

NTT時系列データのKalman filterによる解析

01001600 成蹊大学 上田 徹 UEDA Tohru

02103080 上村 哲志 KAMIMURA Tetsushi

1.はじめに

国鉄(JR)の民営化前のデータをKalman filter^[1]を用いてデータを縮約したが、そのとき民営化の効果がデータ予測に大きな影響を与えたので、電電公社(NTT)についても民営化やその他の効果についてKalman filterを用いて分析する。

2.Kalman filterについて

時系列データを分析する方法には様々な方法があるが、モデルの柔軟性の観点からJRの場合と同様にKalman filter^[2]を用いる。

時点 t でのトレンドなどに関する状態を表す変数 $a(t)$ を推定、予測したいものとする。時点 t から時点 $(t+1)$ への状態変化は、雑音 $u(t)$ を考慮して構造方程式

$$a(t+1) = Fa(t) + u(t)G$$

で与えられるものとする。時点 t での観測値 $b(t)$ は状態 $a(t)$ と関連づけられるが、雑音 $w(t)$ を含んでいるため、観測方程式

$$b(t) = H(t)'a(t) + w(t)$$

で表されるものとする。

時点 t でのデータ $b(t)$ は、民営化効果以外の効果も考慮することとして

$$b(t) = \text{トレンド成分 } \pi(t) + \text{単位料金改定効果 } L(t, D_0) \\ + \text{民営化効果 } L(t, D_1) \\ + \text{NTTデータ分割効果 } L(t, D_2) + \text{不規則成分 } w(t)$$

と表現できるとする。トレンド成分 $\pi(t)$ は2次差分

$$\nabla^2 = \{T(t) - T(t-1)\} - \{T(t-1) - T(t-2)\} \\ = u(n) \sim N(0, \tau^2)$$

を用いる方法の他に、1次、3次差分も取り上げる。時点 D_0 で単位料金改定、時点 D_1 で民営化、時点 D_2 でNTTデータ分割が行われた。このとき、状態方程式は、トレンド成分を2次差分で表現する場合には、

$$a(t+1) = Fa(t) + u(t)G$$

$$a(t) = (T(t), T(t-1), L(t, D_0), L(t, D_1), L(t, D_2))'$$

:状態

$$F = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

で与えられ、

観測方程式は、

$$b(t) = H(t)'a(t) + w(t)$$

$$w(t) \sim N(0, \omega^2)$$

$$H(t)' = [1 \ 0 \ \delta(t, D_0) \ \delta(t, D_1) \ \delta(t, D_2)]$$

$$\delta(t, D_i) = 0 : t < D_i \\ = 1 : t \geq D_i \quad ; \quad i = 0, 1, 2$$

で与えられる。 ω^2 は規格化して1とすることが出来、

未知パラメタは τ^2 だけなので対数尤度

$$l(\tau^2) = \sum_t (-1/2) \log \{2\pi R(t|t-1)\} \\ - \sum_t \{b(t) - b(t-1)\}^2 / \{2\pi R(t|t-1)\}$$

を最大にするように決めればよい。ただし、

$$R(t|t-1) = H(t)'P(t|t-1)H(t) + \omega^2$$

である。誤差分散および初期値 $a(0)$ はKalman filterを逆向きに使うことによって求める。

3.分析結果

文献[3]のデータに基づき電電公社、NTTの時系列データの6項目、すなわち市外通話収入、市内通話収入、その他の収入、総資産、従業員数、加入電話総数についてKalman filterを用いて分析を行った。

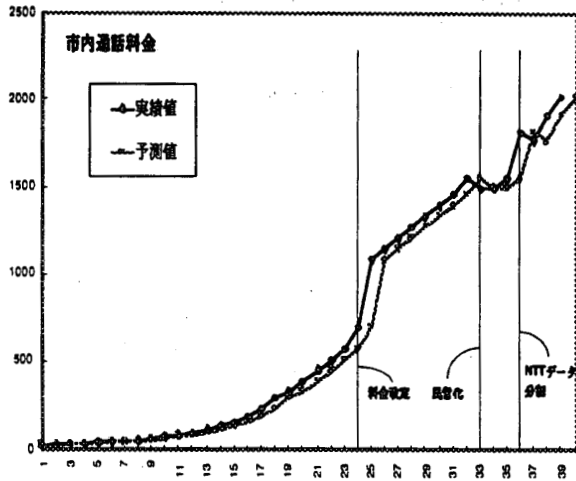
その結果、総資産、従業員数、加入電話総数については単位料金改定、民営化、NTTデータ分割の効果は見られなかった。

また、市外通話収入、市内通話収入、その他の収入については、その効果が見られた。

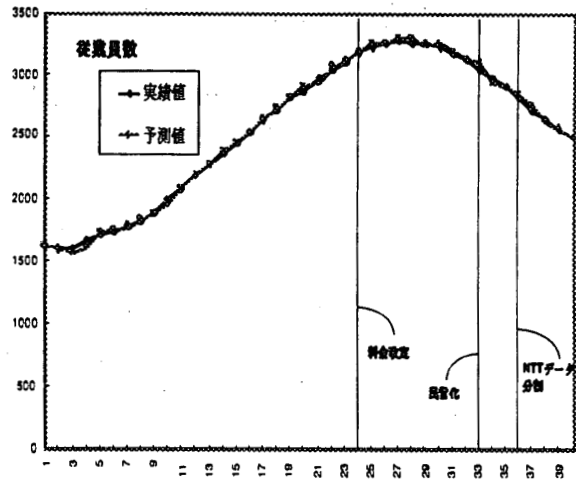
その結果、効果が見られた項目についてグラフを調べていくと、市内通話収入の項目だけ予測が時点 D_0 、 D_1 、 D_2 で大きく外れる傾向があったので、その項目についてだけトレンド成分を2次から1次、3次に変更してみたが、1次については誤差を加えただけになってしまい結局1時点ずれただけのようになり、3次に変更したらより大きく外れるようになってしまった。

また、効果が見られなかった項目でも従業員数のグラフのみ2次曲線のような形だったので、トレンド成分を3次に変更して推定、予測をおこなったが、2次のときとほとんどかわらなかった。以上のことより、結局トレンド成分は2次でおこなえば十分であるということが分かった。

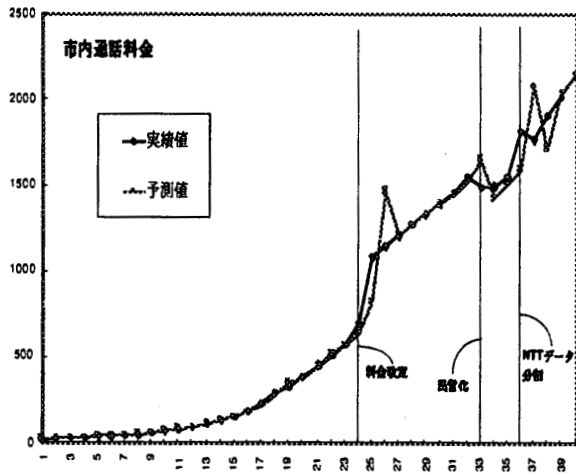
<効果が見られたデータ（市内通話料金）:1次>



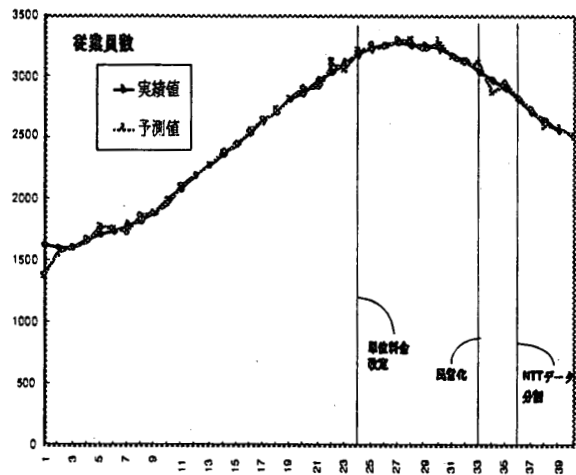
<効果が見られないデータ（従業員数）:2次>



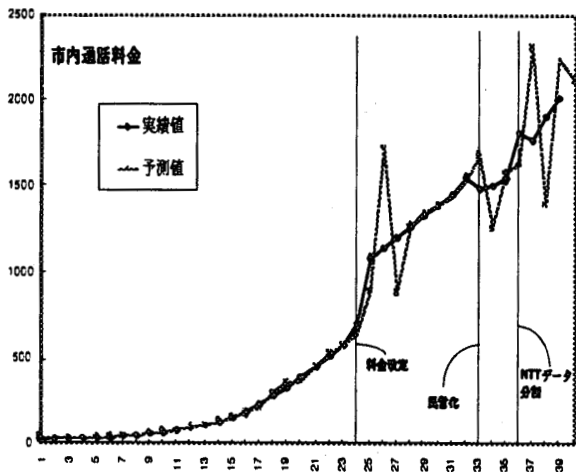
<効果が見られたデータ（市内通話料金）:2次>



<効果が見られないデータ（従業員数）:3次>



<効果が見られたデータ（市内通話料金）:3次>



4. 今後の課題

Kalman filterにより縮約された予測値を入出力値としてもつDMUを想定し、その他のDMUとして各年度ごとの実績値やNTT以外の通信事業者の実績値を対応させてDEAにより業績を評価したい。

- [1]上田、染谷:「DEAにおけるKalman filterの利用」、1998春期OR学会大会
- [2]上田:「予測手法(1):時系列予測法」、オペレーションズ・リサーチ、1994年6月号
- [3]末吉:"Measuring Efficiencies and Returns to Scale of Nippon Telegraph & Telephone in Production and Cost Analyses", management science, Vol.43, No.6, June 1997