Ⅳ-40　予測誤差基準による発生重回帰モデルの評価方法に関する研究

建設省　正員　○桐越　信
北海道大学　正員　佐藤馨一
北海道大学　正員　五十嵐日出夫

1. はじめに

都市交通計画における総合的交通需要予測プロセスでは，発生交通量予測モデルとして重回帰モデルを採用することが多い．しかし，重回帰モデルの評価については，現象をどの程度説明できるかという点からの評価が主であり，予測モデルであるということを意識した場合の評価についてはあまり考慮されていない．そこで本研究では，重回帰モデルによる発生交通量予測モデルを〈発生重回帰モデル〉と名付け，このモデルが予測モデルであるということを意識した場合の評価について，モデルに取り入れられた説明変数のもつ将来予測に際しての不確実性，すなわち説明変数の予測時における予測誤差を考慮した予測誤差基準KCによる発生重回帰モデルの評価方法を提案し，さらにこの方法と従来の方法との関連性を把握することを目的としている．

2. 発生重回帰モデルによる予測

発生重回帰モデルによる予測において，モデル構築時と予測時とで変数間に構造的変化がある場合には，構築された発生重回帰モデルが変数間の構造的変化にどの程度追従できるかが問題となる．このモデルの追従性は，変数間の構造的変化の程度によって影響を受ける．ここではこの追従性をも考慮した予測性を〈広義の予測性〉とよぶことにする．一方，モデル構築時と予測時とで変数間に構造的変化がないと仮定した場合には，モデルに取り入れられた説明変数の予測時における予測誤差による予測精度の低下が問題となる．この説明変数の予測時における予測誤差は説明変数自体の予測の困難さに起因するもので，一般につねに存在するので，発生重回帰モデルによる予測では，予測値が説明変数の予測時における予測誤差にあまり影響を受けないこと，すなわちモデルの頑健性が要求される．この頑健性のみを考慮した予測性を〈狭義の予測性〉とよぶことにする．発生重回帰モデルによる予測では，モデルの時間的移転可能性[1]を予測の基本的前提条件としている．このことはモデル構築時と予測時とで変数間に構造的変化がないということを仮定していることになり，予測といっても狭義の予測であることを示している．

3. 予測誤差基準KCの理論的検討

予測対象y_0と説明変数の予測時における予測誤差Δx_0の関係を示したのが図1である．モデル構築時の被説明変数をy_i，y_iのモデルによる期待値を$\beta(x_i)$，偶然的変動による影響をε_iとすると，

$$y_i = \beta(x_i) + \varepsilon_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i \qquad (1)$$

となる．ここでε_iは$N(0, \sigma^2)$に従うと仮定する．

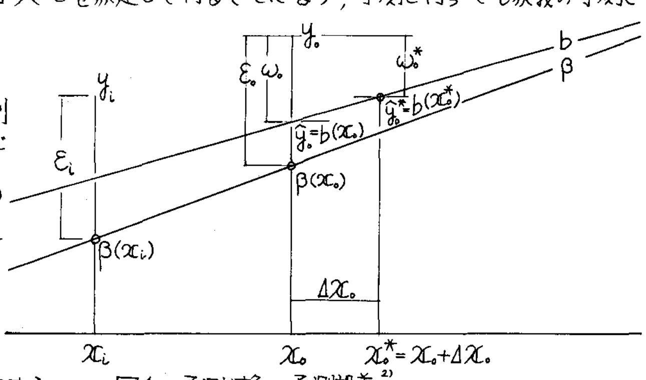

図1　予測対象と予測誤差[2]

(1)式より次式を満たすモデルβの推定モデルbが得られる．

$$E[b(x_i)] = E[b_0 + \sum_{k=1}^{p} b_k x_{ik}] = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} = \beta(x_i) \qquad (2)$$

予測対象をy_0とし，モデルに取り入れられた説明変数の予測時における予測誤差をΔx_0とすると，推定モデルb

による予測誤差ω_0^*は

$$\omega_0^* = y_0 - \hat{y}_0^*$$
$$= \beta(x_0) + \varepsilon_0 - b(x_0 + \Delta x_0)$$
$$= \beta_0 + \sum_{k=1}^{p} \beta_k x_{0k} + \varepsilon_0 - \{ b_0 + \sum_{k=1}^{p} b_k (x_{0k} + \Delta x_{0k}) \} \quad \text{———} (3)$$

となる。ここで，狭義の予測であるので，$E[\Delta x_0] = 0$を仮定すると，（4）式が得られる．

$$E[\omega_0^*] = 0, \quad V[\omega_0^*] = E[\omega_0^{*2}] = \{1 + 1/n + D_0^2/(n-1)\}\sigma^2 + V[\sum_{k=1}^{p} b_k \Delta x_{0k}] \quad \text{———} (4)$$

（4）式でD_0はP次元空間における$x_0 = (x_{01}, \cdots x_{0k} \cdots x_{0p})$と$\bar{x} = (\bar{x}_1, \cdots \bar{x}_k, \cdots \bar{x}_p)$のマハラノビス汎距離である。モデル構築時と予測時とで変数間に構造的変化はないと仮定しているので，予測時のx_0としてモデル構築時のx_iを採用し，x_0^*には次式を満たすx_i^*を採用して，

$$E[x_i^*] = E[x_i + \Delta x_i] = x_i, \quad V[x_i^*] = V[x_i + \Delta x_i] = V[\Delta x_i] \quad \text{———} (5)$$

そこで得られるであろう被説明変数yをy_i^0と表わすと，（4）式より次式が得られる．

$$E[\sum_{i=1}^{n} \omega_i^{*2}] = E[\sum_{i=1}^{n} (y_i^0 - \hat{y}_i^*)^2]$$
$$= (n+p+1)\sigma^2 + \sum_{i=1}^{n} V[\sum_{k=1}^{p} b_k \Delta x_{ik}] \quad \text{———} (6)$$

ここで，$KC = \sum_{i=1}^{n} (y_i^0 - \hat{y}_i^*)^2$とおき，$\Delta x_{ik}$と$\Delta x_{ik'}$が互いに独立であるとすると，（6）式は

$$E[KC] = (n+p+1)\sigma^2 + \sum_{i=1}^{n}\sum_{k=1}^{p} V[\Delta x_{ik}]\{(E[b_k])^2 + V[b_k]\} \quad \text{———} (7)$$

となる。（6）式，あるいは（7）式を予測誤差基準と名付ける。

4. 他の方法との関係

4.1 RSSとの関係

従来発生量回帰モデルの評価は，モデルの現象説明力の点より寄与率R^2，すなわち残差平方和に注目して行なわれてきた。RSSとKCとの関係は，$E[RSS] = (n-p-1)\sigma^2$であるから，（7）式より

$$E[KC] = E[RSS] + 2(p+1)\sigma^2 + \sum_{i=1}^{n}\sum_{k=1}^{p} V[\Delta x_{ik}]\{(E[b_k])^2 + V[b_k]\} \quad \text{———} (8)$$

となる。このことよりKCは残差平方和RSSをも含んだ評価方法であることがわかる。

4.2 PSSとの関係

PSSは次式で定義される．

$$PSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad \text{———} (9)$$

$$\hat{y}_i = b_{0i} + \sum_{k=1}^{p} b_{ki} x_{ik} \quad \text{———} (10)$$

係数b_{ki}は第i番目のデータ（y_i, x_i）を除いた残りの（n−1）組のデータより求める偏回帰係数である。PSSとRSSの関係は$E[PSS] \geqq E[RSS] + 2(p+1)\sigma^2$[3]であるので，（8）式よりPSSとKCの関係は，

$$E[KC] \leqq E[PSS] + \sum_{i=1}^{n}\sum_{k=1}^{p} V[\Delta x_{ik}]\{(E[b_k])^2 + V[b_k]\} \quad \text{———} (11)$$

となる。

5. 予測誤差基準KCの解釈と意義

説明変数の予測時における予測誤差をも考慮した発生量回帰モデルの評価方法である予測誤差基準KCと従来の方法との関連性について把握したが，予測誤差基準KCによる評価では，モデル構築時のデータへの適合性がよく，モデルに取り入れられた説明変数の数が少なく，さらにその説明変数が予測しやすいモデルほど予測モデルとして望ましいモデルということになる。また予測誤差基準KCは，モデルに取り入れられた説明変数自体も予測しなければならないという予測の多段階性を考慮した場合のモデルの簡略化を示唆している。

参考文献　1)杉恵頼寧：交通機関別分担モデルの移転可能性，第2回土木計画学研究発表会講演集，1980年

2)山形耕一，桐越　信：交通需要予測プロセスの予測精度について，第1回土木計画学研究発表会講演集，1979年

3)奥野　忠一他：続多変量解析法，日科技連，1976年