International Journal of Information Technology & Decision Making
Vol. 15, No. 2 (2016) 387–401

DOI: 10.1142/S0219622016500085

Improving Forecasting Performance by Exploiting Expert Knowledge: Evidence from Guangzhou Port

Anqiang Huang*, Han Qiao†,¶, Shouyang Wang‡ and John Liu§

*School of Economics and Management
Beijing Jiaotong University, Beijing 100044, P. R. China

†School of Economics and Management
University of Chinese Academy of Sciences, Beijing 100190, P. R. China

‡Academy of Mathematics and Systems Science
Chinese Academy of Sciences, Beijing 100190, P. R. China

§College of Business
City University of Hong Kong, Kowloon, Hong Kong (SAR)
¶qiaohan@ucas.ac.cn

Published 1 March 2016

Expert knowledge has been proved by substantial studies to be contributory to higher forecasting performance; meanwhile, its application is criticized and opposed by some groups for biases and inconsistency inherent in experts' subjective judgment. This paper proposes a new approach to improving forecasting performance, which takes advantage of expert knowledge by constructing a constraint equation rather than directly adjusting the predicted values by experts. For the comparison purpose, the proposed approach, together with several widely used models including ARIMA, BP-ANN and the judgment model (JM), is applied to forecasting the container throughput of Guangzhou Port, which is one of the most important ports of China. Forecasting performances of the above models are compared and the results clearly show superiority of the proposed approach over its rivals, which implies that expert knowledge will make positive contribution as long as it is used in a right way.

Keywords: Container throughput forecast; expert knowledge; subjective judgment.

1. Introduction

In preparation for the 5th National Contest On Logistics Design by University Students of China (http://wlsjds.clpp.org.cn/NCOLDStatic/pages/main.htm), Dr. Zhang is facing a big challenge of forecasting transportation demand in the coming year. What confuses him is that his colleagues disagree with his prediction generated by the sophisticated statistical model built in the forecasting software package, and even worse, different colleagues suggest various adjustments to his prediction. What is the best decision for Dr. Zhang? Is it better to insist on his

¶Corresponding author.

prediction from the pure statistical model or to rely on his colleagues? If latter, how to deal with the deviation between them? How to take advantage of the judgments of his colleagues, by directly adding them to the prediction or by other methods?

Dr. Zhang is not the lone one who confronts the above problems. Actually, when it comes to the function of expert knowledge, there is no unanimous conclusion. Reference 1 surveys 240 US corporations and suggests that only 11% are using forecasting software, of whom 60% adjust their predictions by using experts' judgment. Reference 2 concludes that models with expert knowledge are preferred in macroeconomic forecast after reviewing a lot of published articles. More studies advocating positive contribution of expert knowledge can be found in Refs. 3 and 4 that made thorough reviews of this issue.

However, application of expert knowledge in the forecasting process is also criticized owing to biases and inconsistence inherent in subjective judgment. Some experimental results show that forecasters tend to make unnecessary judgmental adjustments to statistical projections,[4] even when they do not possess additional contextual information. Even worse, some forecasters persist in making judgmental adjustments, though their adjustments are proved to be harmful.[5]

This paper aims to understand the following issues: (1) should expert knowledge be incorporated in the forecasting process to improve the forecasting performance? If it should, (2) how to deal with the deviation between them? (3) how to exploit expert knowledge, not only making use of its advantages but also avoiding disadvantages?

To answer these questions, this paper proposes a new model taking advantage of expert knowledge by constructing a constraint equation instead of directly making the subjective adjustment by experts, and then the proposed model is compared to several widely used models including ARIMA, BP-ANN and the judgment model (JM). The comparison results clearly show the superiority of the proposed model over its rivals, which implies that expert knowledge can be contributory to higher forecasting performance on the condition that it is used in a right way.

The remainder of this paper is organized as follows: Section 2 reviews typical existing papers on combination of statistical models and expert knowledge. Section 3 describes the methodology; Sec. 4 elaborates the newly proposed forecasting model; Sec. 5 tests the new model by conducting an empirical study; Sec. 6 discusses and concludes.

2. Literature Review

Although statistical models, e.g., linear models, tend to be more precise than subjective judgments, the former will generate larger errors and lead to leptokurtic distributions with the use of unsuitable logical rules.[6] In contrast, human judges perform better in circumstances where objective methods cannot be implemented, such as identifying new prediction variables[7] and assessing variables that cannot be measured objectively.[8] Yet, judgments are likely to suffer from inconsistency and biases.[9,10] Considering that both statistical and judgmental models have their own advantages and disadvantages, the combination of them is an option well worth consideration. According to this idea, some studies, including Refs. 11–17, have been carried out and shown that forecasts based on model–expert combinations are superior to model outputs and expert judgment alone, in conjunction with some opposite findings like Refs. 4, 5 and 18.

Various model-judge(s) combination methods have been proposed, such as Refs. 19–22. Generally, procedures of combining statistical model and judgments range from mechanical methods (e.g., simply weighted average of constituent forecasts) to employing judgment to determine the way to combine the forecasts.[4] Some reviews of the literature argue that mechanical combinations of point forecasts tend to perform better than those based on judgment.[3,23] Especially, Ref. 7 suggests that a simple 50/50 weighting between model outputs and managerial judgment can effectively improve the forecasting accuracy, and Refs. 12, 13 and 24 further confirm this conclusion. Reference 25 makes a thorough review on mechanical combination methods. However, other researches present opposite conclusions. For example, Ref. 26 argues that judgmental combination generates higher accuracy than the simple average when the judges obtain performance feedback of individual forecasts available for combination.

Other researchers have investigated some rules, which should be obeyed in a combination. Reference 27 argues that the forecast generated by the person in charge of the combination should not be included lest that they would overweight their own forecast. Reference 12 argues that the judgmental forecast should be based on contextual information (or domain knowledge), especially where the time series has a high degree of variability. Reference 28 demonstrates a procedure for screening out forecasting methods that are inferior to other methods or do not add any information to the combination. Reference 29 suggests that a larger number of judgments can yield higher forecasting accuracy.

To summarize, both mechanical combination and judgment-based combination directly calibrate point forecasts generated by statistical models according to judgments. This strategy only adjusts the forecasted values and makes no improvement on the forecasting function. Moreover, it is a huge task for experts to provide point forecasts, because human brains function better in processing fuzzy information than precise data. To overcome this problem, new combination methods are proposed in the following sections which calibrates parameters of the forecasting function. By this way, an improved forecasting function with higher accuracy is generated, which can generate more reliable managerial implications and more precise forecast in a new round. Meanwhile, this new proposed method just needs experts to provide interval forecasts other than point forecasts, which is much easier.

3. Methodology

3.1. *Investigation framework*

The investigation is proceeded under the following framework: (1) two pure statistical models are constructed, including ARIMA and BP-ANN; (2) different expert judgments are integrated; (3) the integrated judgment is incorporated into the forecasting process, where two models (i.e., the JM and the proposed model) are generated; (4) conclusions are drawn by comparing the above models in terms of their forecasting performance.

3.2. *Two pure statistical models*

3.2.1. *ARIMA*

The ARIMA model with the order $(p,\ d,\ q)$ is expressed as follows:

$$\begin{cases} \phi_p(B)(1-B)^d x_t = \theta_q(B)\varepsilon_t \\ E(\varepsilon_t)=0, \mathrm{VAR}(\varepsilon_t)=\sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s)=0(t\neq s) \\ E(x_s\varepsilon_t)=0) \quad \forall s<t \\ \phi_p(B)=1-\phi_1 B-\phi_2 B^2\cdots-\phi_p B^p \\ \theta_q(B)=1-\theta_1 B-\theta_2 B^2\cdots-\theta_q B^q, \end{cases}$$

where B is the backshift operator $B(x_t)=x_{t-1}$, x_t denotes t-time forecast value and ε_t denotes t-time error. Although ARIMA has been widely accepted as one of the most effective models for univariate nonstationary time series analysis, the X-12 seasonal adjustment algorithm should be referred to before running ARIMA when the time series contains a strong seasonal component.

3.2.2. *BP-ANN*

The BP-ANN model has been very famous for its powerful ability to solve nonlinear problems. According to Ref. 30, a three-layer BP-ANN with enough nodes in the hidden layer is able to approach to any integrable continuous nonlinear function. The objective function of BP-ANN model can be written as

$$E=\frac{1}{2}\|Y-\widehat{Y}\|^2=\frac{1}{2}\|Y-f(W_{ih}XW_{ho}+\theta)\|^2,$$

where $Y=\{y_1,y_2,\ldots,y_n\}$ is the observed time series, $\widehat{Y}$ is the simulated series, W_{ih} is the weight matrix measuring the links between the input nodes and hidden nodes, W_{ho} is the weight matrix between the hidden nodes and output nodes, X is the input matrix, θ is the bias vector of the hidden layer, and f, frequently the sigmoid function, is the transfer function in the hidden layer. W_{ih}, W_{ho} and θ can be obtained by applying the back propagation algorithm.

3.3. *Expert knowledge integration*

When broad deviation between expert judgment appears, it is necessary to resort to the expert knowledge integration procedure to generate unified actionable knowledge. The Delphi method is one of the most frequently used, whose procedures are composed of the following steps: (1) selecting a delegation of experts appropriate for solving the problem; (2) accurately describing the problem to be solved and providing experts with as much valuable background information as possible; (3) collecting and summarizing judgments of the experts after they respectively give out their ideas; (4) feeding back the statistics to experts and asking them to calibrate their judgments; (5) repeating steps (3) and (4) until the unanimous conclusion is obtained.

For successfully implementing the Delphi method, we can never pay too much attention to two key factors. One is the selection of the delegation of experts. Only those deeply understanding the problem can offer contributory ideas, while those unfamiliar to the problem probably make wrong decisions. The other is adhering to the rule that experts must make their judgments without any communication with others, which avoids the occurrence of blind obedience to authority or majority.

3.4. *Judgment model*

To obtain more accurate predictions, the integrated expert knowledge should be incorporated into the forecasting process. In practice, the JM is frequently used owing to its convenience and understandability.

Given a set of data points $\{x_i, y_i\}^n$ (x_i is the input vector, y_i is the desired value and n is the total number of observations), a pure statistical model $y_i = g(x_i)$ approximating the pattern between x and y, and the expert judgment $\{e_i\}^n$ generated by the Delphi method, the JM is expressed as

$$\hat{y}_i = f(x_i) + e_i. \tag{1}$$

Equation (1) implies that the predicted value $\hat{y}_i$ is composed of the statistical component $f(x_i)$ generated from structured historical data and the expert judgment e_i extracted from expert knowledge.

As presented by Eq. (1), the JM tunes the final prediction by directly adding expert judgment to the pure statistical prediction, which tends to result in subjective biases and inconsistence. Therefore, it is necessary to design a more reasonable approach to making full use of expert knowledge, which is addressed in detail in Sec. 3.

4. A New Forecasting Model Using Expert Knowledge

4.1. *Model specification*

Given a time series denoted by $X = \{x_1, x_2, \ldots, x_L\}$, it is assumed that Eq. (2) is sufficient to represent the historical and future elements

$$x_t = h(Y_t) + \frac{\Phi(B)}{\Psi(B)}\mu_t, \quad t = p+1, p+2, \ldots, L, L+1, \ldots, L+T, \tag{2}$$

where $x_t (t = 1, 2, \ldots, L)$ are historical observations and $x_t \quad (t = L + 1, L + 2, \ldots, L + T)$ are predictions in the projection horizon T, Y is the vector of exogenous variables, $h(Y_t)$ is a linear function of Y_t, $\Phi(B)$ and $\Psi(B)$ are polynomials of B that is the backshift operator, $p > 0$ is the lag order, μ_t is the stochastic term assumed to follow a normal distribution with mean 0.

When provided with the historical observations X, contextual information θ and primary projections Z generated by Eq. (2), experts will make their judgments e to the projections in their special way. Taking X and Z as inputs and e as outputs, a Delphi-based expert system can be considered as a function which is closely dependent on experts' special domain knowledge θ and unique thinking pattern g. Therefore, the relationship between expert judgment e, historical observations X, primary projections Z and contextual knowledge θ can be described by Eq. (3),

$$e_j = g(X_j, Z_j, \theta_j) + \varepsilon_j, \quad j = 1, 2, \ldots, m, \tag{3}$$

where the subscript j indicates the jth period, ε_j is assumed to follow a mean 0 normal distribution, m is the total number of expert judgment.

With the assumption that experts are more competent to estimate the gross, rather than the monthly, container throughput of Guangzhou Port in the forecast horizon T, in that the gross variable moves more smoothly and thus is easier to be understood, we specify Eq. (3) as

$$e_j = a_0 + a_1 \sum_{i=L+(j-m)T+1}^{L+(j-m+1)T} x_i + a_2 \sum_{i=L+(j-m)T+1}^{L+(j-m+1)T} Z_i + \varepsilon_j. \tag{4}$$

Accordingly, a joint model, presented by Eq. (5), can be obtained by combining Eq. (2) with Eq. (4):

$$\begin{cases} x_t = h(Y_t) + \dfrac{\Phi(B)}{\Psi(B)} \mu_t, & \text{(5a)} \\ e_j = a_0 + a_1 \displaystyle\sum_{i=L+(j-m)T+1}^{L+(j-m+1)T} x_i + a_2 \sum_{i=L+(j-m)T+1}^{L+(j-m+1)T} Z_i + \varepsilon_j, & \text{(5b)} \end{cases}$$

where the subscript t ranges from $p + 1$ to $L + T$ and j locates in the interval $[1, m]$.

It is notable that Eq. (5) takes advantage of expert knowledge by constructing a constraint equation (Eq. (5b)), which helps to avoid biases and inconsistence resulted from the direct adjustment to the prediction.

In Eq. (5), there exists a vector of unknown quantities denoted by $\Upsilon = [h, \phi, \varphi, a, \sigma_\mu, \sigma_\varepsilon]$ to be estimated. h is composed of the coefficients in $h(Y_t)$. ϕ and φ respectively indicate coefficients in $\Phi(B)$ and $\Psi(B)$. a is set to be $a = \{a_0, a_1, a_2\}$. σ_μ and σ_ε are respectively standard variance of μ and ε. The logical relationship between the above unknown quantities is simply described as follows:

- h, ϕ and φ are generated by estimating Eq. (5a), based on which, Z is computed.

- Based on Z and expert knowledge, judgmental adjustment e is made by experts.
- With e, Eq. (5a) is re-estimated together with Eq. (5b). Then the value of Z is updated.

To estimate Eq. (5), the MCMC algorithm should be adopted as suggested by Ref. 31, in that Eq. (5) has the hierarchical structure of parameter dependency.

4.2. *Preliminaries for estimating the proposed model*

The core difference between conventional estimation methods and the MCMC algorithm is that the former take parameters as constants and estimate them by minimizing the loss function or maximizing the likelihood function, while the latter takes parameters as stochastic variables and estimates them by drawing samples from their posterior probability density functions.

Before applying the MCMC algorithm, some indispensable assumptions have to be set forth as follows:

- $\mu_t \overset{iid}{\sim} N(0, \sigma_\mu^2)$, $\mathrm{cov}(\mu, Y) = 0$, $\mathrm{cov}(\mu, x) = 0$.
- $\varepsilon_j \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$, $\mathrm{cov}(\varepsilon, Y) = 0$, $\mathrm{cov}(\varepsilon, x) = 0$, $\mathrm{cov}(\varepsilon, Z) = 0$.
- All roots of $\Phi(B) = 0$ and $\Psi(B) = 0$ lie out of the unit circle.
- Prior distributions of parameters are defined as

$$\begin{aligned}\pi_0(\Upsilon) &= \pi_0(h, \phi, \varphi, a, \sigma_\mu, \sigma_\varepsilon)\\ &= N(h_0, \Sigma_h) \times N(\phi_0, \Sigma_\phi) \times N(\varphi_0, \Sigma_\varphi) \times N(a_0, \Sigma_a)\\ &\quad \times IG\left(\frac{v_{\mu_0}}{2}, \frac{\delta_{\mu_0}}{2}\right) \times IG\left(\frac{v_{\varepsilon_0}}{2}, \frac{\delta_{\varepsilon_0}}{2}\right),\end{aligned}$$

where $h_0, \Sigma_h, \phi_0, \Sigma_\phi, \varphi_0, \Sigma_\varphi, a_0, \Sigma_a, v_{\mu_0}, \delta_{\mu_0}, v_{\varepsilon_0}$ and δ_{ε_0} are hyper parameters and already known.

The first two assumptions assure that Eq. (5) will not suffer from problems of serial correlation and heteroscedasticity. Assumption (3) assures that the series x_t is stationary. The last one assumes parameters' prior distributions, where h, ϕ, φ and a follow multivariate normal distributions and σ_μ^2 and σ_ε^2 follow inverted gamma distributions. According to Ref. 32, the prior distributions can be defined by experts when enough relevant domain knowledge is available. Otherwise, noninformative prior distributions, frequently with relatively large precisions, should be employed.

4.3. *Model estimation by the MCMC algorithm*

Based on the preliminaries described in Sec. 3.2, the following estimation method is suggested, composed of four steps:

(i) Equation (5a) is estimated to obtain the primary prediction Z by utilizing Box–Jenkins strategy.
(ii) Expert judgment e is made, where the Delphi method is employed to integrate different expert opinions.

(iii) With expert judgment e, the MCMC algorithm is implemented to jointly re-estimate Eqs. (5a) and (5b).
(iv) With the latest estimate of parameters, Z is recomputed.

The posterior conditional probability density function can be written as

$$P(\Upsilon|Y, X, e) = P(Y, X, e|\Upsilon) \times \frac{P(\Upsilon)}{P(Y, X, e)} \propto P(Y, X, e|\Upsilon) \times \pi_0(\Upsilon), \tag{6}$$

where $\pi_0(\Upsilon)$ is the prior joint probability function of Υ which is defined in Sec. 3.2. With the Bayesian theory and the relationship among variables described by Eq. (5), $P(Y, X, e|\Upsilon)$ can be transformed to

$$\begin{aligned} P(Y, X, e|\Upsilon) &= P(Y, X|e, \Upsilon) \times P(e|\Upsilon) \\ &= P(Y, X|h, \phi, \varphi, \sigma_\mu) \times P(e|h, \phi, \varphi, a, \sigma_\varepsilon). \end{aligned} \tag{7}$$

Substituting Eq. (7) into Eq. (6) generates

$$P(\Upsilon|Y, X, e) \propto P(Y, X|h, \phi, \varphi, \sigma_\mu) \times P(e|h, \phi, \varphi, a, \sigma_\varepsilon) \times \pi_0(\Upsilon). \tag{8}$$

The probability density function presented by Eq. (8) is very complex and we cannot directly draw samples from it, consequently a Metropolis–Hastings (MH) algorithm is designed, composed of six blocks: (1) coefficients of exogenous variables h, (2) AR coefficients ϕ, (3) MA coefficients φ, (4) regression coefficients a, (5) the variance of μ, i.e., σ_μ^2, (6) the variance of ε, i.e., σ_ε^2. The set of unknown quantities Υ can be estimated block by block using the MH algorithm. For more details of application of MH, Refs. 31 and 32 are helpful.

5. Empirical Study

5.1. *Data description*

This study selects the monthly data of Guangzhou Port's container throughput, downloaded from CEIC macroeconomic database, ranging from January 2005 to December 2012. The observations from January 2005 to December 2011 are taken as in-sample data for model estimation, and the remainder are treated as out-of-sample data for forecasting performance evaluation. Figure 1 visually depicted the data.

The motivation of selecting the above data lies in the two facts: (1) to our practical experience, traditional models of Guangzhou Port tend to generate predictions with broad deviation from the observations, therefore it is urgent to design a new forecasting model with higher accuracy; (2) unexpected events of strong influence frequently occurred in the period from January 2005 to December 2012, such as the great snowstorms in southern China, massive earthquake in Wenchuan of Sichuan province, earthquake in northeast Japan, U.S. subprime mortgage crisis and European debt crisis among others. This volatile circumstance will highlight the

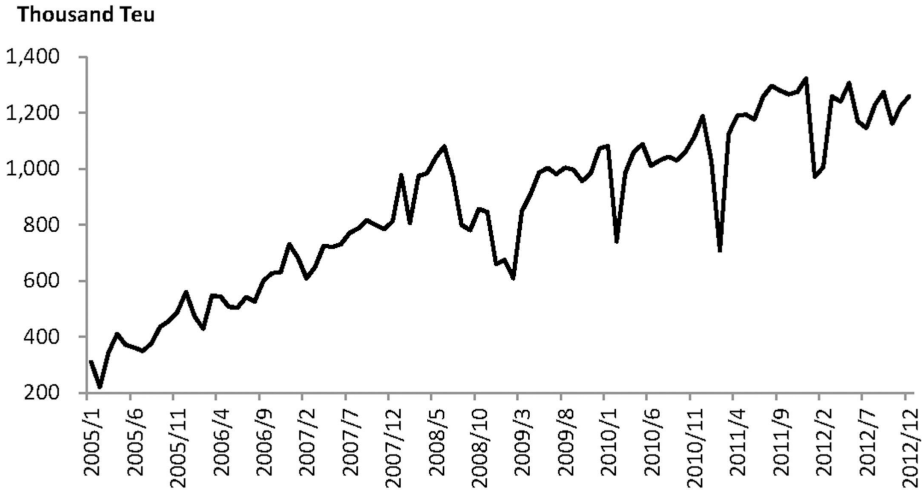


Fig. 1. Container throughput of Guangzhou Port during the period 2005–2012.

advantages of the proposed model over its rivals and that expert knowledge brings higher forecasting accuracy when the task is ill-structured.[29]

5.2. *Composition of the panel*

The panel comprises 30 experts, of whom, 1/3 are econometric modeling experts from Center for Forecasting Science Chinese Academy of Sciences, 1/3 from the frontier staff of Guangzhou Port operating container transportation business, and the remainder from the management staff of Guangzhou Port. A host distributes messages to and collects judgments from the panel by E-mail, but a panelist member cannot communicate with others within the panel, which implies that all panelists have to formulate their judgment individually.

Intuitively, the human brain is more equipped to formulate interval estimates than point estimates, therefore the panelists are required to make interval judgment in this work. Besides, rich background knowledge are provided to the panel to facilitate higher forecasting performance. The background knowledge mainly comprises three aspects, the first is the state of art in Guangzhou Port (e.g., the state of hinterland, investment, infrastructure, the number of shipping lines, historical container throughput, etc.). Besides, information on macroeconomics should be considered, including the growth speed of the world economy, the current state of the international trade, the volume of Chinese imports and exports, the volatility of fuel price, etc.

In order to avoid negative influence of those panelists that have no sufficient domain knowledge about the task, a refining strategy is used to identify and remove those incompetent experts from the panel, and finally 10 experts are remained after 10 iterations. Details of the strategy can be found in Ref. 33.

5.3. *Empirical results*

5.3.1. *Estimated coefficients*

(1) β_i estimated by ARIMA

In this section, a variable t is introduced to capture the temporal trend in that the time series is trend stationary. feb_t and $crisis_t$ are constructed to respectively capture the seasonality and the effect of the crisis, since Fig. 1 shows obvious seasonality on February and a significant decline after August 2008. By using AIC and BIC criteria, the values of p and q are set to be $p = 1$ and $q = 0$. The estimated values of the coefficients are listed in Table 1.

(2) β_i estimated by the proposed model

With $p = 1$ and $q = 0$, Eq. (5) can be rewritten as

$$\begin{cases} x_t = \beta_1 + \beta_2 \,\mathrm{crisis}_t + \beta_3 feb_t + \beta_4 t + \beta_5 x_{t-1} + \mu_t & \text{(9a)} \\ e_j = \alpha_0 + \alpha_1 \displaystyle\sum_{i=85+12\times(j-6)}^{84+12\times(j-6+1)} x_i + \alpha_2 \displaystyle\sum_{i=85+12\times(j-6)}^{84+12\times(j-6+1)} z_i + \varepsilon_j, & \text{(9b)} \end{cases}$$

where $t = 1, 2, \ldots, 96$ and $j = 1, 2, \ldots, 6$, and e_j is the jth yearly judgment made in light of the Delphi method.

Equation (9) is estimated by the proposed approach described in Sec. 3.3. It should be noted that the estimates generated by MCMC is accepted only when the algorithm converges. For the purpose of convergence diagnostic, three chains are generated for each parameter $\beta_i (i = 1, 2, \ldots, 5)$. If the three chains of each β_i come to the same level, convergence is reached. To eliminate auto-correlation and assure stationarity, the first 3000 draws of β_i are burned and 1 in every 10 draws is kept from the remainder. Hundred samples are finally remained, whose average is regarded as the estimate of β_i.

Table 2 presents the details of the estimate of $\beta_i (i = 1, 2, \ldots, 5)$ and Fig. 2 suggests that all of the estimates are converged.

5.4. *Performance comparison*

This section compares four different models in terms of their forecasting performance. ANN and ARIMA are first constructed without considering expert judgment.

Table 1. β_i estimated by ARIMA.

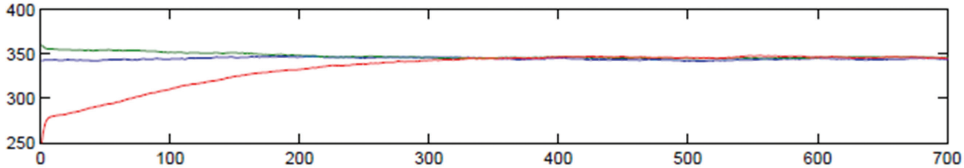
	β_1	β_2	β_3	β_4	β_5
Estimates	346.11	−218.05	−178.36	14.28	0.52
t value	32.12***	47.60***	21.04***	1.05***	0.10***
R^2	0.95				

*** Significant at 1% level.

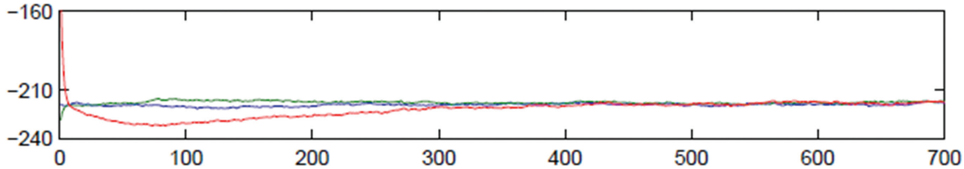
Table 2. β_i estimated by the proposed model.

	β_1	β_2	β_3	β_4	β_5
E	345.01	−218.42	−184.13	12.12	0.31
CI	[341.50, 348.11]	[−219.80, −216.90]	[−215.50, −146.01]	[11.98, 12.25]	[0.30, 0.33]

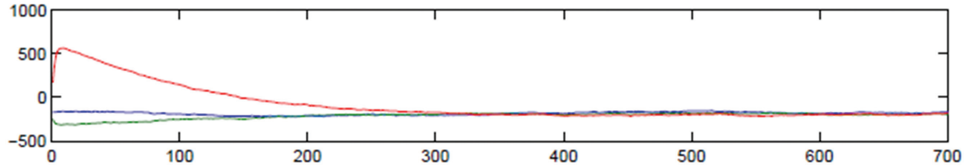
E: The estimated value of β_i; CI: The confidence interval at 1% significant level.



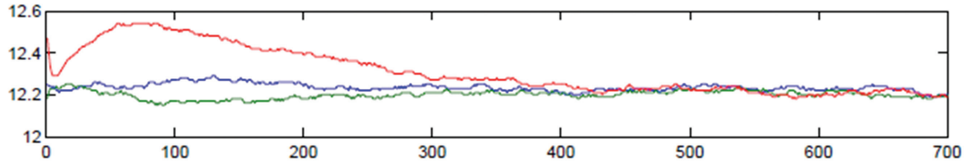
(a) Chains 1:3 for β_1



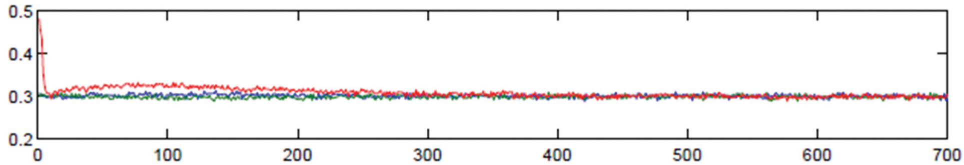
(b) Chains 1:3 for β_2



(c) Chains 1:3 for β_3



(d) Chains 1:3 for β_4



(e) Chains 1:3 for β_5

Fig. 2. Convergence diagnostic for β_1–β_5.

Table 3. The performance comparison among different models.

Model	RMSE	MAPE	TPE
ANN	198.32	15.83%	15.36%
ARIMA	219.57	17.75%	17.25%
JM	232.07	18.61%	16.29%
MBJ	89.11	5.83%	1.32%

JM is the judgment model; MBJ is the proposed model.

Then, a JM is generated based on the constructed ARIMA model. Next, MBJ, the newly proposed model, is constructed.

Three statistics are introduced to evaluate the forecasting performance including RMSE, MAPE and TPE, computed as

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2},$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|,$$

$$\mathrm{TPE} = \left|\frac{\sum_{i=1}^{n}\hat{y}_i}{\sum_{i=1}^{n}y_i}\right|,$$

where y_i is the observation and $\hat{y}_i$ is the simulated value. RMSE and MAPE evaluate the average monthly forecasting accuracy and TPE reveals the accuracy of the total quantity across the whole forecasting horizon. The comparison results are listed in Table 3.

From Table 3, we can draw the following conclusions: (1) ANN outperforms ARIMA in terms of all of the three criteria, thanks to its powerful nonlinear mapping ability; (2) the JM performs even worse than ARIMA in terms of the two monthly accuracy criteria (RMSE, MAPE) owing to experts' subjective biases and inconsistence, but better in terms of the gross accuracy (TPE). One possible reason lies in the fact that it is really difficult for experts to accurately forecast with high frequency, e.g., by month, but they are competent for low frequency forecast, e.g., by year; (3) the proposed model, MBJ, shows significant advantage over its rivals in terms of all of the three criteria, which implies that expert knowledge does improve the forecasting performance on the condition that it is used in an appropriate way.

6. Conclusions

To answer the three questions, including (1) should expert knowledge be considered when forecasting, (2) how to integrate different opinions of experts, (3) how to take advantage of expert knowledge, this paper employs the Delphi method to obtain the unified expert judgment. Then, based on the integrated judgment, a new forecasting method is proposed. Next, for the purpose of validation, the proposed model is compared with two statistical models (i.e., ANN and ARIMA) and a JM.

According to the comparison results, on the one hand, performance of the JM is worse than the two pure statistical models in terms of RMSE and MAPE, which implies that expert knowledge will play a negative role in some cases due to subjective biases and inconsistence. On the other hand, the results show the significant superiority of the proposed model over the others, which means expert knowledge can be helpful when used in a suitable way.

Consequently, we can give answers to the above questions: (1) expert knowledge should be considered when forecasting, but it needs to be used appropriately; (2) the Delphi method is effective on expert opinion integration; (3) exploiting expert knowledge by applying the combination of the Delphi method and constraint equation, as proposed by this paper, can effectively solve the problem of subjective biases and inconsistencies and thus generate more accurate predictions.

It should be noted that we currently apply the proposed model to forecasting Guangzhou Port's container throughput and more validation work involving other ports will be conducted in the future.

Acknowledgments

Financial support from the National Natural Science Foundation of China (Grant Nos. 71373262, 71390330, 71390331, 71390334) and the Fundamental Research Funds for the Central Universities (Grant No. B15JB00040) are gratefully acknowledged.

References

1. N. R. Sanders and K. B. Manrodt, The efficacy of using judgmental versus quantitative forecasting methods in practice, *Omega* **31** (2003) 511–522.
2. R. Fildes and H. Stekler, The state of macroeconomic forecasting, *Journal of Macroeconomics* **24**(4) (2002) 435–468.
3. W. Richard and M. O'Connor, Judgmental and statistical time series forecasting: A review of the literature, *International Journal of Forecasting* **12**(1) (1996) 91–118.
4. M. Lawrence, P. Goodwin, M. O'Connor and D. Onkal, Judgmental forecasting: A review of progress over the last 25 years, *International Journal of Forecasting* **22** (2006) 493–518.
5. J. S. Lim and M. O'Connor, Judgmental adjustment of initial forecasts: Its effectiveness and biases, *Journal of Behavioral Decision Making* **8** (1995) 149–168.
6. P. Meehl, *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (Jason Aronson, Northvale, NJ, 1996).
7. R. C. Blattberg and S. J. Hoch, Database models and managerial intuition: 50% model + 50% manager, *Management Science* **36**(8) (1990) 887–899.
8. H. J. Einhorn, Cue definition and residual judgment, *Organizational Behavior and Human Decision Processes*, **12** (1974) 30–49.
9. D. Kahneman and A. Tversky, Choices, values and frames, *American Psychologist* **39** (1984) 341–350.

10. Z. Shapira, *Organizational Decision Making* (Cambridge University Press, New York, 2002).
11. D. K. McClish and S. H. Powell, How well can physicians estimate mortality in a medical intensive care unit?, *Medical Decision Making* **9** (1989) 125–132.
12. N. Sanders and L. Ritzman, Bringing judgment into combination forecasts, *Journal of Operations Management* **3** (1995) 311–321.
13. T. R. Stewart, P. J. Roebber and L. F. Bosart, The importance of the task in analyzing expert judgment, *Organizational Behavior & Human Decision Processes* **69**(3) (1997) 205–219.
14. I. Yaniv and R. M. Hogarth, Judgmental versus statistical prediction: Information asymmetry and combination rules, *Psychological Science* **4**(1) (1993) 58–62.
15. R. M. Cooke, *Experts in Uncertainty: Opinion and Subjective Probability in Science* (Oxford University Press, New York, 1991).
16. R. M. Cooke and L. L. H. J. Goossens, TU Delft expert judgment data base, *Reliability Engineering & System Safety* **93**(5) (2008) 657–674.
17. A. O'Hagan, C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley and T. Rakow, *Uncertain Judgements: Eliciting Experts' Probabilities* (John Wiley & Sons, 2006).
18. P. Goodwin and R. Fildes, Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy?, *Journal of Behavioral Decision Making* **12** (1999) 37–53.
19. B. M. Ayyub, *Elicitation of Expert Opinions for Uncertainty and Risks* (CRC Press LLC, 2001).
20. R. T. Clemen and R. L. Winkler, Combining probability distributions from experts in risk analysis, *Risk Analysis* **19**(2) (1999) 187–203.
21. A. O'Hagan, Probabilistic uncertainty specification: Overview, elaboration techniques and their application to a mechanistic model of carbon flux, *Environmental Modelling & Software* **36** (2012) 35–48.
22. T. Krueger, T. Page, K. Hubacek, L. Smith and K. Hiscock, The role of expert opinion in environmental modelling, *Environmental Modelling & Software* **36** (2012) 4–18.
23. P. Goodwin and G. Wright, Heuristics, biases and improvement strategies in judgmental time series forecasting, *Omega* **22** (1994) 553–568.
24. S. J. Hoch and D. A. Schkade, A psychological approach to decision support systems, *Management Science* **42**(1) (1996) 51–64.
25. L. M. de Menezes, D. W. Bunn and J. W. Taylor, Review of guidelines for the use of combined forecasts, *European Journal of Operational Research* **120**(1) (2000) 190–204.
26. I. Fischer and N. Harvey, Combining forecasts: What information do judges need to outperform the simple average?, *International Journal of Forecasting* **15** (1999) 227–246.
27. N. Harvey and C. Harries, Effects of judges' forecasts on their later combination of forecasts for the same outcomes, *International Journal of Forecasting* **20** (2004) 391–409.
28. R. T. Clemen, A. H. Murphy and R. L. Winkler, Screening probability forecasts; contrasts between choosing and combining, *International Journal of Forecasting* **11** (1995) 133–146.
29. M. Seifert and A. L. Hadida, On the relative importance of linear model and human judge(s) in combined forecasting, *Organizational Behavior and Human Decision Processes* **120** (2013) 24–36.
30. R. Hecht-Nielsen, Theory of the backpropagation neural network, in *Proc Int Joint Conference on Neural Networks* (Washington, DC, 2007).

31. J. Geweke, Evaluating the accuracy of sampling based approaches to the calculation of posterior moments, in *Bayesian Statistics*, eds. J.M. Bernardo, J. Berger, A.P. Dawid and A.M.F. Smith (Oxford University Press, Oxford, U.K., 1992), pp. 169–193.
32. A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, (Cambridge University Press, New York, 2009).
33. A. Q. Huang, K. K. Lai, H. Qiao, S. Y. Wang and Z. J. Zhang, An interval knowledge based forecasting paradigm for container throughput prediction, *Procedia Computer Science* **55** (2015) 1381–1389.