The Statistical Analysis of the Simple Interrupted Time-Series Quasi-Experiment

INTRODUCTION

(略)[1]

THE PROBLEM WITH ORDINARY LEAST SQUARE REGRESSION

時系列データの分析で OLS (Ordinary Least Square) をやってはいけない。OLS は誤差の独立性を前提するが、時系列データはたいてい自己相関しているから。

自己相関していても、パラメータ推定には問題はない＝回帰式にはバイアスはかからない。バイアスがかかるのは標準偏差の推定。よってパラメータの検定に問題が生じる。[2]

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODELS

interrupted time-series を分析するには ARIMA モデル（Box-Jenkins 伝達関数アプローチ 1976）を使おう！ ←この本はちと古いけどまあ基本。

	誤差が独立でない	誤差が独立
オブザベーションが 50～100 以上	ARIMA	OLS
オブザベーションが 50～100 に満たない	MANOVA or GG 修正した反復測定 ANOVA	反復測定 ANOVA (たぶん古典的乱塊法分析のこと)

The Deterministic and Stochastic Components of a Time Series

「すべての統計手続きはモデルを必要とする」余談 p.235

時系列モデルは２つのコンポーネント（変動を生む要素、成分）を持つ。

- 決定論的コンポーネント
- 確率論的コンポーネント　- ノイズコンポーネント
  - システマティック部分
  - 非システマティック部分

Defining ARIMA (p, d, q)

ARIMA モデルはノイズプロセスを記述するモデル。p, d, q の３つの整数パラメータを持つ。これらの値を特定することを同定 identification と呼び、これが最初のステップ。

[1] ここでは単一系列からの介入あり時系列データの分析だけを扱う。

[2] このインフレを甘く見るな。「t 値が十分大きいから多少インフレが含まれているとしても問題ないはず」という慣習的考えは危険な誤り。例えば ARIMA(1,0,0)で ϕ=.7（これくらいはよくある）だとしたら 256%インフレし、t 値が 6.0 くらいでも有意でないこともある。逆に、ϕが負だとデフレする。(p.235 注)

Stationarity

ARIMA モデルは、定常な時系列か、nonstationary in the homogeneous sense な時系列に対してしか適切でない。

定常な時系列 ＝ 長期的トレンドを持たない時系列

社会科学における多くの時系列は長期的トレンドを持つ＝非定常。しかしそのほとんどは階差をとることで定常になる。そのような非定常を nonstationary in the homogeneous sense と言う。

1, 2, 3, 4, 5, … , N (非定常) →階差をとると→ 1, 1, 1, 1, … , 1 (定常)

階差をとったりしたら、調べたい介入効果の有意性テストに影響が出るのでは？→No.

$$Y_t = Y_{t-1} + \theta_0$$
$$Y_t = \theta_0 t$$

書き方を変えただけで、トレンドを取り除いてはいない。

１回で定常にならなければ、定常になるまで階差をとり続ける。しかし、２回以上階差をとらなければならないような時系列はレア。

パラメータ d は定常にするために階差をとった回数。

Autoregressive Models

パラメータ p は自己回帰の次数

時系列の隣接するオブザベーション間に直接の関係がある場合、p が 0 より大きくなる。

p =1 の場合（ARIMA(1,0,0)モデル）

$$Y_t = \phi_1 Y_{t-1} + a_t$$

p =2 の場合（ARIMA(2,0,0)モデル）

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + a_t$$

φは相関係数みたいなもの。$-1 < \phi_t < +1$ [3]

a_t は誤差項。$a_t \sim NID(0, \sigma^2)$ ホワイトノイズ、あるいはランダムショックとも呼ばれる。＝非システマティック部分。

Moving Average Models

パラメータ q は移動平均の次数

---

[3] でもそう単純でもない。$p = 2$ のときは、$\begin{array}{l} \phi_2 + \phi_1 < 1 \\ \phi_2 - \phi_1 < 1 \\ |\phi_2| < 1 \end{array}$

時系列によってはオブザベーションからオブザベーションへのランダムショックの持続性に違いがある。

ARIMA(0,0,1)モデル

$$Y_t = a_t - \theta_1 a_{t-1}$$

ARIMA(0,0,2)モデル

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

θは相関係数みたいなもの。$-1 < \theta_t < +1$ ・・・φと同様

Mixed Models

ARIMA(1,0,1)モデル

$$Y_t = \phi_1 Y_{t-1} - \theta_1 a_{t-1} + a_t$$

mixed はレア。普通は AR か MA のどっちか。

Noise Model Identification

システマティック部分を同定するために、自己相関関数 ACF と偏自己相関関数 PACF を使う。

「同定されるモデルはその時系列に関して最も倹約的で適切なモデルであり、必ずしも”真の”モデルとは限らない」余談 p.241

ACF はラグとの相関

Lag 0	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	.		
Lag 1	.	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8		
Lag 2	.	.	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	
Lag 3	.	.	.	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8

Lag k の ACF, r_k は

$$r_k = \frac{\sum_{t=1}^{N-k}(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{N}(Y_t - \bar{Y})^2} \quad \text{for } k = 1, 2, 3, .. \qquad (1)$$

上からわかるように、ラグが大きくなるほど ACF の推定は信頼度が低くなるが、時系列の同定には主として初めのいくつかラグの ACF を使うので大丈夫。

PACF, r_{kk}は ACF ほどシンプルに示すことはできず、Yule-Walker 方程式を解いて推定

される。が、値はコンピュータプログラムが計算してくれるからアルゴリズムにとくに関心はなく、我々はその値が意味するところをわかっていればよい。

ACF と PACF による同定というのは、要するに、現実のデータの ACF、PACF と各モデルから想定される ACF、PACF を比較してどれくらい似ているかを見るということ。
全体的特徴をつかみやすいようにACF や PACF をプロットしたものをコレログラムと言う。正しい ARIMA(p,d,q)モデルはプロットだけでわかることもあるが、視覚プロットはたいてい多義的なので、相関がゼロかどうかの検定も利用する。

各プロセスの特徴（Figure 6.1(a～h）も参照)

	Autocorrelations	Partial autocorrelations
Nonstationary Processes	ゆっくりと減衰。持続的。	lag 1 での値が 1 に近い
White Noise Process	すべてゼロ	すべてゼロ
Autoregressive Processes	lag 1 から指数関数的に消えていく	最初の p 個のラグは非ゼロ、p より大きいラグはゼロ
Moving Average Processes	最初の q 個のラグは非ゼロ、q より大きいラグはゼロ	lag 1 から指数関数的に消えていく
Mixed Processes	q - p より大きなラグでは指数関数的に消えていく	p - q より大きなラグでは指数関数的に消えていく

すべてのラグの ACF の有意性を同時にテストするには（ノイズかどうかをテストするには）Q統計量を使う。[4] $Q = N\sum_{j=1}^{k} r_j^2$

同定の手順

1. 非定常かどうかを調べて、そうなら非定常でなくなるまで階差をとる。dが決定。
2. ACF、PACF を見て、どっちが指数関数的に減衰しているかで自己回帰モデル、移動平均モデルを見分ける。
3. 自己回帰か移動平均かがわかったら、もう一方のスパイク（有意なラグ）の数で p または q の値を決める。必ずいつも、あり得るうちの最も小さい p と q からスタートすること。[5]
4. 両方とも指数関数的に減衰していたら、混合モデルを考える。

---

[4] これは本質的に χ^2 による適合度検定。臨界値は χ^2 の表から 自由度 = ラグの数 - パラメータの数で引く。

[5] もし underestimate だったら診断段階でエラーが出るが、overestimate だったとしても診断段階でのエラーは出ない。

Estimation of φ and θ Values

ARIMA(p,d,q)モデルを同定したら、そのパラメータ（φとθ）を推定する。たいていのプログラムは信頼区間も一緒に計算してくれる。

ここでときどき不適切な結果を得る。

その１つは、φあるいはθの推定値が定常性や可逆性の境界（-1 << +1）から外れているとき。多くの場合その原因は階差をとりすぎているか、とれていないか。

もう１つは、φやθが統計的に有意でないとき。それらのパラメータは必要ないのでモデルから落とす。

Diagnosis

システマティックな部分を記述して相関しない誤差だけ残すことが目的だったのだから、推定が終わったらモデルの残差をとって、残差がホワイトノイズであるかどうかを確かめる。もしそうだったら、誤差モデリングは完了。そうでなかったら、同定からやり直し。

Q 統計量は、ACF プロット中のスパイクの位置には敏感でない。だが、このスパイクの位置は重要で、たとえ Q 統計量が非有意だったとしても、スパイクの位置によってはそのモデルは却下される。

例えば残差の ACF の lag 1 と lag 2 にスパイクがあれば、モデル化されていない移動平均コンポーネントが存在する可能性を示しているし、lag 12 にスパイクがあってデータが月々のオブザベーションなら季節性コンポーネントがあることを示唆している。

ARIMA SEASONAL MODELS

Multiplicative ARIMA Seasonal Models

例. Figure 6.2

季節性を扱うにはいくつか方法があるが、より良いのは ARIMA(p,d,q)に季節性構造を組み込むこと。

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}
Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}	Y_{21}	Y_{22}	Y_{23}	Y_{24}
Y_{25}	Y_{26}	Y_{27}	Y_{28}	Y_{29}	Y_{30}	Y_{31}	Y_{32}	Y_{33}	Y_{34}	Y_{35}	Y_{36}
…											
$Y_{N\text{-}11}$	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	Y_N

月毎の季節性コンポーネントならたとえば

$$Y_t = \phi_{12} Y_{t-12} + a_t$$

自己相関には通常の自己相関 regular autocorrelation と seasonal autocorrelation がある。前者はこれまでに述べたもの、後者は周期で区切ったオブザベーション間の従属性を表す。
seasonal autocorrelation のパラメータは大文字で表すことにすると、

$$\text{ARIMA }(p, d, q)\ (P, D, Q)\ s$$

P は季節性の自己回帰
D は季節性の階差
Q は季節性の移動平均
S は周期の長さ

regular 構造と seasonal 構造は掛け算される。
理解のために ARIMA $(1,0,0)(1,0,0)_{12}$ を考えてみる。単純に足し算だと、

$$Y_t = \phi_1 Y_{t-1} + \phi_{12} Y_{t-12} + a_t$$

しかし実際は違う。

$$Y_t = \phi_1 Y_{t-1} + \phi_{12} Y_{t-12} - \phi_1 \phi_{12} Y_{t-13} + a_t$$

掛け算の項で調整される。
ただ、$\phi_1 \phi_{12}$ はいつも小さくなるし、無視できるほどになる（＝足し算だと見なせる）こともある。

Identifying the Seasonal Model

季節 ARIMA モデルは時系列に季節性があるという証拠がないなら使うべきではない。同定段階や診断段階での ACF、PACF が季節的ラグ（月々のデータの 12,24 … 、四半期データの 4,8 … など）で有意な相関を示したなら、試してみる。

一般的ルールとして、seasonal コンポーネントは regular コンポーネントと同じタイプである＝自己回帰なら自己回帰、移動平均なら移動平均。また、seasonal コンポーネントの次数が 1 より大きくなることは滅多にない。最もよくある季節 ARIMA モデルは

ARIMA $(1,0,0)(1,0,0)_s$
ARIMA $(2,0,0)(1,0,0)_s$
ARIMA $(0,0,1)(0,0,1)_s$
ARIMA $(0,1,1)(0,0,1)_s$
ARIMA $(0,1,1)(0,1,1)_s$

各プロセスの特徴

	Autocorrelations	Partial autocorrelations
Seasonal Nonstationary	季節ラグの ACF はゆっくり減衰する。regular コンポーネントが非定常かどうかは関係ない。	
Seasonal Autoregressive	季節ラグの相関は指数関数的に減衰し、各季節ラグから始まるパターンも指数関数的に減衰する。もし $p = 0$ なら最初のラグの減衰パターンはない。	$p+SP$ までにスパイク、それより後ではゼロ。もし $p = 0$ なら最初のラグのスパイクはない。
Seasonal Moving Average	$Q = 1$ なら lag 1〜q と $S-q$ 〜$S+q$ にスパイクを示す。Q が 2 以上なら周期でパターンが繰り返される。例えば q =1, Q = 2, S =12 なら 1, 11, 12, 13, 23, 24, 25 にスパイク。	季節ラグの相関が指数関数的に減衰する。

同定の手順

1. 季節モデルが必要であるという確信を得る。証拠は同定段階の ACF、PACF か、診断段階の ACF、PACF。季節コンポーネントはいつも後から足されるべき。
2. 季節コンポーネントが非定常かどうかを調べる。もしそうなら階差をとる。非定常の証拠としては、例えば月毎データなら lag 12 が ACF、PACF ともにかなり大きく、lag 24 の ACF も有意でなければならない。
3. 季節コンポーネントは regular コンポーネントとふつうは同じタイプ。自己回帰なら PACF が季節ラグでスパイクするし、移動平均なら ACF が季節ラグでスパイクする。
4. いつも低次の季節コンポーネントから試す。

THE INTERVENTION COMPONENT

ARIMA モデルを時系列にフィットさせる＝ノイズを記述するということ。すなわち、

$$Y_t = noise$$

ノイズモデルは中断時系列準実験において介入がない場合を予測する。次はこれに介入コンポーネントを加える。

$$Y_t = noise + intervention$$

結局、調べたいのは、介入コンポーネントを足すと時系列の振舞いの予測が有意に増すかどうか、である。

「統計的解析は、それ自体で”原因”をテストしてはいない。それは、統計的に有意な変化が、系列の特定のポイントで発生しているかどうかを尋ねるだけである。変化についてのどんな説明も評価されない。」余談 p.262

Box-Jenkins において介入コンポーネントは伝達関数 transfer functions[6]と呼ばれる。ここでは３つのシンプルな伝達関数を紹介する。

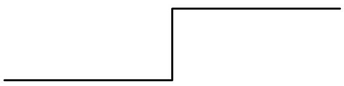
Abrupt, Constant Change

$$Y_t = \omega I_t + noise \qquad (2)$$

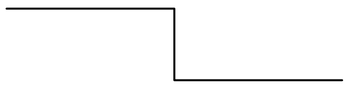
$I_t = 0$　*before the intervention,* $t < i$

$= 1$　*after the intervention,* $t \geqq i$

ωは変化のマグニチュード



pretreatment : posttreatment



pretreatment : posttreatment

Gradual, Constant Change

$$Y_t = \delta Y_{t-1} + \omega I_t + noise \qquad (3)$$

(2)ほど簡単ではない。ノイズを 0、介入前の値を 0 と仮定して振舞いを見てみる。
介入のときの値は、

$$\begin{aligned} Y_i &= \delta Y_{i-1} + \omega I_i \\ &= \delta(0) + \omega(1) \\ &= \omega \end{aligned}$$

次の時点では、

$$\begin{aligned} Y_{i+1} &= \delta Y_i + \omega I_{i+1} \\ &= \delta(\omega) + \omega(1) \\ &= \delta\omega + \omega \end{aligned}$$

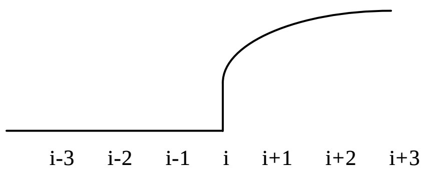
次の時点では、

$$\begin{aligned} Y_{i+2} &= \delta Y_{i+1} + \omega I_{i+2} \\ &= \delta(\delta\omega + \omega) + \omega(1) \\ &= \delta^2\omega + \delta\omega + \omega \end{aligned}$$

介入から n 番目の時点では、

$$Y_{i+n} = \delta^n\omega + \delta^{n-1}\omega + \ldots + \delta^2\omega + \delta\omega + \omega$$

パラメータδは 1 より小さいという制限があるので、パターンは次のようになる。



[6] 工学の用語に由来。

$$change\ in\ level = \frac{\omega}{1\text{-}\delta}$$

この種のパターンはよくある。Box & Tiao(1975) 大気汚染データ、Hibbs(1977)非雇用率データなど。

Abrupt, Temporary Change

(3)の I_t をステップ関数でなくパルス関数として定義

$I_t = 0$　*before the intervention, $t < i$*

$= 1$　*at the moment of intervention, $t = i$*

$= 0$　there*after, $t > i$*

前節と同様に振舞いを見ると、介入のときの値は、

$$\begin{aligned} Y_i &= \delta Y_{i-1} + \omega I_i \\ &= \delta(0) + \omega(1) \\ &= \omega \end{aligned}$$

次の時点では、

$$\begin{aligned} Y_{i+1} &= \delta Y_i + \omega I_{i+1} \\ &= \delta(\omega) + \omega(0) \\ &= \delta\omega \end{aligned}$$

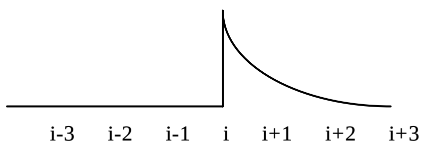
次の時点では、

$$\begin{aligned} Y_{i+2} &= \delta Y_{i+1} + \omega I_{i+2} \\ &= \delta(\delta\omega) + \omega(0) \\ &= \delta^2\omega \end{aligned}$$

介入から n 番目の時点では、

$$Y_{i+n} = \delta^n \omega$$

パラメータδは 1 より小さいという制限があるので、パターンは次のようになる。



Which Transfer Function Should Be Used?

効果についてアプリオリな考えがあるなら、それに合った伝達関数を使う。

効果がもっと複雑な様相を示すと考えるなら、もっと手の込んだ伝達関数を使えばよい。

単に（性質はわからないが）何らかの効果があると考えているだけなら、

1.　まず、I_t をパルス関数にした(3)を試す。それでδが小さければ、分析終了。効果は急

激で一時的だということ。

2. δが大きければ（例えば.9 以上）持続的効果だという証拠なので、I_tをステップ関数にして試す。そこでδが大きければ、分析終了。徐々に上がる効果。
3. δが小さければ（例えば.1 以下）持続的で急激だという証拠なので、さらに(2)を試す。

SUMMARY OF THE MODELING STRATEGY

（略）

Intervention Hypothesis Test

伝達関数を ARIMA モデルと結合させたら、そこで再び適合度を確かめるべきである。[7] そして、最後に、介入パラメータの有意性検定から、介入効果についての帰無仮説の棄却/保留を決定する。

THREE EXAMPLES

（要点のみ）

電話番号案内への電話に課金を導入したことの効果

介入効果が大きいため、介入前のデータのみで ARIMA モデルの同定を行った。結果として採用したモデルは ARIMA(0,1,0)(0,1,1)$_{12}$。もし全オブザベーションでモデル同定を行っていたら、ARIMA(0,1,0)を採用するところだっただろう。介入効果が大きい場合は ACF、PACF の推定値を歪める。

介入効果が大きいため、伝達関数結合後のモデルの診断も行っている。

導入の効果あり。

地域経済に対する洪水の効果

季節的に非定常なので、季節コンポーネントについても階差をとっている。

洪水は有意なインパクトなし。

武装強盗の発生率に対する新しい銃規制法の効果

Deutsch and Alt(1977)は効果ありと結論したが、効果ありとは言えない。

モデルが診断段階で却下された結果、対数変換をしてからモデル同定している。

ARIMA モデルは介入効果を対比する確率論的ベンチマークである。

---

[7] この点については著者間の意見は一致していない。一方は診断が常に必要だと考えるが、もう一方は介入効果が大きいときだけ必要だと考える。

$$t = \frac{transfer\ \ function}{"\,unexplained"\ variance}$$

説明される分散の量を増やすことで t 値は任意に大きくすることができてしまう。よって著者の勧めるモデリング方略を守ることで、バイアスのかかっていない倹約的なモデルに行き着くことが重要。

CONCLUSION

学生が時系列の問題でおびえているのは、実践的スキルを得るには高度な数学的洗練を要すると彼らが思っているからでないだろうか。

他の統計手法にくらべて時系列分析が格段にそうだということはない。例えば、多くの社会科学者は、たとえ基盤となる数学を理解していなくても分散分析を使っている。[8] 実践に必要なのは主に、時間と、忍耐と、コンピュータプログラムへのアクセスである。

理論的基盤をマスターすることは幅広い実践的利益があるだろう。が、同時に、初心者には実践の経験を積むことを勧める。自信も高まるし、理論的側面への動機づけも高まるだろう。

[8] オイオイ--;