Forecasting container transshipment in Germany

Schulze, Peter M.; Prinz, Alexander

Forecasting container transshipment in Germany

Journal:	*Applied Economics*
Manuscript ID:	APE-07-0249
Journal Selection:	Applied Economics
Date Submitted by the Author:	05-Apr-2007
Complete List of Authors:	Schulze, Peter; Johannes Gutenberg-University, Institute of Statistics and Econometrics Prinz, Alexander; Johannes Gutenberg-University, Institute of Statistics and Econometrics
JEL Code:	C22 - Time-Series Models < C2 - Econometric Methods: Single Equation Models < C - Mathematical and Quantitative Methods, C53 - Forecasting and Other Model Applications < C5 - Econometric Modeling < C - Mathematical and Quantitative Methods, L99 - Other < L9 - Industry Studies: Transportation and Utilities < L - Industrial Organization
Keywords:	Globalization, Container transshipment, Time Series Analyses, Forecasting

Forecasting container transshipment in Germany

Authors:
Univ.-Prof. Dr. Peter M. Schulze
Institute of Statistics and Econometrics
Johannes Gutenberg University
Jakob-Welder-Weg 4
55128 Mainz
Germany

Dipl.-Kfm. Alexander Prinz
Institute of Statistics and Econometrics
Johannes Gutenberg University
Jakob-Welder-Weg 4
55128 Mainz
Germany

Abstract

In this paper, we examine container transshipment at German ports using the seasonal ARIMA (SARIMA) model and the Holt-Winters exponential smoothing approach. Our models are designed especially to take account of the seasonal behavior of the quarterly data used. We consider the dynamic development in this sector for the whole container throughput and also the destinations Asia, Europe and North America, which are the world's three main economic regions. Our data runs from the first quarter of 1989 to the fourth quarter of 2006. We provide detailed quarterly forecasts for the year 2007 and 2008.

According to forecasting error measures such as Mean Square Error and Theil's U, the SARIMA-approach yields slightly better values of modeling the container throughput than the exponential smoothing approach.

Our forecast results indicate further strong growth for German container handling in total and especially for the destinations Asia and Europe. Only the container transshipment between Germany and North America shows rather small increases up to the end of 2008.

Correspondence:
Univ.-Prof. Dr. Peter M. Schulze
Institute of Statistics and Econometrics
Johannes Gutenberg University
Jakob-Welder-Weg 4
55128 Mainz

I. Introduction

Seaborne trade laid the foundations for the rapid and mostly steady growth of world trade during the past decades. Especially the rapidly increasing container transport, measured in TEU (Twenty Foot Equivalent Unit), is an important indicator to represent the dynamics of world trade. Despite the fact that, in 2005, container ships only accounted for 14% (as share of the world gross tonnage) of world seaborne trade (Heymann, 2006: 4) the value of goods transported in containers is much higher – oil and gas tankers as well as bulkers are more significant according to weight.

World wide container transshipment grew with an annual rate of about 11% between 1995 and 2005. Forecasts predict annual growth rates of about 10% until 2010 (Heymann, 2006: 6).

In 2005, Germany held about 9% of the worlds' real exports (in prices and exchange rates of the year 2000) (Deutsche Bundesbank, 2006: 33) and more than 3% of the rapidly growing market of container transshipment. Furthermore, Germany controls one third of the worlds' container fleet. Put into figures, German shipping companies controlled 1,157 of the 3,499 container ships in 2006. (ISL, 2006a: 23)

Quantitatively analytical literature on international container throughput, as well as on container transport, is rather scant: On the one hand, there is no freely accessible broad supranational database for international maritime shipping. On the other hand, the theoretical side of this theme is mostly integrated into general logistics, at least in scientific publications. Some institutions like Drewry Shipping Consultants Ltd., London (Drewry), Institut für Seeverkehrswirtschaft und Logistik, Bremen (ISL) and Ocean Shipping Consultants Ltd., Surrey (OSC) regularly publish analyses for certain market segments on a descriptive-empirical base, and sometimes medium term forecasts without giving the methodology.

We present quantitative short-run forecasts by using quarterly data for the German container throughput, separated by the world´s main economic areas Asia, Europe and North America, the so-called triad of globalization. These three regions cumulate around 80 percent of global container shipment. As well, we provide forecasts for the German container throughput in total. These forecasts are important for logistics companies, e.g. shipping companies, port authorities and shipyards. Using quarterly data and therefore catching seasonal effects, it is possible to give a detailed outlook for a manageable time period. Given attention to the seasonal structure of the data, forecasts can account for this level of detail. This can be a great advantage for companies when allocating capacities. The methodologies we use are two univariate time series approaches: the traditional exponential smoothing analysis of Holt-Winters and the ARIMA framework adapted to seasonal data (SARIMA). Certain model specifications of these two approaches are formal equivalents, so another question here is to see whether, when applied, they lead to similar results.

Next, we outline the main estimation and forecast equations. This theoretical part is followed by some remarks on the database. The principle part contains estimation, diagnosis and interpretation of the main results. Finally, using forecast error measures, we compare the results of the two approaches, and present our forecast results.

II. Estimation and Forecast Techniques

High frequency data often exhibit seasonal behavior, that is there are returning characteristics within certain periods of a year. For data with higher than annual frequency it is therefore often appropriate to make seasonal adjustments. Traditionally, seasonality is seen as some kind of pollution of the data. In modern time series analysis, trend and seasonality is seen as a chance to lead to more exact forecasts. Especially in business matters, knowledge of seasonal changes can help to improve outcomes or reduce costs. From the theoretical point of view, taking account of seasonality may reduce forecasting errors.

In the SARIMA methodology (Newbold/Bos, 1994: 319-325), the original data series are differenced by linear transformation until the data seems to be stationary and the estimated autocorrelation function (ACF) and partial autocorrelation function (PACF) show only a few significant, easily interpretable autocorrelations. The linear transformation is done by differencing the data. Non-seasonal and seasonal differencing filters are applied. Seasonal differencing consists of subtracting the values of two observations that are L periods apart. For quarterly data, the double differencing filter, which is a combination of a non-seasonal and a seasonal differencing filter, amounts to

(1)

$$\Delta_1\Delta_4 = (y_t - y_{t-1}) - (y_{t-4} - y_{t-4-1})$$

The forecasting equation for the one-step-ahead out-of-sample forecast of the often used so-called airline model (Box/Jenkins, 1970: 300-322) SARIMA(0,1,1)(0,1,1) applied to quarterly data, is

(2)

$$\hat{y}_{T+1} = y_T + y_{T-3} - y_{T-4} + \beta_1 \cdot e_T + \beta_4 \cdot e_{T-3} + \beta_1\beta_4 \cdot e_{T-4}$$

with the β's being the estimated coefficients of the MA-terms e.
For h = 6, 7, … we get

(3)

$$\hat{y}_{T+h} = \hat{y}_{T+h-1} + \hat{y}_{T+h-4} + \hat{y}_{T+h-5}$$

which indicates that after h = 5, the estimated MA-parts have disappeared.

We then use the Holt-Winters algorithm for seasonal time series based on the exponential smoothing approach (Newbold/Bos, 1994: 199-210). This technique is capable of producing short-term forecasts for a large collection of time series with a trend and additive or multiplicative seasonal variation. The model of the time series y_t $(t = 1, 2, ..., T)$ with linear trend and multiplicative seasonality consists of the following three equations:

(4)

$$a_t = \alpha_1 \frac{y_t}{s_{t-L}} + (1-\alpha_1)(a_{t-1} + b_{t-1})$$

(5)

$$b_t = \alpha_2 (a_t - a_{t-1}) + (1-\alpha_2) b_{t-1}$$

(6)

$$s_t = \alpha_3 \frac{y_t}{a_t} + (1-\alpha_3) s_{t-L}$$

In these equations, α_1, α_2 and α_3 $(0 \le \alpha \le 1)$ are smoothing coefficients, a_t in (4) yields the current level estimate, b_t in (5) shows the estimate of current slope, and s_t in (6) determines the current seasonal factor. L denotes the number of periods per year in the data, e.g. for quarterly data $L = 4$.
Forecasts in the case of a linear trend and multiplicative seasonal factor can be made by

(7)
$$\hat{y}_{t+h} = (a_t + b_t \cdot h)\, s_{t-L+h}$$

with $h = 1, 2, \ldots$. The additive and multiplicative algorithm may yield quite different forecasting values. So the concrete data set leads to considerations about the combination of the model parts as well as the choice of values for the smoothing terms. This decision has been taken by using the model with the smallest forecasting error.

It can be shown that under certain conditions the additive Holt-Winters model is theoretically equivalent to the seasonal ARIMA model (SARIMA) (Mills, 1990: 180-183) and the interesting question to answer is whether the two approaches lead to equivalent forecasting results in applications. The estimations are compared by using the forecast error measures (Greene, 2003: 113) Root Mean Square Error (RMSE), and Theil´s U which is defined as

(8)
$$U = \frac{RMSE_{(forecast)}}{RMSE_{(naive)}}$$

In our model, for Theil´s U we use the value of the prevailing last period y_{t-4} as forecasting values of the naive model.

III. Database

The worldwide container transshipment has increased with impressing two digit growth rates during the past years. Figure 1 visualizes this dynamic obviously exponential growth for the years 1985 to 2005.

This development is driven by the high growth rates of world trade, further globalization with the division of labor, the decentralization of production, the liberalization of trade and also the increasing economic importance of China. Especially the Asian market is the driving force for global container shipments. Many Asian countries depend on seaborne trade because of their insular location or missing landside infrastructure. Figure 2 gives the distribution of the container handling in 2005, itemized by continent. Asia handles nearly 57% of the world container transshipment. Also, the worlds' six biggest container ports (Singapore, Hongkong, Shanghai, Shenzhen, Pusan, and Kaohsiung) are located in Asia. In 2005, these ports cumulated a volume of 101,220 Mln. TEU which was 25.5% of the worlds´ container throughput, and these ports are still experiencing dynamic growth. Thirteen of the twenty biggest container ports of the world (by means of container handling) are located in Asia. Europe handles a proportion of about 20%. The biggest ports here are Rotterdam, Hamburg, and Antwerp (the worlds´ seventh, eighth and twelfth largest container ports), which cumulate a volume of 23,862 Mln. TEU. American container harbors, of which Los Angeles and Long Beach are the biggest, handle 18% of the worlds´container throughput.

The following analysis is based on quarterly time series data of German container transshipment in Mln. TEU, provided by the Federal Statistical Office of Germany (Statistisches Bundesamt). The analysed time period is 1/1989 to 2/2006. We focus on possible seasonal variations of container transshipment.

IV. Model Analysis

The analysis will concentrate on the container transshipment at German ports with destination to one of the three main regions of container shipping Asia, Europe, North America. We will give forecasts for each of these destinations separately and for the container throughput at German ports in total. The development of the latter for the time period 1/1989 to 2/2006 (in 1,000 TEU) is shown in figure 3.

The graph in figure 3 exhibits a non-linear trend with seasonal variations. The seasonal variations show a strong downward movement at the beginning of each year. Graphs for the three regions look similar and are not presented here.

Using the double differencing filter according to equation (1) on German container handling leads to estimated ACF and PACF values visualized in figure 4.

The PACF shows two significant values at lags 4 and 8. The significant ACF value at lag 4 ($\hat{\rho} = -0.454$) is a typical feature of many doubly differenced seasonal time series. It takes values near –0.5 which may suggest overdifferenced seasonal data. Therefore we additionally apply a unit root test for quarterly series to identify the most appropriate differencing filter for y_t.

To confirm the use of the correct differencing filter we carry out the HEGY unit root test for quarterly data with different auxiliary regressions (Hylleberg et al., 1990). Table 1 gives the results for the double differenced container handling in Germany (SDTEU).

The series are tested with different specifications: without a deterministic term (0), with a constant (C), constant and seasonal dummies (C, SD), constant and linear trend (C, Tr), and constant, seasonal dummies and linear trend (C, SD, Tr). If $\pi_1 = 0$, the null hypothesis "presence of a non-seasonal unit root" cannot be rejected; if $\pi_2 = 0$, a half-yearly unit root is present, and the corresponding tests are designed as $t(\pi_1)$ and $t(\pi_2)$. To test for seasonal (quarterly) unit roots, one has to test the complex unit root $\pi_3 \cap \pi_4 = 0$ by an F-statistic $F(\pi_3, \pi_4)$. If both π_3 and π_4 are zero, the joint null hypothesis cannot be rejected, and a seasonal unit root is present.

Nearly all values at $T = 48$ and the conventional significance level of 5% are smaller than the corresponding critical values (Hylleberg et al., 1990: 226-227). The value for π_4 is greater than the critical values, but that does not restrict the results in general. So the nulls are rejected and stationarity by double differencing is achieved. The last column of table 1 indicates the p-values of the LM-test on autocorrelation in the error term which indicates no autocorrelation.

These considerations lead to the identification of a SARIMA(0,1,0)(2,1,0) model. Furthermore, we identified several other possible SARIMA models, but the model mentioned above proved to be the best one.

Analogously, for the three destinations Asia, Europe and North America we identified similar models. Table 2 shows the details of the estimation results of the respectively best model for each destination and total German container handling (world), concerning significance and autocorrelations of the error term. Estimations were done with WinRATS 6.3 using the Gauss-Newton-algorithm.

The first row gives information about the chosen model. The values in parenthesis in the coefficient rows give the p-values of the t-tests for the coefficient estimators. "Q-sign" indicates the p-values of the Box-Pierce statistic on no autocorrelation in the error term. Constants have been generally omitted, because of non-significance.

As an alternative to the SARIMA analysis, the Holt-Winters method provides an easy way to analyze and predict seasonal time series. For the smoothing coefficients in equations (4) to (6) we apply, as commonly made in

applications, $\alpha_1 = 0.20$, $\alpha_2 = 0.25$ and $\alpha_3 = 0.30$. Other α-values were tested, though they did not change the results substantially, but rather led to unstable outcomes.
Investigating the diagrams of the original series for the different regions, we chose the following specifications: an exponential trend with a multiplicative seasonal factor for the destination Asia and the whole German container throughput, while for the destinations Europe and North America we assume a linear trend combined with a multiplicative season.

To compare the two methods – SARIMA and Holt-Winters exponential smoothing – concerning their predictability we computed two measures for ex-post forecast errors, the RMSE and Theil's U. Table 3 gives details on the results.
The RMSE show a mean deviation between 22 and 30 thousand TEU for the three destinations, but 57 thousand TEU for the total German container transshipment. These differences arise from the overall strongly different and heterogeneous development of the worlds´ economic regions. Due to RMSE there are remarkable differences between the two types of models: the RMSE values of the SARIMA approach are about 10% smaller for the destinations Asia and Europe, and more than 30% for the whole container handling. For the destination North America the exponential smoothing approach obviously led to better results.
For Theil's U we use the observed value with a lag of four quarters as 'naive forecasts'. Experience shows that values smaller than 0.5 for Theil's U lead to useful forecasting results (in the past). Again, only the value for North America leads to a value greater than 0.5, which means that the two approaches produce better results than the 'naive model', but it should not be used for ex ante forecasts. For all datasets except North America the SARIMA U-values are smaller than those of the Holt-Winters model. Therefore, the SARIMA approach should be used for ex ante forecasting rather than the Holt-Winters approach. The identified SARIMA models will lead to better and more stable forecasts, given stationarity. Visualizing these findings, figure 5 compares the forecasting results of the SARIMA and the Holt-Winters model for the whole German transshipment from 3/2006 until the end of 2008. It can be seen that the Holt-Winters approach overestimates the development of the container throughput in the future.

In figure 6, the forecasts of the SARIMA-models are shown for the three regions Asia, Europe and North America. Table 4 contains the numerical forecasting values. Taking a look at the graphical illustration, the seasonal variation which is inherent in the quarterly data can be found in the forecast values as well.

Our forecasts give a continuingly positive outlook for the container industry in Germany. Especially the destinations Asia and Europe are the driving forces for development. We forecast that within the forecasting period the German container throughput will rise by 22.8 percent in total, for the destination Asia we predict a gain of 25.3 percent and for Europe 21.5 percent. A moderate increase of 9.5 percent is given for destination North America.
Given such detailed forecast information as in table 4, port authorities can do their planning more efficiently. Furthermore, capacity planning of shipping companies can be positively influenced.

V. Conclusions

This analysis discusses SARIMA and exponential smoothing models for estimating and forecasting quarterly data of the German container transshipment (measured in TEU). Different models are identified and tested for the whole German container throughput and for the throughput with the destinations Asia, Europe and North America. The SARIMA models are examined by the ACF and PACF, the HEGY test on stationarity for quarterly data, the Box-Pierce statistic and t-tests.

Although there is, theoretically, equivalence between the two types of models, the SARIMA models perform slightly better than the Holt-Winters approach – expressed by the forecasting measures MSE and Theil's U.

The quarterly forecasts up to the end of the year 2008 show a permanent dynamic growth of German container throughput, especially with the destinations Asia and Europe. A moderate increase can be anticipated for the container handling concerning North America.

The methodologies we use can be easily applied to many kinds of time series data. Given more detailed data, for example on smaller aggregates such as a small group of countries, valuable information on the further development in a special sector can be produced. For Germany, container handling with the destination of Baltic Sea abutters could be of particular interest.

References

Box, G. E. P., Jenkins, G. M. (1970), Time Series Analysis, Forecasting and Control, San Francisco, Holden Day

Deutsche Bundesbank (2006) Monatsbericht 12, Frankfurt/Main

Greene, W.H. (2003) Econometric Analysis, 5th ed., Upper Saddle River (N.J.), Prentice Hall

Heymann, E., (2006) Containerschifffahrt, Deutsche Bank Research, Nr. 347

Hylleberg, S., Engle R. F., Granger, C. W. J., Yoo, B. S. (1990), Seasonal Integration and Cointegration, Journal of Econometrics, 44, 215-238

Institut für Seeverkehrswirtschaft und Logistik (ISL) (2006a), Shipping Statistics and Market Review (SSMR) – World Merchant Fleet by Ownership Patterns, Vol. 51, Iss. 4

Institut für Seeverkehrswirtschaft und Logistik (ISL) (2006b), Shipping Statistics and Market Review (SSMR) – General Cargo and Container Shipping, Vol. 51, Iss. 6

Institut für Seeverkehrswirtschaft und Logistik (ISL) (2006c), Shipping Statistics and Market Review (SSMR) – World Port Development, Vol. 51, Iss. 11/12

Mills, T. C. (1990), Time Series Techniques for Economists, Cambridge University Press

Newbold, P., Bos, Th. (1994), Introductory Business and Economic Forecasting, 2nd ed., Cincinnati (Ohio), South-Western Publ.

Statistisches Bundesamt, Seeschifffahrt, Fachserie 8 Reihe 5, different sets, Wiesbaden

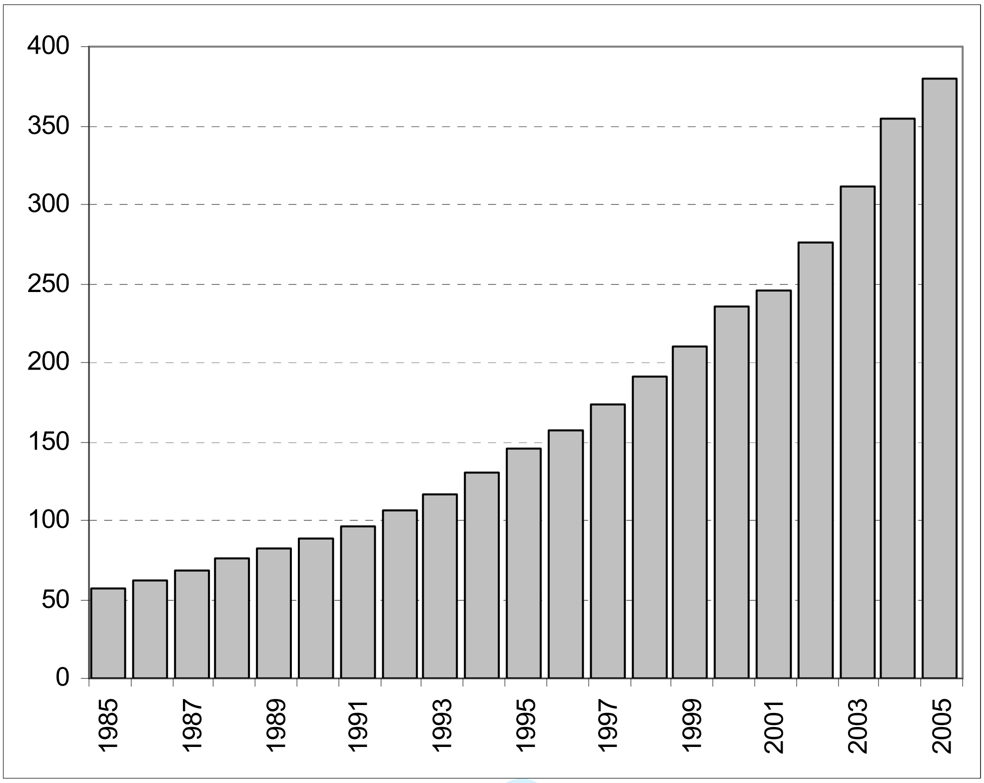

Figure 1: Container transshipment 1985 - 2005 (in Mln. TEU) worldwide
Source: ISL (2006b)

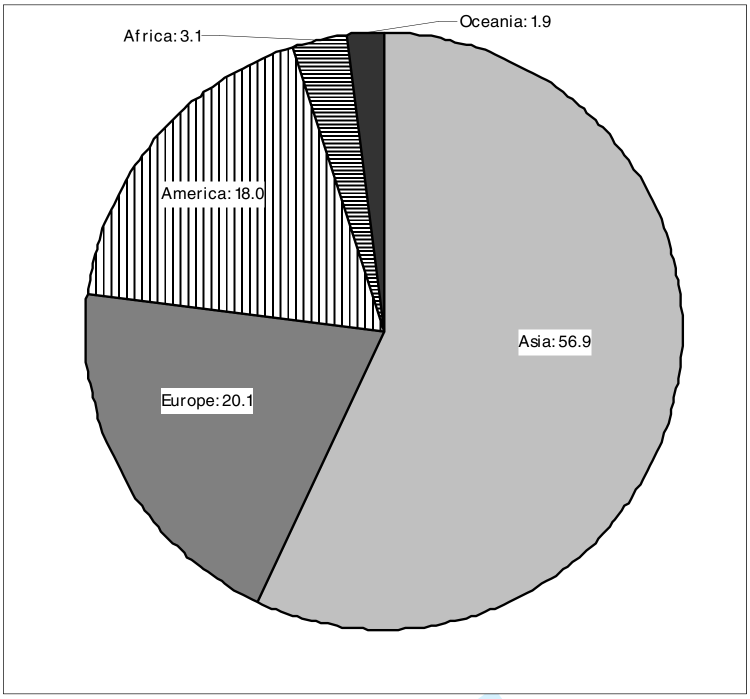

Figure 2: Container handling 2005 divided to continents.
Source: ISL (2006c)

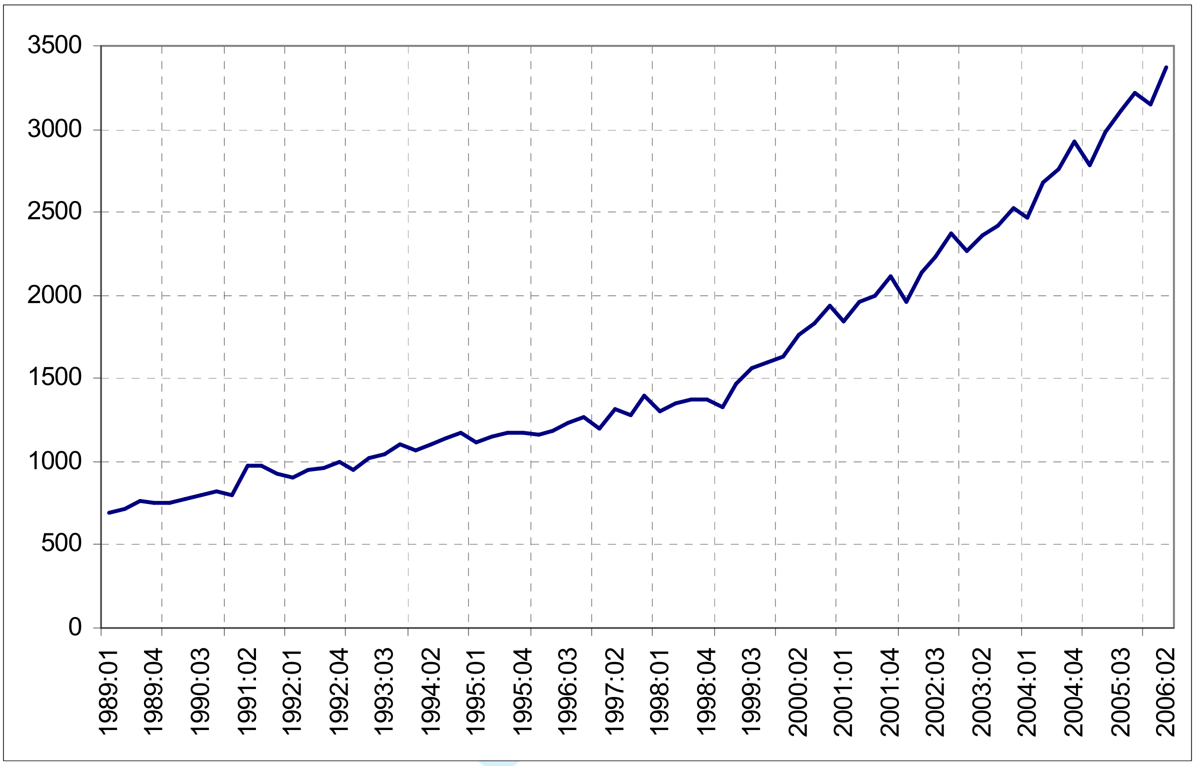

Figure 3: Container transshipment in Germany 1/1989-2/2006 (in 1,000 TEU)
Source: Statistisches Bundesamt

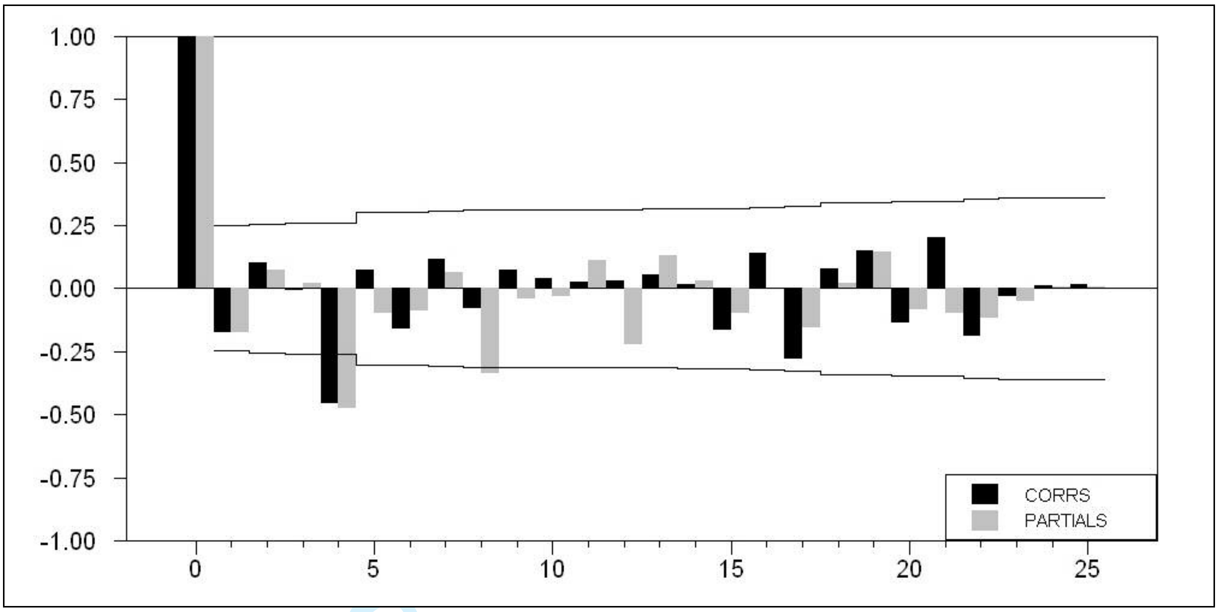

Figure 4: Estimated ACF and PACFvalues of the double differenced German container throughput

HEGY unit root test: testing for seasonal integration in the series SDTEU Effective sample: 1992:03 to 2006:02						
Auxiliary regressions	$t(\pi_1)$	$t(\pi_2)$	$t(\pi_3)$	$t(\pi_4)$	$F(\pi_3, \pi_4)$	LM-sign
0	-5,816	-5,565	-8,543	0,752	37,733	0,259
C	-5,949	-5,562	-8,537	0,780	37,732	0,160
C, SD	-5,839	-5,535	-8,482	0,722	37,193	0,118
C, Tr	-5,946	-5,527	-8,519	0,741	37,540	0,101
C, SD, Tr	-4,375	-3,495	-4,789	1,185	12,931	0,108

Table 1: Results of the HEGY test for German container throughput

	Asia	Europe	North America	World
SARIMA	(0,1,0)(1,1,0)	(0,1,0)(0,1,1)	(1,1,0)(1,1,0)	(0,1,0)(2,1,0)
Iteration steps	2	2	7	2
AR			-0.4916 (0.0001)	
SAR	-0.5724 (0.0000)		-0.4465 (0.0003)	-0.5909 (0.0000) -0.3313 (-0.0092)
SMA		-0.7192 (0.0000)		
Q-sign	0.4397	0.6419	0.6122	0.5969

Table 2: Estimation results for the different SARIMA models

	Holt-Winters exponential smoothing				SARIMA		
Region	Trend	Season	RMSE	U	Type	RMSE	U
Asia	Exponential	Multiplicative	26.37	0.3118	(0,1,0)(1,1,0)	24.72	0.2923
Europe	Linear	Multiplicative	29.99	0.4132	(0,1,0)(1,1,0)	27.39	0.3666
North America	Linear	Multiplicative	22.57	0.7885	(1,1,0)(1,1,0)	24.12	0.8070
World	Exponential	Multiplicative	57.08	0.3116	(0,1,0)(2,1,0)	43.16	0.2356

Table 3: Model types and forecast errors

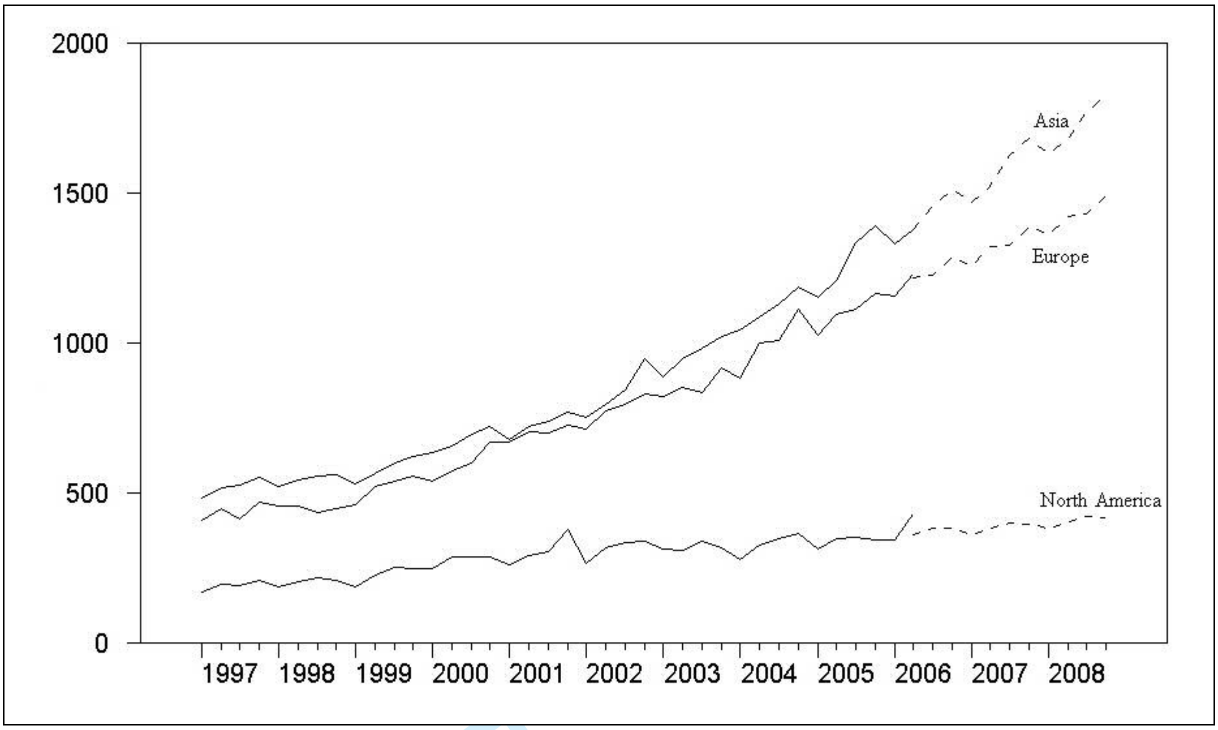

Figure 5: SARIMA and Holt-Winters forecasts for the whole container throughput in Germany (in 1,000 TEU)

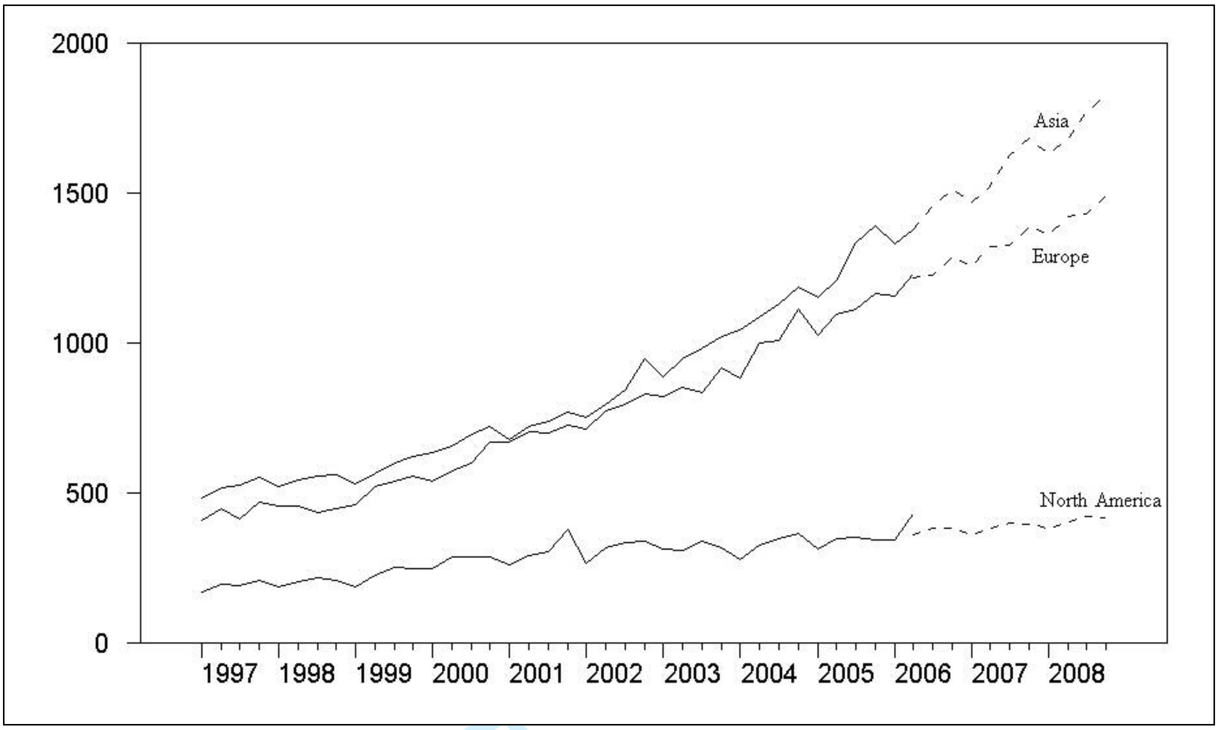

Figure 6: SARIMA forecasts for destinations Asia, Europe and North America (in 1,000 TEU)

	World	Asia	Europe	North America
2006:03	3416.163	1457.342	1225.300	382.019
2006:04	3536.961	1512.480	1285.345	381.409
2007:01	3450.250	1466.651	1257.379	360.268
2007:02	3645.487	1520.959	1319.417	384.123
2007:03	3744.720	1626.133	1326.679	397.893
2007:04	3876.864	1682.909	1386.725	394.591
2008:01	3774.378	1627.252	1358.758	381.691
2008:02	3967.375	1679.497	1420.796	403.133
2008:03	4075.730	1769.483	1428.058	420.350
2008:04	4195.274	1825.322	1488.104	418.233

Table 4: Forecasting values for container throughput (in 1,000 TEU)