時系列解析

証券アナリスト

Takayuki Suzuki

This is institude of the author

時系列解析目標

- トレンドの推定
- 定常性
- 自己回帰モデルで推定する
- 単位根仮定、ランダムウォークを理解
- 構造変化、季節調整
- ボラティリティ変動モデル
- 非定常時系列の見せかけの回帰と共和分
- 応用する

自己相関

自己相関 (Autocorrelation) は、系列相関 (Serial Correlation) ともいう。
相関でなく分散を指す場合、自己分散とか、自己共分散とかいう場合もある。
基本的には同一の概念。
時系列データ

$$Y_t = \{y_1, y_2, \ldots, y_t\}$$

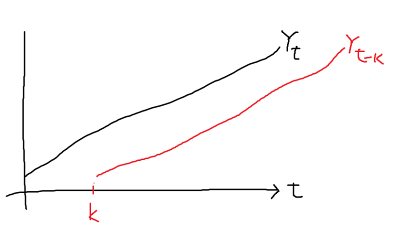
があった時、たとえば、Y と、Y を k 時刻だけずらした系列

$$Y_{t-k} = \{y_{1-k}, y_{2-k}, \ldots, y_{t-k}\}$$

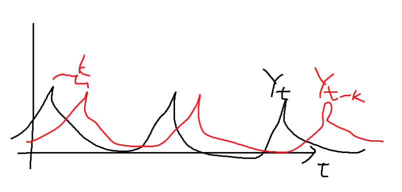
を作り、Y_t と Y_{t-k} の相関 $\mathrm{Corr}(Y_t, Y_{t-k})$ を考える。
このように時刻をずらした自分自身との相関のことを、自己相関という。
自己相関はラグ k の関数であるが、どのようなときに自己相関が大きくなるか？

自己相関

1. トレンドがあり、増加を続ける、または減少を続ける場合。この場合ラグ k の値に関わらず相関が大きくなる



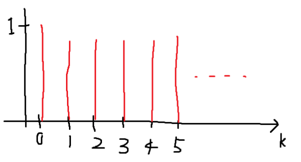
2. 周期性がある場合。この場合ラグ k が周期と一致した場合に相関がおおきくなる



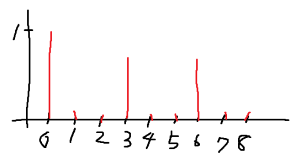
コレログラム

コレログラムとは、自己相関をラグ K の関数とみて、横軸にラグ k、縦軸に相関係数の図を書いたものである。前ページの例では、下記のようになる。

1. ラグ k の値に関わらず相関は大きい



2. ラグ k が周期と一致した場合に相関が大きくなる (下記図は周期が 3 の場合)

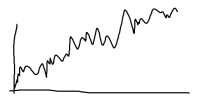


k=0 の時は相関=1 である点に注意。

定常性

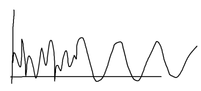
時系列データが下記３つの条件を満たすとき、そのデータは定常性を持つ、あるいは定常過程であるという。図はダメな例。

- 平均 $E(y_t)$ がすべてのｔに対して等しい（トレンドが有る場合は平均がｔによって変わるのでダメ）



- 分散 $V(y_t)$ がすべてのｔに対して等しい（ボラティリティが変わったらダメ）

- 自己分散がラグｋのみに依存し、時刻ｔに依存しない（周期性が変わったらダメ）



トレンドモデル

トレンドモデルは、データの期待値が、時刻の一次関数で変化すると仮定したモデルである。数式にすると、下記のようなモデルである。

$$y_t = \alpha + \beta t + \epsilon_t$$

ここで、t は時刻、α, β は回帰係数, ϵ は誤差項である。
トレンドモデルは、左辺を自然対数にした、対数トレンドモデルも採用されることがある

$$\ln(y_t) = \alpha + \beta t + \epsilon_t$$

推定には、ふつう最小二乗回帰 (OLS, Ordinary Least Squares) で行う。
OLS は二乗誤差を最小化していることだけ理解しておけばよい。
推定に利用する期間を「インサンプル」、推定に利用しない期間を、「アウトサンプル」という。
アウトサンプルではモデルの当てはまりを評価するために利用されることが多いようだ。

トレンドモデルでの予測計算

試験では、推定された予測式が与えられ、それに基づき次期の予測を行う問題が出題されそう。できるようにしておく。（協会テキスト４章の練習問題、問１および問２）

例題 1

あるデータ Y について、期間 $T = 1, 2, \ldots, 100$ を用いて、トレンドモデルに当てはめた結果、下記のような結果を得た

$$y_t = \alpha + \beta t + \epsilon_t$$

$$\text{ただし } \alpha = 1000, \beta = 20.0,$$

このとき、$T = 101$ および $T = 200$ の予測値を求めよ.

計算すると、T=101 のとき、3020, T=200 のとき、5000 になる。
誤差項は予測には不要であることに注意。

トレンドモデルでの予測計算２

例題２

あるデータ Y について、期間 $T = 1, 2, ..., 100$ を用いて、対数トレンドモデルに当てはめた結果、下記のような結果を得た

$$\ln(y_t) = \alpha + \beta t + \epsilon_t$$

$$\text{ただし } \alpha = 5.0, \beta = 0.003,$$

このとき、$T = 101$ および $T = 200$ の予測値を求めよ.

計算すると、T=101 のとき、3071, T=200 のとき、59874 になる.
左辺が自然対数になっているので、計算は $\exp\{5.0 + 0.003 \times 101\}$ などを計算すればよい。

例題１と比べると、T=101 の時の予測値はほぼ同じだが、T=200 の予測値は大きく異なる。つまり時間が進むと誤差は大きくなることがある。

誤差の評価：Durbin-watson 比

残差の評価には、よく Durbin-watson 比という値が用いられる。（式は覚えなくてよい）

$$DW = \frac{\sum_{i=1}^{t}(e_i - e_{i-1})^2}{\sum_{i=1}^{t} e_i^2} \quad \text{ただし } e \text{ は残差（誤差）}$$

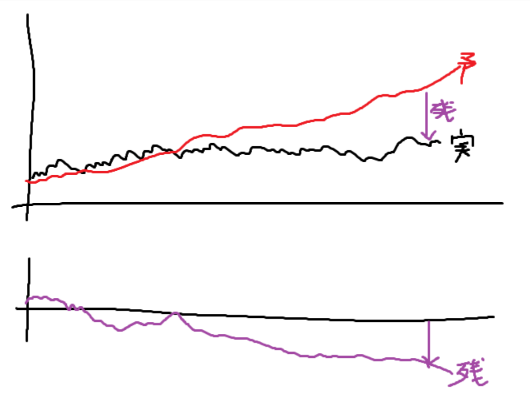
これは考え方としては、残差のラグ１の自己相関を推定している。
Durbin-watson 比 (DW) の値は 0 から 4 までの値を取る。

- DW が 2 であった場合、残差は自己相関（系列相関）を持たない
- DW が 0 であった場合、残差は正の 100%自己相関を持つ
- DW が 4 であった場合、残差は負の 100%自己相関を持つ

残差が正／負の自己相関をもう、とはどういう意味か？

誤差の評価：Durbin-watson 比

例えば、下記の図のような予測を行った場合、DW は２よりも０に近い値になる。



なぜなら、残差が時間が増えると増加するトレンドがあり、つまり前の時刻の残差が大きいと次の時刻の残差も大きくなる傾向がある。

誤差の評価：Durbin-watson 比

つまり下記のように評価できる。

- DW が２より０に近い
  → 残差に正の系列相関があり、誤差が一度発生するとなかなか消えない傾向がある。
- DW が２より４に近い
  → 残差に負の系列相関があり、残差に周期性がある事になる。
- DW が２に近い
  → 残差に系列相関はない。この場合のみ、残差は良い性質を持っていると言える。

トレンドモデルではしばしば、DW が４に近い値になり、残差が正の系列相関を持つ場合がある。なぜならトレンドモデルは一定の増加／減少トレンドがずっと続くと仮定しているモデルであるが、現実は様々な出来事がインパクトを与え、そのインパクトの影響は後々まで影響を与えるものであるから。だから DW 比での誤差評価は、特にトレンドモデルの場合は重要である。

⇒ ここまでの内容を用いて、協会テキスト例題４－１を解け

自己回帰（AR) モデル

自己回帰（AutoRegression）モデル は、下記のようなモデルを仮定している。

$$y_t = \alpha + \beta y_{t-1} + \epsilon_t$$

トレンドモデルとの違いは、第二項の変数が時刻 t ではなく、前期の自分の値 y_{t-1} である点である。自己回帰モデルは前期だけでなく前前期、前前前期、。。。と好きなだけ変数を増やすことが原理上可能である。特に前記だけを用いたモデル（上記に書いたモデル）のことを、AR(1) モデルという。 k 期前までの値を利用したモデルは、AR(k) モデルという。

AR モデルでの予想

AR(1) モデルの予測計算も出題されそう。

例題 3

あるデータ Y について、期間 $T = 1, 2, \ldots, 100$ を用いて、AR(1) モデルに当てはめた結果、下記のような結果を得た

$$y_t = \alpha + \beta y_{t-1} + \epsilon_t$$

ただし $\alpha = 1000, \beta = 0.2,$

また、$y_{100} = 1200$ であった。

このとき、$T = 101$ および $T = 102$ の予測値を求めよ.

計算すると、T=101 のとき、$y_{101} = 1240$, T=102 のとき、$y_{102} = 1248$ になる。

平均回帰水準

前記の予測を先の時刻まで続けてゆくと、だんだんある値に漸近してゆくことが分かる。
この値のことを、平均回帰水準という。平均回帰水準は、AR モデルの y をすべて同一の x として一次方程式を解けば求まる。
前ページの例の平均回帰水準は 1250 である。

$$\begin{aligned} x &= 1000 + 0.2x \\ &\to x = 1250 \end{aligned}$$

⇒ ここまでの内容で、協会テキスト例題４－２を解け。

ランダムウォーク

実は全ページの計算では、平均回帰水準について言及していないことがあった。
AR モデルの係数によっては、平均回帰水準が存在しないケースがある。実際、

$$y_t = \alpha + \beta y_{t-1} + \epsilon_t$$
$$\text{ただし } \alpha = 1000, \beta = 1.0,$$

のような場合、平均回帰水準のための方程式（特性方程式）を立てると、$x = 1000 + x$ となってしまい、解が一意に定まらない。このような場合をランダムウォークといい、特に重要な意味を持つ。

ランダムウォーク

$$y_t = \alpha + y_{t-1} + \epsilon$$

なお定数項 α のことをドリフト項と呼ぶ。

単位根検定

- ある時系列データが、$\beta = 1$ の自己回帰モデルに従っている場合、その時系列データは「単位根を持つ」という。
- 単位根をもつなら平均回帰水準が存在せず、つまり平均値が時間によって変化する。つまり定常性を持たない
- 定常性を持たないデータはそのままトレンドモデルや AR モデルを当てはめても、予測に利用することはできない。

したがって、あるデータが単位根を持つか否かは、重要である。
なお単位根を持つデータを分析するには、前期との差を取り「変化率」のデータにしてから分析すればよい。

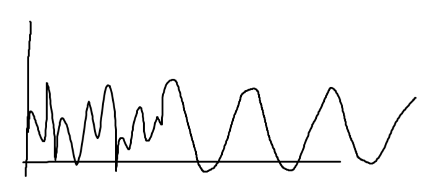
Dickey-Fuller 検定

単位根を持つか否かは、基本的には $\beta = 1$ の帰無仮説を検定すればよいが、通常の t 検定ではダメで、特別な検定方式を行う。その検定を Dickey-Fuller 検定という。
特別ではあるが、t 値や p 値を閾値と比べて棄却する／有意とすることには変わらない。
⇒ 協会テキスト例題４－３を読め（解けなくてもよい）

ARCH モデルや GARCH モデルは出題されるとは思えないので割愛する。

構造変化と季節性

AR モデルの実践としては、データを時系列プロットして、構造変化のある点を境にして、分析を分割する方法がある。



他の実践として、データに季節性（つまり１年単位の周期性）が有る場合は、月次データの場合、ラグ k=12 の項を追加した AR モデルで分析するケースがある。

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-12} + \epsilon_t$$

こうすることで、年単位の周期性の因子をモデルに導入することができる。

分析手順のフロー

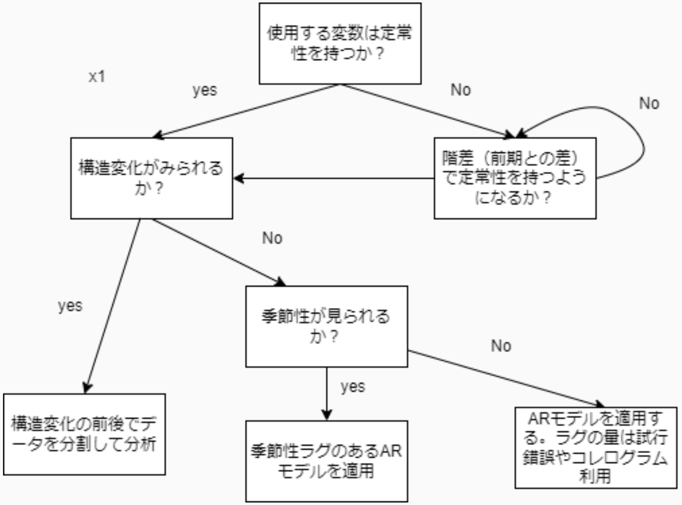


Figure 1: 分析手順のフロー

見せかけの回帰と共和分

単位根過程の重要な性質として、見せかけの回帰がある。
これは、まったく関係のない二つの単位根過程で回帰分析を行うと、まったく無関係にもかかわらず、回帰係数が統計的に有意にゼロでない回帰式が導かれてしまう、というもの。
原理はともかく、

- 二つの単位根過程同士の回帰分析をすると、見せかけの回帰が起こるために普通は正しく分析ができない。

ということを覚えておくこと。これには例外があり、二つの単位根過程が「共和分関係」にある場合は見せかけの回帰が発生しない。共和分関係についてのこれ以上の内容はおそらく出題されないので割愛する。

⇒ 協会テキスト 第４章の練習問題を解け。