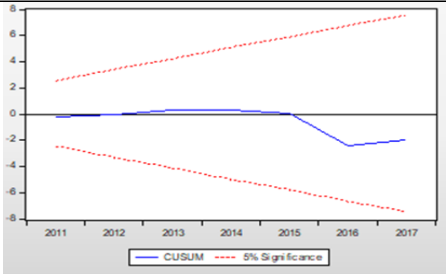
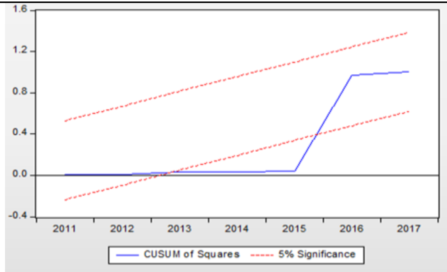
A short-run and long-run analysis were conducted for the equations upon stability examination. For the test, we relied on : cumulative sum of Recursive Residual (CUSUM),and cumulative sum of squares Recursive Residual (CUSUMSQ)

To test the stability of the long-run coefficients ,The tests applied to the residuals of the ECM model.

In Figures (1), it can be seen that the plot of the CUSUM for all models falls within the critical 5% bounds, supporting (the null hypothesis that is the coefficients in the given regression are stable cannot be rejected) and thus shows the stability of coefficient.

Figures (1) plot the CUSUM and CUSUM of squares statistics

Country	(A) CUSUM Test	(B) CUSUM Of Squares
Germany		
Saudi		

UK		

Source: Researcher's preparing using EViews 10

11. Basic Information About the Main Delphi Survey

Visitors from various countries (Germany, Saudi Arabia and the UK) served by Egypt's tourism industry were invited to modify their annual forecasts. The ARDL-ECM method was used to produce statistical forecasts using the sample 1993-2017. For the three source markets, IMF provides annual GDP growth projections and exchange rate forecasts. Due to a lack of distinct criteria and a lack of knowledge of the Delphi method, some of the articles did not meet the scientific guidelines for Delphi research. Ensure the reliability and validity of Delphi data by referring to scientific principles and procedures.

Ten academic researchers (59%) and seven industry practitioners (41%) made up the final panel. The first round of the Delphi survey was completed by over half (58%) of the panellists; the second round saw a lower positive response rate (54.8%).On a 7-point Likert scale, each panelist rated his/her level of expertise in tourism forecasting. Panelists were asked to adjust the economic forecasts of visitor arrivals to Egypt for three countries (Germany, Saudi Arabia, and the United Kingdom) for 1992-2017.

A. Evaluation of Forecasting Performance:

Visitor arrivals over the period 2013-2017 and the corresponding forecasts,based on the results tabulated in Tables (5), a smaller value denotes a greater reliable for APE, MAPE, and RMSPE measures. U statistics indicate statistical or subjective forecasts are better than Naïve forecasts when their value is less than one. With accuracy measures, we compare statistical forecasts with judgmental forecast.

- **Statistical forecasting accuracy**

Our first step should be to examine the efficiency with which the econometric model produces forecasts. A model's fit can definitely be used as a guide to forecast accuracy. It is not likely to be useful to

forecast Most long-term changes cannot be described by a model. It is not only possible use R^2 value for judging forecast accuracy and model fit, but it also provides a measure of the model fit.

The assessment of the likely forecasting ability of econometric models of tourism demand on the basis of common criteria such as goodness of fit, statistical significance of the coefficients (Witt & Witt,1992). Even if there is a good fit and a high proportion of statistically significant coefficients, this is not enough to guarantee accurate forecasting at a high level.

The fit of a model can definitely be used to determine forecast accuracy. If a model can't explain the majority of historical variations, it won't be useful for forecasts. In addition to providing a measure of model fit, R2 value can also be used to judge forecast accuracy.

In Table (5) one of the three models using R^2, Developed extremely exact arrival forecasts, based on their high R 2. It is useful to evaluate their R^2 for evaluating forecasts since they did not refer to actual results, Value of 0.91 for Germany.

Within the three source markets, one reported a mean MAPE of less than 10 % (UK) The greatest forecast errors determined by mean MAPE occurred in the Germany model(18.23) and Saudi (18.23).The U statistics indicate that two models the Germany (2.13) & Saudi (1.68) exceeded the Naive 1 model .

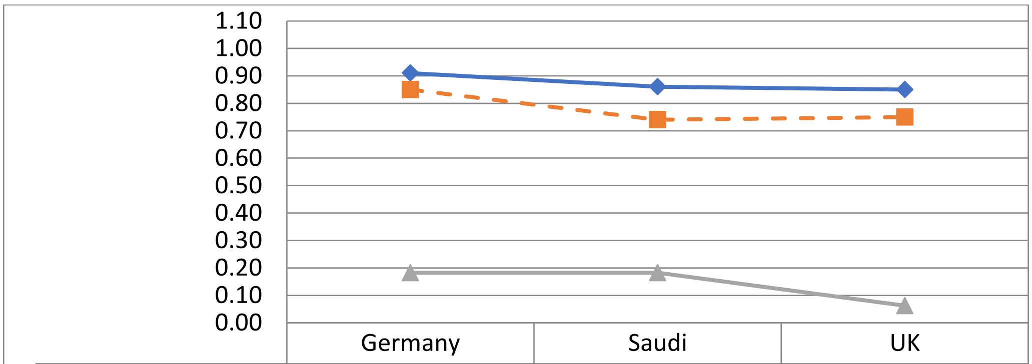
Table (5)Accuracy of statistical forecasts 2013-2017

Test statistic	Germany	Saudi	UK
R^2	0.91	0.86	0.85
Adjusted R^2	0.85	0.74	0.75
APE (%)	18.28	16.32	6.33
MAPE (%)	18.23	18.23	6.23
RMPSE (%)	18.23	18.23	6.23
Theil's U	2.13	1.68	0.56
Pearson Correlation (APE, *R*2)		-0.695**	

****Correlation is significant at the 1% level (1-tailed).**

Table (5) shows Pearson Correlation (APE, R^2) r = (- 0.695) , p one-tailed (< 0.01) the forecast errors evaluated by APE a significant relationship existed between forecasts and goodness-of-fit,: the lower the APE, the higher the R^2 are likely to produce forecasts that are more accurate ,This implies that statistical models with a higher R^2 .

Figures (2) The relationship between R^2 and MAPE by market 2013-2017



	Germany	Saudi	UK
R2	0.91	0.86	0.85
adjust R2	0.85	0.74	0.75
Mean MAPE %	0.1823	0.1823	0.0623

B. Basic distributional properties of forecast errors

Unbiasedness, efficiency, and accuracy are three dimensions, In prior analyses, these three dimensions were evaluated, standard statistical procedures were applied for forecast evaluation along , In order to get a basic understanding of forecasting performance, it is important to present an overall statistical summary of forecast errors. The distribution of forecast errors (measured by PE) for the arrival series view .

There is statistically significant indication that the percent errors are not normally distributed as the Sig values of S-W statistic are higher than 0.05, the student one-sample t test and the one-sample Wilcoxon signed-rank test were applied to test mean unbiasedness These two test results indicate that the percentage errors were unbiased: (mean : 0.92, 0.83)، the P value was greater than 0.

C. Results of hypothesis testing

To determine whether the difference between forecast accuracy and forecast performance was statistically significant, A set of error measures was employed conducted, A traditional comparison to a Naive forecast was carried out and by conducting regression analysis the investigation of potential forecast bias and efficiency was made.

(1) An improvement in the accuracy of judgmental forecasts based on statistical forecasts

- **Accept the Null hypothesis , reject the alternative hypothesis**

In Table (6) Based on the econometric model, the Delphi panellists evaluated the forecasts based on judgment the MAPE and RMSPE of the forecasts generated. By comparing the MAPE and RMSPE Forecast accuracy was evaluated. An analysis of the statistical difference between the two groups of forecasts was conducted to determine whether there was a significant difference, it can be seen that the largest improvement in accuracy over statistical forecasts has been found in the prediction of visitor arrivals from Germany, followed by Saudi Arabia.

As a result of the judgmentally adjusted forecasts, the MAPE decreased from 9.66 to 8.29% in round 1 and to 5.73% in round 2 when compared with the statistical modeling alone. MAPE showed reductions from 14.18 to 40.68 %, The results generated by RMSPE were mostly comparable to the ones generated with MAPE; the mean RMSPE decreased from 14.11 to 8.31 % in the round one (R1) and to 5.76 % in the round two (R2). A range of 41.11 to 59.18% was observed in the reduction of RMSPE .

Table (6) Overall forecasting performance 2013-2017

Country	MAPE(%)			RMSPE(%)			U		
	SF	GF1	GF2	SF	GF1	GF2	SF	GF1	GF2
All	9.66	8.29	5.73	14.11	8.31	5.76	1.03	0.99	0.89
Germany	18.23	12.32	10.15	18.23	12.06	9.94	2.13	1.32	1.03
Saudi	18.23	17.56	15.2	18.23	16.67	14.5	1.68	1.63	1.55
UK	6.23	5.9	5.44	6.23	7.08	5.95	0.56	0.6	0.54
mean	14.23	11.93	10.26	4.42	4.32	5.30	0.31	0.43	0.54
Percentage reduction (%)	GF1-SF	GF2-SF	GF2-GF1	GF1-SF	GF2-SF	GF2-GF1	GF1-SF	GF2-SF	GF2-GF1
All	-14.18	-40.68	-30.88	-41.11	-59.18	-30.69	-3.88	-13.59	-10.1
Germany	-32.42	-44.32	-17.61	-33.85	-45.47	-17.58	-38.03	-51.64	-21.97
Saudi	-3.68	-12.68	-7.8	-8.56	-4.49	-15.96	-2.98	-3.57	-10
UK	-5.3	-7.8	7.54	13.64	-15.96	-6.55	7.14	-10	-3.7
mean	-18.44	-18.44	-14.86	-5.00	-6.55	-11.86	-2.30	-6.21	-11.34

Note: SF, GF1 and GF2 represent the econometric (statistical) forecasts, group 1 forecasts and group 2 forecasts, respectively. MAPE = mean absolute percentage error; RMSPE = root mean square percentage error.

Table (7) Wilcoxon signed rank test results evaluated by APE

*H*0: test if 0 *H*1 : test if <0	Test 1 $(APE_{GF1}-APE_{SF})$	Test 2 $(APE_{GF2}-APE_{SF})$	Test 3 $(APE_{GF2}-APE_{GF1})$
Positive ranks (*T*)	**16**	**11**	**6**
Z	**- 0.405**	**-1.214**	**-1.310**
Asymp. Sig. (1-tailed)	**0.113**	**0.083**	**0.005**
Effect size (*r*)	**-.02** **(Small effect)**	**-0.18** **(Small effect)**	**-0.89** **(high effect)**

Source: Preparing Researcher's by using SPSS Program

The MAPEs exceeded high percentage after the experts' judgments were made, indicating a significant increase in accuracy. The results of evaluating the forecast accuracy using MAPE and RMSPE found no significant differences, the forecast adjustments improve the overall forecast accuracy across markets and over different rounds of Delphi

We test in cases where forecasts were significant different before and after adjustment by using Wilcoxon signed-rank tests in table (7) As compared to the group forecast, the statistical forecast did not outperform (Z = 0.405 ,Sig.(1-tailed) 0.113 ,T =16, r = -.02). In contrast to the initial statistical forecasts (Round 1 forecasts), the forecast accuracy evaluated by APE was found to generate more accurate Round 2 forecasts: the Sig values was 0.083 for test 2 and 0.005 for test 3, both statistically significant at 10%.

Forecast adjustments improve forecast accuracy and have been apparent across markets and over the many iterations of Delphi, As (Table 5), According to the experts, Over the statistical forecasts, the greatest accuracy improvement was recorded in the prediction of visitors coming from Germany, Saudi Arabia. When similar comparisons were made using APE to those shown in table (6), the results were found to be similar in most cases.

(2) Naive forecasts do not perform as well as judgmentally adjusted forecasts.

The U statistic was used to analyze the performance of forecasts made by the Naive 1. The general accuracy was comparable to the Naive 1 forecast in estimating Egypt 's tourism since the U statistic 1,03; therefore, the results supported hypothesis H2 that unbiased forecasts are generally more accurate than Naive forecasts,in (Table 6) show examination of the U statistic results markets were, in general, better than the Naive forecasts. by The markets have a great value of the U stat primarily a result of the Germany market, with a

high value (SF 2.13, GF1 1.32, and GF2 1.03) The other two markets Saudi and the UK,0.56) ,(0.31) U statistics under one, shows These are the revised and unadjusted forecasts for the two regions.

This finding could have been determined based on two factors: the inclusion of three market sources, a mix of multiple step forecasts, and a mix of multiple step forecasts.

- **Accept the Null hypothesis , reject the alternative hypothesis**

(3) Forecasts adjusted for judgment are biased.

For the sample period 2013 to 2017 we used a regression model of Equation

$$PE_t = \alpha_0 + \beta_0 PE_{t-1} + \mu_t \qquad \text{Equ 5}$$

Where $PE_t = (A_t - F_t)/ A_t$ **. It must be positive (or negative).**

**** A positive** α_0 **coefficient shows that the average forecast error is greater than zero, indicating that forecasters are overpessimistic.**

**** A negative** α_0 **coefficient means that the average forecast error is less than zero that there is overforecasting.**

- **The null hypothesis reject if** α **=zero, experts' forecasts display a level bias.**

Unbiasedness requires that α = 0 and efficiency requires that β = 0 in the above equation. An acceptance of a = 0, indicates that the forecast is unbiased. An acceptance of β = 0 indicates that forecasters use all available information at the time of forecast efficiently in their forecast revisions. If a forecast is efficient, its forecast error should be independent of its past forecast revisions. Nordhaus (1987), If there is no bias in the forecasts, α_0**.is expected to be zero (Harris, 1999).**

The judgmentally adjusted forecasts was estimated to test for the bias of, Table (8) presents regression analysis by the source markets. null hypothesis of no bias was used to build the statistical model of forecast errors. In the first regression model, individual forecasts from both rounds, G1 and G2, were averaged to determine group levels,The result α **was insignificant so the adjustment forecasts for R1 and R2 were unbiased, Both the first and second round of forecasts were not biased,Two of three countries tested in Round 1 (Germany and the UK) and two of three countries tested in Round 2 (Germany and the UK) the intercept term was significantly different from zero,A model's intercept (R1,R2 for Germany, UK),(R2 in Saudi) were significantly greater than zero, according to data from these markets, the forecasts were underestimated, there was a negative intercept in R1 for the Saudi model, but it was not significant.**

- **Reject the Null hypothesis , accept the alternative hypothesis**

(4) Forecasts based on judgemental adjustments are inefficient.

Analyzing all the effects of the past forecasts and the impact of forecast errors, this hypothesis studies forecast efficiency, table (8) shows that the forecasts for one market (Germany) was efficient but the group forecasts were found to be inefficient with β_0 was significantly different from unity was rejected for all three markets in the initial round (indicating that the 3 sets of forecasts were not efficient significantly different from zero) ,

For two rounds β_0 was significantly different from zero at the 5% level , meaning that the group forecasts were inefficient forecast errors in the experts' adjustments, according to judgmentally adjusted forecasts were unbiased and were inefficients as they neglected to incorporate all of the data from their previous forecasts and forecast errors. Results support the H3 ,H4 hypothesis that forecasts that are unbiased usually show inefficiency as they give little information to predict the future systematic biases must be detected during the process of judgmental forecasting.

(5) Forecast accuracy improves via the Delphi approach.

- **Compared with the average of member estimates, the final Delphi forecasts are generally more accurate .In order to test H5, a series of statistical tests were conducted examining group forecasting performance and individual expert forecasting performance using three error measures (APE, MAPE, and RSMPE).**
- **The accuracy of MAPE, RMSPE, and the U statistic also occurred over rounds in spite of the consensus measure employed (table 6).The results from the previous two tests indicated that the efficiency of the group panellists significantly improved when using the Delphi technique.**

Based on the regression analysis, we gathered further evidence into the relative performance of the forecasts in Rounds 1 and 2. As we found in testing hypothesis H1, the forecast accuracy has improved over rounds in MAPE and RMSPE.

As shown in table (6) the highest mean improvement in accuracy in the two rounds was observed in the Germany market Italy and Russia (with % 17.61,%15.97,%14,22) decrease in MAPE. During the evaluation period 2018−2019, the experts' adjustments from Round 1 to Round 2 in one (Russia) out of the six markets recorded an unimproved performance in MAPE and RMSPE; this market, accuracy was decreased during both rounds of adjustments, according to table (6) (Germany, Saudi, and the UK). Over rounds, the forecast accuracy has improved for three source markets.

Table (8) Regression coefficients for bias and inefficiency *(Dependent variable:PE_t)*

Market		Constant	t	PE_{t-1}	t	Results	Preference on bias
All (group forecasts)	SF	-0.023	-0.212	0.164	0.806**	Unbiased, inefficient	over
	R1	0.017	0.414	0.092	0.455**	Unbiased, inefficient	under
	R2	0.029	0.942	0.244	1.233**	Unbiased, inefficient	under
All (individual forecasts)	SF	-0.023	0.129	0.164	(12.945)**	Unbiased, inefficient	Under
	R1	0.170	-0.158	0.092	(10.422)**	Unbiased, inefficient	Over
	R2	0.029	0.802	0.244	(8.713)**	Unbiased, inefficient	Over
Germany	R1	0.018	0.582**	-0.010	-0.052	Biased, efficient	under
	R2	0.086	2.243**	0.122	0.615**	Biased, inefficient	under
Saudi	R1	-.091	-0.761	0.299	0.368**	Biased, inefficient	over
	R2	.005	0.171	0.407	2.108**	Biased, inefficient	under
UK	R1	.027	1.014*	0.360	2.137**	Biased, inefficient	under
	R2	.026	0.945**	0.348	2.075**	Biased, inefficient	under

Source: Researcher's Preparing by using spss

The one sample t-test results in table(8) A significant lower MAPE was found in the experts (t (12) = -3.064, sig= 0.005 < 0.05) and lower RMSPE (t (12) = -3.564, sig = 0.002< 0.05) in the second round.

We use a Wilcoxon signed-rank test with the null that the median of gapmape/gaprmspe equals zero further confirmed , significantly lower MAPE (Z= -2.691, sig= 0.0035 < 0.01) and lower RMSPE Z = -2.353 , sig = 0.0095< 0.01) in the second round.

Table (9) Results for one-sample t test and Wilcoxon signed rank test

Test	One sample *t* test		Wilcoxon signed rank test	
	T	*sig*. (one-tailed)	Z	*sig*. (one-tailed)
Gapmape	-3.064	.005**	-2.691	0.0035***
Gaprmspe	-3.564	.002**	-2.353	0.0095***

Note: *, ** and * indicate significance at the 1%, 5% and 10% level, respectively.**
Source: Preparing Researcher's by using SPSS pregramm

We use the Wilcoxon signed-rank test to examine whether there was statistical significance in such performance difference .Table(10) show that there was a significant decrease in MAPE among the 13 experts (z = -2.760, sig = 0.006 < 0.01) and lower RMSPE (z = -2.830, sig = 0.005 < 0.01) in the second round. Utilizing the Delphi approach, this study demonstrated a significant difference in performance between individual experts.

Table(10) Result Wilcoxon signed rank test 13 individual experts

Test	Wilcoxon signed rank test	
	Z	*sig*. (1-tailed)
MAPE	-2.760	0.006 **
RMSPE	-2.830	0.005 ***

Note: *, ** and * indicate significance at the 1%, 5% and 10% level, respectively.**
Source: Researcher's preperationusing SPSS

These results indicate that either the result of the information they received in the second round allowed them to better predict, or the experts probably regarded Delphi to be a helpful method for getting them to improve their forecasts, hypothesis (H5) asserting that the Delph technique could contribute to stronger forecasts than the average of their initial judgments.

- **Accept the Null hypothesis , reject the alternative hypothesis.**

12 . Conclusion:

In this study we found three main Findings(Effectiveness of investigation judgmental adjustments ,Bias and inefficiency of judgmental adjustments, Effectiveness of implementing judgmental adjustments, Conditions for using judgmental adjustments, Usefulness of applying the Delphi procedure).

In the next part, we will present these conclusions :

1. <u>The effectiveness of investigation judgemental adjustments</u>

The results of the hypothesis tests obtained by APE were about the same as those obtained by MAPE and RMSPE, the judgmental adjustments made based on statistics improved accuracy. By comparing judgmentally adjusted forecasts to the initial statistical forecasts, the effectiveness of judgmental adjustments can be evaluated. Results indicate that During Rounds 1 and 2 of the consensus group forecasts for all three source markets, the forecast adjustments helped improve the overall and market forecast accuracy. we found that mathematically calculated and judgementally adjusted forecasts were statistically equivalent. .

U statistics for Germany did not decrease much after the experts' judgmental adjustments, but they were still above unity .Germany& USA markets were the only cases reporting a U statistic larger than one. Two of the three markets(Saudi,and the UK) were below unity for the two rounds by using Theil's U statistics it appears that the Naive forecasts for these markets were less accurate than the unadjusted and adjusted forecasts. However, the U statistic still fell below unity after the experts' judgmental adjustments, but the Germany forecast still exceeded it.

There may be a number of reasons for such a difference in accuracy improvement :

a. **Judgmentally adjusted forecasts proved more accurate than the statistical predictions this suggests that human interventions may be more effective for improving accuracy when a series has high volatility,**
b. **Improved forecast accuracy is unlikely to be possible with expert judgment,**
c. **the accuracy of tourism forecasts will be unlikely to be significantly improved .**

In the case of highly accurate statistical forecasts,These changes would have no benefit on accuracy but would have a negative effect on it,

Moreover, results of the hypothesis tests showed that the Delphi method improves judgemental adjustments to statistical forecasts. Also, quantitative

methods Statistical tests, for example can be used to examine the changes in error measures and raw forecasts.

An analysis of regression data provided further analyze performance from a comparative perspective of the statistically and forecasts that are adjusted for judgement , which indicated that , In the final Delphi forecast, actual arrivals in Egypt were better predicted than in the initial group forecast . Expert exchanges will likely make Delphi participants more effective forecasters. The conclusions above were derived from the evaluation period 2013−2017, Compared with the initial consensus forecast and statistical forecast, the final Delphi forecasts were more accurate, indicating that reducing forecasting risk might be possible with structured group techniques given that they produce more accurate forecasts.

As a result of examining the accuracy criteria (APE, MAPE, and RMSPE) and assessing the degree of accuracy improvement, the overall accuracy averaged from the six source markets decreased and experts' judgemental forecasting ability decreased over time .

2. <u>Bias and inefficiency of judgmental adjustments</u>

Delphi group forecasting is biased as a judgmental method, although structured procedures mitigate this. A Delphi method may be used to aggregate and structure the experts' adjustments to improve the efficiency of the adjusted forecasts, but bias cannot be eliminated. According to the tests of hypotheses H2 and H3, estimates made by the experts are biased for some individual source markets, however, the consensus group forecasts are, on average, unbiased. On the other hand, the experts have different tendencies in predicting different markets.

Study results indicate that statistical forecasts are incomplete and are not incorporating past data when based on judgmentally adjusted forecasts and errors. Since judgmentally adjusted forecasts are biased to the individual market.

13.References

The first English References:

(1) Andrey Davydenko (2012) Integration of Judgmental and Statistical Approaches for Demand Forecasting: Models and Methods.

(2) Bunn,D.& Wright, G.(1991) Interaction of judgmental and statistical forecasting methods: issues and analysis. Management Science, 37 (5), 501-518.

(3) Caniato, F., Kalchschmidt, M., & Ronchi, S. (2011) Integrating quantitative and qualitative forecasting approaches: organizational learning in an action research case. Journal of the Operational Research Society,62 (3), 413-424.

(4) Emeka Nkoro and Aham Kelvin Uko , Autoregressive Distributed Lag (ARDL) cointegration technique: application and interpretation, Journal of Statistical and Econometric Methods, vol.5, no.4, 2016, 63-91

(5) Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. International Journal of Forecasting , 25 (1), 3-23.

(6) Goodwin,(2000a).Improving the voluntary integration of statistical forecasting and judgement, International Journal of Forecasting , 16 (1), 85-99.

(7) Goodwin, P. (2000). Correct or Combine? Mechanically integrating judgmental forecasts with statistical methods. International Journal of Forecasting , 16 (2), 261-275.

(8) Goodwin, P. (2001). Integrating management judgment and statistical methods to improve short term forecasts. The International Journal of Management Science , 30 (2), 127-135.

(9) Goodwin, P., & Fildes, R. (1999). Judgmental forecasts of time series affected by special events: does providing a statistical forecast improve accuracy? Journal of Behavioral Decision Making , 12 (1), 37-53.

(10) Haile Ademe Ayalew (2013) Determinants of domestic saving in Ethiopia: An autoregressive distributed lag (ARDL) bounds testing approach.

(11) Haiyan Song, Bastian Z. Gao, Vera S. Lin, Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system, International Journal of Forecasting 29 (2013) 295–310.

(12) Hanyuan Zhang, Haiyan Song, Long Wen, Chang Liu (2021), Forecasting tourism recovery amid COVID-19.

(13) Haroon Rasool et al.(2021)The relationship between tourism and economic growth among BRICS countries:a panel cointegration analysis
(14) Harris, R. (1999). The accuracy, bias and efficiency of analysts' long run earnings growth rates. Journal of Business Finance & Accounting, 26 (5–6), 725–755.
(15) Lawrence, M., O'Connor, M., & Edmundson, B. (2000). A field study of sales forecasting accuracy and processes. European Journal of Operational Research , 122 (1), 151-160.
(16) Lim, J.S.,& O'Conner,M.(1995).Judgmental adjustment of initial forecasts its effectiveness and biases. Journal of Behavioral Decision Making, 8 (3),149 168.
(17) R. Santos Alimi, ARDL bound testing Approach to Cointegration: A Reexamination of Augmented Fisher Hypothesis in an open economy, Asian Journal of Economic Modelling, 2014, 2(2): 103-114.
(18) Song, H., Gao, Z, Zhang, X., & Lin, S. (2012). A Web-based Hong Kong tourism demand forecasting system. International Journal of Networking and Virtual Organizations, 10 (3/4), 275-291.
(19) Vera Shanshan Lin, & Song, H. (2012). A review of qualitative forecasting in tourism,In C. Cooper (Ed.). Contemporary Tourism Reviews, Oxford: Goodfellow Publishers.
(20) Vera Shanshan Lin, (2013) Improving forecasting accuracy by combining statistical and judgmental forecasts in tourism. Journal of China Tourism Research,(3),325-352.
(21) Vera Shanshan Lin, Paul Goodwin , Haiyan Song (2016), accuracy and bais of exports'adjusted forecasts, Journal of Statistical and Econometric Methods, vol.5, no.4, 2016, 63-91 ISSN: 1792-6602 (print), 1792-6939 (online) Scienpress Ltd.
(22) Vera Shanshan Lin, 2019. "Judgmental adjustments in tourism forecasting practice: How good are they?," Tourism Economics, vol. 25(3), pages 402-424, May.
(23) Zegers, R.J.G.,2013 The best of two worlds integrating statistical and judgemental forecasting, Master Eindhoven University of Technology.

ثانياً المراجع العربية

(١) مالك علام عفت عودة, (2018), قياس وتحليل محددات الطلب على النقود فى الإقتصاد العراقى
(2) خضير عباس الوائلى ,(2017) ,استعمال أسلوب ARDL فى تقدير أثر سياسات الاقتصاد الكلى على بعض المتغيرات الاقتصادية