

## 論文

# 長期統計を用いた中国人口の時系列モデルの推定 と短期将来人口の予測

白砂 堤津耶\*

中国国家统计局の推計によれば、2021 年末現在の中国の総人口は 14 億 1260 万人であり、前年から 48 万人だけ増加したに過ぎない。出生率は 7.52% で、建国以来最低となった。中国では少子高齢化が急速に進行しており、人口減少がいつ始まっても不思議ではない。本稿の目的は、国家统计局が公表している建国以来の長期人口統計 (1949-2021 年) に基づき、人口の時系列モデルを Levenberg-Marquardt 法の非線形最小 2 乗アルゴリズムを用いて推定し、短期 (2022-28 年) 人口の予測値と 95%・68% 予測区間を求めることである。分析の結果、定数項を含まない ARIMA(3,1,1) が良好に推定され、将来人口を表 3、表 4 のように予測することができた。中国において COVID-19 の影響は急激かつ深刻であり、早ければ 2022 年から人口減少に転じる可能性は否定できない。

キーワード：中国の総人口 時系列モデル 人口予測 長期統計  
人口減少

---

\* 東京女子大学現代教養学部国際社会学科経済学専攻教授

## 1. はじめに

中国の総人口は、2021年末現在14億1260万人<sup>1</sup>であり、世界一の人口大国である。中国の総人口は、中華人民共和国の成立の1949年には、すでに5億4167万人<sup>2</sup>に達していた。図1は、1949年以降の総人口の推移を示している。厳しい自然災害と政治的な混乱から総人口が減少した1960、61年を除くと一貫して増加してきたが、近年はその増加傾向が非常に緩やかになってきている。

図2は、1949年以降の出生率、死亡率、人口の自然増加率(出生率-死亡率)の推移を表したものであり、図1以上に中国の人口動態の変化を知ることができる。日本の中国人口研究の第一人者であった若林敬子(1989)は、建国後の人口動態を4段階に区分した。第1段階(1949-57年)が第1次人口増加期、第2段階(1958-61年)が先にも触れた自然災害期、第3段階(1962-71年)が第2次人口増加期、そして第4段階(1971年以降)が79年の一人っ子政策に象徴される人口抑制政策が本格的に導入された時期にあたる<sup>3</sup>。

一人っ子政策の導入によって、人口の急増に歯止めがかかり、出生率と自然増加率は緩やかに低下していった。一方、出生率の低下と同時に人口の高齢化が進展し、労働力人口の減少が厳しくなってきた<sup>4</sup>。そのため、2015年にはついに一人っ子政策は廃止され、16年からは夫婦1組で2人の子供を持つことが許され(「二人っ子政策」)、更にその政策効果が乏しいため21年には3人の子供をもつこと(「三人っ子政策」)が容認された。2021年の出生数は1062万人、出生率は7.52%と、いずれも建国以来最小となった<sup>5</sup>。人口の自然増加率も、5.58%(2017年)、3.78%(18年)、3.32%(19年)、1.45%(20年)、0.34%(21年)と急速に低下を続けている<sup>6</sup>。このトレンドだけを眺めると、22年から中国の総人口が減少に転じていても不思議ではない。これから数年の中国の人口動態には大きな関心が集まっている。

本稿の目的は、こうした現状を踏まえた上で、短期(2022-28年)の中国の総人口を、伝統的なコーホート要因法ではなく、時系列モデル(ARIMAモデル)を用いて推計することである。このモデルでは、中国の社会経済要因はモデルに組み入れないが、49年の建国以来の総人口データをすべて使用することで、過去の人口の動きによってこれから7年間(2022-28年)の総人口の推計を試みた

1 2022年1月17日、国家統計局発表。『人民日報』2022年1月18日。

2 国家統計局編『中国統計年鑑2021』31頁。

3 若林敬子(1989)9-12頁。

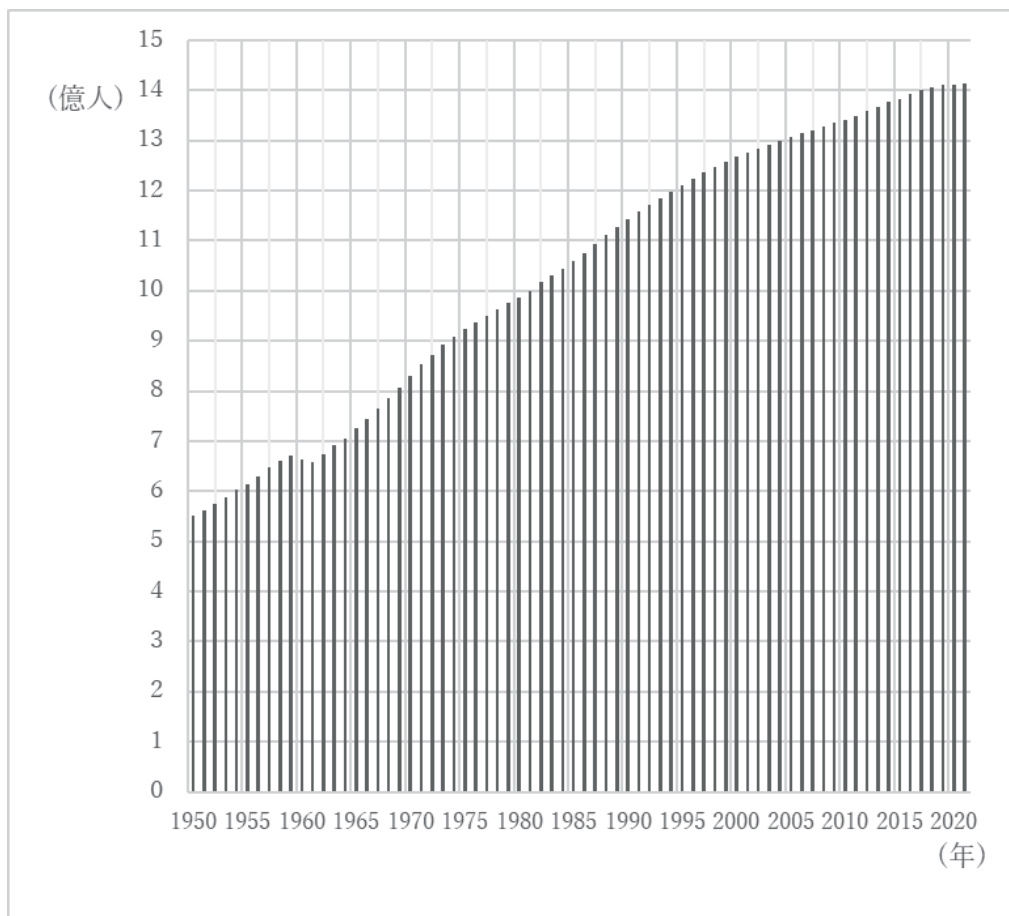
4 中国の高齢化率(総人口に占める65歳以上人口の割合)は、2010年には8.9%であったが、2020年現在13.5%であり、急上昇している。一方、生産年齢(15-64歳)人口の割合は、2010年の74.5%をピークに下降しており、2020年現在68.6%である。『中国統計年鑑2021』33頁。

5 (注1)に同じ。

6 『中国統計年鑑2021』32頁。『人民日報』2022年1月18日。

い。本稿では総人口の予測値を推計すると共に、95%と68% (2標準偏差と1標準偏差の範囲) の予測区間も併せて算出し、人口減少への転換の可能性から予測が非常にむずかしい期間を分析対象とする。なお、予測期間を7年としたのは、ARIMAモデルの高い予測精度が全データの10%程度の「未来」(短期)であること<sup>7</sup>、そして繰り返すが中国の人口減少の可能性を含む期間であるからである。

図1 総人口の推移(1949-2021年)



(注) 台湾、マカオ、香港を除く大陸人口  
 (出所) 国家統計局編『中国統計年鑑 1987』89頁。  
 同 『中国統計年鑑 2021』31頁。  
 『人民日報』2022年1月18日。

7 予測期間が長期になると、制度や政策が大きく変化する可能性があり、精度の高い予測を行うことが難しくなる。

表1 時系列モデルの推定に用いるデータ(中国総人口)

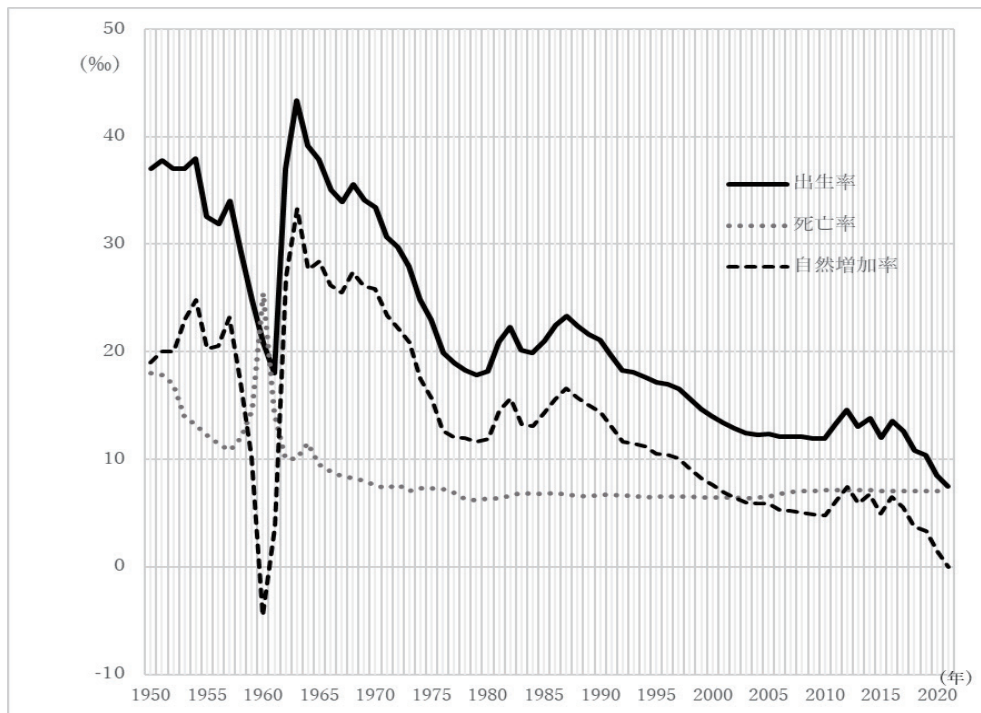
(単位:万人)

年	人口	年	人口	年	人口
1949	54167	1975	92420	2001	127627
1950	55196	1976	93717	2002	128453
1951	56300	1977	94974	2003	129227
1952	57482	1978	96259	2004	129988
1953	58796	1979	97542	2005	130756
1954	60266	1980	98705	2006	131448
1955	61465	1981	100072	2007	132129
1956	62828	1982	101654	2008	132802
1957	64653	1983	103008	2009	133450
1958	65994	1984	104357	2010	134091
1959	67207	1985	105851	2011	134916
1960	66207	1986	107507	2012	135922
1961	65859	1987	109300	2013	136726
1962	67295	1988	111026	2014	137646
1963	69172	1989	112704	2015	138326
1964	70499	1990	114333	2016	139232
1965	72538	1991	115823	2017	140011
1966	74542	1992	117171	2018	140541
1967	76368	1993	118517	2019	141008
1968	78534	1994	119850	2020	141212
1969	80671	1995	121121	2021	141260
1970	82992	1996	122389		
1971	85229	1997	123626		
1972	87177	1998	124761		
1973	89211	1999	125786		
1974	90859	2000	126743		

(注) 図1に同じ。

(出所) 図1に同じ。

図2 出生率・死亡率・自然増加率の推移



(注) 台湾、マカオ、香港を除く大陸人口に関する変化率。

(出所) 国家統計局編『中国統計年鑑 1987』90頁。

同 『中国統計年鑑 2021』32頁。

『人民日報』2022年1月18日。

## 2. 時系列モデルの定式化

本稿では、Box-Jenkins(1976)によって提唱された自己回帰和分移動平均モデル (autoregressive integrated moving average model: ARIMA モデル) を用いて、中国人口の時系列モデルを構築し、その推定結果に基づいて将来推計を行う。ARIMA モデルの特徴は、連立方程式体系のモデルと異なり、変数自身をその変数のラグと誤差項のラグによってのみ説明する点にある。非常にシンプルな統計学的思想に立脚した方法である。現在中国で進みつつある少子化や人口増加率の急激な低下は、様々な要因が時系列にわたって複雑にからみ合った結果生じている<sup>8</sup>。筆

8 少子化の要因としては、以下の諸点があげられる。①養育費・教育費の高騰、②一人っ子政策が長く続いたため、夫婦2人で4人の親の老後の面倒をみる必要がある(したがって、年金や介護サービスなどの社会保障の充実が現在求められている)、③不動産価格の高騰と男性が結婚時に住宅を用意する慣習、④女性の社会進出と機会費用の上昇、⑤経済成長による価値観の多様化、⑥ COVID-19の影響で2020、21年に関しては計画的に結婚・出産を延期するケースが見られた。

者は、連立方程式モデルやコーホート要因法を否定するものではなく、古典的ではあるが短期予測において高い評価のある ARIMA モデルを採用して、シンプルに今後7年間の人口予測を試みたい。

さて、図1でみたように、中国の総人口には1949年から2021年までの73期にわたって増加トレンドがある。ここで、 $t$ 期の中国の総人口を  $X_t$  とすると、 $X_t$  自体は明らかに定常過程 (stationary process) ではない。本稿ではトレンドのない系列にデータを変換するため、下式のように1次の階差をとり分析を進める<sup>9</sup>。

$$\Delta X_t = X_t - X_{t-1} \quad (1)$$

一般に ARIMA モデルのタイプは、以下のように表記する。

ARIMA( $p, d, q$ )

$p$ : 自己回帰 (AR) 部分の次数

$d$ : 階差の次数

$q$ : 移動平均 (MA) 部分の次数

今、階差の次数が1であるから、本稿の4節で推定する ARIMA( $p, 1, q$ ) を定式化すると以下ようになる。

$$\Delta X_t = \sum_{i=1}^p \phi_i \times \Delta X_{t-i} + \sum_{j=1}^q \theta_j \times u_{t-j} \quad (2)$$

$\Delta X_t$ :  $X_t$  の1次の階差

$u$ : 誤差項

$t$ : 1, 2, 3, …, 73

$\phi_i$ : 推定するパラメータ

$\theta_j$ : 推定するパラメータ

### 3. データ

ARIMA モデルの推定に用いる中国総人口のデータは、建国以来の1949-2021年の年末データ(12月末)であり、中国国家统计局編『中国統計年鑑1987』89頁、『中国統計年鑑2021』31頁に公表されている数値による。最新の2021年末の総人口のデータは、2022年1月17日に国家统计局によって発表されたデータを用

9 本分析では、階差の次数が2及び3のモデルや、データを自然対数変換し階差をとるモデルも推定したが、良好な結果はえられなかった。

いる(『人民日報』2022年1月18日)。

『中国統計年鑑2021』30頁の人口調査の「簡要説明」によると、総人口(年末)のデータは、1981年までは戸籍統計に基づき、1982、1990、2000、2010、2020年は人口センサス(全数調査)を基準とした年末の推計値であり、それ以外の年は毎年国家が実施する人口変動サンプル調査とセンサスのデータから修正を加えた数値である。<sup>10</sup>

#### 4. 推定結果

本節では、2節で定式化したARIMA(p,1,q)、(2)式を推定する。このモデルのパラメータを推定するためには、非線形式の解を数値的に近似推計する必要がある。本稿ではLevenberg-Marquardt法による非線形最小2乗アルゴリズムを用いて、残差が最小になるように反復計算しパラメータを推定する。推定するARIMAモデルのタイプは、表2の1列目に示したARIMA(1,1,0)からARIMA(5,1,2)までの15通りであり、いずれも定数項を含まないタイプである。<sup>11</sup> 5節の人口予測に使用するモデルの選択には、(3)式の赤池情報量基準(Akaike information criterion:AIC)を中心に、推定したパラメータのt値(とりわけ最長のラグ数のt値)を判断の基準とする。<sup>12</sup>

$$\text{赤池情報量基準(AIC)} = -2 \times (\text{対数最大尤度}) + 2 \times (\text{推定したパラメータ数}) \quad (3)$$

表2は、推定したARIMAモデルのAICと順位を整理したものである。AICは最適なモデルの選択に用いる統計量であり、小さいほど良いモデルであると判断する。表2の3列目はAICの小さい順位を示しており、1位はARIMA(2,1,2)、2位はARIMA(1,1,2)、3位はARIMA(3,1,1)となった。本稿では、推定したパラメータのt値も考慮に入れ、予測に使用する最終モデルとしてARIMA(3,1,1)を選択する

10 中国では建国以来7回の人口センサスが実施されている。第1回人口センサス(1953年7月1日)、第2回(1964年7月1日)、第3回(1982年7月1日)、第4回(1990年7月1日)、第5回(2000年11月1日)、第6回(2010年11月1日)、第7回(2020年11月1日)、である。なお、中国の人口統計を利用するうえで留意すべき点については、若林敬子(2005)「第六章 人口動態の推移と人口統計」(153-184頁)を参照。

11 本分析のデータ入力・推定・検定・予測は、筆者自身がTSP(Version 5.1)でプログラムを組み実行したため、出力結果にミスがあった場合には、筆者の責任である。繰り返し計算(イタレーション)の回数の上限は100、収束判定条件は0.00001でプログラミングした。Levenberg-Marquardt法については、Marquardt(1963) pp.431-441参照。

12 定数項を含むタイプを15通り推定したが、多くで定数項のパラメータのt値が有意ではなかった。また、原系列の差分をとったARIMAモデルでは、定数項を含まないタイプの選択が一般的である。

こととした。以下が推定結果であり、( )内の数値はt値、[ ]内の数値はp値である。<sup>13</sup>

Levenberg-Marquardt 法による非線形最小 2 乗推定 (イタレーション回数 = 21 回)

$$\begin{aligned} \Delta X_t = & 1.7357 \Delta X_{t-1} - 0.95859 \Delta X_{t-2} + 0.21999 \Delta X_{t-3} + 0.85492u_{t-1} \quad (4) \\ & (11.110) \quad (-4.368) \quad (1.723) \quad (7.401) \\ & [0.000] \quad [0.000] \quad [0.085] \quad [0.000] \end{aligned}$$

決定係数 = 0.5915

自由度修正済み決定係数 = 0.5735

回帰の標準誤差 = 380.78

AIC=1063.892

推定されたパラメータは、 $\Delta X_{t-3}$  のパラメータ (10%水準で有意) を除くと、1%水準で有意である。誤差項の不均一分散の LM 検定は 0.431 [P 値 0.512] であり、誤差項の不均一分散の問題はないといえる。次に誤差項の系列相関の有無について、このモデルが定数項を含まないタイプであるから Durbin-Watson test が行なえないので、誤差項の自己相関係数を計算すると 0.026 ときわめて低く、1 階の系列相関は認められない。<sup>14</sup>

本稿では、元のデータが非定常であり定常化するために 1 次の階差をとったので、単位根検定を行う必要があり、拡張 Dickey-Fuller 検定 (augmented Dickey-Fuller (ADF) test) を実施しておく。<sup>15</sup> 本稿では定数項のない ARIMA モデルを推定した

13 定数項を含まない ARIMA(3,1,1) を最尤法で推定した結果も示しておく。最尤法 (イタレーション回数 = 10 回)

$$\begin{aligned} \Delta X_t = & 1.6813X_{t-1} - 0.92653 \Delta X_{t-2} + 0.24097 \Delta X_{t-3} + 0.77781u_{t-1} \\ & (4.933) \quad (-3.121) \quad (1.620) \quad (2.110) \\ & [0.000] \quad [0.002] \quad [0.105] \quad [0.035] \end{aligned}$$

決定係数 = 0.5887

自由度修正済み決定係数 = 0.5705

回帰の標準誤差 = 383.53

AIC = 1067.758

14 ただし、誤差項の正規性の検定 (Jarque - Bera test) には問題が残った。予測 (5 節) に関しては、この点を十分留意する必要がある。一人っ子政策導入後の 1979-2021 年に絞ったデータで ARIMA(3,1,1) を推定したところ (イタレーション回数 = 88 回)、この正規性の問題は解決できたが、本稿では建国以来のすべてのデータ (長期統計) を利用する推定と予測に拘った。参考までに 1979-2021 年の推定結果に基づく人口予測値は、以下の通りである。141295.1(2022 年)、141332.6(23 年)、141366.4(24 年)、141402.2(25 年)、141435.0(26 年)、141469.2(27 年)、141501.0(28 年)。2021-28 年の増加率 (年率) は 0.24%。

15 Dickey-Fuller(1999)pp.427-431 参照。



ので、ADF 検定においても定数項のない ADF 回帰式を OLS で推定する。( ) 内の数値は、t 値である。

$$\Delta X_t = 0.000567X_{t-1} + 0.96125 \Delta X_{t-1} - 0.30421 \Delta X_{t-2} + 0.26434 \Delta X_{t-3} \quad (5)$$

(0.645)            (7.986)            (-1.844)            (2.134)

決定係数 = 0.5898

自由度修正済み決定係数 = 0.5708

回帰の標準誤差 = 393.09

ADF 検定では、推定した (5) 式の  $X_{t-1}$  のパラメータについて、「帰無仮説： $X_{t-1}$  のパラメータはゼロ」(単位根はある) という検定を行うことになる。ADF 統計量は 0.645 であり、有意水準 10% で帰無仮説は棄却されない。したがって、1 次の階差をとったことに問題はないといえる。<sup>16</sup>

表2 推定した ARIMA モデルの赤池情報量基準 (AIC)

モデル	赤池情報量基準 (AIC)	順位
ARIMA(1.1.0)	1066.246	8
ARIMA(1.1.1)	1068.170	13
ARIMA(1.1.2)	1063.048	2
ARIMA(2.1.0)	1068.214	14
ARIMA(2.1.1)	1068.974	15
ARIMA(2.1.2)	1062.942	1
ARIMA(3.1.0)	1063.918	4
ARIMA(3.1.1)	1063.892	3
ARIMA(3.1.2)	1064.940	6
ARIMA(4.1.0)	1065.312	7
ARIMA(4.1.1)	1064.676	5
ARIMA(4.1.2)	1066.662	10
ARIMA(5.1.0)	1066.924	11
ARIMA(5.1.1)	1066.658	9
ARIMA(5.1.2)	1068.094	12

16 定数項を含む ADF 統計量は -2.049、定数項と時間トレンドを含む ADF 統計量は 0.122 であり、いずれも有意水準 10% で「帰無仮説： $X_{t-1}$  のパラメータはゼロ」(単位根はある) は棄却されない。

## 5. 短期の人口予測

ARIMA(3,1,1)の推定結果である(4)式を用いて、2022年から2028年の短期の総人口予測を行う。表3が総人口の予測値と95%予測区間(2標準偏差の区間)、表4が予測値と68%予測区間(1標準偏差の区間)、そして図3が表3と表4の予測結果をグラフ化したものである<sup>17</sup>。

表3の予測値(万人)より、141486.2(22年)、141877.7(23年)、142351.0(24年)、142847.0(25年)、143340.3(26年)、143825.2(27年)、144303.2(28年)と点推定に関しては、この期間の人口増加が予測される。しかしながら、表3の95%予測区間の下限値および表4の68%予測区間の下限値に着目すると、2022年より人口減少に転じる可能性は否定できない。特にCOVID-19は日本と同様中国においても深刻な問題であり、出生率の低下に及ぼす影響は非常に大きい。一方で本分析のモデルでは、COVID-19の影響は2020、21年のデータに反映されたに過ぎない。したがって、2022年以降中国人口が減少に転じる可能性は十分にありうる。

表3 総人口の予測値と95%予測区間

(単位：万人)

年末	総人口の予測値	95%予測区間	
		下限値	上限値
2022	141486.2	140739.8	142232.6
2023	141877.7	140287.9	143467.5
2024	142351.0	139927.4	144774.6
2025	142847.0	139638.7	146055.2
2026	143340.3	139385.5	147295.0
2027	143825.2	139141.6	148508.8
2028	144303.2	138892.4	149714.1

(注) 推定したARIMA(3,1,1)に基づき、全期間(1949-2021年, n = 73)の約1割を目安として総人口の予測(2022-28年)を行う。

17 95%予測区間は厳密な2標準偏差の確率(95.44997%)ではなく、あくまで95.0%の確率である。同様に68%予測区間も厳密な1標準偏差の確率(68.26895%)ではなく、68.0%の確率である。

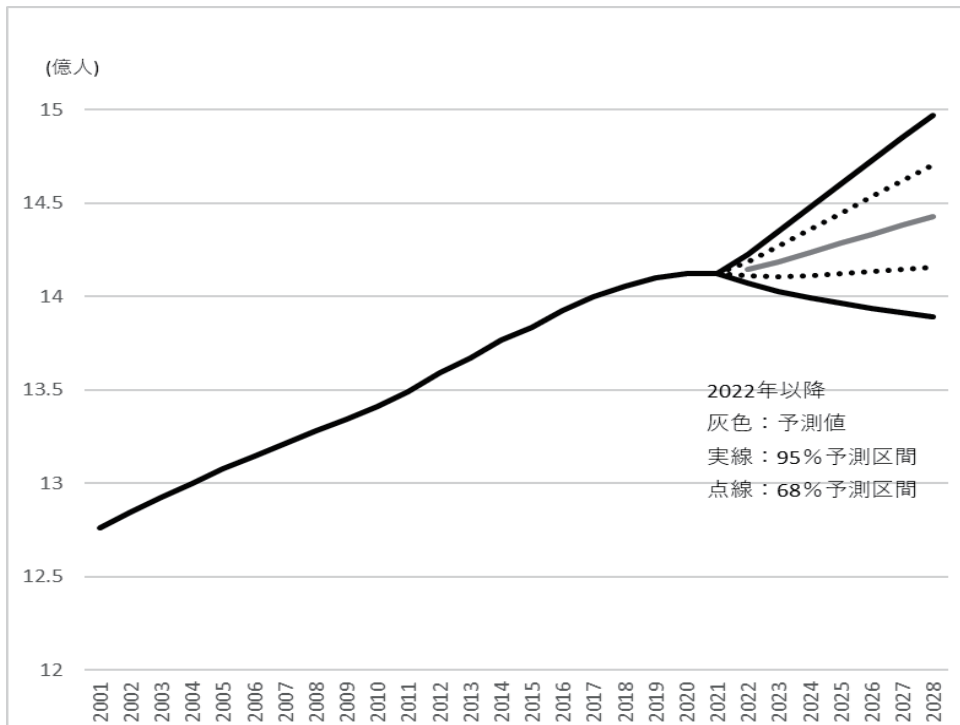
表4 総人口の予測値と68%予測区間

(単位：万人)

年末	総人口の予測値	68% 予測区間	
		下限値	上限値
2022	141486.2	141107.5	141864.9
2023	141877.7	141071.0	142684.4
2024	142351.0	141121.3	143580.7
2025	142847.0	141219.4	144474.8
2026	143340.3	141333.7	145346.9
2027	143825.2	141448.8	146201.6
2028	144303.2	141557.8	147048.6

(注) 表3に同じ。

図3 総人口の予測値と予測区間(2022-28年)



(注1) 表1のデータ、および表3、表4の推計結果より作図。

(注2) 台湾、マカオ、香港の人口は除く。

## 6. おわりに

本稿の目的は、国家統計局が公表している長期人口統計(1949-2021)に基づいて、中国総人口の時系列モデル(ARIMAモデル)を推定し、短期(2022-28年)の人口予測を行うことであった。15通りのARIMAモデルを設定し、Levenberg-Marquardtの非線形最小2乗アルゴリズムを用いて推定した。最終的には定数項を含まないARIMA(3,1,1)が、収束のための21回の繰り返し計算によって良好に推定され、この結果に基づいて中国総人口の予測を行った。

表3が総人口の予測値と95%予測区間、表4が同じく予測値と68%予測区間、そして図3が2つの表をグラフ化したものである。点推定においては、この期間(2022-28年)の人口増加率(年率)は3.05%と予測される。しかしながら、表3の95%予測区間の下限値および表4の68%予測区間の下限値を見ると、2022年より人口減少に転じている。とりわけCOVID-19は中国において非常に深刻な問題であり、2020年以降出生数の低下に及ぼす影響はきわめて大きい。一方で本分析で推定したモデルではCOVID-19の影響は2020、21年のデータに反映されただけである。したがって、早ければ2022年以降、中国の人口が減少に転じる可能性は十分にありうると考えられる。

参考文献

Box, G.E.P. and G.M. Jenkins (1976) *Time Series Analysis: Forecasting and control*, Revised edition, Holden-Day.

Dickey, D. A. and W. A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74 (366), 427-431.

Hamilton, J. D. (1994) *Time Series Analysis*, Princeton University Press.

Marquardt, D.W (1963) "An Algorithm for Least-squares Estimation of Nonlinear Parameters," *Journal of the Society for Industrial and Applied Mathematics*, 11 (2), 431-441.

沖本竜義 (2010) 『経済・ファイナンスデータの計量時系列分析』朝倉書店。

増田幹人 (2007) 「ARIMA モデルに基づく出生率の将来推計」『人口学研究』41,75-85。

縄田和満 (2006) 『TSP による計量経済分析入門 第2版』朝倉書店。

若林敬子 (1989) 『中国の人口問題』東京大学出版会。

——— (2005) 『中国の人口問題と社会的現実』ミネルヴァ書房。

(2022年3月26日 提出)

## Estimating time-series models of the Chinese population using long-term statistical data and forecasting the short-term future population

SHIRASAGO, Tetsuya

The total population of China at the end of 2021 was 1412.6 million. The total population has only increased by 480,000. The birth rate was 7.52 ‰, the lowest since the country was founded. With China's declining birthrate and aging population becoming more serious, it is no wonder that the population begins to decline at any time. The purpose of this treatise is to estimate a time-series model (ARIMA model) of the Chinese population using long-term statistical data (1949-2021) and to predict the short-term (2022-2028) population. Finally, ARIMA (3,1,1) without a constant term could be well estimated by the nonlinear least squares method. As a result of the analysis, the predicted values of the population and the 95% / 68% forecast intervals in China are estimated as shown in Tables 3 and 4. Since the impact of COVID-19 in China is rapid and serious, it is undeniable that the population may decline from 2022 at the earliest.

Keyword : total population of China, time-series model, long-term statistical data, population forecast, population decline