Pang G, Gebka B.

Forecasting container throughput using aggregate or terminal-specific data?

The case of Tanjung Priok Port, Indonesia.

International Journal of Production Research 2016

DOI: <http://dx.doi.org/10.1080/00207543.2016.1227102>

Copyright:

This is an Accepted Manuscript of an article published by Taylor & Francis in *International Journal of Production Research* on 29/08/2016, available online:

<http://www.tandfonline.com/10.1080/00207543.2016.1227102>

DOI link to article:

<http://dx.doi.org/10.1080/00207543.2016.1227102>

Date deposited:

25/08/2016

Embargo release date:

29 August 2017

This work is licensed under a

[Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Forecasting container throughput using aggregate or terminal-specific data? The case of Tanjung Priok Port, Indonesia.

Gu Pang^a, Bartosz Gebka^{a*}

^a, *Newcastle University Business School, 5 Barrack Road, Newcastle upon Tyne, UK*

This version: 11 August 2016

Forthcoming in: International Journal of Production Research

Abstract: We propose a new approach to forecasting total port container throughput: to generate forecasts based on each of the port's terminals and aggregate them into the total throughput forecast. We forecast the demand for total container throughput at the Indonesia's largest seaport Tanjung Priok Port, employing SARIMA, the additive and multiplicative Seasonal Holt-Winters, and the Vector Error Correction Model on the monthly port and individual terminal container throughput time series between 2003 and 2013. The performance of forecasting models is evaluated based on Mean Absolute Error and Root Mean Squared Error. Our results show that the multiplicative Seasonal Holt-Winters model produces the most accurate forecasts of total container throughput, whereas SARIMA generates the worst in-sample model fit. The Vector Error Correction Model provides the best model fits and forecasts for individual terminals. Our results report that the total container throughput forecasts based on modelling the total throughput time series are consistently better than those obtained by combining those forecasts generated by terminal-specific models. The forecasts of total throughput until the end of 2018 provide an essential insight into the strategic decision-making on the expansion of port's capacity and construction of new container terminals at Tanjung Priok Port.

Keywords: container throughput, forecasting, SARIMA, Holt-Winters method, Vector Error Correction Model

* Corresponding author, email: b.t.gebka@ncl.ac.uk.

1. Introduction

In this study, we propose a novel approach to forecasting total port container throughput: rather than forecasting total throughput directly, we suggest to utilize data from individual port's terminals, both individually and with account for interdependencies between terminals' throughputs, to obtain the aggregate, total port's throughput. This approach complements the prevailing practice of forecasting total throughput directly, which by its design ignores the potential predictive content of subcomponent (individual terminals') throughputs as well as of interactions among those subcomponents.

As an empirical, policy-relevant illustration of our approach, we forecast the demand of total container throughput at Tanjung Priok Port. Four forecasting models are employed: SARIMA, the additive Seasonal Holt-Winters (ASHW), the multiplicative Seasonal Holt-Winters (MSHW), and the Vector Error Correction Model (VECM). We are also interested in producing the most accurate forecasts of the port's total container throughput until the end of 2018; this provides an essential insight into the strategic decision-making on the recent capacity expansion project and construction of new container terminals at Tanjung Priok Port. Our results suggest that the maximum capacity planned by Indonesian Port Corporation seems to be sufficient for the expected volume of container traffic, with some margin to accommodate unexpectedly high seasonal throughputs.

As the world biggest archipelago country, Indonesia's sea transportation is considered to be the engine of its economic growth and international trade. Because of Indonesia's strategic and geographical location, over 90% of the country's international trade is transported by sea. In 2012, Indonesia's container ports handled total container of 9.3 million TEUs with the rank position of 15 from the world container port traffic (The World Bank 2014). This trend continues due to the country's economic development and rising share of

containerised cargo for international and domestic trade. As a result, Indonesia's port authorities and port corporations are urged to expand port capacity by the construction of new ports and terminals of existing ports.

Prior to the Tanjung Priok Port's expansions, which commenced in early 2013, the port had four container terminals. These are the Jakarta International Container Terminal (JICT), the KOJA Container Terminal, the MTI Container Terminal, and the conventional terminal. The port handles approximately 75% of Indonesia's total container traffic. In 2013, it handled approximately 6.4 million TEUs, up from 6.2 million TEUs in 2012 (Port Finance International 2014). These figures suggest that the port's capacity of 5 million TEUs have already been substantially exceeded. Due to the increase in container traffic and lack of major port development over a number of decades, Tanjung Priok port has become severely congested.

The Indonesian Port Corporation's expansion project is being implemented in two phases, with the completion of phase one and two in year 2018 and 2023, respectively. The project will expand Tanjung Priok Port by adding seven new container terminals (three of which with total capacity of 4.5 million TEUs in phase one, and four of which with total capacity of 8 million TEUs in phase two). This will result in a total increase of the port's container handling capacity to approximately 18 million TEUs by 2023 (see Table 1). The decision to be made on developing port facilities and the construction of new ports/ terminals is strategic. Such a decision-making exercise requires a thorough investigation and determination of accurate port's container throughput forecasts. This information does not only provide vital indication on the future trend of total number of containers handled over a period of time, but also assists port management in identifying emerging port under/ over capacity issues. In addition, efficient utilisation of existing and new capacities, including yard

storage space and yard-to-berth planning, is another area of managerial optimisation (Li and Yip 2013, Roy et al. 2016).

Insert Table 1 about here

This study contributes to the literature on forecasting container throughput in several ways. First, to the authors' best knowledge, we are the first to examine whether directly forecasting the total throughput using an univariate model (for the total throughput series) is superior to the forecasts of the total throughput determined by combining the best individual terminal forecasts. In contrast, existing studies focus exclusively on the former approach (see Section 2 for more details).¹ Second, in the container throughput literature, we are the first to propose to use the VAR/VECM method on terminal data, which also accounts for cross-correlations among individual components (i.e., container terminals). Third, this study presents the port senior management strategic implications aimed at ensuring whether the recently initiated capacity expansion can indeed accommodate the expected rising container throughput in the near future. Hence, our forecasts are not only of academic interest, but can also be used to assess a concrete, real-life port expansion plan. Last but not least, we believe that academic scrutiny of public infrastructure projects is important, especially that public funding may be insufficient and private investors' participation can be discouraged by political risk and bureaucratic complexities.

The structure of this paper is organised as follows. Section 2 reviews the literature on single and hybrid forecasting methods for container throughput. Section 3 presents our methodology. Section 4 describes the data and discusses the solution procedure. The analysis

¹ Terminal-level analysis has been implemented in the strand of the literature dealing with historical efficiency of ports (see, e.g., Cullinane et al., 2006, for a review). We thank an anonymous reviewer for pointing it out.

of the data and the results are discussed in Section 5. Section 6 provides the robustness checks. The forecasts of total throughput until the end of 2018 are presented in Section 7. Section 8 concludes the study.

2. Literature review

2.1 Single forecasting methods for container throughput

Fung (2002) developed a simple oligopoly model to study the competitive relationship among three players: the Hong Kong terminal, the Hong Kong midstream, and the Singapore terminals. The structural error-correction model (and the structural vector autoregressive model for the purpose of robustness check) was proposed to forecast the container throughput at Hong Kong's container terminals. The structural model produced a higher and more accurate growth path for future container throughput. The resulting forecasts suggested an earlier construction of new container terminals to meet future demand in Hong Kong.

Schulze and Prinz (2009) studied container transshipment at German ports to three destinations Asia, Europe, and North America. They also forecast the total container throughput in Germany. The authors analysed quarterly time-series data between 1989 and 2006. The data showed a nonlinear trend with seasonal variations of container transshipment, i.e., a strong downward movement at the beginning of each year. Schulze and Prinz (2009) proposed two approaches: the SARIMA model and the Holt-Winters exponential smoothing approach. Based on two forecasting measures, RMSE and Theil's U, the SARIMA models performed slightly better than the Holt-Winters approach. The out-of-sample forecasts between 2007 and 2008 indicated a permanent dynamic growth of German container throughput to destinations such as Asia and Europe, and a moderate increase for North America.

Peng and Chu (2009) analysed monthly data between 2003 and 2006 on container throughput volumes for each of the three major ports in Taiwan: Keelung, Taichung, and Kaohsiung. The time series data exhibited a seasonal pattern, with a sharp reduction in the throughputs in February of each year for every port. Peng and Chu (2009) split the data into two sets: an in-sample dataset for the estimation from January 2003 to December 2005, and an out-of-sample dataset for the prediction between January 2006 and December 2006. The authors proposed and compared six univariate forecasting models. Based on MAE, MAPE and RMSE, the classical decomposition was the best forecasting model for two ports out of three. In these cases, the authors indicated that SARIMA was not as good as the classical decomposition method, possibly due to overfitting with limited number of observations. The classical decomposition and SARIMA model provided equally good predictions for one port out of all three. This led to the conclusion that different methods were better (or worse) for different ports. In line with Makridakis *et al.* (1982), Peng and Chu (2009) agreed that no single forecasting model could be claimed as the best for all situations under all circumstance. Due to the scope of their study, Peng and Chu (2009) did not provide out-of-sample predictions; rather, they concentrated on short-term forecasting and aimed at finding a model that was relatively simple, practical, and highly accurate for their dataset.

Chen and Chen (2010) also studied monthly data on container throughput volumes generated by three major ports in Taiwan: Keelung, Taichung and Kaohsiung, between 1978 and 2006. Their nonstationary time series had significant upward trends and seasonal cycles. The authors suggested that Chinese New Year had a significant impact on container throughput at all ports. Three models were proposed, including genetic programming, decomposition approach, and SARIMA. All methods produced MAPE below 4%. The genetic programming approach produced good results compared with other two methods with

MAPE being 35% lower. Chen and Chen (2010) evaluated the 2007 container throughputs of ports and predicted port throughputs for 2008 to experience a steady increase.

To overcome the nonlinearity in a time series, nonlinear methods such as artificial neural networks (ANNs) (see, e.g., Sarvareddy *et al.* 2005, Lam *et al.* 2006, and Gosasang *et al.* 2011) were proposed to produce container and freight forecasts. According to Hornik *et al.* (1989), ANN models are capable of approximating a nonlinear mapping with any degree of complexity and without prior knowledge of the underlying functional form. In the literature, ANN models are often compared with traditional linear forecasting models such as ARIMA. The studies of Lingras *et al.* (2000) and Vlahogianni *et al.* (2005) conclude that ANNs provide better forecasting performance than ARIMA models. However, the findings of Taskaya-Temizel and Casey (2005) provided an opposite conclusion. Kamariannakis *et al.* (2010) suggest the reason might be that most single forecasting methods are not able to capture and model the nonlinear complexity and constant transitions to other conditions of the time series. Both ARIMA and ANNs demonstrate high quality forecasts in their domains. However, both models show weaknesses when nonlinear and linear patterns are present simultaneously in the time series. This is often the case for a real-life time series containing both nonlinear and linear patterns.

Tsai and Huang (2015) apply ANNs to predict container flows by considering GDP, interest rates, the value of export and import trade, the numbers of export and import containers and the number of quay cranes. The ANNs is developed mainly for data mining purposes, and the model can simultaneously predict container flows between the major Asian ports. Their reported forecasts indicate that the forecasting errors are relatively small in most studied ports. Therefore, the authors suggest that their study is beneficial to shipping companies who can use the container flow forecasting model to make operational decisions.

2.2 Hybrid forecasting methods for container throughput

According to Xie *et al.* (2013), though single methods are not unconditionally superior in container throughput forecasting, their combinations, i.e., hybrid approaches, generate better forecasts. Single methods are capable of capturing only linear component of time series. On the contrary, hybrid approaches can produce better forecasting performance, especially when the container throughput series has complex characteristics such as seasonality, nonlinearity, and complexity.

Several hybrid models combining (S)ARIMA and Artificial Neural Networks (ANNs) are proposed to forecast time series achieve good prediction performance, for example, Tseng *et al.* (2002), Zhang (2003), Hansen and Nelson (2003), Khashei and Bijari (2010). These studies report the hybrid models are able to outperform each component model used in isolation. For the recent development and good descriptions of the hybrid ARIMA-ANN forecasting approach, readers may refer to Aburto and Weber (2007), Jain and Kumar (2007), Ömer Faruk (2010), Shi *et al.* (2012), and Wang *et al.* (2012). Vlahogianni *et al.* (2005) and Wei and Chen (2012) combined ANNs with generic approaches, clustering methods and empirical mode decomposition techniques respectively to improve final forecasts.

Most recently, Xie *et al.* (2013) studied monthly data between January 2001 and August 2012 on container throughput volumes for Shanghai Port and Shenzhen Port. These authors proposed three hybrid approaches: SARIMA-LSSVR, SD-LSSVR, and CD-LSSVR. Based on Diebold-Mariano (DM) statistic test, Xie *et al.* (2013) reported that a single method could be occasionally more competitive than a proposed hybrid approach. The proposed hybrid approaches achieved the highest accuracy from the level measurement such as the RMSE, MAE and MAPE. Among other single forecasting methods, SARIMA's forecasting performance was excellent. This is because SARIMA makes use of seasonal adjustment, which is not considered by single LSSVR and SVR models. To evaluate the robustness of

different methods, three different ratios (85%, 90% and 95% respectively) of training dataset to sample sizes were used. The authors suggest that the description of the seasonality, complexity and nonlinear characteristics of container throughput series is indeed important for good forecasting performance. Their study shows that decomposition and the ‘divide and conquer’ principle can improve performance in the case of container throughput forecasting at ports. Moreover, Ruiz-Aguilar, Turias and Jiménez-Come (2014) investigated the daily records of number of goods subject to inspection in the Algeciras Border Inspection Posts between January 2010 and December 2012. They proposed hybrid SARIMA-ANN models with 16 combinations. Based on the forecasting performance criteria such as R^2 , d , MSE, and MAE, a hybrid model (SARIMA-ANN) outperformed either of the SARIMA and ANN models used separately. However, in several instances the traditional approaches perform better than the hybrid ones, hence the general superiority of the later cannot be clearly established, especially that ex-ante the best hybrid model would be unknown.

In summary, although the literature on container throughput forecasting employs a variety of techniques, we identify the following gaps. Firstly, existing studies tend to employ historical total throughput series to forecast its future values, rather than utilizing information potentially contained in terminal-level data. Secondly, correlations among terminal throughputs are not analyzed in the literature but can potentially provide additional information for improved port-level forecasts. Thirdly, VAR/VECM methodology on terminal data has not been employed in this branch of the literature. These gaps are addressed in the following sections.

3. Methodology

3.1 Unit root properties of time series

The existence of stochastic trends in time-series variables poses a potential problem for

statistical inference. Traditional significance tests are not valid under conditions of non-stationarity when a series contains unit roots. In addition, using non-stationary time series in multiple regression models will lead to the problem of spurious regressions. This is an incorrect identification of causality between variables. Hence, any time series has to be tested for the existence of unit roots. If it is non-stationary, amendments to the variable and/ or the model should be applied. Specifically, in the augmented Dickey-Fuller test (Said and Dickey, 1984), a variable X_t is tested for stationarity (unit roots) using the following regression:

$$\Delta X_t = \alpha_0 + \alpha_1 t + \delta X_{t-1} + \sum_{p=1}^P \beta_p \Delta X_{t-p} + \varepsilon_t ,$$

where ΔX_t denotes the first difference of X_t (i.e., $\Delta X_t = X_t - X_{t-1}$). The Null hypothesis is that of unit roots/ non-stationarity, which is the case for $\delta = 0$, while the alternative hypothesis is that of stationarity of X_t , which is given if the parameter δ is significantly lower than zero (i.e., $\delta < 0$). The value of the test statistic has to be compared to the critical values tabulated for this test, as the t-statistic does not follow the t-distribution under the Null. If a variable is non-stationary in levels but the first (k-th) difference is stationary, the variable is said to be integrated of order one (order k): $X_t \sim I(k)$.

3.2 Cointegration of non-stationary variables and the VECM

If two or more variables are non-stationary but of the same order of integration, they might follow one or more common stochastic trends (i.e., be cointegrated). Specifically, if $X_t^1 - X_t^4$ represent container throughput for each of four terminals (see Section 1 for more details), they can share up to $4-1=3$ common trends (i.e., the cointegrating vectors). If those variables are $I(1)$ and stacked in a 4-dimensional vector $\bar{X}_t$, a Vector Error Correction Model (VECM) can be estimated and used for forecasting:

$$\Delta \bar{X}_t = \delta + \sum_{p=1}^P \Gamma_p \Delta \bar{X}_{t-p} + \Pi \bar{X}_{t-1} \varepsilon_t$$

The rank of the matrix Π is indicative of the number of cointegrating relationships between $X_t^1 - X_t^4$ (Johansen 1991)

3.3 SARIMA

The SARIMA models are based on the autoregressive moving average models (e.g., Box, Jenkins and Reinsel 2008). An autoregressive model assumes that a variable X_t depends on its own p past values (i.e., $X_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_p X_{t-p} + \varepsilon_t$). This is called an AR(p) model. The autoregressive order p is not known *a priori*, and has to be established empirically for each time series analysed. In addition, the error term ε_t can follow an autoregressive model itself. For example, in an moving average model of order q , MA(q): $Y_t = \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \dots + \gamma_q \varepsilon_{t-q}$. By combining those two models, we obtain an ARMA(p, q) model. If the variable of interest X_t is non-stationary, it cannot be used directly in the ARMA model but has to be transformed into a stationary one. For example, if $X_t \sim I(1)$, then its 1st difference ($\Delta X_t = X_t - X_{t-1}$) is stationary and can be modelled by an ARMA process. The ARMA(p, q) becomes an ARIMA(p, d, q), with d denoting the order of integration of the original variable ($d = 1$ in our case). Both AR and MA components can be characterised by seasonal effects, at lag S . For example, S could equal to 12 for monthly data to capture reoccurring annual effects. Taking those effects into account leads to the SARIMA(p, d, q)(P, D, Q) $_S$ model, where p , d , and q are defined as above, S denotes the number of seasonal lags ($\Delta_S X_t = X_t - X_{t-S}$, $S=12$ for monthly data), D denotes how many times that original time series have been seasonally differentiated, P (Q) stands for the number of seasonal AR(MA) terms (Box and Jenkins 1970, Brockwell and Davis 1991,

Harvey 1993, Franses 1996, Box, Jenkins and Reinsel 2008, Enders 2014). More formally, the SARIMA formula is:

$$\rho(L^p)\rho_S(L^P)\Delta^d\Delta_S^D X_t = \theta(L^q)\theta_S(L^Q)\varepsilon_t,$$

where L stands for the lag operator, i.e. $L^k X_t = X_{t-k}$. The value of parameter d is established via the unit root tests (see Section 3.1). For monthly data, $S=12$ is set. The exact values of parameters p, g, P, D, Q are established in our study based on a grid search, with models having all possible combinations of parameter values $(p, g, P, D, Q) \in \{0,1,2\}$ estimated. The model with the best fit, based on the Bayesian Information Criterion (Schwarz, 1978), is selected for further analysis and forecasting. This procedure is repeated for each time series used.

3.4 Additive Seasonal Holt-Winters (ASHW)

The additive Seasonal Holt-Winters approach is based on a simple exponential smoothing filter (Gardner 2006, Box, Jenkins and Reinsel 2008). It allows for the time trend and seasonal effects in the data. An exponential filter generates a prediction of the mean of X_t for the next period as $\hat{X}_{t+1} = \alpha X_t + (1 - \alpha)\hat{X}_t$, where α is the smoothing parameter. For the ease of comparability, this can also be presented as $\hat{X}_{t+1} = a_t$, where a_t is the updating equation: $a_t = \alpha(X_t) + (1 - \alpha)a_{t-1}$.

The additive version captures the processes of a form: $X_t = (\beta_0 + \beta_1 t) + s_t + \varepsilon_t$, where the magnitude of the seasonal effect does not increase over time. A τ -step forecast can be then computed as $\hat{X}_{t+\tau} = a_t + b_t \tau + s_{t+\tau-L}$, where L is a seasonality order and a_t, b_t , and s_t are updating equations, and α, β and γ are smoothing parameters, for the level, trend, and seasonal factor, respectively, which can be computed by minimising the sum of squared forecast errors.

$$a_t = \alpha(X_t - s_{t-L}) + (1 - \alpha)(a_{t-1} + b_{t-1})$$

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$$

$$s_t = \gamma(X_t - a_t) + (1 - \gamma)s_{t-L}$$

3.5 Multiplicative Seasonal Holt-Winters (MSHW)

The multiplicative version of the Seasonal Holt-Winters filter captures the processes of a form: $X_t = (\beta_0 + \beta_1 t)s_t + \epsilon_t$, where the magnitude of the seasonal effect increases over time (Gardner 2006, Box, Jenkins and Reinsel 2008). A τ -step forecast can be then computed as $\hat{X}_{t+\tau} = (a_t + b_t \tau)s_{t+\tau-L}$, and the updating equations for the level, trend, and seasonal effects are:

$$a_t = \alpha(X_t/s_{t-L}) + (1 - \alpha)(a_{t-1} + b_{t-1})$$

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$$

$$s_t = \gamma(X_t/a_t) + (1 - \gamma)s_{t-L}$$

4. Data and solution procedure

In this study, we collected aggregated monthly container throughput data of Tanjung Priok Port between January 2003 and November 2013. The monthly container throughput data of each of the four container terminals (JICT, KOJA, MTI container terminals, and the conventional terminal) was also obtained (see Figure 1). We note an unprecedented decline of container volume in the last quarter of year 2008, and it continues to worsen in early 2009 due to the recent global recession.

Insert Figure 1 about here

This total sample was divided into two subsamples: the estimation period containing all but the last year of data, and the in-sample forecasting period, comprising observations in the last year of the sample. Each model (SARIMA, SWHA, SWHM and VECM) is fitted in an optimal way to each time series (the total throughput as well as throughputs in each terminal, where applicable) in the estimation period. The model fit for each model and time series is obtained. Based on those estimates, each optimal model is used to generate dynamic predictions for the last year of the sample. Those predictions are then compared with the actual observed values for each time series.

In addition, to forecast the total throughput, rather than utilising models fitted to total time series we also combine forecasts for individual terminals. For instance, the total throughput can be predicted using a sum of SARIMA forecasts (for each individual terminal). This sum can be equally weighted (with each terminal/ prediction carrying a weight of 1), or the weights are calculated based on the fit of each model/ terminal in the estimation sample. More specifically, the weights for terminal i , $i=1, \dots, 4$, are calculated according to the following formula:

$$w_i = \frac{4 \frac{E^{max}}{E^i}}{\sum_{i=1}^4 \frac{E^{max}}{E^i}} \times 100\%$$

Here, E^i denotes the forecasting error in the estimation period, measured either by RMSE or MAE, and E^{max} stands for the maximum error across all four methods, measured either by RMSE or MAE, for any given aggregated/ individual terminal time series. Results based on weights using RMSE (MAE) are denoted as “weighted sum 1” (“weighted sum 2”). For every series, these error measures are standardised by dividing them by the average of the variable (throughput) being predicted, to assure comparability of errors across terminals of varying throughput magnitudes. Those weights add up to four, as in an unweighted sum of

four predicted throughputs, ensuring comparability, and are higher for terminals with better model fits (lower forecast errors) in the estimation sample. Hence, the overall contribution of each terminal forecast to the total throughput forecast is determined, firstly, by the magnitude of this terminal's throughput (as large terminals will have higher observed and, hence, predicted throughputs), and the relative precision of its throughput forecast (unless weights w_i are set to equal one for unweighted sums of forecasts).

To fulfil the aim of this study, we analyse whether individual terminal forecasts can be combined by selecting the best fitting model for each terminal separately. For instance, if the best fit for terminal JICT is observed for SARIMA, for terminal KOJA: SWHA, for terminal MTI: SWHM, and for conventional terminal: VECM, those individual forecasts are combined to obtain a total container throughput. It can be noted that this approach only accounts for possible cross-correlations among individual throughputs when the VECM method is used. This is a further argument as to why the use of VAR/VECM models could be fruitful in container forecasting, which constitutes one of our contributions to the literature.

In a related literature on forecast combinations, where multiple forecasts of the same target variable are available, explicit account of correlations between forecast errors has been proposed (Bates and Granger 1969). However, as this approach requires estimation of more complex combination weights, it introduces further estimation errors into the analysis, potentially resulting in inferior forecasts (Smith and Wallis 2009). Hence, as Timmermann (2006) notices, the literature tends to advise to ignore correlations, especially when sample size is small. In our sample, cross correlations between forecast errors are not excessive and lay in the range .19-.40. Moreover, even using individual errors to construct weights appears to be problematic, as hundreds of empirical studies find that using simple weights (averages) instead generates superior forecasts, a phenomenon termed “forecasting combination puzzle” (Clemen 1989, Stock and Watson 2004, Timmermann 2006, Wang and Petropoulos 2016). In

addition, in our context the future fractions/contributions of terminals to the total throughput would have to be optimally forecasted as well, something for which an analytical solution does not exist, and which would induce existence of additional estimation errors. Hence, when combining forecasts we use simple sums and errors-based weights as described above. The literature on combining forecast of the same variable, i.e., a set-up similar but simpler than ours, overwhelmingly documents the superiority of combined versus individual forecasts (e.g., Clemen 1989, Stock and Watson 2004, Timmermann 2006, Wang and Petropoulos, 2016).

On theoretical grounds, it is not clear whether forecasting the total throughput using an univariate model for the historical total throughput series is superior to the forecasts of the total throughput obtained by summing up best individual terminal forecasts. On the one hand, even when applying the same model (e.g., SARIMA) to terminal time series, different specifications (parameter values) of that model might be optimal for different terminals. Hence, forecasts for individual terminals obtained using terminal-specific models might be more precise (as based on better-fitting models) and their sum might provide a better prediction for the total throughput than a forecast of the total throughput series based on one specific model only, applied to that series. In other words, the latter implicitly imposes the same model (with identical parameters) to all terminals (i.e., components of the total throughput), whereas in reality the time series behaviour of each terminal might be unique and different from that of other terminals. Hence, we hypothesise that estimating an optimal model different for each terminal and summing up the obtained forecasts could provide a better forecast. Potentially superior forecasts of total throughput could be obtained if one allows each terminal to be described, and each forecast to be generated, by a best-fitting model (e.g. SARIMA for terminal JICT, VECM for terminal KOJA, etc.).

On the other hand, arguments can also be made in favour of modelling the total rather than individual throughputs. With the absolute terminal throughputs being smaller than the total sum of throughputs, these values are more susceptible to random shocks. Hence, a forecasting model fitted to each of those series, each potentially largely affected by shocks, could suffer from poor fit and generate error-ridden forecasts. If those individual shocks are random, then modelling and forecasting total throughput can result in better fit and forecasts, as shocks will cancel out each other, at least to an extent. Moreover, patterns in throughput of individual terminals, with volumes smaller than the total port's throughput, can be more prone to structural changes over time. Hence, even if terminal-level model fit is good in the estimation period, the optimal model might not be stable over time, resulting in sub-optimal forecasts out-of-sample. Again, modelling and forecasting the total throughput can help to alleviate this instability of models through aggregation of data across terminals, as long as not all terminals change their time series behaviour at the same time. Therefore, constructing forecasts based on individual terminals rather than the aggregated throughput might result in superior forecasting performance. This reasoning is in line with findings in the literature on top-down versus bottom-up forecasting (e.g., Schwarzkopf *et al.* 1988). However, arguments to the contrary can also be raised.

5. Results of full sample

5.1 Results of unit root tests and Johansen cointegration tests

As discussed in Section 3, we first analysed the order of integration. The results of Augmented Dickey-Fuller (ADF) unit root tests are presented in Table 2, Panel A. The optimal lag length p is determined in each case based on the Schwarz's Bayesian Information Criterion (Schwarz 1978). The left-hand side of Table 2 contains the results obtained from tests on original, non-differentiated time series (i.e., X_t). Test statistics for all series are lower

than the relevant critical values. This indicates that the throughput data is non-stationary at 1% level and even at 5% level for all but one case. However, when we take the first difference of each series (i.e. $\Delta X_t = X_t - X_{t-1}$), those newly created variables are stationary at any conventional significance level, as demonstrated on the right-hand side of Table 2. Hence, given that the original (non-differentiated) data is non-stationary but the first differences are, we can conclude that each throughput series is integrated of order one, $X_t \sim I(1)$.

Insert Table 2 about here

Furthermore, given that our terminal series are non-stationary and integrated of the same order (one), they could share common stochastic trends (i.e., be cointegrated). Hence, we test for cointegration among all four terminals by means of the Johansen test (Johansen 1991). The total series is excluded here, as it is the sum of all four terminals' throughputs. Hence, including the total series into the VAR/VECM models would result in perfect multicollinearity and a failure to estimate any model. These results are presented in Table 3, Panel A. There exists one cointegrating relationship among those four time series (i.e., they are cointegrated). This is apparent as the test statistic for the hypothesis that the rank is 0 (or lower) can be rejected. However, the hypothesis of the rank being 1 at the maximum cannot be rejected in favour of an alternative hypothesis that the rank is equal to 2 at maximum. Hence, there exists one cointegrating vector. Therefore, to capture the full dynamics of causality among the terminal series, a VECM, rather than a VAR, model should be used to account for this cointegration.

Insert Table 3 about here

5.2 Results of in-sample model fits

Table 4 presents model fits for all time series (the total throughput and four individual terminals' throughputs) and forecasting methods, obtained for the estimation period up to and including year 2012. We focus on two measures of fit, the root of mean square error (RMSE) and the mean absolute error (MAE), rather than “traditional” fit measures such as R-squared, information criteria, and the like. The same fit measures will be applied for the in-sample forecasts for the year 2013. Hence, the choice of RMSE and MAE allows for comparability and consistency. All values in Table 4 are in percentage points, representing the in-sample errors as compared to the average value of the variables. For example, 15.9048 indicates that using the SARIMA model for the total data, the “average” monthly error equals 15.9048% of the average monthly total throughput in the same period.

Insert Table 4 about here

We compare the model fit when the methods are applied to the total versus individual terminal series. For both RMSE and MAE, modelling total throughput is less error-ridden than modelling individual terminal throughputs (except the KOJA terminal modelled using SHW methods and model fit using the RMSE criterion). This indicates that terminal throughputs generally suffer from more noise than the total throughput. The best model fit is obtained from the MSHW method.

Despite its popularity and good empirical performance reported in the literature, the SARIMA model tends to generate largest errors in-sample. This indicates, at the very least, that conclusions from one study/ port/ period regarding the choice of the optimal forecasting method should not be automatically applied to another port/ time period. For methods where model fit could be obtained for both terminal-level and total series (all except for the VECM), at the terminal level the ASHW method seems to perform best in most cases, followed closely by its multiplicative counterpart.

Moreover, the VECM, which has been largely ignored in this branch of the literature, seems to generate the most precise model fits. The reported errors are the smallest, for most terminals and both model fit measurements. However, it is outperformed by the ASHW method in a few cases. We predict that if this superior model fit in-sample translates into superior forecasts out-of-sample, the VECM could yield the most precise predictions on terminal level as well as for the total throughput. Furthermore, the empirical results are in line with our theoretical conjecture that one model/ method should not be expected to generate superior fit for each terminal considered. In other words, there is a heterogeneity in the time series behaviour of terminal throughputs, as well as in comparison with the total throughput. Even for one method, the optimal model parameters differ across terminal/ total series, for example, the SARIMA model (see Table 5). Hence, one model/ specification does not best fit all cases. Nevertheless, terminal-level data seems to suffer from more noise or structural breaks than total throughput, resulting in superior model fit for the later.

Insert Table 5 about here

5.3 Results of out-of-sample forecasts

Table 6 presents prediction errors from the in-sample forecasting exercise (i.e., using models estimated optimally for the pre-2013 period), which are employed for each time series to generate dynamic predictions in year 2013. We can observe that the forecasting performance of models using the total throughput is generally still superior to that of the models applied to individual terminal throughputs (with some exceptions for the JICT terminal and Seasonal Holt-Winters methods). Hence, an important implication would be that in-sample model fit can be a good indicator of how well (relatively) one method will perform out-of-sample, as compared to alternative methods (at least over a short forecasting horizon).

Insert Table 6 about here

Contrary to the poor in-sample fit of SARIMA-based models, the forecasts they generate for out-of-sample periods for the total throughput are the most precise ones. However, this does not apply to forecasts for individual terminals. The ASHW method performs best instead when VECM is ignored. Again, the forecasting performance is in line with the good in-sample model fit of the ASHW method reported in Table 4. When we consider the forecasting performance of all methods on terminal level, including VECM, the good in-sample fit of the latter method does not result in all-encompassing best forecasting performance out-of-sample. This finding suggests the existence of time-series instability in the data, as best models change over even a short period of time (of one year). VECM- or ASHM-based terminal-level forecasts appear to be the most promising ones to combine into an aggregated forecast for the total throughput. However, their errors are still larger than those of the best forecasts for total throughput.

Moreover, when we predict the total throughput by summing up the predictions for individual terminals, those “hybrid” predictions are less accurate (generate larger errors) than forecasts using the total series. Even in case of the VECM model which yields best fits for individual terminals, when its forecasts are summed up (unweighted or weighted sums²), the resulting forecasts for total throughput have larger errors than the best forecasts based on total throughput series (i.e., those using SARIMA). The same applies to ASHM method which also performed rather well in forecasting individual terminals: the sums of forecasts tend to generate larger errors than the best forecasts based on total throughput. Therefore, even if

² “Weighted sum 1” refers to sums using weights where the prediction error is measured using the RMSE, whereas, “Weighted sum 2” refers to sums using weights where the prediction error is measured using the MAE.

there is some merit in this approach, for our dataset it fails to improve upon the best model using total throughput time series. As discussed above, this could indicate that terminal-level time series suffer from higher levels of noise than total throughput, but these individual noise components partially cancel out each other, leading to more predictable total throughput time series. This result might also indicate that terminal-level data suffers more from structural breaks than the aggregated total throughput. Nevertheless, if one wanted to forecast terminal-level rather than total throughput, the VECM models yield the best in-sample fits and both VECM and ASHM generate best forecasts in our case.

Furthermore, weighting terminal-level forecasts by model fit in the estimation period does not seem to yield superior forecasts. The errors from weighted approaches are higher than those from unweighted sums. This further suggests that there might be structural breaks in the time-series behaviour of data and, hence, in the optimal model for each time series. Specifically, we combine individual terminal forecasts based on the method with superior in-sample fit for each terminal (as reported in Table 4). For example, when RMSE is used as the model fit precision criterion in the estimation sample, the best forecast for JICT is generated by the ASHM method. For the remaining terminals, the best forecast comes from the VECM model. The results presented at the bottom of Table 6 indicate that this cherry-picking approach to forecasting total throughput still fails to generate superior total forecasts. Most errors are larger than their counterparts' (when using terminal specific forecasts from one method across all terminals). Even when this combination of methods performs better than one-method approach (for weights based on in-sample MAEs ("Weighted sum 2") and using MAE as a forecast performance measure), those errors are still larger than the ones from the best (SARIMA) model applied to total throughput series.

6. A robustness check: truncated sample

One approach to test for time-invariability of conclusions regarding model fits and forecasting performance is to shift the date separating the estimation and the forecasting subsamples within a fixed sample of data (e.g., Xie *et al.* 2013). For instance, for the total sample ranging from January 2003 to November 2013, one could start with the estimation period ending in December 2012 (as we do), leaving one with the forecasting period of the remaining data in 2013. Subsequently, one could shift the cut-off point by one (or more) year backwards, resulting in the estimation period of 2003-2011 and the forecasting period of 2012-2013, and so on. This approach allows for an analysis of robustness over time. However, as it changes the absolute length of the forecasting period, the resulting forecast errors obtained for “out-of-sample” forecasts of different length would not be directly comparable to each other. Forecasts over longer horizons tend to have larger errors. This approach does not mirror the reality as faced by decision makers, either. If the decision has to be made at the beginning of year 2013 (with all the data available up to that point in time) and a forecast for the next 12 months of year 2013 is required, the relevant question is how reliable this one-year ahead forecast is. This can be addressed by looking backwards and assessing the forecasting performance of one-year ahead forecasts, yet not the performance of forecasts over different time horizons. Hence, in our robustness check, we shift our decision time back to the beginning of 2012. We repeat the whole analysis using observations prior to year 2012 to find the best-fitting models (in-sample analysis) and those in 2012 and 2012 only (i.e., we do not utilise year 2013) for out-of sample forecasting evaluation.

6.1 In-sample model fits

As with the full sample, we test our data in the truncated sample ending in 2012 for stationarity and cointegration and the results are qualitatively identical (Tables 2 and 3, Panel

B). All series are $I(1)$ and there is an evidence for cointegration (with two cointegrating vectors). Therefore, the order of integration of one is used in SARIMA estimations and a VECM model is employed to account for the existence of the cointegrating relationships.

Table 7 presents model fits for all time series (total throughput and four individual terminals' throughputs) and forecasting methods, obtained for the estimation period up to year 2011. First, we compare the model fit when the methods are applied to the total versus individual terminal series. As was the case for the full sample, in most cases modelling total throughput is less error-ridden than modelling individual terminal throughputs, with an exception of KOJA terminal modelled using SHW methods and model fit using the RMSE criterion. This confirms our previous conjecture that terminal throughputs tend to suffer more from noise than the total throughput.

Insert Table 7 about here

Moreover, as was the case for the full sample, the SARIMA model tends to generate largest errors, for both total and terminal-specific data. Rather, as in the full sample, the best model fit stems from the multiplicative Seasonal Holt-Winters methods (or the additive version when MAE is used to assess errors). For methods where model fit could be obtained for both terminal-level and total series (all except for the VECM), additive Seasonal Holt-Winters method seems to perform best in most cases. This is again in line with the findings for the full sample. Furthermore, the VECM method generates the most precise models for individual terminals. The reported errors are the smallest for each terminal and model fit measurement. The superiority of VECM for data fit on terminal level is even more pronounced than for the whole sample.

6.2 Out-of-sample forecasts

Table 8 presents prediction errors from the in-sample forecasting exercise. This is conducted

using optimal models estimated for the pre-2012 period, which are employed for each time series to generate dynamic predictions in year 2012. We can observe that the forecasting performance of the models using the total throughput is still superior to that of the models applied to individual terminal throughputs, with some exceptions. Hence, an important implication is that in-sample model fit can be a good indicator of how well (relatively) one method will perform out of sample, as compared to alternative methods.

Insert Table 8 about here

Completely contrary to the findings for the whole sample, the SARIMA methods yield the worst forecasts for total throughput time series. Instead, the additive Seasonal Holt-Winters method generates the most precise forecasts. The VECM forecasts for individual terminals seem to weakly outperform forecasts based on other models, as they yield smallest errors in five out of eight cases. Again, their relative forecasting performance is in line with the good model fit of VECM models in the estimation sample, but their superiority is less clear in the forecast sample.

When we predict the total throughput by summing up the predictions for individual terminals, those “hybrid” predictions are again less accurate (generate higher errors) than forecasts using the total series. Even in case of the VECM model which yields best forecasts for individual terminals, when those forecasts are summed up (unweighted or weighted sums), the resulting forecasts for total throughput have higher errors than best forecasts based on total throughput series. Therefore, even if there is some advantage of this approach, it fails to improve upon the best models using total throughput time series. Similar to the observations we presented in Section 5.3, this could indicate that terminal-level time series suffer from higher levels of noise than total throughput. These individual noise components partially offset each other, leading to more predictable total throughput time series. This result might also indicate that terminal-level data suffers more from structural breaks than the

aggregated total throughput. In addition, weighting by model fit in the estimation period does not seem to yield superior forecasts. The resulting errors are larger than those from forecasting total time series, as in the case for the whole sample. However, when weights are computed based on RMSE in-sample (“Weighted sum 1”), the resulting errors are generally lower than those from unweighted sums, except for those based on VECM. This suggests that a weighting scheme, maybe putting even more (less) emphasis on terminals with good (bad) model fits in-sample, could potentially generate more accurate forecasts. Finally, as the VECM yield best fit for all terminals and precision criteria, we cannot devise a “cherry-picking” strategy equivalent to the one used for the whole sample, where different terminals were best described in-sample by different methods.

7. The 2018 forecasts of container throughput

Based on the abovementioned results, we select what would have been judged the best method to forecast total throughput by the end of year 2012 and of 2013. We apply the additive and multiplicative Seasonal Holt-Winters methods, respectively, to generate forecasts for total throughput until the end of year 2018. This forecast horizon of five years is dictated by the planned completion of phase one of the port expansion; one should bear in mind that such a long-term forecast will be more susceptible to larger errors than a shorter-term forecast would be. The estimated forecasts, as well as their 95% confidence intervals (calculated using quantile regression technique and assuming that the estimates’ standard error will develop over time in the forecasting period as they did in the respective sample period) are presented in Figure 2.

Insert Figure 2 about here

The horizontal line represents the monthly equivalent of the annual throughput capacity of 9,500,000 TEUs, which underlies the port expansion plans (see Table 1 for more

details). As can be seen, for the in-sample period, both forecasting models fit the data very well. For the forecast period, i.e., observations starting in 2013 and 2014 for the model based on truncated and full sample, respectively, the latter turns out to have more volatile forecasts. This is not necessarily a sign of its inferior performance, as these forecasts would have to be compared with the actual data from 2014 onwards, which are not available to us. In addition, this higher volatility comes from more pronounced seasonal variations. Hence, it might be more accurate for the actual future throughput varying in a seasonal way. However, both models generate the same qualitative and very similar quantitative conclusions. Their point estimates for the end of 2018 seem to suggest that the planned capacity is somehow excessive. However, given the 95% confidence interval one would not be able to reject the hypothesis that the predicted throughput is significantly different from the planned capacity. Hence, the planned maximum capacity seems to be sufficient for the expected volume of traffic, with some margin to accommodate unexpectedly (but not excessively) high throughputs. Only a major crisis, such as the one the world economy experienced following the 2007-8 financial turmoil, could potentially decrease the traffic volume significantly, thus rendering some newly added capacities excessive and superfluous.

8. Conclusions

In this paper, we forecast the demand of total container throughput at Tanjung Priok Port using a novel approach of utilising the potential forecasting power of terminals' throughputs. We employ four univariate forecasting models, including SARIMA, the additive Seasonal Holt-Winters (ASHW), the multiplicative Seasonal Holt-Winters (MSHW) and, newly to this literature, the Vector Error Correction Model (VECM). Our aim is to provide insights into whether forecasting the total container throughput obtained by historical aggregated port throughput time series is superior to the forecasts of the total container throughput obtained

by aggregating individual terminal forecasts.

Our results show that both in-sample model fit and out-of-sample forecasts for the total throughput are more precise, i.e., have smaller errors than those obtained for aggregating individual terminals' forecasts. SARIMA is found to generate the worst model fit in-sample, whereas the total throughput is best described by the multiplicative Seasonal Holt-Winters models. The VECM approach appears to generate best model fits and forecasts for individual terminals. However, forecasts of the total throughput based on modelling the total throughput time series are consistently better (with smaller errors) than those obtained by combining those forecasts generated by terminal-specific models, even when the latter are weighted to allow more precise models to affect the total forecast most. These findings appear to be qualitatively robust against a shift in the timing of estimation vs. forecasting period. The proposed method of aggregating individual terminal forecasts may be data-dependent and could well generate superior forecasts for other ports and periods. However, we also observe that even a relatively minor shift in the timing of the end of the estimation period resulted in a reversal of the relative forecasting performance out-of-sample of the SARIMA method. Specifically, whereas it was the worse model for the whole sample, it turned out to generate the best forecasts when the last year of data was not utilised. Lastly, our forecasts for the total container throughput at the Tanjung Priok Port for year 2018 indicated that the planned port's expansion will result in almost full utilisation of its capacities.

These results suggest that researchers and policy makers should wary of assuming that a model which fits the data well in one terminal/ period will also fit the data well for another terminal/ period, and that superior in-sample fit will necessarily result in superior out-of-sample forecasting performance. This might be due to instability of the data generating process, and necessitates repetitions of modelling and forecasting exercises over time to capture the time-varying nature of data and, hence, the best models, rather than relying on the

initial results and assuming them to be fixed in time. Furthermore, even if our specific sample rejects the idea of forecasting total throughput by aggregating terminal-specific forecasts, some theoretical arguments speak in favour of this approach and it could be successfully applied to other ports.

Acknowledgements

Earlier versions of this research were presented at the XIII International Conference on Strategic and Operational Management. 2015, Tokyo, Japan (under the title: “Forecasting Container Throughput: Using Aggregate or Terminal-Specific Data?”) and the 20th Conference of the International Federation of Operational Research Societies. 2014, Barcelona, Spain (under the title: “Forecasting Container Throughput at Tanjung Priok Port, Indonesia, using Univariate Forecasting Models”). We thank participants at these events for their comments. We also thank anonymous referees for their helpful comments and suggestions.

References

- Aburto, L. and R. Weber. 2007. "Improved supply chain management based on hybrid demand forecasts". *Applied Soft Computing* 7 (1): 136-144.
- Bates, J. M. and C. W. J. Granger. 1969. "The combination of forecasts". *Operational Research Quarterly* 20 (4): 451-68.
- Box, G. E. P., and G. M. Jenkins. 1970. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 2008. *Time Series Analysis: Forecasting and Control*. Forth edition. John Wiley & Sons, New Jersey.
- Brockwell, P. J., and R. A. Davis. 1991. *Time Series: Theory and Methods*. Second edition. Springer-Verlag, New York.
- Chen, S-H. and J-N. Chen 2010. "Forecasting container throughputs at ports using genetic programming". *Expert Systems with Applications* 37 (3): 2054-2058.
- Clemen, R. T. 1989. "Combining forecasts: a review and annotated bibliography". *International Journal of Forecasting* 5 (4): 559-583.
- Cullinane, K., T.-F. Wang, D.-W. Song, and P. Ji. 2006. "The technical efficiency of container ports: Comparing data envelopment analysis and stochastic frontier analysis". *Transportation Research Part A: Policy and Practice* 40(4): 354-374.
- Enders, W. 2014. *Applied Econometric Time Series*. John Wiley & Sons, Chichester, UK
- Franses, P. H. 1996. *Periodicity and Stochastic Trends in Economic Time Series*. Oxford University Press, Oxford.
- Fung, M.K. 2002. "Forecasting Hong Kong's container throughput: an error-correction model". *Journal of Forecasting* 21 (1): 69-80.
- Gardner, E. Jr. 2006. "Exponential smoothing: The state of the art--Part II". *International Journal of Forecasting* 22 (4): 637-666.
- Gosasang, V., W. Chandraprakaikul, and S. Kiattisin. 2011. "A comparison of traditional and neural networks forecasting techniques for container throughput at Bangkok port". *Asian Journal of Shipping Logistics* 27 (3): 463-482.
- Hansen, J. V. and R.D. Nelson. 2003. "Time-series analysis with neural networks and ARIMA-neural networks and ARIMA-neural network hybrids". *Journal of Experimental & Theoretical Artificial Intelligence* 15 (3): 315-330.
- Harvey, A. C. 1993. *Time Series Models*. Second edition. The MIT Press.
- Hornik, K., M. Stinchcombe, and H. White. 1989. "Multilayer feedforward networks are universal approximators". *Neural Networks* 2 (5): 585-592.
- Indonesia Port Corporation. 2012. Detailed development plan. Available at: <http://www.indonesiaport.co.id/newpriok/sub/development-details.html> (accessed on 30 09 2014).
- Jain, A. and A. M. Kumar. 2007. "Hybrid neural network models for hydrologic time series forecasting". *Applied Soft Computing* 7 (2): 585-592.
- Johansen, S. 1991. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models". *Econometrica* 59 (6): 1551-1580.

- Khashei, M. and M. Bijari. 2010. "An artificial neural network (p, d, q) model for time series forecasting". *Expert Systems with Applications* 37 (1): 479-489.
- Lam, W., Y. Tang, K. Chan, and M. Tam. 2006. "Short-term hourly traffic forecasts using Hong Kong annual traffic census". *Transportation* 33 (3): 291-310.
- Li, M. K. and T. L. Yip. 2013. "Joint planning for yard storage space and home berths in container terminals". *International Journal of Production Research*, 51 (10): 3143-3155.
- Lingras, P., S. C. Sharma, P. Osborne, and I. Kalyar. 2000. "Traffic volume time-series analysis according to the type of road use". *Computer-Aided Civil Infrastructure Engineering* 15 (5): 365-373.
- Makridakis, S. and M. Hibon. 2000. "The M3-Competition: results, conclusions and implications". *International Journal of Forecasting* 16 (4): 451-476.
- Ömer Faruk, D. 2010. "A hybrid neural network and ARIMA model for water quality time series prediction". *Engineering Applications of Artificial Intelligence* 23 (4): 586-594.
- Peng, W-Y. and C-W. Chu. 2009. "A comparison of univariate methods for forecasting container throughput volumes". *Mathematical and Computer Modelling* 50 (7-8): 1045-1057.
- Port Finance International. 2014. New Jakarta port expected to cut Indonesian logistic costs. Available at <http://portfinanceinternational.com/features/item/1485-new-jakarta-port-expected-to-cut-indonesian-logistic-costs> (accessed on 30 09 2014)
- Roy, D., A. Gupta, and R. B. M. De Koster. 2016. "A non-linear traffic flow-based queuing model to estimate container terminal throughput with AGVs". *International Journal of Production Research* 54 (2): 472-493.
- Ruiz-Aguilar, J. J., I. J. Turias, and M. J. Jiménez-Come. 2014. "Hybrid approaches based on SARIMA and artificial neural networks for inspection time series forecasting". *Transportation Research Part E* 67: 1-13.
- Said, S. E. and D. A. Dickey. 1984. "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order". *Biometrika* 71 (3): 599-607.
- Schulze, P. M. and A. Prinz. 2009. "Forecasting container transshipment in Germany". *Applied Economics* 41 (22): 2809-2815.
- Schwarz, G. E. 1978. "Estimating the dimension of a model". *Annals of Statistics* 6 (2): 461-464.
- Schwarzkopf, A., R. Tersine, and J. Morris. 1988. "Top-down versus bottom-up forecasting strategies". *International Journal of Production Research* 26 (11): 1833-1843.
- Smith, J. and K. F. Wallis. 2009. "A simple explanation of the forecast combination puzzle". *Oxford Bulletin of Economics and Statistics* 71 (3): 331-355.
- Stock, J. H. and M. W. Watson. 2004. "Combination forecasts of output growth in a seven-country data set". *Journal of Forecasting* 23 (6): 405-430.
- Taskaya-Temizel, T. and M. C. Casey. 2005. "A comparative study of autoregressive neural network hybrids". *Neural Networks* 18 (5-6): 781-789.

- The World Bank. 2014. Container port traffic (TEU: 20 foot equivalent units).
http://data.worldbank.org/indicator/IS.SHP.GOOD.TU?order=wbapi_data_value_2012+wbapi_data_value+wbapi_data_value-last&sort=desc (accessed on 30 09 2014).
- Timmermann, A. 2006. "Forecast combinations". In: Elliott, G., C. Granger, and A. Timmermann. Eds. *Handbook of Economic Forecasting*. Elsevier, Amsterdam.
- Tsai, F-M. and L. J. W. Huang. 2015. "Using artificial neural networks to predict container flows between the major ports of Asia". *International Journal of Production Research* In press. DOI:10.1080/00207543.2015.1112046
- Tseng, F., H. Yu, and G. Tzeng. 2002. "Combining neural network model with seasonal time series ARIMA model". *Technological Forecasting and Social Change* 69 (1): 71-87.
- Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. 2005. "Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach". *Transportation Research Part C: Emerging Technologies* 13 (3): 211-234.
- Wang, X. and F. Petropoulos. 2016. "To select or to combine? The inventory performance of model and expert forecasts". *International Journal of Production Research* 54 (17): 5271-5282.
- Wang, J., J. Wang, Z. Zhang, and Z. Guo. 2012. "Stock index forecasting based on a hybrid model". *Omega* 40 (6): 758-766.
- Wei, Y. and M. Chen. 2012. "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks". *Transportation Research Part C: Emerging Technologies* 21 (1): 148-162.
- Xie, G., S. Wang, Y. Zhao, and K. K. Lai. 2013. "Hybrid approaches based on LSSVR model for container throughput forecasting: a comparative study". *Applied Soft Computing* 13 (5): 2232-2241.
- Zhang, G. P., 2003. "Time series forecasting using a hybrid ARIMA and neural network model". *Neurocomputing* 50: 159-175.

Table 1. New container terminals: capacity and planned operation (Indonesia Port Corporation, 2012).

Container Terminals	Capacity (TEUs/year)	Planned operation	
		Commencement (year)	Completion (year)
Phase one			
Terminal 1	1,500,000	2013	2018
Terminal 2	1,500,000	2013	2018
Terminal 3	1,500,000	2013	2018
Phase two			
Terminal 4	2,000,000	2018	2023
Terminal 5	2,000,000	2018	2023
Terminal 6	2,000,000	2018	2023
Terminal 7	2,000,000	2018	2023

Table 2. Results of unit root tests using the augmented Dickey-Fuller tests (ADF).

Time series	Test	Critical Values			Test	Critical Values		
	Statistic	1%	5%	10%	Statistic	1%	5%	10%
Panel A: Estimation period results for the full sample (ends November 2013)								
		Levels				1st Differences		
Total	-3.302	-4.034	-3.448	-3.148	-10.469	-4.034	-3.448	-3.148
JICT	-3.124	-4.035	-3.448	-3.148	-8.468	-4.035	-3.448	-3.148
KOJA	-3.713	-4.034	-3.448	-3.148	-11.852	-4.034	-3.448	-3.148
MTI	-3.027	-4.034	-3.448	-3.148	-7.193	-4.035	-3.448	-3.148
Conventional	-2.402	-4.106	-3.480	-3.168	-7.495	-4.108	-3.481	-3.169
Panel B: Estimation period results for the truncated sample (ends December 2012)								
Total	-2.869	-4.038	-3.449	-3.149	-13.830	-3.508	-2.890	-2.580
JICT	-2.980	-4.039	-3.450	-3.150	-15.404	-3.508	-2.890	-2.580
KOJA	-3.423	-4.038	-3.449	-3.149	-16.059	-3.508	-2.890	-2.580
MTI	-2.282	-4.039	-3.450	-3.150	-13.503	-3.508	-2.890	-2.580
Conventional	-2.208	-4.130	-3.491	-3.175	-8.793	-3.569	-2.924	-2.597

Table 3. Result of Johansen cointegration tests.

Maximum rank	Number of Parameters	LL	Eigenvalue	Trace statistic	5% critical value
Panel A: Estimation period results for the full sample (ends November 2013)					
0	20	-2940.65	.	79.3888	62.99
1	28	-2921.11	0.42788	40.3005	42.44
2	34	-2911.46	0.24093	21.0044	25.32
3	38	-2904.26	0.1859	6.6070	12.25
4	40	-2900.96	0.09007		
Panel B: Estimation period results for the truncated sample (ends December 2012)					
0	4	-2492.12	.	81.9054	62.99
1	12	-2474.20	0.45523	46.0693	42.44
2	18	-2461.37	0.35279	20.3992	25.32
3	22	-2454.18	0.21638	6.0130	12.25
4	24	-2451.17	0.09689		

Note: LL stands for the value of the log-likelihood function.

Table 4. Forecast errors in the full estimation sample (pre-2013).

	Total	JICT	KOJA	MTI	Conventional
RMSE					
SARIMA	15.9048	17.7146	16.5363	38.2498	26.7701
ASHM	11.4366	11.6373	10.4548	23.5926	15.3447
MSHW	11.3884	11.7140	10.5567	20.9158	17.4812
VECM		11.9530	10.3918	17.8528	13.8237
MAE					
SARIMA	11.4926	13.1829	13.1482	27.1560	17.4409
ASHM	7.7171	9.2685	8.0118	17.3405	9.8555
MSHW	7.6770	9.3477	8.1647	15.1338	10.9791
VECM		9.5207	7.7678	14.4466	10.0387

Table 5. Optimal parameter values for best SARIMA(p,d,q)(P,D,Q)_s models.

Series/Parameter	p	q	P	D	Q
Panel A:					
Estimation period results for the full sample (ends November 2013)					
Total	0	1	1	2	2
JICT	0	1	0	2	2
KOJA	0	1	1	2	1
MTI	0	1	0	2	2
Conventional	1	0	1	2	1
Panel B:					
Estimation period results for the truncated sample (ends December 2012)					
Total	1	1	0	2	2
JICT	0	1	0	2	2
KOJA	0	1	1	2	1
MTI	1	1	0	2	2
Conventional	1	0	1	2	2

Note: the value of d is set to d=1 based on unit root test results, and the value of S is set to S=12 due to the monthly frequency of data.

Table 6. Forecast errors in the full forecast sample (2013).

	Total	JICT	KOJA	MTI	Conventional	Unweighted sum	Weighted sum 1	Weighted sum 2
RMSE								
SARIMA	3.4785	12.5467	12.3585	15.5032	11.8088	4.0595	5.8197	7.6972
ASHM	7.8412	6.7538	10.2292	16.8394	10.1308	7.7467	9.9030	11.7402
MSHW	8.7708	7.0149	11.0000	17.1700	11.7416	8.3087	8.3398	9.9793
VECM		12.6915	10.0222	16.8414	9.1930	9.2705	11.0500	11.7670
MAE								
SARIMA	3.0537	11.2602	9.4245	11.1414	9.6596	3.6150	4.6403	6.5430
ASHM	5.2577	4.9852	8.7745	13.9801	6.3033	4.4090	6.5729	8.8656
MSHW	6.3901	5.2657	9.3200	14.3379	6.9318	4.4606	4.5682	6.3320
VECM		9.7119	8.1993	14.4633	6.2583	6.1059	7.9970	8.9643
Combination of terminals								
Based on best in-sample model fit (RMSE)						RMSE	7.2026	8.2127
						MAE	5.1310	4.9640
Based on best in-sample model fit (MAE)						RMSE	7.8636	7.7757
						MAE	4.8673	4.5856

Note: “Unweighted sum” denotes a sum of forecasts generated for each terminal separately, “Weighted sum 1” (“Weighted sum 2”) denotes a sum of individual terminals’ forecasts, weighted using formula shown in Section 4 and RMSE (MAE) for each terminal’s model fit from the estimation period. Highlighted are those results with the lowest errors.

Table 7. Forecast errors in the truncated estimation sample (pre-2012).

	Total	JICT	KOJA	MTI	Conventional
RMSE					
SARIMA	16.5352	18.5033	17.1024	41.7524	31.4604
ASHM	11.5750	12.0240	10.5499	26.0193	17.7959
MSHW	11.5663	12.0652	10.5715	22.5162	24.7733
VECM	./.	10.4034	10.0204	19.8414	13.6154
MAE					
SARIMA	11.8554	13.9052	13.6921	30.5240	17.0055
ASHM	7.8955	9.5880	8.1970	18.8038	11.3571
MSHW	7.9101	9.7221	8.3518	16.4521	13.5467
VECM	./.	8.3155	7.5429	15.9036	10.2379

Table 8. Forecast errors in the truncated forecast sample (2012).

	Total	JICT	KOJA	MTI	Conventional	Unweighted sum	Weighted sum 1	Weighted sum 2
RMSE								
SARIMA	14.0131	13.3761	16.3301	47.6620	78.9441	43.1006	36.8647	51.2054
ASHM	8.2681	11.0155	10.2916	26.6375	31.9625	15.9401	15.5171	19.4207
MSHW	9.0168	10.4938	12.2281	32.9178	52.1559	26.2981	14.3873	22.3233
VECM		10.5053	11.6517	17.3909	16.2382	10.6560	14.0034	14.3068
MAE								
SARIMA	10.9786	10.1562	13.6253	36.7378	73.3543	39.7642	33.8563	48.4084
ASHM	5.4956	10.3452	7.6683	23.1762	28.8308	12.7035	12.4432	16.9005
MSHW	6.1429	9.3435	8.7526	30.4334	47.1151	22.3109	10.2649	18.3563
VECM		9.3078	9.5073	13.6164	13.0428	7.5515	11.1935	11.5676

Note: “Unweighted sum” denotes a sum of forecasts generated for each terminal separately, “Weighted sum 1” (“Weighted sum 2”) denotes a sum of individual terminals’ forecasts, weighted using formula shown in Section 4 and RMSE (MAE) for each terminal’s model fit from the estimation period. Highlighted are those results with the lowest errors

Figure 1. Monthly container throughput data of total and individual terminals (2003-2013).

Figure 2. Actual and predicted total throughputs