

電話通話量の予測

星合 擁湖, 上田 徹

1. はじめに

電電公社1社で独占的に電話サービスを提供していた時代と異なり、マルチキャリア化（新通信業者〔以下、NCC：New Common Carrier〕参入とサービス提供地域の拡大）、各種新サービスの出現、情報通信ネットワークの拡大などのため、通話量（発信回数、総通話時間などのトラヒック需要）が単純には推定できなくなっている。しかし、最適な電話設備の設計・建設、適切なサービス料金の設定などのためには、将来の通話量を精度よく把握しておかなければならない。

本稿では、マルチキャリア環境および電話料金の改定を考慮した電話通話量の予測について述べる。

まず、第2章で、通話量の予測に用いる基本モデルを定義し、第3章で、モデルを比較評価するために用いる指標（自由度2重調整済み決定係数、AIC）を紹介する。第4章では、基本モデルの各説明変数に関して検討した結果、および、その結果選択した予測モデルについて述べる。第5章で、選択したモデルによる将来の通話量の予測例を示す。

2. 基本モデルと説明変数 [1]

電話料金が発着段階ごとに電話料金と通話量との関係について分析する。

通話量の変化の要因として、電話料金、トレンド、マルチキャリア環境の拡大、景気動向および加入者数の増減などが挙げられる。これらの要因を説明変数にもつ、量的変数と質的変数（0, 1の値をとるダミー変数）の混在する重回帰モデルを基本モデルとする。

一般の重回帰モデルは、線形回帰型モデルであるが、通話量分析では、経済学の弾力性分析の考え方が利用できる乗法型モデル(1)を用いる[2]。

[基本モデル]

特定の距離段階における第*i*地域の第*j*四半期（通

しで第*t*四半期）における総通話時間 Y_{it}

$$Y_{it} = C \cdot S_j \cdot A_{it} \cdot B_t \cdot T_t \cdot M_{it} \cdot E_t \cdot SUB_{it} \quad (1)$$

ただし、 C ：定数； S ：季節（四半期）；

A ：地域； B ：料金； T ：トレンド；

M ：マルチキャリア環境；

E ：経済指標； SUB ：NTT加入者数

乗法型モデルは、対数変換により線形回帰型モデル(2)に変換し、線形型モデルと同様、最小二乗法により分析することができる。

$$\log Y_{it} = \log C + \log S_j + \log A_{it} + \log B_t + \log T_t + \log M_{it} + \log E_t + \log SUB_{it} \quad (2)$$

この基本モデルの各説明変数を追加、削除、変更しつつ分析・比較し、最もあてはまりのよいモデルを選択する。以下に、各説明変数（成分）について詳しく記す。

2.1 季節成分： S_j

ダミー変数を用いる。 s_j は第*j*四半期の影響を表わす。

$$[\text{モデル}S] \log S_j = \sum_k s_k \delta_{kj} = s_j,$$

$$\text{ただし, } \delta_{kj} = \begin{cases} 1, & \text{if } k = j; \\ 0, & \text{otherwise.} \end{cases}$$

ただし、第4四半期を基準（ $s_4 = 0$ ）とし、推定すべき偏回帰係数を s_1, s_2, s_3 の3個とする。

2.2 地域成分： A_i （または A_{it} ）

季節成分と同じダミー変数、地域別NTT加入者数などを用いた以下の3通りの地域成分を検討する。 a_i は第*i*地域の影響を表わす。モデルA3は、モデルA2の弾力性モデルである。

$$[\text{モデル}A1] \log A_i = \sum_k a_k \delta_{ki} = a_i$$

ただし、地域数が*I*のとき、第*I*地域を基準（ $a_I = 0$ ）とし、推定すべき偏回帰係数を a_1, a_2, \dots, a_{I-1} の*I*-1個とする。

$$[\text{モデル}A2] \log A_{it} = \sum_k a_k \delta_{ki} sub_{it} = a_i sub_{it}$$

ほしあい ようこ, うえだ とおる

NTT通信網総合研究所

〒180 武蔵野市緑町3-9-11

地域 i の時点 t における加入者数 sub_{it} を考慮する。
 $I-1$ 個の偏回帰係数 a_i を推定する。

[モデルA 3] $A_{it} = sub_{it} \sum_k \delta_{ki} = sub_{it}^{a_i}$

2.3 料金成分: B_t

料金改定とそれが認定されるまでのタイムラグを考慮した「料金認定モデル」を適用する。 P_t は時点 t での認識されている料金、 e は料金改定の影響を表わす。

[モデルB 1] $B_t = (P_t)^{-e}$

NTT料金そのもの(名目値)の他に、NTT料金のNCC料金に対する相対価格(=NTT料金/NCC料金)、実質値(=NTT料金/消費者物価指数(C.P.I))などを検討する。

2.3.1 料金認識モデル

料金改定は徐々に認識されるものとし、以下に記す料金認識モデルを検討する。改定前後の料金を、それぞれ、 P_a 、 P_b とし、改定後 t ヶ月経過したときに認識されている料金を P_t とする。改定が認識されるまで(モデルP 3では改定が90%認識されるまで) N ヶ月かかるとする。(図1参照)

[モデルP 1] $P_t = \begin{cases} P_a, & \text{if } t < N; \\ P_b, & \text{if } t \geq N. \end{cases}$
 [モデルP 2] $P_t = P_a + (P_b - P_a) \times t / N$
 [モデルP 3] $P_t = P_a + (P_b - P_a) \times \{1 - \exp(-\lambda t)\}$
 ただし、 $\lambda = (\log 10) / N$

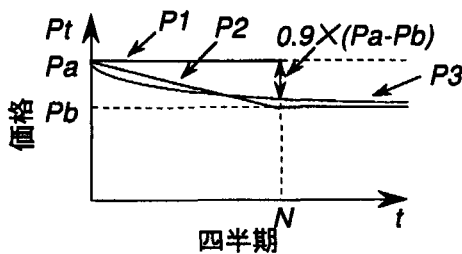


図1 料金認識モデル

2.4 トレンド成分: T_t

トレンドの効き具合を考慮した以下の3通りのトレンド成分を検討する。 t は通しの四半期を、 c_T はトレンドの影響を表わす。(図2参照)

[モデルT 1] $\log T_t = c_T t$
 [モデルT 2] $\log T_t = c_T \log(t) = \log(t^{c_T})$

[モデルT 3] $\log T_t = c_T \sqrt{t}$

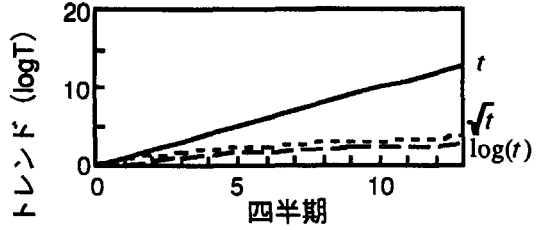


図2 トレンド成分

2.5 マルチキャリア環境成分: M_{it}

各地域内のPOI (Point of Interface; NCC網とNTT網との接続点)の数、NCCサービス提供地域の加入者数、ID(発信加入者番号)登録(NCC契約)加入者数などを用いた以下の10通りのマルチキャリア環境成分を検討する。他社網を利用している通話量は把握できないため、考慮の対象から外した。 i は地域を、 t は通しの四半期を、 c_M はマルチキャリア環境の影響を表わす。

[モデルM 1] 地域 i における時点 t でのPOIの有無

$\log M_{it} = c_M \xi_{it}$
 $\xi_{it} = \begin{cases} 1, & \exists \text{ POI} \in \text{area of } i, \text{ time } t; \\ 0, & \forall \text{ POI} \notin \text{area of } i, \text{ time } t. \end{cases}$

ただし、以下のモデルでも ξ_{it} の定義は同じ。

[モデルM 2] 地域 i における時点 t でのPOI数

$\log M_{it} = c_M n_{it}$
 n_{it} : 地域 i における時点 t でのPOI数

[モデルM 3] 全国における時点 t でのNCCのサービス提供地域率

$\log M_{it} = c_M \xi_{it} \times \log m_t$
 m_t (NCCサービス提供地域の加入者数) / (全国加入者数)

[モデルM 4] 地域 i における時点 t でのNCCのサービス提供地域率

$\log M_{it} = c_M \xi_{it} \times \log(m_t \times m_{it})$
 m_{it} : (地域 i 内のNCCサービス提供地域の加入者数) / (地域 i 内加入者数)

[モデルM 5] 地域 i における時点 t でのID登録率

$\log M_{it} = c_M \xi_{it} \times \log(m_t \times d_{it})$
 d_{it} : (地域 i 内のID登録加入数) / (地域 i 内加入数)

[モデルM 6] モデルM 3に対応するモデル

$\log M_{it} = c_M \xi_{it} \times m_t$

[モデルM7] モデルM4に対応するモデル

$$\log M_{it} = c_{M1} \xi_{it} \times m_t \times m_{it}$$

[モデルM8] モデルM5に対応するモデル

$$\log M_{it} = c_{M1} \xi_{it} \times m_t \times d_{it}$$

[モデルM9] モデルM5の成分を2つに分けたモデル

$$M_{it} = m_t^{CM1} \xi_{it} \times d_{it}^{CM2}$$

[モデルM10] モデルM8の成分を2つに分けたモデル, モデルM9に対応

$$\log M_{it} = c_{M1} \xi_{it} \times m_t + c_{M2} \xi_{it} \times d_{it}$$

マルチキャリア環境を考える成分として, このほかに, 料金成分と同様のNCC料金成分を検討する。

2.6 経済指標成分: E_t

経済指標として, 名目国民総支出を採用し, この成分を検討する。経済の影響のタイムラグも考慮する。 t は通しの四半期を, c_E は経済の影響を表わす。モデルE2はモデルE1の弾力性モデルである。経済指標として, 名目値だけでなく実質値(=名目値/消費者物価指数(C.P.I.))も検討する。

[モデルE1] $\log E_t = c_E gne_{t-k}$

gne_t : 名目国民総支出(原系列),

k : タイムラグ

[モデルE2] $E_t = (gne_{t-k})^{c_E}$

2.7 加入者数成分: SUB_{it}

i は地域, t は通しの四半期, c_s は加入者数の影響を表わす。モデルSUB2はモデルSUB1の弾力性モデルである。この成分は, 地域成分A2, A3と酷似しているが, 偏回帰係数が c_s の1個だけである点異なる。

[モデルSUB1] $\log(SUB_{it}) = c_s \times sub_{it}$

sub_{it} : 地域別, 四半期別加入者数

[モデルSUB2] $SUB_{it} = (sub_{it})^{c_s}$

3. モデルの評価指標

モデルのあてはまり具合を検討する指標として, 決定係数と赤池の情報量基準を用いた。

また, 各モデルを詳細に検討するため, 重回帰式の各説明変数の分析を行なった。重回帰式の各説明変数の偏回帰係数, 標準偏回帰係数, F 値, t 値, 標準誤差, 偏相関係数および単相関係数を求め, 特に, 偏回帰係数の符合が成分の意味づけと合致しているかによりモデルの捨取選択を行ない, また, 偏回帰係数の t

値により当該成分の有意性を判断した。

3.1 自由度2重調整済み決定係数 [3]

決定係数 R^2 はモデルのあてはまりのよさを評価するための統計量であり, その値は, $0 \leq R^2 \leq 1$ である。 R^2 が大きいほどデータに対して仮定したモデルのあてはまりはよくなるが, あてはまりは変数の数を増やすほどよくなる。しかし, 変数の数が増えるほど予測精度は下がるため, 変数の数を考慮した自由度2重調整済み決定係数 R^{**2} を用いてモデルを評価した。 R^{**2} が大きいほどよいモデルである。

決定係数 R^2 , および, 自由度2重調整済み決定係数 R^{**2} は以下の式により算出される。

$$R^2 = 1 - \frac{\sum(Y_{it} - \dot{Y}_{it})^2}{\sum(Y_{it} - \bar{Y})^2}$$

ただし, \dot{Y}_{it} : Y_{it} の推定値, \bar{Y} : Y_{it} の平均値。

$$R^{**2} = R^2 - \frac{2np}{(n+1)(n-p-1)}(1-R^2)$$

ただし, n : サンプル数, p : パラメータ数, $\sqrt{R^2}$ を重相関係数という。

3.2 赤池のAIC [4]

赤池の情報量基準AICは以下の式により算出される値で, モデルの相対的なよさを評価するための統計量である。この値が小さいほどよいモデルである。

$$\begin{aligned} AIC &= -2 \times \log(\text{最大尤度}) + 2(p+2) \\ &= n(\log 2\pi + 1) + n \log(S_E/n) + 2(p+2) \end{aligned}$$

ただし, p : パラメータ数, n : サンプル数, S_E : 残差平方和(= $\sum(Y_{it} - \dot{Y}_{it})^2$)

ここで, $(p+2)$ の2は, モデルの誤差項の平均と分散である。ただし, AICの絶対値には意味がなく比較するモデル間のAICの差だけが問題となるので, $(p+2)$ の代わりに一貫して p を用いることにしてよい。

4. 分析例(説明変数およびモデルの選択)

今回の分析には, 四半期別地域別総通話時間を用いた。各分析に用いたサンプル数は, 121(=11四半期×11地域)である。

期間: 1988年度第4四半期~1991年度第2四半期(11四半期)

地域: 11地域

事住別：事務用／住宅用／事住総合の3種類
 時間帯：平日昼間／曜日総合1日計の2種類
 距離段階：20～30 km／320 km以遠の2種類

(注) 現在、320 km以遠の距離段階は、
 160 km以遠として統合されている。

4.1 各説明変数に関する分析結果

各説明変数に関する分析結果を示す。

4.1.1 季節成分： S_j

季節成分は、どのモデルにおいても有効である。その偏回帰係数の例を図3に示す。

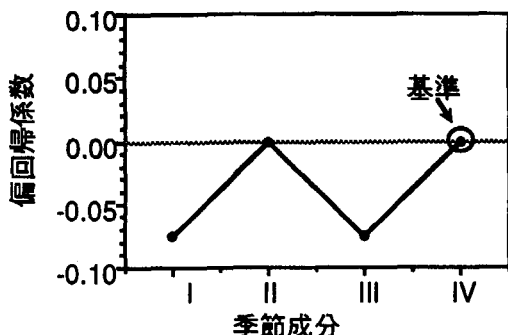


図3 季節成分の偏回帰係数の例

4.1.2 地域成分： A_i (または A_{it})

節2.2で示した3通りの成分を検討した結果、ダミー変数を用いたA1が最もよい。過去の分析例(米国[5], NTT)においても、地域性は加入者数よりも人口、世帯数などの方に強く現われており、また、事務用、住宅用通話量と相関が高かったのは、それぞれ、3次産業事業所数、世帯数であった。これは地域成分を加入者数だけでは説明できないためと思われる。地域成分の偏回帰係数の例を図4に示す。

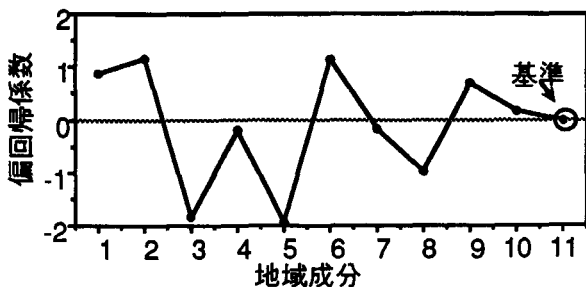


図4 地域成分の偏回帰係数の例

4.1.3 料金成分： B_t

節2.3.1.で示した3通りの料金認識モデルについて、タイムラグを変化させて検討した結果、3ヵ月後

に90%認識されるモデルP3

$$P_t = P_a + (P_b - P_a) \times \{1 - \exp(-\lambda t)\}$$

ただし、 t ：料金改定後の月数； $\lambda = (\log 10 / 3)$ が最もよい。

距離段階20～30 kmの総通話時間に関して、NTT料金のNCC料金に対する相対価格を用いて分析すると、NTT料金そのものを分析結果より、AICが若干小さくなる。また、NTT料金成分を追加して分析すると、AICは大きくなる。したがって、このNCC料金成分の追加は不要である。最後に、名目値と実質値の両方に関して分析したが、どちらを用いても大差はない。

4.1.4 トレンド成分： T_t

節2.4の3通りのものを検討した結果、モデルT2

$$\log T_t = c_T \log(t) = \log(t^{c_T})$$

が最もよい。ただし、距離段階20～30 kmに関しては、トレンド成分を考慮しない方がよい。

4.1.5 マルチキャリア環境成分： M_{it}

節2.5の10通りのマルチキャリア環境成分を検討した結果、事務用通話量においては、NCCのID登録加入数を考慮したモデル(距離段階320 km以遠ではモデルM8、20～30 kmではモデルM5)がよい。

$$[\text{モデルM5}] M_{it} = (m_t \times d_{it})^{c_{M5it}}$$

$$[\text{モデルM8}] \log M_{it} = c_{M8} \xi_{it} \times m_t \times d_{it}$$

また、住宅用通話量に関してはマルチキャリア環境成分を考慮しない方がよい。

4.1.6 経済指標成分： E_t

モデルE1を用いて経済指標成分を検討した。経済指標成分に関しては、タイムラグを考慮してもAICの観点での改善はない。また、料金成分同様、名目値と実質値の両方に関して分析したが、どちらを用いても大差はない。

$$\log E_t = c_E gne_t$$

4.1.7 加入者数成分： SUB_{it}

モデルSUB2を用いて加入者数成分を検討した。

$$SUB_{it} = (sub_{it})^{c_S}$$

ただし、地域成分として加入者数を考慮したモデル(節2.2A2, A3)を選択したときには、この成分は削除する。

4.2 選択モデル

距離段階320 km以遠の事務用の総通話時間に関して検討した結果を記す。当てはまりのよいモデルは表1に示す3モデルである。この中でも、モデル1が自由度2重修正済み決定係数が最大、AICが最小であ

表1 基本モデルの各成分の有効性

成分	モデル1	モデル2	モデル3
季節(ダミー変数)	○	○	○
地域(ダミー変数)	○	○	○
料金(名目値)	○	○	○
トレンド(T2)	×	○	×
マルチキャリア(M8)	○	○	○
経済指標(名目値)	×	×	×
地域別NTT加入者数	○	×	×
R^{**2}	0.99727	0.99712	0.99684
AIC	-385.016	-378.561	-367.436

(○:有効である成分; ×:有効でない成分)

り, 最もよい.

選択モデル

$$\log Y_{it} = \log C + \sum_k s_k \delta_{ki} + \sum_k a_k \delta_{ki} - e \cdot \log(P_i) + c_M \xi_{it} \times m_i \times d_{it} + c_s \cdot \log(sub_{it})$$

5. 総通話時間の予測例

節4.2で選択したモデルにより, ある地域のNTTの総合通話時間を予測する.

まず, シナリオを設定する. 4回目の料金改定▲の有無とNCCのID登録加入率の推移(図5のID1またはID2)を仮定し, それぞれの場合の通話量を予測する. なお, NTT加入者数は直線外挿で算出する.

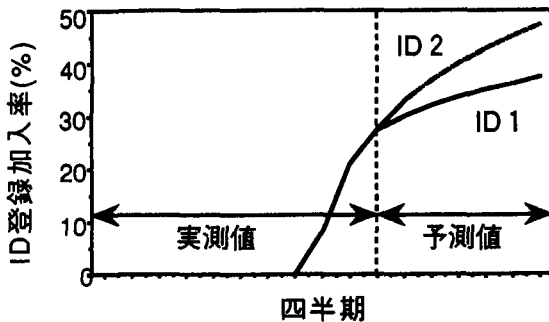


図5 ID登録加入率のシナリオ

以上の4通りのシナリオで予測した結果を, 図6に示す. シナリオを変更することにより, さまざまな条件下で予測することができる.

6. あとがき

各距離段階ごとに適切なモデルを選ぶため試行錯誤を繰り返した. そのようなプロセスのいらない分析ツールがあればよいが, 環境の変化に充分追従できるようなツール化を通じて, 予測担当者の負担を軽減して

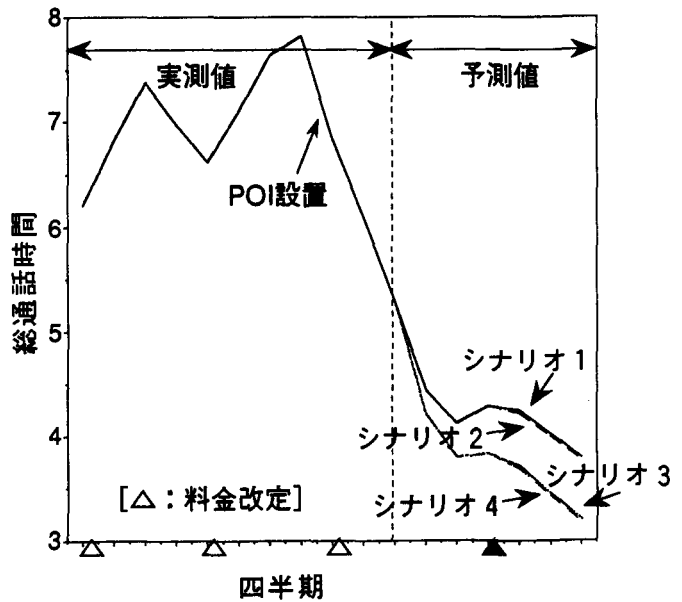


図6 A地域発総通話時間(距離段階X:事務用)

——	シナリオ1	(ID1:料金改定▲あり)
- - - -	シナリオ2	(ID1:料金改定▲なし)
— · — ·	シナリオ3	(ID2:料金改定▲あり)
- · - · - ·	シナリオ4	(ID2:料金改定▲なし)

いるが, 基本モデルを越える環境変化については, やはりツールの改造が必要である.

参考文献

- [1] 星合, 上田: 通話量に関する重回帰モデルの検討, 1992年電子情報通信学会秋季大会, B-503, 1992.
- [2] 円山由次郎: 需要予測と計量経済分析, 日本生産性本部, 1970.
- [3] 奥野忠一他: 続多変量解析法, 日科技連出版社, 1976.
- [4] 赤池弘次: 情報量基準AICとは何か—その意味と将来への展望—, 数理科学, No. 153, pp. 5-11, 1976.
- [5] Gatto, J. P.ほか: "Interstate Switched Access Demand Analysis", *Information Economics and Policy*, 3, pp. 333-358, 1988.