Forecasting Container Throughput within Multi-Port Region

Using CDE-MPR Methodology: the Case of PRD Region

Zhang Jiawei\textsuperscript{a}, Xu Lizhi\textsuperscript{b}, Zhang Xun\textsuperscript{a}, Wang Shouyang\textsuperscript{a}

\textsuperscript{a} Chinese Academy of Sciences, 100190, Beijing, China

\textsuperscript{b} BeiHang University, 100190, Beijing, China

Abstract:
This paper proposes a CDE-MPR methodology designed to forecast and analyze container throughput volumes. The proposed methodology includes: a container throughput module which considers both interactions among ports and other influence factors, a demand demand-oriented module which provides additional information and chance events effect had on the maritime industries. It is applied to forecast container throughput volumes of the ports in the Pearl River Delta (PRD) region of South China after global financial crisis. To evaluate the forecasting performance of CDE-MPR methodology, we compare its performance with those of ARIMA, VEC and VAR. The empirical results show that CDE-MPR method substantially outperforms other single-model methods. It also demonstrates the consequences of the global financial crisis on each port in PRD region and the weakening effects of the financial crisis on each port during the second half of year 2009, especially on Guangdong port. The study for PRD region also confirms the necessity of analyzing and forecasting from a multi-port region perspective. It is also feasible and effective to apply and expand the CDE-MPR methodology to other similar issues.

Keywords: Forecasting; Container throughput; CDE-MPR (Multi-port region) methodology; Global financial crisis; Foreign trade.

\textsuperscript{*}Corresponding author, e-mail: zhangjiawei@amss.ac.cn
1. Introduction

Globalization and containerization have spatially and functionally reshaped international ports and have generated new multi-port systems in a great many of regions around the world. As pointed by Wang, et al. (2010), dual hub port systems appear in megacity regions. The Pearl River Delta (PRD) region of South China is a notable case of multi-port region with three hub ports-Hong Kong port, Shenzhen port and Guangzhou port. It handled a (world) record 56.72 million TEUs (20-foot equivalent units) in 2008, an increase of 2.36 million TEUs (4.35%) over the previous year, despite the global financial crisis. This accounted for 45% of China’s total throughput of 126 million TEUs, making PRD ports China’s and even the world’s most significant container trade ports. In 2008, the Hong Kong port handled 24.3 million TEUs, ranking third among the world’s top ten ports and the second among China’s coastal ports. Shenzhen port ranked the fourth, handling 21.42 million TEUs and Guangzhou port leapt from 12th to seventh rank in terms of container traffic, reaching 11 million TEUs. With development of the international trade, PDR is becoming more open and integrated into the world economy. Furthermore, the Pearl River Delta has the world’s largest export production base centered in Shenzhen Special Economic Zone set up in 1980 when export-oriented industries started to be built up in Shenzhen (Wang et al., 2000). Its manufactures, including electronics, chemicals, processed foods, textiles, construction materials, and pharmaceuticals, contribute significantly to the region’s spectacular economic growth and development of ports serving this region. In China today, ports in PRD region are significant gateways between China and the rest of the world. Such a tremendous volume of container traffic deserves an in-depth investigation.

There are few existing literatures on the structural and spatial development of seaport systems in PRD region. From a port geography point of view, Wang et al. (2000) investigated the progress of container port system development in South China. They point out that the particular situation of the PRD is unique in many respects, and the features emerging there, i.e. its dominant hub, its network of feeder ports and its emerging direct-service non-hub terminals, are being replicated elsewhere. Along with the process of globalization in PRD region, business activities intensified
and the dual hub port systems appeared in this region, i.e. the ports of Hong Kong and Shenzhen. Notteboom (2005) conceptualized this trend as port regionalization. The development of Hong Kong has illustrated this trend. Hong Kong, the world’s top container port until 2005, has launched many large-scale construction projects in order to connect Hong Kong to the rest of the PRD region, such as the Shenzhen Western Corridor, the Hong Kong-Zhuhai-Macao Bridge and the Guangzhou-Shenzhen-Hong Kong Express Rail Link. So, with port regionalization, quantitative predict and qualitative assessment are necessary, and that can form a foundation on the basis of which the government and related parties can make decisions and postulate the future of ports development. More and more literatures have investigated transport issues from a geographic perspective. Loo (2009) pointed out that the transport industry’s geography with a spatial focus on Asia is an exciting hallmark, and perhaps a reflection of the emerging importance of China, not only in the world transport industry but in the entire global economy. In investigations of airports, Multi-airport regions (MAR) are well documented. Several scholars have studied airports from this perspective. Hansen (1990) studied the airline competition in Multi-airport regions. Loo (2008) studied passengers’ airport choice within multi-airport regions (MARs). In respect of port logistics, many scholars have realized the necessity studied it from a geographical point of perspective. A new study of Imai, et al. (2009) demonstrated that the multi-port calling is superior in terms of total cost for the Asia–Europe and Asia–North America trade lanes in most scenarios based on a theoretical model and varieties of numerical experiments. Theoretically, it is efficient to develop a multi-port region in Asia trade lanes, and rapid development of container throughput in PRD has confirmed this. Thus, it is necessary and also possible to analyze the port cargo throughput in PRD as well as other similar ports, like in North America, Japan, and South-East Asia in terms of a multi-port region (MPR).

However, most of previous literatures only focus on one port, like Hong Kong, considering the effects of other ports as influencing factors. For example, Fung (2002) forecasted Hong Kong’s container throughput using an error-correction model. Hui et al. (2004) also forecasted Hong Kong’s container throughput. Chou et al. (2008) proposed a modified regression model for forecasting volumes of Taiwan’s import containers. Chen & Chen (2010) attempted to forecast container throughputs at ports using genetic programming. At the same time, they analyzed the relationships among a region’s ports, which have attracted much attention recently. Wang & Slack
(2000) investigated the progress of container port systems between Hong Kong and other ports in the Pearl River Delta (PRD) region in 2000. Yap & Lam (2006) analyzed the competition between major container ports in East Asia. The one exception is the research of Seabrooke et al. (2003). They realized the necessity of this when forecasting the cargo volume growth and the development of the port of Hong Kong, which is a good start for multi-port region analysis. The PRD is considered as a distinct and independent entity, economically and geographically, although the PRD is not an official Chinese province. All factors affecting port cargo volume in Hong Kong, namely, macroeconomic conditions, liberalization of cross-strait trade, market structure and power of terminal operators, etc., essentially affect ports in the rest of PRD. Finally, they only forecasted the container throughput in Hong Kong. Thus, there is still little literature on forecasting the cargo throughput within multi-port region simultaneously in one conceptual framework. This paper aims to analyze and forecast container throughput in a multi-port region (MPR).

Furthermore, despite the Asia-Pacific Rim’s port cities, as emphasized by Rimmer (1999), having been reshaped since the mid-1980s as major global trading centers, trade, as one of the most important factors, attracted relatively little attention. Janssens et al. (2003) also pointed out that nobody can or will deny the role international trade has played in boosting throughput of ports. By using the correct methodology to analyze historical data, we can come to a more realistic and reliable conclusion. They investigated the relationships between container throughput and international trade in the port of Antwerp using an error correction model. However, according to recent studies, there is little discussion to investigate the effects of trade when analyzing and forecasting container throughputs. In this paper, we attempt to emphasize the role the trade plays in container throughput analysis, and a separate trade module is introduced in the model to illustrate its role.

Because of effects of the global financial crisis, the world's trade volume during the first quarter of 2009 fell more than 11% below the last quarter of 2008. As a derived demand, it is inevitable that trade and market declines directly and negatively impact the development of maritime industries. Due to the export-oriented economy of the PRD region, its cargo and container traffic experienced a sharp decline. The global financial crisis is a chance event, which cannot be captured by classic model, predictions involving chance events are usually very difficult but extremely significant. During the Asian crisis, a great many of literatures forecasted the economic performance and
evaluated the effect of the crisis (Summers, 2001; Choy, 1998; Prideaux et al., 2003; Barro (2001)). After this global financial crisis, literatures related to global financial crisis sprung up, and some researchers investigated the impacts of the financial crisis on maritime industries. Slack (2010) identified four major issues, namely, the changing patterns of international trade, the importance of being green, the changing government–industry relations and the need for transparency, which play key roles in shaping maritime industries. Ng & Liu (2010) introduced challenges and opportunities of port and maritime industries in the post-2008 world. Focusing mostly on developments in Europe, Pallis & De Langen (2010) also analyzed the structural effects the crisis had on seaports. To the best of our knowledge, little has been done to forecast the performance and evaluate the effect of the crisis for maritime industries after the global financial crisis. In this paper, we take the economic crisis into consideration and measure and analyze the effect on ports cargo traffic.

This paper presents a CDE-MPR methodology for forecasting and analyzing cargo volume in multi-port region, like Hong Kong port, Shenzhen port and Guangzhou port in PRD. This paper also highlights the dependence of the forecasts on the nature of interaction between the multi-port region and other multi-port regions, such as the interaction among the regions in PRD and other major ports in East and South-East Asia. At the same time, we also emphasize the important role the trade plays in determining port throughput. A specific module for trade is modeled in the system to take the impact of the demand into account. Further, chance event like financial crisis is also measured and take into consideration in the system, which is expected to help reduce the prediction error. This paper proposes a new methodology for forecasting and analyzing the container throughput in a multi-ports region, which combine the container throughput and trade module as well as including sudden effect in one model.

The organization of the rest of this paper is as follows. Section 2 presents a new methodology for forecasting port cargo volume in Multi-port region; Section 3 takes PRD region as an example to illustrate how to forecast using this new approach; In section 4, we discuss the data and estimation result, the forecasting performance are also examined; and the concluding remarks are given in Section 5.
2. CDE Multi-Port Region Forecasting Methodology

In this section, we introduce a new methodology, CDE multi-port region forecasting (CDE-MPR) methodology for container throughput forecasting. There are two modules and events in this methodology, where C denotes container throughput module which is our forecasting target; D denotes the demand module of the container throughput, which is the explanatory factors for container throughput, like trade for some ports; And E denotes the chance events, which cannot be capture by other two modules. Variety of research may include the two modules (C and D) in there model when forecasting the container throughput, less literature considered E (events) which also play a key role to obtain a better forecasting result. Consideration of the sudden events is of great help to reduce forecasting error.

Further, along with the development of the ports, many ports in one region increasingly interact with each other. And the relationships of the multi-port in one region also have gained more and more attraction (Notteboom (2010); Lam(2010); Yap & Lam (2006), etc.). Under such circumstance, it is not sufficient to forecast one port even including all the CDEs. The multi-ports in one region often share the common factors, for example the hinterland, the policy, the labor cost, etc. Thus, the factors which influence one port in the region is probability also influence another port in the same region. And the influencing factors of these ports may also have high correlations. As a result, we propose a new model which combines the multi-port in one region with CDE modules. The model framework of our methodology is illustrated in Fig.1.
We suggest including a demand module in the model to gain better prediction results. Choose the proper demand-oriented influencing factors of all the ports according to their instincts relations as the explanations of each container throughput. Because the demand variables of the ports in one region also interact with each other, it is feasible to model them in one module and the results are employed in the C-module. Most importantly, chance-event effect should also be taken into account. A simple but effective method to represent that is to introduce a dummy variable, which is preferred by many economists in many research areas. The dummy variable approach may be used to tackle the problem of a model structure that is unstable over time if the instability is caused by one-off type changes (Song, et al. (2003)). Barro (2001) introduced a dummy variable to measure the effect of the Asia financial crisis for economic growth. Song, et al. (2003) comprised two oil crisis dummy variables in their models. For simplicity, we use a dummy variable to capture the chance-event effect in this paper. Definitely, other methods which can be used to measure the event are all appreciate in E-module.

3. A Case in PRD Using CDE-MPR Methodology

3.1 Model Framework

In this section, to illustrate how to forecast using the CDE-MPR methodology, we take the ports,
namely, Hong Kong, Shenzhen and Guangzhou port, in PRD region as an example. The systematic model built by using this methodology not only generates projections of container throughput of three busiest ports - Hong Kong, Shenzhen and Guangzhou port - but it also explores their internal and external relationships and the interplay among the three ports in the PRD region, and the port traffics dynamic structure being formed in South China. Considering the objective law under circumstances specific to the PRD, here we consider the demand denoted by trade separately in this system, by dividing the models into two parts: one is the container throughput module (C), and the other is trade module (D). In cognizance of each port's particular circumstances, different models are formulated. In this system, both the effect of other ports and the effect of the demand are considered. The explanatory variables of container throughput are similar, including foreign trade, effect of other ports, and their respective lag variables.

In this system, we aim to investigate container throughput of each port under one conceptual framework, as follows (Fig 2).

![Fig. 2 The model framework of forecasting container throughput](#)

In the port module, four ports (Hong Kong, Guangzhou, Shenzhen and Shanghai) are jointly investigated. In this paper, a VEC model with co-integration restriction is proposed to investigate the relationships of multi-ports in the PRD region. The volume of container throughput of the Shanghai port is used as an exogenous variable. However, Guangzhou port is a little different from Hong Kong and Shenzhen ports, a factor we will analyze with an ECM model using the volume of container throughput and the demand denoted by the total trade volume in Guangdong province. In addition, Shanghai port is modeled based on the trade in Shanghai, using a PDL model.
Several of literatures have investigated the relationships between the trade and the container throughput. Hui et al. (2004) emphasized that demand for port services is derived from demand for imports and exports and introduced trade values in their model for forecasting cargo throughput of the Hong Kong Port. Fung (2001) confirmed the interdependency between trade and container throughput by including various external trade variables as exogenous variables in a vector error correction model. This interdependency is the intrinsic reason and the foundation for multi-port regions (MPRs). Just like the key role passenger demand plays in multi-airport regions (MARs), trade volume is important for analysis of MPRs. Considering the fact that cargo hinterlands of Hong Kong and Shenzhen ports are overlapped, we introduce the sum of the trade in Hong Kong and Guangdong province into the equation, which is denoted by demand 1. Demand for Guangzhou port services is relatively different from demand for Hong Kong and Shenzhen ports services, as the latter are more export-oriented than Guangzhou. Demand for Guangzhou port’s services is mainly derived from intra-trade activities. Therefore, we only introduce the total trade in Guangdong province as a representation of demand for Guangzhou Port’s services. For Shanghai port, we use trade statistics from Shanghai Custom (demand 3) as representation of demand in the Shanghai port.

3.2 Research Methodology

The basic model under consideration is the Error-Correlation model, which is the most widely used in analysis and forecast of container throughput (Fung, 2002; Hui et al., 2004; Yap & Lam, 2006) However, the co-integration relationship is often considered in two folds; one is the relationships among two or more ports, and the other is the relationship between trade and throughput of container ports. Yap & Lam (2006) used the former, while Janssens (2003) analyzed the latter aspect. But when analyzing PRD’s ports in one uniform framework, the forecasting accuracy has little benefit from using only one type of co-integration, neither relationships among the ports nor those between trade and throughput. So we select the proper type of error-correlation model for each port and the variables, use restricted co-integration model and then estimate and forecast them simultaneously. Thus, we use a different and proper form for every variable to ensure each equation’s accuracy, and we estimate them simultaneously to maximize the information. We also use other models to forecast port container throughput to compare their accuracy, namely, the ARIMA model, which is used as a benchmark model, vector auto-regression
(VAR) model, vector error correction (VEC) model.

### 3.2.1 Unit Root Tests

First, unit root tests are used to test stationary properties of data sets. The authors apply Augmented Dickey Duller (ADF) (Dickey & Fuller, 1981; Dickey & Wuller, 1981) tests to test stationary of the variables. Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC) are used to select the appropriate lag length for testing.

### 3.2.2 Co-integration Tests

To determine the existence of long run relationship between the variables, co-integration tests are employed in the non-stationary variables with the same order. We implement VAR-based co-integration tests using the methodology developed in Johansen (1991, 1995).

Consider a VAR of order p:

\[
y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + Bx_t + \epsilon_t
\]

Where \( y_t \) is a k-vector of non-stationary I(1) variables, \( x_t \) is a d-vector of deterministic variables, and \( \epsilon_t \) is a vector of innovations.

Rewriting equation (1) as,

\[
\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + Bx_t + \epsilon_t
\]

Where

\[
\Pi = \sum_{i=1}^{p} A_i - I, \quad \Gamma_i = -\sum_{j=i+1}^{p} A_j
\]

Johansen’s method is used to estimate the \( \Pi \) matrix from an unrestricted VAR, and to test whether we can reject the restriction implied by the reduced rank of \( \Pi \).

### 3.2.3 VEC Model and EC Model

EC models can be constructed from the co-integration equation for short-run adjustment dynamics. Engle & Granger (1987) pointed out that a valid error-correction representation of the data exists for a set of variables that are co-integrated. Consider a two-variable with one co-integration equation and no lagged difference term. The co-integrating equation is

\[
y_{1,t} = \beta y_{2,t} + \epsilon_t
\]
The corresponding EC model is

$$\Delta y_{1,t} = \alpha_1 (ec_{1,t-1}) + \varepsilon_{1,t} = \alpha_1 (y_{1,t-1} - \beta y_{2,t-1}) + \varepsilon_{1,t}$$  \hspace{1cm} (5)$$

Equation (4) is the co-integration model, denoting the long run equilibrium, and $ec_{1,t}$ is error correction term, which is stationary. Equation (5) measures short-run adjustment, and $\alpha_{i}$ measures the speed of adjustment of $y_{1,t}$ towards the equilibrium.

A vector error correction (VEC) model is a restricted VAR designed for use with non-stationary series that are known to be co-integrated.

Based on model(5), the VEC model is

$$\Delta y_{1,t} = \alpha_1 (ec_{1,t-1}) + \varepsilon_{1,t}$$
$$\Delta y_{2,t} = \alpha_2 (ec_{1,t-1}) + \varepsilon_{2,t}$$

The general VECM (p) form is written as

$$\Delta y_{t} = \delta + \Pi y_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta y_{t-i} + Bx_t + \varepsilon_t$$

where $\Delta$ is the differencing operator, $y_t$ is a k-vector of non-stationary I(1) variables.

### 3.2.4 System Estimation

Many studies criticize using a VEC model without theory-based as a pure mathematical convenience (Bonham et al., 2009; Pesaran & Shin, 2001, 2002), while testing and imposing theory-based models is welcomed.

For port cargo container throughput in PRD region, we find that there is long-term equilibrium between Shenzhen and Hong Kong, although Guangzhou and Shanghai ports also affect them, and the co-integration relationship is significant statistically. But the forecasting accuracy benefits little from this because the equilibrium between Shenzhen, Hong Kong is not as steady as that among Guangzhou, Shenzhen and Hong Kong, in practice. It is not proper to treat all the ports the same. As an alternative, it will be beneficial to use a restricted co-integration for Guangzhou port.

Therefore, we employ different types of models for each port, and estimate them simultaneously.

### 3.2.5 Forecast Evaluation

Additionally, three main evaluation criteria are used in this study to assess the model (De Gooijer and Hyndman, 2006). Let $Y_t$ denote the observation at time $t$ and $F_t$ denote the forecast of $Y_t$. [10]
Define the forecast error as \( e_t = F_t - Y_t \), and the percentage error as \( p_t = 100 \cdot e_t / Y_t \).

Commonly used measures MSE and MAPE are used, which are defined as

\[
MAPE = mean(|p_t|)
\]

sMAPE, which is discussed in Goodwin & Lawton (1999), has a heavier penalty when forecasts are high compared to when forecasts are low for the same value of \( Y_t \). So sMAPE is also used in order to compare with measure of MAPE.

\[
sMAPE = mean(2|Y_t - F_t|/|Y_t + F_t|)
\]

4. The Estimation and Prediction Results

4.1 The Model

In this system, there are four ports: Hong Kong, Shenzhen, Guangzhou and Shanghai. We focus on the PRD ports. Firstly, we consider the relationships among the ports. When considering one port, we introduce other ports’ container throughputs as explanatory variables. Many literatures have predicted cargo throughput for Hong Kong. Hui et al. (2004) forecast it using the ECM model. Seabrooke et al. (2003) propose a regression model that consists of 43 equations grouped into four different modules and predicts cargo growth and development of the port of Hong Kong by means of regression analysis.

Based on the model discussed in Section 2.1, we use the model as follows.

\[
\Delta Y_t = A \begin{pmatrix} ec_{t-1}(HKP, SZP, GZP) \\ ec_{t-1}(HKP, SZP, GZP) \\ ec_{t-1}(HKP, SZP, GZP) \end{pmatrix} + \sum_{i=1}^3 \phi_i \Delta Y_{t-i} + \beta \cdot SHP_{t-2} + B \cdot \begin{pmatrix} HKT_{t-1} \\ SHP_{t-1} \\ GZP_{t-1} \end{pmatrix}
\]

Where \( Y_t = (HKP_t, SZP_t, GZP_t)' \), and ec (HKP, SZP) denotes the error correction term of variables HKP and SZP. We find that forecasting of Guangzhou port does not benefit from the co-integration, so a restriction on the co-integration of the coefficient of Guangzhou port’s error correction term is imposed. LR test is used to test whether it is proper to impose the restriction or not (Boswijk, 1996; Boswijk & Jurgen, 2004). The test does not reject the imposed restriction at conventional levels, and the result is listed in Table 7 in Appendix.

Further, Shanghai port, as one representation of the mainland, serves as an explanatory variable.
We only introduce the two-step lagged variables according to their corrections. As a representation of demand, trade is also introduced in the model. The structure of the three ports is a little different, and their hinterlands are not the same. For Hong Kong and Shenzhen, the hinterlands are more similar than Guangzhou, so the sum of the total trade value of Hong Kong and Guangdong province is used to denote the demand of Hong Kong and Shenzhen ports. For Guangzhou port, only the total trade value of Guangdong province is used to represent the demand. It is to be noted here that we use Shanghai port in PRD ports’ equations as a representation of other Chinese ports outside the PRD region. It does not mean that other Chinese ports do not have any effects on container throughput in the PRD. It is just that we could not consider all variables in one model. However, Shanghai port is the most typical one, which is considered representative. As representation of the demand, trade affects container throughput by nature. But it is not proper to treat all ports’ demand as the same even within the PRD region. As mentioned before, we use the sum of the total trade volume of Hong Kong and Guangdong as the representation of demand of Hong Kong and Guangzhou ports, the total trade volume of Guangdong province as the representation of demand of Guangdong, and the total trade volume of Shanghai as the representation of the Shanghai port. When forecasting the container throughput, value of the trade is obtained from a VEC model (Equation(7)).

\[
\Delta \text{Trade}_t = \delta + \Pi \text{Trade}_{t-1} + \sum_{i=1}^{3} \phi_i \Delta \text{Trade}_{t-i} + \epsilon_t
\]

(7)

Where Trade= (HKT, SZT, GZT).

The recent financial crisis has seriously affected all ports; some ports container throughput declined sharply, by nearly fifty percent. Consequently, container throughput has undergone a structure change. However, the relationships among the ports or other explanatory variables may not have been changed. How to deal with this abnormal data and how to estimate the effects of the crisis? That would be an interesting issue.

According to practical situations, the relationships in the system changed a little. The decline is mainly attributed to the unexpected shock inflicted by the financial crisis. So we introduce a dummy variable to denote the financial crisis in each equation in Equation(6).

Thus, the model takes the form:
\[ \Delta Y_t = A \left( ec_{t-1}(HKP, SZP, GZP) + \beta \cdot SHP_{t-2} + B \cdot \begin{pmatrix} HKT_{t-1} \\ SZP_{t-1} \\ GZP_{t-1} \end{pmatrix} + \mu \cdot \text{dummy} \right) \]

This will be helpful in the following aspects. Firstly, it takes the structure change into account in case of misspecification of the coefficient. Secondly, influence of the financial crisis will be estimated by the coefficient of the financial crisis.

Thus, we extend our models in the following aspects:

First of all, based on Equations (7) and (8) for container throughput and trade in the PRD region, a simultaneous system is established and forecasting is made simultaneously. This is helpful to enlarge the sample and to use all available data to estimate the coefficient. At the same time, it also builds a bridge between container throughput and the trade. This will be of great help to investigate the relationships between them.

This financial crisis has had great effects on container throughput. It is impossible to analyze the container throughput without considering the effect of the financial crisis. However, there is a lack of literatures which consider the financial crisis simultaneously while analyzing or predicting port traffic. Under this circumstance, it is impossible, and it makes no sense, to hope to provide an accurate prediction while ignoring the great shock caused by the financial crisis. This paper is the first attempt to introduce a dummy variable which denotes the financial crisis in each equation of the system. We analyze the effect of the financial crisis for each port. In this paper, all ports’ container throughputs and trade are investigated under one conceptual framework for the first time.

It is convenient to update the system and use it to make forecasts in practical situations.

### 4.2 Data Preparation and Estimation Results

The Hong Kong port’s container throughput data used in this study are monthly and are obtained from Census and Statistics Department (C&SD) of the Hong Kong Government (http://sc.info.gov.hk). Furthermore, there are also some other monthly data from Census and Statistics department of the Hong Kong Government (http://sc.info.gov.hk) and the CEIC database, covering the period from January 2000 to December 2009. Statistical analyses in this section will consider the following variables:

The total container throughput is measured in thousand TEUs for all the four ports considered.
The total value of foreign trade of all the three regions (Hong Kong, Guangdong and Shanghai is measured in constant million US$).

First, all data are nature logarithm transformed. Next, the unit root test results of the transformed data are obtained, as listed in Table 1.

<table>
<thead>
<tr>
<th>Region</th>
<th>Level t-stats</th>
<th>P-value</th>
<th>1st difference t-stats</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HKP</td>
<td>0.647</td>
<td>0.855</td>
<td>-17.108</td>
<td>0.000</td>
</tr>
<tr>
<td>GZP</td>
<td>0.960</td>
<td>0.910</td>
<td>-12.514</td>
<td>0.000</td>
</tr>
<tr>
<td>SZP</td>
<td>3.172</td>
<td>1.000</td>
<td>-17.102</td>
<td>0.000</td>
</tr>
<tr>
<td>SHP</td>
<td>4.293</td>
<td>1.000</td>
<td>-2.204</td>
<td>0.027</td>
</tr>
<tr>
<td>HKT</td>
<td>1.259</td>
<td>0.947</td>
<td>-1.731</td>
<td>0.079</td>
</tr>
<tr>
<td>GDT</td>
<td>3.402</td>
<td>1.000</td>
<td>-19.037</td>
<td>0.000</td>
</tr>
<tr>
<td>SHT</td>
<td>2.915</td>
<td>0.999</td>
<td>-2.752</td>
<td>0.006</td>
</tr>
</tbody>
</table>

The data set contains monthly data from 2000.1 to 2009.12. All data are seasonally adjusted by X12-ARIMA methods.

The estimated system appears to be an adequate model for forecasting container throughput of ports in PRD. All equations perform reasonably well, explaining 58%, 45% and 49% of variations in Hong Kong, Guangzhou and Shenzhen ports, respectively, when the estimation sample is from 2000.1 to 2009.6.

First, we compare the forecast performance using the methodology proposed in this model with other models like ARIMA, VAR and VEC model. The estimation sample is from 2000.1 to 2009.6, and the prediction sample is from 2009.7 to 2009.12.

Coefficients of month-wise impacts of the financial crisis on Hong Kong, Guangzhou and Shenzhen port are listed in Table 6 in the Appendix. It is obvious that the coefficient of the financial crisis dummy variable does not directly and totally reflect the impact of the financial crisis, but it is a representation of the impact from the financial crisis that other explanatory variables cannot explain. Comparing the coefficients of Jan. 2009 to Jul. 2009, we draw the conclusion that the consequences of financial crisis weakened remarkably PRD’s ports, especially in Guangzhou Port. By May 2009, the effect of the crisis had become non-negative, that is to say, compared with other ports, Guangzhou port had recovered from the crisis. At the same time, coefficients of effect of crisis in Hong Kong and Shenzhen ports also declined gradually. As seen in Fig. 3, Guangzhou port recovered much faster than the other two ports in PRD region, a trend
which could be detected quite early.

![Fig. 3 The impact from the financial crisis](image)

### 4.3 Forecasting performance

This model initially established, in early 2009, the sharp decline of cargo throughput in PRD’s ports. The estimation of the crisis effect has helped reduce the errors. Comparisons of forecast errors in 2009 are shown in Fig. 4. The MAPE of each port is summarized in Appendix.

![Fig. 4 Comparison of forecast errors](image)

The forecasting MAPE and sMAPE in period 2009.2 - 2009.12 are summarized in Table 2, the forecasting sample is from 2000.1 to 2009.1, and the prediction error is calculated based on the bias between the prediction value and the actual value in 2009.2 -2009.12. MAPE and sMAPE are calculated using the method illustrated in Section 2.2.5.

<table>
<thead>
<tr>
<th></th>
<th>Current</th>
<th>VEC</th>
<th>VAR</th>
<th>Current</th>
<th>VEC</th>
<th>VAR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Guangzhou</strong></td>
<td>0.116</td>
<td>0.161</td>
<td>0.113</td>
<td>0.106</td>
<td>0.178</td>
<td>0.122</td>
</tr>
<tr>
<td><strong>Hong Kong</strong></td>
<td>0.033</td>
<td>0.035</td>
<td>0.088</td>
<td>0.033</td>
<td>0.036</td>
<td>0.084</td>
</tr>
<tr>
<td><strong>Shenzhen</strong></td>
<td>0.127</td>
<td>0.170</td>
<td>0.388</td>
<td>0.117</td>
<td>0.157</td>
<td>0.323</td>
</tr>
</tbody>
</table>
The performance of the current method is much better than the other two, especially for Shenzhen and Guangzhou ports. For Guangzhou port, although the MAPE of the current method is a little larger than VAR, the sMAPE shows that it’s better than VAR. And the yearly forecasting results confirmed this. It is an extraordinary challenge to make forecasts during a period of financial crisis due to the fluctuating and non-stationary data which has been structurally changed. In this paper, we try to consider the effect of the crisis in the system. And it seems helpful for analysis and forecast.

However, the percentage error is a little bigger for Guangzhou and Shenzhen ports at the beginning of the crisis shock. When this effect has become steady, the error has reduced significantly. In the last three months of 2009, the forecasting performance is shows in Table 3.

Table 3  Performance of various methods of forecasting in 2009.10-2009.12

<table>
<thead>
<tr>
<th></th>
<th>MAPE</th>
<th></th>
<th></th>
<th>sMAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Current</td>
<td>VEC</td>
<td>VAR</td>
<td>Current</td>
</tr>
<tr>
<td>Guangzhou</td>
<td>0.056</td>
<td>0.073</td>
<td>0.062</td>
<td>0.056</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>0.004</td>
<td>0.032</td>
<td>0.024</td>
<td>0.004</td>
</tr>
<tr>
<td>Shenzhen</td>
<td>0.017</td>
<td>0.037</td>
<td>0.024</td>
<td>0.017</td>
</tr>
</tbody>
</table>

For examining robustness, we also roll the estimation sample. The sample starts from 2000 Jan., and the end time of the sample rolls from 2009.2 to 2009.6. There are six months left for testing the forecasting performance. The forecasting MAPE(Mean absolute percentage error) for the period 2009.3 - 2009.7 is summarized in Tables 5 and 6; ARIMA(1,1,1) model is used as a benchmark model due to its good performance in short term forecasts.

Table 4 One-month ahead forecasting results

<table>
<thead>
<tr>
<th></th>
<th>CDE-MPR</th>
<th>ARIMA(1,1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HK</td>
<td>SZ</td>
</tr>
<tr>
<td>2009.3</td>
<td>0.014</td>
<td>0.038</td>
</tr>
<tr>
<td>2009.4</td>
<td>0.012</td>
<td>0.016</td>
</tr>
<tr>
<td>2009.5</td>
<td>0.020</td>
<td>0.049</td>
</tr>
<tr>
<td>2009.6</td>
<td>0.020</td>
<td>0.041</td>
</tr>
<tr>
<td>2009.7</td>
<td>0.028</td>
<td>0.019</td>
</tr>
</tbody>
</table>

Table 5 Two-month ahead forecasting results

<table>
<thead>
<tr>
<th></th>
<th>CDE-MPR</th>
<th>ARIMA(1,1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HK</td>
<td>SZ</td>
</tr>
<tr>
<td>2009.3</td>
<td>0.014</td>
<td>0.038</td>
</tr>
</tbody>
</table>
According to the forecasting results, most of the percentage errors for one-month ahead forecasts are below 3%. Most of the two months ahead forecasting percentage errors are below 5%, and for Hong Kong port, all of them are below or near 3%.

### 5. Conclusion

This paper proposed a new methodology- CDE-MPR methodology to forecast the container throughput. A specific case using this methodology to forecast the container throughput within ports in PRD region is provided. This is a systematic methodology to forecast cargo throughputs of multi-ports in one region, like PRD. Under this new methodology, a systematic model comprises two modules - the container throughput module and the trade module in multi-port region (MPR) and one event- the global financial crisis. Based on the vector auto-regression (VAR) model, the container throughput module explores internal and external relationships among the three ports in the PRD region and its dynamic structure being formed in South China. The trade module, based on the vector error correction (VEC) model, is of great help to provide information of the demand of container throughput. In addition, the financial crisis effect is also considered in the system, which is helpful for analyzing the consequences of the financial crisis and reducing the error of prediction.

The results from this study suggest that there is competition among the three ports in PRD, except the effect from Shenzhen port to Hong Kong port. According to the results, the influence of financial crisis is sharply weakened, especially for Guangdong port. The financial tsunami has had significant consequences on maritime industries. As a result, patterns of global and local trades have been reshaped and maritime industries performances have been changed significantly. So it is illogical and incorrect to forecast container throughput without taking the effects of financial crisis into consideration. In this system, we consider the effect of the financial crisis by introducing dummy variables, which is helpful for reducing the error of forecasts during this period. Five-month ahead forecasting MAPE of Hong Kong is below 3 percent. And other forecasting
results of the system are also promising.

It is clear that this paper has provided an invaluable insight into the problem of addressing an increasingly important issue in planning the future development of ports in PRD. The findings form a basis for further research on what really can contribute to strategic choices that the regional governments and the port managers make. Furthermore, although the systematic model concentrates on the particular issue of prediction of ports container throughputs in PRD region, it is clear that the idea and method could be easily generalized and used elsewhere.

This paper proposes a CDE-MPR methodology for forecasting and analyzing container throughput volumes in multi-port regions and applies it to investigate various relationships between ports in the PRD region of South China. The empirical studies demonstrate that it is a feasible method to effectively forecast and analyze maritime logistics in multi-port region. It investigates the Pearl River Delta’s ports under one single conceptual framework and forecasts their respective container traffics simultaneously. At the same time, the measure of a chance event - global financial crisis is also available in this methodology. Future research could expand and even refine the each module in CDE-MPR methodology in practical studies. For instance, in D-module (demand), it is also doable to include other important factors, such as the finance and energy markets, into this systematic model for better understanding of the dynamic structure and for more accurate forecasting results. More events and other measure for the chance events are also feasible in this method. For example, Iceland’s volcanic eruption in 2010 may be taken into account when forecasting the ports cargo throughput in Europe.
References


20

Physica Verlag, pp. 91-113.


Appendix

Table 6 The impact from the financial crisis

<table>
<thead>
<tr>
<th></th>
<th>Hong Kong</th>
<th>Guangzhou</th>
<th>Shenzhen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan-09</td>
<td>-0.14***</td>
<td>-0.13***</td>
<td>-0.09***</td>
</tr>
<tr>
<td>Feb-09</td>
<td>-0.13***</td>
<td>-0.11***</td>
<td>-0.11***</td>
</tr>
<tr>
<td>Mar-09</td>
<td>-0.12***</td>
<td>-0.06**</td>
<td>-0.10***</td>
</tr>
<tr>
<td>Apr-09</td>
<td>-0.12***</td>
<td>-0.03*</td>
<td>-0.10***</td>
</tr>
<tr>
<td>May-09</td>
<td>-0.12***</td>
<td>0.00</td>
<td>-0.08***</td>
</tr>
<tr>
<td>Jun-09</td>
<td>-0.12***</td>
<td>0.01</td>
<td>-0.07***</td>
</tr>
<tr>
<td>Jul-09</td>
<td>-0.10***</td>
<td>0.00</td>
<td>-0.06***</td>
</tr>
<tr>
<td>Aug-09</td>
<td>-0.10***</td>
<td>0.01</td>
<td>-0.05**</td>
</tr>
<tr>
<td>Sep-09</td>
<td>-0.10***</td>
<td>0.00</td>
<td>-0.04**</td>
</tr>
<tr>
<td>Oct-09</td>
<td>-0.10***</td>
<td>0.00</td>
<td>-0.04**</td>
</tr>
<tr>
<td>Nov-09</td>
<td>-0.09***</td>
<td>0.00</td>
<td>-0.03</td>
</tr>
<tr>
<td>Dec-09</td>
<td>-0.09***</td>
<td>0.01</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

Table 7 Co-integration restriction LR test

Co-integration restrictions:

Convergence achieved after 6 iterations.

LR test for binding restrictions (rank = 1):

<table>
<thead>
<tr>
<th>Chi-square(1)</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.698434</td>
<td>0.100447</td>
</tr>
</tbody>
</table>

Table 8 The percentage error of each model in 2009

<table>
<thead>
<tr>
<th></th>
<th>Guangzhou</th>
<th>Hong Kong</th>
<th>Shenzhen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>8.44%</td>
<td>-2.56%</td>
<td>11.82%</td>
</tr>
<tr>
<td>VAR</td>
<td>-15.84%</td>
<td>-3.39%</td>
<td>15.42%</td>
</tr>
<tr>
<td>VEC</td>
<td>-9.94%</td>
<td>6.33%</td>
<td>30.92%</td>
</tr>
</tbody>
</table>