International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:8, No:7, 2014

Forecasting Models for Steel Demand Uncertainty Using Bayesian Methods

Watcharin Sangma, Onsiri Chanmuang, Pitsanu Tongkhow

***Abstract*—**A forecasting model for steel demand uncertainty in Thailand is proposed. It consists of trend, autocorrelation, and outliers in a hierarchical Bayesian frame work. The proposed model uses a cumulative Weibull distribution function, latent first-order autocorrelation, and binary selection, to account for trend, time-varying autocorrelation, and outliers, respectively. The Gibbs sampling Markov Chain Monte Carlo (MCMC) is used for parameter estimation. The proposed model is applied to steel demand index data in Thailand. The root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) criteria are used for model comparison. The study reveals that the proposed model is more appropriate than the exponential smoothing method.

***Keywords*—**Forecasting model, Steel demand uncertainty, Hierarchical Bayesian framework, Exponential smoothing method.

I. Introduction

THE steel industry in Thailand is experiencing growing demands as the economy continues to expand and government infrastructure spending increases. Steel is a major raw material that is widely used in downstream industries, such as those in the automotive, electrical appliance and electronic, petrochemical, machinery, and packaging sectors. The Iron and Steel Institute of Thailand by [1] has reported in Fig. 1 the overall steel market has increased on average 7.31% percent per year during 1997 to 2011. Since the steel demand is uncertain, forecasting is useful for the stake holders. The steel demand index monthly time series data in [2] motivated us to find an appropriate model for forecasting the steel demand time series data in Thailand.

Times series data usually consist of trend (long term direction), seasonal (systematic, calendar related movements) and the irregular (unsystematic, short term fluctuations) components.

Outliers and autocorrelation can also be implicit in time series data. Outliers are values far from most others in a data set. The autocorrelation specifies that the values of a series of data at particular points in time are highly correlated with the value which precede and succeed them.

W. Sangma is with the Department of Industrial Engineering, Faculty of Engineering, Rajamangala University of Technology Phra Na Khon, Bangkok, Thailand (phone: 662-9132424 ext. 225 ; fax: 662-9132424 ext. 234; e-mail: watcharin.s@ rmutp.ac.th).

O. Chanmuang is with the Department of Industrial Engineering Faculty of Engineering, Rajamangala University of Technology Phra Na Khon, Bangkok, Thailand (phone: 662-9132424 ext. 226 ; fax: 662-9132424 ext. 209; e-mail: onsiri.c@rmutp.ac.th).

P. Tongkhow is with the Department of Industrial Engineering, Faculty of Engineering, Rajamangala University of Technology Phra Na Khon, Bangkok, Thailand (phone: 662-9132424 ext. 228 ; fax: 662-9132424 ext. 228; e-mail: pitsanu.t@rmutp.ac.th).

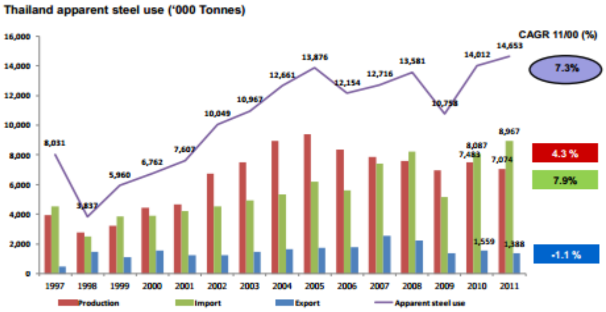

Fig. 1 Overall steel market increasing on average 7.3% per year

Some classical forecasting models are exponential smoothing, moving average, autoregressive integrated moving average (ARIMA). However, there are other methods trying to detect those components. Reference [3] described a binary selection model to account for the outliers. Reference [4] used a cumulative Weibull distribution to detect the trend. Reference [5] added a latent autocorrelation in the model. Reference [6], combining the models proposed by [3]-[5], proposed a Bayesian model for forecasting parts demand consisting of trend, autocorrelation, and outliers, and found that the performance of the proposed model was better than the exponential smoothing model.

Reference [7] modified the model of [6] by reducing the number of parameters to avoid over fitting, using noninformative proper priors instead of noninformative improper priors, and adjusting the constant in the outlier term. The noninformative priors commonly used are a normal distribution with zero mean and large variance and an inverse gamma distribution with a large scale parameter [8]. For instance, $N(0,1.0E06)$ is used for each fixed effect and the mean of each prior, and inverse gamma (IG), $IG(0.1,0.001)$, is used for the variance of each prior. Their modified model was applied to the vegetable prices data in Thailand used in their previous study [9]. They found that their modified model was the most appropriate, compared to the exponential and Seasonal ARIMA models.

The proposed model uses a cumulative Weibull distribution function, a latent first-order autocorrelation, and a binary selection, to account for trend, autocorrelation, and outliers, respectively. The proposed model is applied to steel index data. It is compared with the exponential smoothing method. The root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) criteria are used for model comparison.

This paper is organized as follows. Section II briefly describes the hierarchical Bayesian method, Gibbs sampling, cumulative Weibull distribution, binary selection, and autocorrelation. The proposed model, parameter estimation and the application are also explained in this section. In Section III, the result of the study is presented. Lastly, in Sections IV and V the discussion and conclusion are drawn.

II. Methodology and Application

A. A Hierarchical Bayesian Model

For a vector of data $\mathbf{y}=(y_1,...,y_m)^T$ and a vector of parameters $\boldsymbol{\theta}=(\theta_0,\theta_1,...,\theta_n)^T$, a hierarchical Bayesian model [10] is expressed as (1):

$$\mathrm{p}(\boldsymbol{\theta}\,|\,\mathbf{y})=\frac{\mathrm{f}(\mathbf{y}\,|\,\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\theta})}{\mathrm{f}(\mathbf{y})} \tag{1}$$

where $\mathrm{f}(\mathbf{y}\,|\,\boldsymbol{\theta})$ is a likelihood, $\mathrm{p}(\boldsymbol{\theta}\,|\,\mathbf{y})$ is a posterior distribution which stands for the marginal probability density of the parameter vector $\boldsymbol{\theta}$ given the data $\mathbf{y}$, $\mathrm{p}(\boldsymbol{\theta})$ is a prior distribution of $\boldsymbol{\theta}$, which summarizes any priori or alternative knowledge on the distribution of $\boldsymbol{\theta}$, and $\mathrm{f}(\mathbf{y})$ is the marginal distribution of data $\mathbf{y}$. The most common hierarchical Bayesian model has three levels [10]. Level 1 specifies the distributions for the data given parameters; level 2 specifies the prior distributions for parameters given hyper-parameters; and level 3 specifies the distribution for hyper-parameters. The hierarchical Bayesian model is call "hierarchical" since it has several levels.

The Gibbs sampling, a particular MCMC method, is widely used for parameter estimation. The MCMC algorithms which include random walk Monte Carlo methods are the class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution [11]. A set of vectors $\boldsymbol{\theta}$ with density $\mathrm{p}(\boldsymbol{\theta}\,|\,\mathbf{Y})$ in which the model parameters can be estimated is the final result of the MCMC.

Sampling from the posterior $\mathrm{p}(\boldsymbol{\theta}\,|\,\mathbf{y})$, $\boldsymbol{\theta}=(\theta_0,\theta_1,...,\theta_n)$, the Gibbs sampler requires a random starting point of parameters of interest, $\boldsymbol{\theta}^{(0)}=\left(\theta_1^{(0)},\theta_2^{(0)},...,\theta_n^{(0)}\right)$.

The steps of Gibbs sampling are

1) Sample $\theta_1^{(1)}$ from $\mathrm{p}(\theta_1\,|\,\theta_2^{(0)},...,\theta_n^{(0)},\mathbf{y})$.
2) Sample $\theta_2^{(1)}$ from $\mathrm{p}(\theta_2\,|\,\theta_1^{(1)},\theta_3^{(0)},...,\theta_n^{(0)},\mathbf{y})$.
   Use updated value of $\theta_1^{(1)}$.
3) Sample $\theta_3^{(1)}$ from $\mathrm{p}(\theta_3\,|\,\theta_1^{(1)},\theta_2^{(1)},...,\theta_n^{(0)},\mathbf{y})$.
   Use updated value of $\theta_1^{(1)}$ and $\theta_2^{(1)}$.
4) Sample $\theta_4^{(1)},...,\theta_n^{(1)}$ similarly to step 1 to 3
5) Sample $\boldsymbol{\theta}^{(2)}$ using $\boldsymbol{\theta}^{(1)}$ as a starting point and continually using the most updated values.
6) Repeat until we get M samples, with each sample being a vector of $\boldsymbol{\theta}^{(1)},\boldsymbol{\theta}^{(2)},...,\boldsymbol{\theta}^{(M)}$, where M is the number of samples .
7) Monte Carlo integration on those draws to the quantity of interest can be done. For example, the mean of θ_0 results from (2):

$$\mathrm{E}(\theta_0)=\frac{1}{M}\sum_{i=1}^{M}\theta_0^{(i)} \tag{2}$$

B. A Proposed Model

Let y_t be time series data at time t, $t=1,...,n$. The y_t are assumed to have a normal distribution whose mean can detect trend and autocorrelation and whose variance can detect outliers. The proposed model is created from (3) and is defined as (4)-(6):

$$y_t=\mu_t+\varepsilon_t \tag{3}$$

$$\mu_t=\gamma(\Delta w(t\,|\,\alpha,\delta)+a_t) \tag{4}$$

$$\varepsilon_t\sim\mathrm{N}\left(0,\left[\gamma(1+\zeta_t)\sigma_y\right]^2\right) \tag{5}$$

$$y_t\sim\mathrm{N}\left(\gamma(\Delta w(t\,|\,\alpha,\delta)+a_t),\left[\gamma(1+\zeta_t)\sigma_y\right]\right) \tag{6}$$

The mean of y_t is (7),

$$\mathrm{E}(y_t)=\gamma(\Delta w(t\,|\,\alpha,\delta)+a_t)\,, \tag{7}$$

and the variance of y_t is (8),

$$\mathrm{Var}(y_t)=\left[\gamma(1+\zeta_t)\sigma_y\right]^2. \tag{8}$$

where γ is the expectation of z, the sum of time series data within the study period, $w(t\,|\,\alpha,\delta)$ is a cumulative Weibull trend and $\Delta w(t\,|\,\alpha,\delta)=w(t\,|\,\alpha,\delta)-w(t-1\,|\,\alpha,\delta)$, a_t is a latent autocorrelation at time t, ζ_t is an outlier at time t, and σ_y^2 is the common variance of y_t.

A cumulative Weibull distribution function to is defined as (9):

$$\omega(t,\alpha,\delta)\begin{cases}=1-e^{-(t/\alpha)^{\delta}} & ,t\geq 0\\ =0 & ,t<0\end{cases} \tag{9}$$

Following [3], [6], [7], we use a binary selection for outliers. The observation Y_t is associated with a latent binary variable $\varsigma_t\in\{0,1\}$. The Y_t is identified as an outlier if $\varsigma_t=1$ and is not an outlier if $\varsigma_t=0$. The prior distribution for ς_t is a Bernoulli distribution such that the probability that $\varsigma_t=1$ is assumed to be about 0.05 since the outlier occurrence is a rare event.

Following [5]-[7], we use first-order autocorrelation AR(1) model to detect an autocorrelation. The AR(1) is defined as (10):

$$a_t \sim \mathrm{N}(\lambda a_{t-1}, \sigma_a^2) \quad (10)$$

Hence, the proposed hierarchical Bayesian model is as follows:

For a Bayesian method, a prior distribution is assigned to each parameter. A non informative prior assigned to each fix effect coefficient and the mean of each prior is $N(0,1.0\text{E}06)$ and to the variance of each prior is $\mathrm{IG}(0.1, 0.001)$ [8]. For an outlier, a Bernoulli (Bern), Bern(0.05), is assigned to the outlier variable, since the outlier occurrence is assumed to be about 5 %. The first-order autocorrelation model, AR(1), is used for the latent autocorrelation.

The details of priors are as follows:

Data:

$\mathrm{p}(\sigma_y^2) \sim \mathrm{IG}(0.1, 0.001)$

Trend:

$\Delta w(t \mid \alpha, \delta) = w(t \mid \alpha, \delta) - w(t-1 \mid \alpha, \delta)$, where $w(t \mid \alpha, \delta)$ is a cumulative Weibull distribution,

$\alpha \sim \mathrm{N}_{[0,\infty)}(\mu_\alpha, \sigma_\alpha^2)$, $\mathrm{p}(\mu_\alpha) \sim \mathrm{N}(0,1.0\text{E}06)$,

$\mathrm{p}(\sigma_\alpha^2) \sim \mathrm{IG}(0.1, 0.001)$, $\delta \sim \mathrm{N}_{[0,\infty)}(\mu_\delta, \sigma_\delta^2)$,

$\mathrm{p}(\mu_\delta) \sim \mathrm{N}(0,1.0\text{E}06)$, $p(\sigma_\delta^2) \sim \mathrm{IG}(0.1, 0.001)$

Latent autocorrelation AR(1):

$a_t \sim \mathrm{N}(\lambda a_{t-1}, \sigma_a^2)$, $\mathrm{p}(\sigma_A^2) \sim \mathrm{IG}(0.1, 0.001)$

$\lambda \sim \mathrm{N}(0,1.0\text{E}06)$, $a_0 = 0$

Outliers:

$\zeta_t \sim \mathrm{Bern}(0.05)$

Expectation of total observed data:

$\gamma \sim \mathrm{N}_{[0,\infty)}(\mu_\gamma, \sigma_\gamma^2)$, $\mathrm{p}(\mu_\gamma) \sim \mathrm{N}(0,1.0\text{E}06)$

Total observed data:

$z \sim \mathrm{N}(\gamma, \sigma_z^2)$, $\mathrm{p}(\sigma_z^2) \sim \mathrm{IG}(0.1, 0.001)$

Forecast one step ahead:

$$\mathrm{f}(y_{t+1} \mid y_1, \ldots, y_n) = \int \ldots \int \mathrm{f}(y_{t+1} \mid \boldsymbol{\theta}) \mathrm{f}(y_1, \ldots, y_n \mid \boldsymbol{\theta}) \mathrm{p}(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

C. Application to Steel Demand Index Data

The steel demand index monthly time series data from January, 2000 to September, 2013 (165 months) have been extracted from the database of the Iron and Steel Institute of Thailand [1]. The last year (9 months) data have been reserved for model validation, so the data set of 156 observations is used for model estimation.

A time series plot is used to explore the data. The proposed Bayesian model is applied to those data. The Gibbs sampling MCMC is used for parameter estimation and prediction via the programming in Open BUGS. We found that the Gibbs sampling MCMC was converged when it was run for 15,000 iterations after discarding the first 1,000 iterations. The rest of them were used to compute the posterior means and standard errors. Visual analysis via trace plots, history plots, kernel density plots, and autoregressive plots is used for MCMC convergence diagnostics. For the exponential smoothing, the parameters are estimated via the SPSS for Windows software. The RMSE, MAPE, and MAE are criteria for a model comparison.

III. Results

The trace plots of some response means, for instance, in Fig. 2 show that the Gibbs sampling MCMC is convergent. The parameters estimate is presented in Table I.

The RMSE, MAPE, and MAE in the estimation period (156 months) and in the validation period (9 months) are shown in Table II. Since all error measurements of the proposed model are smallest in both estimation period and validation period, the proposed model has better perform than the exponential smoothing method.

The graphs of the actual and predicted values from the proposed model in the estimation period and validation period of are shown in Figs. 3 and 4, respectively. It is evident that the predicted values from the proposed model are very close to the actual values.

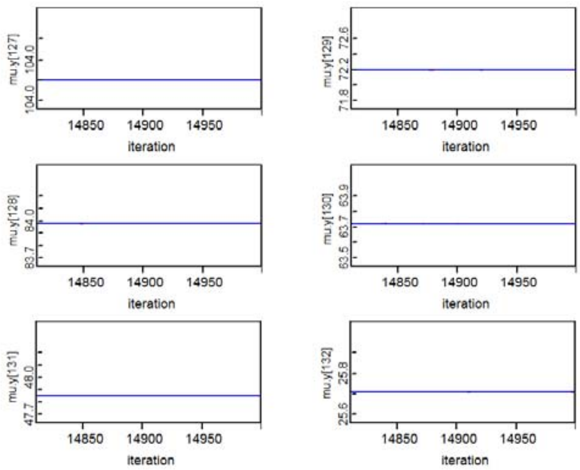

Fig. 2 Trace plots of some parameters indicating that the MCMC is converged

IV. Discussion

The proposed model can detect trend, outliers and autocorrelation components. It is appropriate for steel demand in Thailand because those extraordinary events usually occur in time series data while the traditional times series such as exponential smoothing method cannot account for all those events. The advantage of a Bayesian method is that it can solve a problem of complicated models via MCMC. The proposed model can be applied to other industrial uncertain products demand time series consisting of those extra-ordinary components. This would be valuable to anyone who would like to obtain an appropriate forecasting model for those kinds of data.

TABLE I
PARAMETER ESTIMATES SHOWING THAT μ_γ AND σ_z^2 ARE QUITE LARGE

Parameter	Value	Parameter	Value
α	149.50	μ_γ	1083.00
λ	10.59	σ_Y^2	10.19
δ	240.31	σ_α^2	33.60
γ	515.00	σ_δ^2	41.35
μ_α	57.44	σ_γ^2	30.48
μ_δ	88.63	σ_z^2	4400.00

TABLE II
MODEL COMPARISON SHOWING THAT THE PROPOSED MODEL HAS BETTER PERFORMANCE IN BOTH ESTIMATION AND VALIDATION PERIODS

Period	Method	Error Measurement		
		RMSE	MAPE	MAE
Estimation period	1. Proposed Bayesian model	1.01	5.24	1.12
	2. Exponential Smoothing (Simple Seasonal)	11.60	6.93	8.73
Validation period	1. Proposed Bayesian model	7.92	5.10	6.63
	2. Exponential Smoothing (Simple Seasonal)	8.51	5.94	7.37

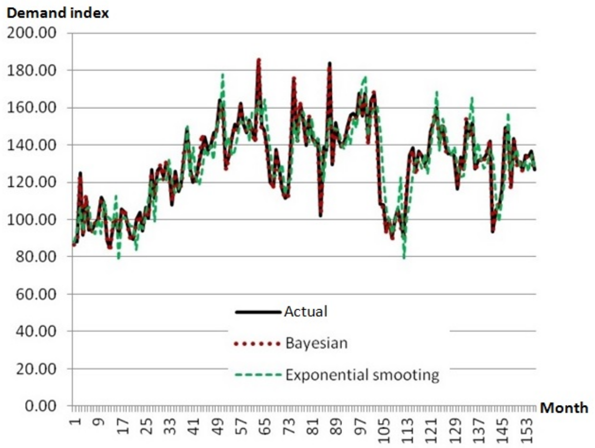

Fig. 3 Actual values and predicted values in the estimation period showing that they are very close

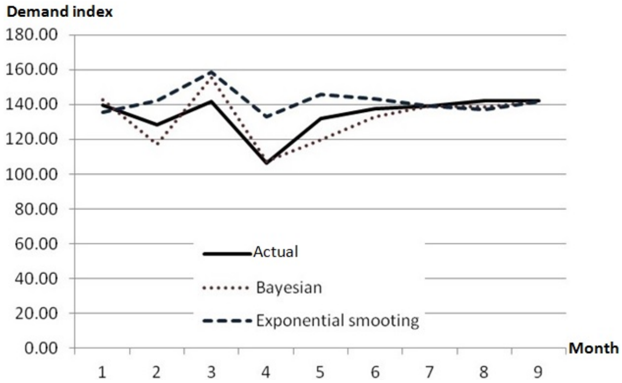

Fig. 4 Actual values and predicted values in the validation period showing that they are very close

V. CONCLUSION

The proposed model using a cumulative Weibull distribution function, first-order autocorrelation, and a binary selection, to account for trend, outliers, and autocorrelation, respectively is well-suited to forecasting the uncertain steel demand. It takes into account all main components which usually occur in time series data, resulting in more superior than the exponential smoothing.

ACKNOWLEDGMENT

Authors gratefully acknowledge the Rajamangala University of technology Phra Nakhon, Institute of Research and Development Rajamangala University of Technology Phra Nakhon and the faculty of Engineering for their technical and financial support.

REFERENCES

[1] ISIT. (2014, Jan 22). "Thailand economic and steel industry". Available: http://www.oecd.org/sti/ind/50494696.pdf.
[2] O. I. E., (2013, Sep 29). "Manufacture of basic iron and steel". Available: http://www.oie.go.th/en/academic/index.
[3] P. Congdon, *Applied Bayesian Modelling*, John Wiley & Sons: New York, 2003.
[4] W.W. Moe and P.S. Fader, "Using advance purchase orders to forecast new product sales," *Marketing Science*, Vol. 21, No. 3, 2002, pp. 347-364.
[5] M. West and P.J. Harrison, Bayesian Forecasting and Dynamic Models, 2nd ed., New York: Springer-Verlag, 1997.
[6] P.M. Yelland, "Bayesian Forecasting of Part Demand," International Journal of Forecasting, Vol. 26, 2010, pp. 374-396.
[7] P. Tongkhow and N. Kantanantha, "Bayesian Model for Time Series with Trend, Autoregression and Outliers," Proceedings of the ICT&KE 2012, IEEE, Bangkok Thailand, 2012, pp. 90-94, 2012.
[8] A. Gelman, "Prior distributions for variance parameters in hierarchical models," Bayesian Analysis, Vol. 1, No. 3, 2006, pp. 515–533.
[9] P. Tongkhow and N. Kantanantha, "Appropriate forecasting models for fluctuating vegetable prices in Thailand," IEEE Proceedings, The CET 2011, Shanghai, 2011, pp.162-166.
[10] P. Congdon, Bayesian Statistical Modelling, 2nd ed., John Wiley & Sons: New York, 2006.
[11] C.P. Robert and G. Casella, Monte Carlo Statistical Methods, 2nd ed., Springer: New York, 2004.