

東アジア積米国揚コンテナ荷動き予測における SARIMA モデルの適用性

An Applicability of SARIMA Model for Forecasting Container Movement
from East Asia to U.S.

川崎智也(正会員：日本大学)、松田琢磨(非会員：日本海事センター)、

花岡伸也(正会員：東京工業大学)

Tomoya KAWASAKI (Nihon University), Takuma MATSUDA (Japan Maritime Center),

Shinya HANAOKA (Tokyo Institute of Technology)

要旨

東アジア積米国揚コンテナ航路は、2011 年の実績で 1 千万 TEU を超える基幹航路である。基幹航路のコンテナ荷動き動向は船社の配船計画などに大きな影響を与えるため、荷動き予測のニーズは高い。しかし、当航路の月次レベルの荷動き予測について、時系列モデルの適用性を検証した研究は少ない。そこで本研究では、季節変動自己回帰移動平均 (SARIMA) モデルを用いてコンテナ荷動き量の予測モデルを開発し、その適用性を検証した。検証の結果、同定されたモデルの残差は自己相関を有さず、実績値と再現値による相関係数は 0.983 となり、SARIMA モデルが当航路の荷動き予測の利用に耐えうることを実証した。

Abstract

A container sea route from East Asia to U.S. is one of the most important trunk lines due to its high volume. Since container movement of trunk line significantly affects vessel assignment plans by shipping liners, demand on container forecasting is fairly high. However, forecasting method using time-series model on its sea route is not examined so far. In this study, an applicability of seasonal autoregressive integrated moving average model for forecasting container movement is inspected. The residual has no autocorrelation and defined model can reproduce volume of container movement with high accuracy that correlation coefficient is as 0.983.

1. はじめに

海運業界では、基幹航路のコンテナ荷動き予測は最重要関心事項の一つである。将来予測は適切な配船戦略や荷主との運賃交渉¹に有益な情報となり得るほか、船社の株価に見通しを付けることができる。

一般的に、コンテナ荷動き量は輸入側の実質国内総生産 (GDP) と密接な関係にある。例えば、2000 年第 1 四半期～2010 年第 2 四半期におけるアジア積米国揚コンテナ荷動き量は、米国の実質 GDP との相関係数は 0.89 と

強い正の相関を有している⁽¹⁾。また、同期間のデータを基に回帰分析を行うと、米国の実質 GDP が 1% 伸びると、コンテナ荷動き量は 3.8% 押し上げられる⁽¹⁾。しかし、実質 GDP は四半期データであるため、月次レベルでの予測が行えないという欠点がある。

荷動き需要予測手法としては、四段階推定法に依拠した需要予測モデルなどがある。しかし、当手法の適用には、荷動きの純流動データのほか、ゾーン間の所要時間や費用など、膨大な入力データが必要となる。

一方、計量時系列モデルは入力データの量が比較的少なく、手法も比較的簡便であることから、経済、金融工学分野を中心に、実務

¹ 例えば、2007～2012 年における月次の中国発米国向け運賃 (出所：中国国際海運網) とコンテナ荷動き量 (出所：PIERS) の相関係数は 0.44 とやや弱い正の相関を有している。

的に幅広く利用されている^{例えば、(2)、(3)}。計量時系列モデルの中でも、自己回帰移動平均 (ARMA) モデルは予測対象の時系列データのみでモデルを同定できるため、簡便である。しかし、ARMA モデルをコンテナ荷動き予測に適用した事例は少ない。予測対象となる時系列データに、説明変数となるその他の時系列データを加えたベクトル自己回帰 (VAR) モデルを用いて、インドネシアにおける取扱貨物量の将来予測を行った研究⁽⁴⁾がある程度である。

本研究では、実務での利用を念頭に置き、ARMA モデルに季節性を考慮した季節変動自己回帰移動平均 (SARIMA) モデルを用いて、基幹航路である東アジア (中国、韓国、日本、台湾) 積米国揚航路のコンテナ荷動き量の月次予測を行う。SARIMA モデルは、被説明変数である荷動き量の時系列データのみを入力データとして必要とし、過去の荷動き変化の特徴から将来荷動き量を予測するモデルである。また、各種統計量を用いて SARIMA モデルの適用性を検討する。

本論文の構成は以下の通りである。まず 2. において、アジア積米国揚では、東アジア積荷動き量が多いことに言及し、研究対象航路にした理由を説明する。次に 3. では時系列データを取扱う際に必要なデータの定常性を確認し、4. において SARIMA モデルを同定する。5. では同定したモデルの妥当性を検証し、6. で SARIMA モデルの適用性について総括する。

2. 東アジア積米国揚コンテナ荷動き概要

本研究の対象航路である東アジア積米国揚航路のコンテナ荷動き動向について概説する。本研究で用いる荷動きデータは、全て Port of Import/Export Reporting Service (PIERS) のデータベースを用いており、アジアとは、日本、韓国、台湾、中国、香港、マカオ、シンガポール、フィリピン、マレーシア、インドネシ

ア、タイ、ベトナム、カンボジア、ミャンマー、インド、パキスタン、スリランカ、バングラデシュの 18 ヶ国・地域を含む。

PIERS データベースによると、2011 年のアジア積米国揚 (往航) コンテナ荷動き量は前年比 0.4% 増の 1,314 万 TEU で、ほぼ横ばいであった。その一方、同年の米国積アジア揚 (復航) コンテナ荷動き量は前年比 7.3% 増の 670 万 TEU であり、アジアー米国コンテナ航路は、伝統的に往航の荷動き量の方が多い。これは、アジアで生産したモノを米国で消費する構図を反映している。

2011 年の往航における仕出し国別コンテナ荷動き量と構成比を表 1 に示す。近年、中国積は貨物量を増やし続けており、約 849 万 TEU でアジア積全体の約 65% と圧倒的なシェアを占め、アジア積全体への影響力は大きい。しかし、2011 年の中国積は、前年比 0.2% の減少となるなど、2000 年代前半に記録した高い成長率は陰りを見せている。中国積では、家財関連品が 190 万 TEU で中国積全体の 14.4% と最多で、当品目が中国積コンテナ荷動き量に大きな影響を与えている。2011 年の米国住宅市場は新築の住宅販売が 364.2 万戸で前年比 5.5% 減⁽⁵⁾と住宅市場が低調であった

表 1 2011 年の往航における仕出し国別コンテナ荷動き量とシェア

順位	仕出し国	荷動き量 (万 TEU)	構成比
1	中国	848.8	64.6%
2	韓国	69.9	5.3%
3	日本	62.8	4.8%
4	台湾	52.5	4.0%
5	香港	50.7	3.9%
6	ベトナム	49.9	3.8%
7	インド	41.5	3.2%
8	タイ	35.1	2.7%
9	インドネシア	30.9	2.4%
10	マレーシア	23.7	1.8%
	その他合計	47.8	3.5%
	合計	1,313.6	100.0%

PIERS データベースを基に筆者作成

ことが、2011年の中国積コンテナ荷動きが低調であった一因と考えられる。

2011年におけるアジア積のコンテナ荷動き量の上位5ヶ国は、上位から中国積（構成比64.6%）、韓国積（同5.3%）、日本積（同4.8%）、台湾積（同4.0%）、香港積（同3.9%）となっている。上位5ヶ国の顔ぶれは10年以上不変だが、2011年は台湾積が香港積を上回って4位となった。香港積は、2010年こそリーマンショック後の反動でプラス成長となったものの、過去10年は減少傾向が続いており、10年間で約50万TEU減少した。一方、ベトナム積は香港積とは対照的に、過去10年間で約50万TEU伸びた。この傾向が継続すれば、2012年には香港積が上位5ヶ国から脱落する可能性が高い。香港港は長年中国のゲートウェイ港としての役割を担ってきた。しかし、近年の中国の経済成長に伴う港湾整備により香港港を利用する必要性が弱まっているに加え、香港港の高い港湾料金と近隣華南諸港湾との競争により利用量が減少している⁶⁾。韓国積については、近年のウォン安を背景に輸出が好調で、09、10年に続き日本積を上回って2位を維持した。

以上より、近年減少傾向が続いている香港積を除く東アジア4ヶ国・地域（中国積、韓国積、日本積、台湾積）が往航コンテナ荷動き予測において重要であることが分かる。これが、本研究が東アジア積に着目した理由である。

3. データの定常性

本研究では、過去15年分（1997～2011年）の東アジア積米国揚の月次データを用いて、SARIMAモデルの適用性を検討する。過去15年は、1997年のアジア通貨危機や香港の中国返還、2001年の中国のWTO加盟、2000年代の中国の経済成長や米国のIT・住宅バブル、2008年のリーマンショックなど、数多くの荷動き変動要因があった。このような荷動き変

動要因は、将来においても発生することは十分考えられる。したがって、予測モデルはこれらのような荷動き変動要因に対して耐えるものである必要がある。これより、荷動き変動要因が多く含まれる過去15年分のデータを用いた。

図1に1997年1月～2011年12月の計180ヶ月間の東アジア積米国揚コンテナ荷動き量の推移を示す。2006年10月の106万TEUをピークに微減もしくは横ばい傾向が続いている。本研究では、図1に示すコンテナ荷動き量について、SARIMAモデルの適用性を検討する。SARIMAモデルを適用する場合、分析に利用する時系列データの定常性を満たす必要がある⁷⁾。時系列データ y_t の任意の時点 t の荷動き量 (y_t) と任意の時点 t から k 期前の荷動き量 (y_{t-k}) について、式(1)～(3)に示すように平均 $[E(y_t)]$ 、分散 $[Var(y_t)]$ 、自己相関 $[Cov(y_t, y_{t-k})]$ が時間に対して一定であるとき、時系列過程は定常と定義される⁷⁾。なお、自己相関とはある時点の変数とその変数の過去の値との相関である⁷⁾。

$$E(y_t) = \mu \quad (1)$$

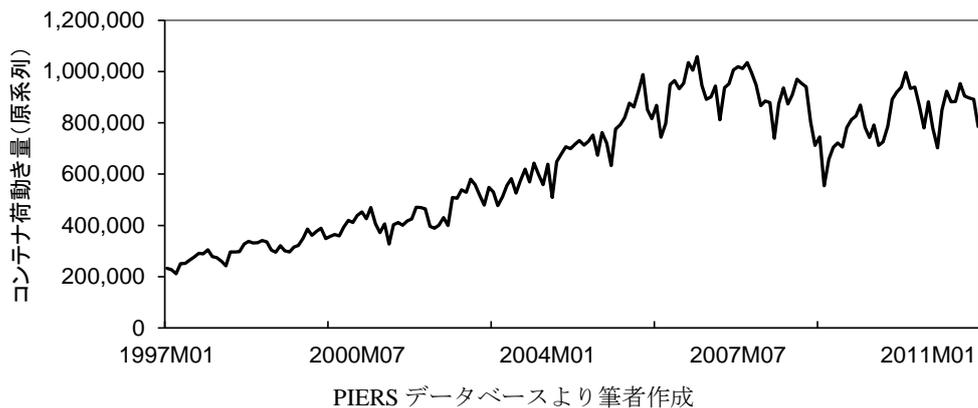
$$Var(y_t) = E[(y_t - \mu)^2] = \gamma_0 \quad (2)$$

$$Cov(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)] = \gamma_k \quad (3)$$

μ 、 γ_0 、 γ_k は、それぞれ平均値、分散、自己相関係数を示している。図2は図1の差分系列であり、視覚的に平均値、分散、自己相関係数が時間に対して一定であると分かる。時系列過程の定常性を統計的に判断するには、単位根検定が用いられる⁸⁾。単位根検定にはいくつか方法があるが、本研究では、一般的に用いられている拡張ディッキーフラー(ADF)検定⁸⁾によって単位根検定を行う。単位根検定では、式(4)に示すモデル式が用いられる。

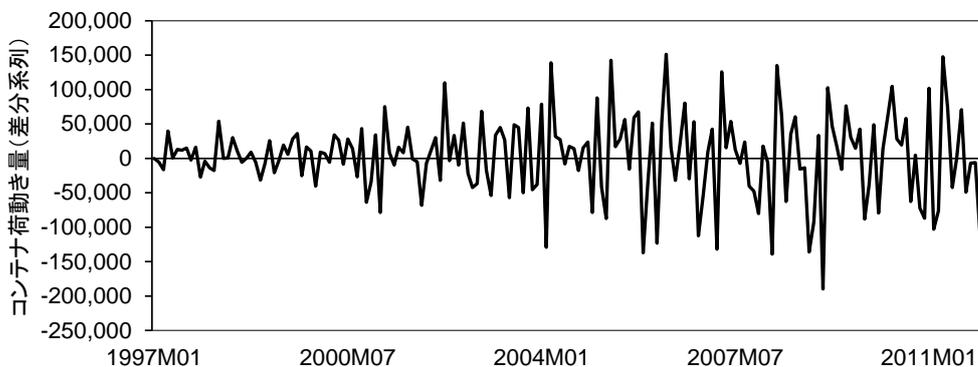
$$y_t = \alpha \cdot y_{t-1} + \sum_i \beta_i \Delta y_{t-i} + e_t \quad (4)$$

ここで、 i はラグ数と呼ばれ、 $i=n$ であれば、



注：東アジアには中国、韓国、日本、台湾が含まれる

図1 東アジア発北米航路におけるコンテナ荷動き量の推移（原系列）（単位：TEU）



注：東アジアには中国、韓国、日本、台湾が含まれる

図2 東アジア発北米航路におけるコンテナ荷動き量の推移（差分系列）（単位：TEU）

n 期前までの荷動きデータが説明変数として用いられることを示している。 i は、赤池情報量基準（AIC）が最小となるラグ数が選択される。 e_t は t 期における誤差項である。

ADF 検定では、式(4)において、帰無仮説を $\alpha=1$ として単位根の存在の有無を検定する。つまり、 $\alpha=1$ であれば単位根は存在する。計算の便宜上、式(4)を式(5)に変形し、 y_{t-1} の係数である“ $\alpha-1$ ”がゼロかどうかを検定する方法が取られることが一般的である。本研究においても、式(5)を用いて時系列過程に定常性が存在するか確認する。

$$\Delta y_t = (\alpha - 1)y_{t-1} + \sum_i \beta_i \Delta y_{t-i} + \mu + e_t \quad (5)$$

表2に単位根検定の結果を示す。データは原系列（図1）と差分系列（図2）の2種類で

行った。原系列に加えて差分系列を用いた理由は、原系列が非定常である場合、差分系列をとることによりデータが定常になることが多い⁽⁹⁾ためである。月次データであることを鑑み、ラグ数は最大12としてAICが最小となる i を選択した。その結果、原系列と差分系列の両系列においてラグ数 i は12が選択された。原系列の単位根検定の結果をみると、 $\alpha-1$ の係数の t 値は-1.101と低く、単位根は存

表2 単位根検定の結果

系列	原系列	差分系列
ラグ数（最大12）	12	12
AIC	24.566	24.500
$\alpha-1$	-0.045	-2.894
t 値	-1.101	-9.368
p 値	0.928	0.000

在しなかった。差分系列の場合は、 t 値が-9.368と99%有意水準でも帰無仮説が棄却されない。これにより、差分系列の定常性が確認された。以上の結果を踏まえ、差分系列を用いて SARIMA モデルを開発し、その適用性を検証することとする。

4. モデル

4-1 モデル概要

ここでは、SARIMA モデルの概要を説明する。時点 t における実績値を y_t とすると、SARIMA モデルは一般的に式(6)のように表すことができる。

$$\phi(L)\Phi(L^s)\nabla_s^D\nabla^d y_t = \theta(L)\Theta(L^s)\varepsilon_t \quad (6)$$

ここに、

$$\begin{aligned} \phi(L) &= 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p \\ \Phi(L^s) &= 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_P L^{Ps} \\ \theta(L) &= 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q \\ \Theta(L^s) &= 1 - \Theta_1 L^s - \Theta_2 L^{2s} - \dots - \Theta_Q L^{Qs} \end{aligned}$$

で、 ϕ_i 、 Φ_i 、 θ_i 、 Θ_i は係数である。 L はラグ演算子 ($L^n y_t = y_{t-n}$)、 ∇ は連続階差を示す差分演算子 ($\nabla^n y_t = y_t - y_{t-n}$)、 ∇_s は季節階差を示す差分演算子 ($\nabla_s^n y_t = y_t - y_{t-n}$) である⁹⁾。このように演算子を定義することにより、 L 、 ∇^n 、 ∇_s を一般の変数とみなして計算できるというメリットがある。なお、 ε_t は誤差項である⁹⁾。

通常、式(6)は式(7)のように要約される。本研究においても、式(7)を用いて SARIMA モデルを表現する。

$$\text{SARIMA}(p,d,q)(P,D,Q)_s \quad (7)$$

- p : 自己回帰項の次数
- d : 階差の次数
- q : 移動平均項の次数
- P : 季節自己回帰項の次数
- D : 季節階差の次数
- Q : 季節移動平均項の次数
- s : 季節変動の期間

自己回帰項の次数 (p) は、自身 (荷動き量) の過去の値について、どこまで遡って説明変数として用いるかを示している。例えば、 p が2であれば1期及び2期前の荷動き量をモデルの説明変数として考慮することになる。階差の次数 (d) については、分析に利用する時系列データが定常性を有するまでに必要となった階差数を示す。本研究では、3. で述べたように、一階の差分系列が定常性を有しているため、 d は1となる。なお、原系列が定常性を有している場合は、 d は0となる。移動平均項の次数 (q) については、自身の過去の値の誤差について、どこまで遡って説明変数として用いるかを示している。

なお、本研究では月次データを用いているため、季節変動の期間 (s) は12である可能性が高い¹⁰⁾。このとき、季節自己回帰項の次数、季節階差の次数、季節移動平均項の次数はそれぞれ P 、 D 、 Q 、に入る数値の12倍前の値が説明変数として用いられる。例えば P が2であれば、12期及び24期前の荷動き量を説明変数として考慮する。

SARIMA モデルは、式(7)で示した p 、 d 、 q 、 P 、 D 、 Q 、 s の7つのパラメータを特定 (特定手順は4-2で示す) することにより規定される。SARIMA モデル同定後は、一般的には gretl や EViews などの計量経済ソフトを用いて式(6)のパラメータ (ϕ_i 、 Φ_i 、 θ_i 、 Θ_i) が推定される。

なお、ここでは SARIMA モデルの概要を説明した。モデルの詳細については沖本⁹⁾や山澤¹⁰⁾に詳しいため、興味のある読者はそれらを参照いただきたい。

4-2 モデルのパラメータ特定

ここでは、式(7)で示した SARIMA モデルのパラメータ特定方法について説明する。

図3に差分系列の自己相関係数、図4に同偏自己相関係数をコログラムとして示す。

前述の通り、自己相関は y_t と y_{t-k} の相関で

あるのに対し、偏自己相関は y_t と y_{t-k} の自己相関関係から $y_{t-1}, \dots, y_{t-k+1}$ の影響を取り除いた相関を示している。破線内の領域は、標本の自己相関がゼロであるという検定の 95% 棄却域である。領域の外側に (偏) 自己相関係数がある場合は、(偏) 自己相関が少なくとも有意水準 95% で存在する。図 3 を参照すると、12 次、24 次、36 次において特に強い自己相関を有していることが分かる。これは 1 年周期の季節性が存在していることを示唆しているため、季節階差 s は 12 とするのが適切と考えられる。そこで、本研究では、季節階差 s は 12 とする。また、前節の単位根検定によって、1 階の差分系列が定常性を有していることが分かった。そのため、 d 及び D は 1 とする。

次に、 p, q, P, Q の特定について検討する。図 3 及び図 4 を参照すると、自己相関と偏自己相関ともに 3 期ラグの相関が比較的高

く、4 期ラグで初めて自己相関係数、偏自己相関係数ともに 95% 水準で棄却される。SARIMA モデルでは、(偏) 自己相関係数のコレログラムを参照するだけではラグ数を決定できない⁽¹⁰⁾。しかし、図 3、4 のコレログラムを参照することにより、 $0 \leq p, q, P, Q \leq 3$ としてパラメータの目安を立てることができる⁽⁹⁾。 p, q, P, Q について、0~3 を各変数に当てはめ、256 通りの (p, q, P, Q) について、 $\phi, \Phi, \theta, \Theta$ の値を最尤法で推定し、AIC が最小となるモデルを同定する。その結果、SARIMA(2,1,1)(0,1,1)₁₂ が同定され、式(6)の $\phi_1, \phi_2, \theta_1, \Theta_1$ は表 3 のように推定された。季節自己回帰 (Φ_i) はゼロで、定数項と 1 期前の移動平均 (θ_1) 以外のパラメータは 99% 有意水準で統計的に有意と推計された。

本モデルより、 p は 2 であるのに対し、 q は 1 と特定された。そのため、ある時期 t の荷動き量予測には、1 ヶ月前 ($t-1$) のデータ

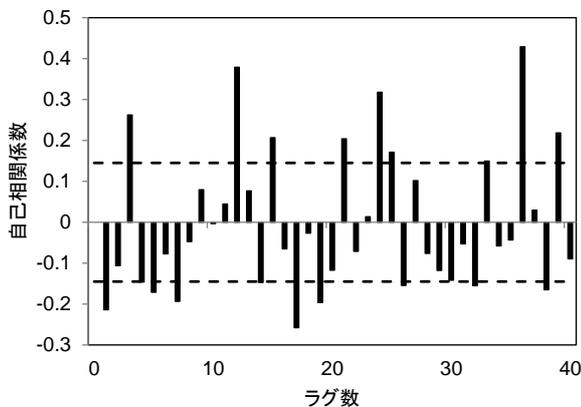


図 3 差分系列の自己相関係数

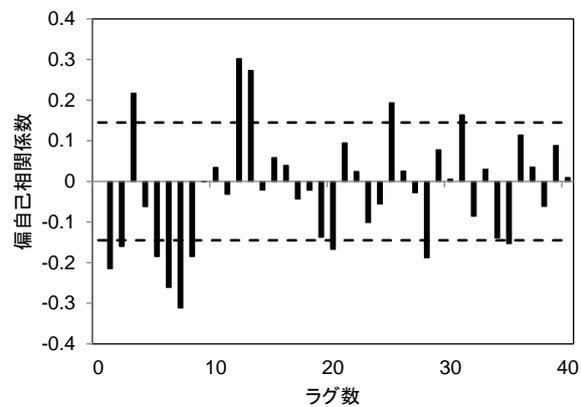


図 4 差分系列の偏自己相関係数

表 3 SARIMA(2,1,1)(0,1,1)₁₂ によるパラメータ推定結果

パラメータ	係数	t 値	p 値
ϕ_1	-0.789	-3.050	0.002
ϕ_2	-0.354	-2.897	0.004
θ_1	0.240	0.872	0.383
Θ_1	-0.699	-11.730	0.000
定数項	-546.375	-0.773	0.440
AIC	4053.782		
対数尤度	-2020.891		
Q 統計量 (p 値)	54.296 (0.065)		

が必要であることが分かる。したがって、本モデルでは月次単位で1ヶ月後の荷動き予測が可能ということになる。ただし、図3、4のコレログラムを参照する限り、 p 及び q は0~3であっても、ある程度の予測精度を有することが示唆される。そのため、2、3ヶ月後の荷動き量予測においても、SARIMAモデルは使用に耐えうる可能性が高いものと考えられる。

5. モデルの診断

5-1 残差の定常性

前章までの検討の結果、SARIMA(2,1,1)(0,1,1)₁₂がAICを最小化するモデルとして選択された。このモデルがコンテナ荷動き予測に耐えうるものかどうかを診断するため、同

定されたモデルによる残差の(偏)自己相関の値がゼロであるかどうか検証する。同定されたモデルが適切であれば、残差は定常性を持ち、(偏)自己相関を持たないことが分かっている⁹⁾。

図5及び図6に同定されたモデルの残差の自己相関係数及び偏自己相関係数のコレログラムを示す。破線の内側の領域は自己相関の値がゼロであるという検定の95%棄却域を示している。つまり、(偏)自己相関係数が破線の内側に留まっていれば、モデルは適切であると判断される。両図より、(偏)自己相関係数は破線の内側にあり、同定されたモデルによる残差が(偏)自己相関を有さないことが確認できる。

以上より、残差の定常性という点からモデ

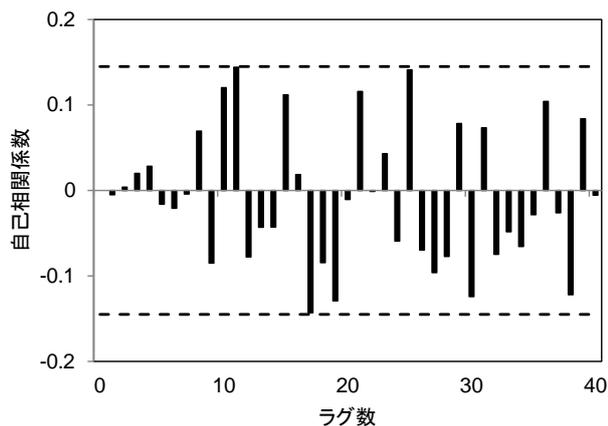


図5 差分系列の自己相関係数(残差)

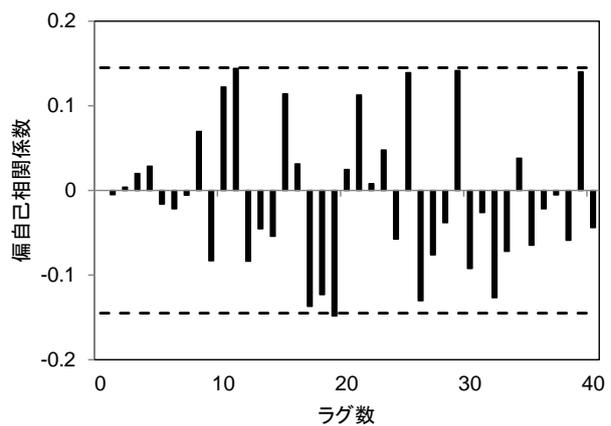


図6 差分系列の偏自己相関係数(残差)

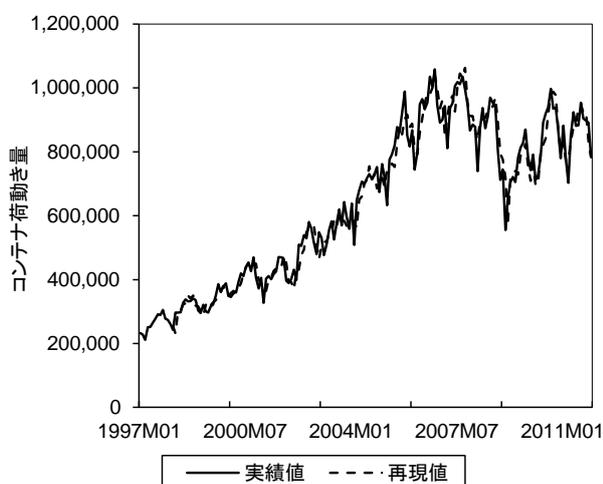


図7 実績値と再現値の比較

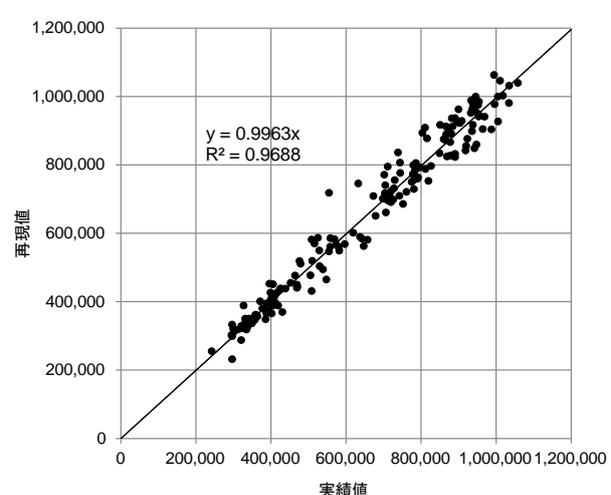


図8 実績値と再現値による散布図

ルの妥当性を確認できた。

5-2 実績値と再現値の比較

図7に1997年～2011年の180ヶ月の実績値を実線で示し、SARIMA(2,1,1)(0,1,1)₁₂により生成した再現値を破線で示す。両者は類似の動きを示しており、モデルによるコンテナ荷動き量の再現性は高いものと考えられる。両者の相関係数は0.983と強い相関関係にある。また、図8に実績値を横軸、再現値を縦軸にした散布図を作成し、原点を通過する回帰直線を引くと、 x の係数は0.996とほぼ1となった。なお、 x の係数が1に近ければ近いほどコンテナ荷動き量の再現性は高い。

以上より、同定されたモデルによる再現性は高く、SARIMAモデルは当航路における将来のコンテナ荷動き予測に耐えうるものと判断できる。

6. おわりに

東アジア積米国揚コンテナ航路は荷動き量の多さから重要な基幹航路として認識されている。そのため、船社を中心に当航路でのコンテナ荷動き予測のニーズは高い。

本研究は東アジア積米国揚航路に着目し、時系列分析手法の一つであるSARIMAモデルを用いてコンテナ荷動き予測を行い、様々な統計量を用いて同定されたモデルの適用性を実証した。モデルはSARIMA(2,1,1)(0,1,1)₁₂と同定され、残差が自己相関を有さないことを確認した。また、実績値と再現値による相関係数は0.983となり、モデルの再現性の高さを示し、実用に耐えうることを示した。

本研究の結果は、当航路において物流事業を展開している船社の配船計画や荷主との運賃交渉などの参考にすることができる。また、船社の株価は船社の収益と関連性を有すると考えられるため、コンテナ荷動きの予測値は海運企業の株価の先行指標として活用できるものと考えられる。

本研究で開発したSARIMAモデルでは、月次単位で1ヶ月後の荷動き予測が可能であるため、短期予測に適している。一方、GDPを利用したモデル⁽¹⁾は四半期ベースであるため、比較的長期の荷動き予測に適している。以上のように、予測対象期間によって両モデルを使い分ける必要がある。

参考文献

- (1) 松田琢磨：北米往航荷動き量の月次予測、2010アジア—北米間コンテナトレード[®]、日刊カーゴ、2010年12月17日
- (2) Tsenga, F. M., Yub, H. C. and Tzeng, G. H. : Combining Neural Network Model with Seasonal Time Series ARIMA Model, Technological Forecasting and Social Change, Vol. 69, Issue 1, pp. 71–87, 2002.
- (3) Zhang, G. P. : Time series forecasting using a hybrid ARIMA and neural network model, Time series forecasting using a hybrid ARIMA and neural network model, Vol. 50, pp. 159–175, 2003.
- (4) Syafi'i, Kuroda, K. and Takebayashi, M. : Forecasting the Demand of Container Throughput in Indonesia, Vol.47, Memoirs of Construction Engineering Research Institute, November 2005.
- (5) 米国商務省経済統計局ホームページ：New Residential Construction
- (6) 池上寛：アジアにおける海上輸送と主要港湾の現状、調査研究報告書、アジア経済研究所、2012
- (7) Hamilton, J. : Time Series Analysis, Princeton University Press, 1994.
- (8) Dickey, D. A. and Fuller, W. A. : Distribution of Estimators for Autoregressive Time Series With A Unit Root, Journal of the American and Statistics Society, Vol. 74, pp. 427-431, 1979.
- (9) 沖本竜義：経済・ファイナンスデータの計量時系列分析、朝倉書店、2010
- (10) 山澤成康：実践計量経済学入門、日本評論社、2004