第 6 章　古典的回帰モデルの拡張 ― その 1

扱うトピック：**分散不均一性，系列相関，ラグ付き変数，分布ラグ・モデル**

1　分散不均一性

単回帰モデル

$$y_i = \alpha + \beta x_i + \varepsilon_i \qquad (i = 1, \cdots, n) \tag{1}$$

において，誤差項 ε_i が平均 0 で互いに独立であるとする．分散が一定でなく，

$$V(\varepsilon_i) = E(\varepsilon_i^2) = \sigma_i^2$$

となる場合，誤差項は**分散不均一**であるという．この場合でも，α と β の LSE $\hat{\alpha}$ と $\hat{\beta}$ は不偏性をもつが，BLUE ではなくなる．

ここで，各分散がある既知の変数 w_i $(\neq 0)$ を使って，$\sigma_i^2 = w_i^2 \sigma^2$ と表すことができる場合を考えよう．ただし，σ^2 は未知である．このとき，(1) を

$$\frac{y_i}{w_i} = \frac{\alpha}{w_i} + \beta \frac{x_i}{w_i} + \frac{\varepsilon_i}{w_i} \tag{3}$$

と変形すれば，新たな誤差項 ε_i / w_i は分散均一となるので，(3) は古典的仮定をみたす．そして，

$$\sum_{i=1}^{n} \frac{1}{w_i^2} (y_i - \alpha - \beta x_i)^2 \tag{4}$$

を最小にする α と β は，

$$\hat{\alpha}_W = \bar{y}_W - \hat{\beta}_W \bar{x}_W \qquad \hat{\beta}_W = \sum_{i=1}^{n} \frac{1}{w_i^2} \left(x_i - \bar{x}_W\right)^2 \Big/ \sum_{i=1}^{n} \frac{1}{w_i^2} \left(x_i - \bar{x}_W\right)\left(y_i - \bar{y}_W\right)$$

となることがわかる．ただし，$\bar{x}_W$ と $\bar{y}_W$ は，それぞれ x と y の加重平均であり，

$$\bar{x}_W = \sum_{i=1}^{n} \frac{x_i}{w_i^2} \Big/ \sum_{i=1}^{n} \frac{1}{w_i^2} \qquad \bar{y}_W = \sum_{i=1}^{n} \frac{y_i}{w_i^2} \Big/ \sum_{i=1}^{n} \frac{1}{w_i^2}$$

で定義される．推定量 $\hat{\alpha}_W$ と $\hat{\beta}_W$ は，変換されたモデル (3) に対して BLUE となる．

(4) のようなウェイト付きの 2 乗和を最小にすることにより未知パラメータを推定する方法を**加重最小 2 乗法**といい，推定量 $\hat{\alpha}_W$，$\hat{\beta}_W$ を α と β の**加重最小 2 乗推定量**（**WLSE**：Weighted LSE）という．分散が大きなデータにはより少ないウェイトがかけられることがわかる．なお，(3) のように変換すると，一般には切片を 0 とした回帰を考えることになる．

$\boxed{\text{応用例 I}}$

図 6-1 ：**付録 1** の都道府県別データにおいて，行政投資額 y（単位：兆円）と人口 x（単位：百万人）の関係をプロット．

行政投資は，国や地方公共団体等が行う公共投資であり，道路，住宅，港湾，環境衛生，治山・治水などの事業を含む．また，図中の実線は回帰直線

$$y = 0.060 + 0.385x \qquad R^2 = 0.908, \bar{R}^2 = 0.906, \hat{\sigma} = 0.305$$

(0.91)　(21.05)

であり，点線は回帰曲線

$$y = 0.270 + 0.231x + 0.0153x^2 \qquad R^2 = 0.921, \overline{R}^2 = 0.917, \hat{\sigma} = 0.286$$

(2.72)　(3.90)　　(2.70)

を表している．これらの結果を見るかぎり，回帰のあてはまりなどの問題はなさそうである．しかし，残差をプロットしてみると**図 6-2**（84 ページ）のようになり，回帰直線の残差（・ 表示）も回帰曲線の残差（＋ 表示）も，ともに分散不均一を示唆している．そこで，次の例題を考えよう．

〔**例題 6.1**〕**付録 1** の都道府県別データにおいて，行政投資額 y を人口 x で説明する単回帰モデルの推定において，ウェイトとして $w_i = x_i$ を使って WLSE を求めよ．また，同一のウェイトを使って，y を x と x^2 で説明する重回帰モデルの WLSE を求め，両者を比較せよ．

（解）まず，単回帰モデルに対しては，(3) で $w_i = x_i$ として，

$$\frac{y_i}{x_i} = \frac{\alpha}{x_i} + \beta + \frac{\varepsilon_i}{x_i}$$

と変換される．同様にして，重回帰モデルに対しては

$$\frac{y_i}{x_i} = \frac{\alpha}{x_i} + \beta + \gamma x_i + \frac{\varepsilon_i}{x_i}$$

のように表すことができる．$x' = 1/x$，$y' = y/x$ とおいて，これら 2 つのモデルは次のように推定される．

$$y' = 0.354 + 0.126x' \qquad R^2 = 0.322, \overline{R}^2 = 0.307, \hat{\sigma} = 0.067$$

(18.04)（4.62)

$$y' = 0.316 + 0.159x' + 0.00652x \qquad R^2 = 0.339, \overline{R}^2 = 0.309, \hat{\sigma} = 0.067$$

(7.60)（3.79)　(1.05)

図 6-3（**加重回帰による残差**）は，これら 2 つの回帰から得られる残差を x（人口）に対してプロットしたものである．今度は，単回帰の残差（点表示）にも重回帰の残差（＋表示）にも，分散不均一を示唆するようなパターンは見られない．また，両者のパターンは非常に似ており，自由度修正済決定係数の観点からも，モデルとしてより単純な前者を選択しても差し支えない，と結論することができる．

応用例 II

分散不均一が必然的に起きる例として，第 3 章で説明したロジット変換を含む回帰モデルを導出しよう．今，ある属性の有無の割合が何らかの水準とともに単調に変動しているものとする．例えば，持ち家世帯の割合が世帯主の年令とともに増加する場合，自家用車の保有率が所得とともに増加する状況などが考えられる．そこで，第 i 水準 x_i において，ある属性をもつものの母集団における比率を p_i，標本における割合を $\hat{p}_i$ とする．ここで，p_i は x_i とともに単調に変化するから，その関係がロジスティック関数

$$p_i = \exp(\alpha+\beta x_i) \Big/ \left[1+\exp(\alpha+\beta x_i)\right] \tag{5}$$

で表されるものとする．このとき，p_i のロジット変換は，

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha+\beta x_i \tag{6}$$

となる．また，$\hat{p}_i$ のロジット変換を

$$y_i = \log(\frac{\hat{p}_i}{1-\hat{p}_i}) \tag{7}$$

で定義することにしよう．このとき，テーラー展開により，

$$y_i = \log(\frac{\hat{p}_i}{1-\hat{p}_i}) \cong \log\left(\frac{p_i}{1-p_i}\right) + \frac{1}{p_i(1-p_i)}(\hat{p}_i - p_i)$$

と表すことができる．ここで，$n_i \times \hat{p}_i$ はパラメータ p_i の二項分布に従うから，$\hat{p}_i$ の分散は $p_i(1-p_i)/n_i$ となり，水準ごとに異なってしまうことになる．

以上のことから，分散不均一の単回帰モデル

$$y_i = \alpha+\beta x_i + \varepsilon_i \qquad \varepsilon_i = \frac{1}{p_i(1-p_i)}(\hat{p}_i - p_i) \tag{9}$$

が得られる．実際，誤差項の分散は

$$V(\varepsilon_i) = \frac{V(\hat{p}_i)}{\{p_i(1-p_i)\}^2} = \frac{1}{n_i p_i(1-p_i)}$$

となり，一般には低水準および高水準における値の方が高くなることがわかる．

(9) を**ロジット・モデル**という．ロジット・モデルは必然的に分散不均一モデルであり，パラメータを推定するためには WLSE を使うのが望ましいことになる．ウェイトとしては，$w_i = \sqrt{V(\varepsilon_i)} = 1\Big/\sqrt{n_i p_i(1-p_i)}$ を使うべきであるが，p_i は未知なので，推定量を使ったウェイト $\hat{w}_i = 1\Big/\sqrt{n_i \hat{p}_i(1-\hat{p}_i)}$ で代用するのが普通である．

〔**例題 6.2**〕**表 3-2 の**世帯主の年令階級（x）と持ち家率（p），月収総額の平均（z，単位：万円）のデータにおいて，持ち家率をロジット変換して，

（i）x で説明する単回帰

（ii）x と z で説明する重回帰

を，それぞれ WLSE を使って推定して，通常の LSE の結果と比較せよ．

（解）WLSE のウェイトとしては，上で述べた推定量 $\hat{w}_i = 1\Big/\sqrt{n_i \hat{p}_i(1-\hat{p}_i)}$ を使うことにする．

(i)　最小 2 乗法　$y = -1.811 + 0.524x$　$R^2 = 0.944, \bar{R}^2 = 0.933$

(-7.11) (9.19)

加重最小 2 乗法　$y/\hat{w} = -1.63/\hat{w} + 0.501x/\hat{w}$

(-5.86)　(8.04)

今の場合，加重最小 2 乗法は切片を 0 とみなして推定することになり，決定係数が本来の意味合いをもたないのでここには記してない．最小 2 乗法による結果と比べると，t 値が若干低くなっているが，分散不均一の仮定が正しければ最小 2 乗法から得られ

る t 値は本来の意味の t 値ではないので，比較自体は無意味である．

(ii) 最小 2 乗法　$y = -4.51 + 0.285x + 0.0347z$　$R^2 = 0.992, \overline{R}^2 = 0.989$

(-8.26) (5.38) (5.04)

加重最小 2 乗法　$y/\hat{w} = -4.43/\hat{w} + 0.278x/\hat{w} + 0.0344z/\hat{w}$

(-7.01)　(4.89)　(4.52)

図による比較：**図 6-4(i)，(ii)**

2　系列相関

時系列データの場合には，被説明変数の時点間の独立性や無相関性を仮定することが一般に困難になる．時間的な相関が存在するとき，**系列相関**があるという．

系列相関を明示的に取り込む一つの方法は，誤差項が**自己回帰（AR）モデル**

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \tag{10}$$

に従うものと仮定することである．(10) の場合には，p 時点前までの過去で表される AR モデルであることから，**AR(p) モデル**と呼ばれる．

以下では，単回帰モデルの誤差項を u_t として，u_t が AR(1) モデルに従うものとする．すなわち，

$$y_t = \alpha + \beta x_t + u_t \qquad u_t = \rho u_{t-1} + \varepsilon_t \qquad (t = 1, \cdots, n) \tag{11}$$

を考える．ここで，ρ（ロー）は未知のパラメータである．もちろん，$\rho = 0$ ならば古典的な場合に帰着するが，一般には ρ の絶対値は 1 より小さいものと仮定される．この条件は AR(1) モデルの**定常性**の条件と呼ばれ，このもとで，u_t の平均と分散は，

$$E(u_t) = 0 \qquad V(u_t) = \sigma^2 \Big/ \left(1 - \rho^2\right)$$

となり，各時点で同一の値となる．さらに，**自己共分散**は，

$$Cov(u_s, u_t) = \sigma^2 \rho^{|s-t|} \Big/ \left(1 - \rho^2\right)$$

で与えられ，時差のみに依存した値となる．$\rho = 0$ でない限り，u_t は相関をもつことになり，その値は時差が大きくなるにつれて指数的に減少していくこともわかる．

図 6-5： AR(1) モデルの実現例（$\rho = 0.1$ と $\rho = 0.8$）

> 系列相関を考慮した (11) の単回帰モデルにおいては，
> $E(y_t) = \alpha + \beta x_t \qquad Cov(y_s, y_t) = \sigma^2 \rho^{|s-t|} \Big/ \left(1 - \rho^2\right)$

系列相関があるにもかかわらず，それを無視した α と β の LSE は，分散不均一の場合と同様に，不偏ではあるが BLUE ではなくなる．しかし，最小 2 乗法の性質から，LSE で得られたモデルの方が決定係数が常に大きく，もっともらしいモデルが得られたかのような印象をもたらす．

ダービン＝ワトソン検定

系列相関を考慮した回帰モデル (11) の分析の出発点は，ρ が 0 でないかどうかを

検定することである．そのためには，右片側検定

$$H_0\!:\!\rho = 0 \qquad \text{vs} \qquad H_1\!:\!\rho > 0$$

を行うのが普通である．この検定には，**ダービン＝ワトソン統計量**

$$DW = \sum_{t=2}^{n}\left(e_t - e_{t-1}\right)^2 \Big/ \sum_{t=1}^{n} e_t^2 \tag{12}$$

が使われる．パラメータ ρ が $0 \le \rho < 1$ の範囲にあれば，$\hat{\rho}$ もほぼ同じ範囲にあると考えられるので，DW はほぼ 0 と 2 の間で分布して，帰無仮説のもとでは大きく，対立仮説のもとでは小さくなる．すなわち，DW が小さいときには系列相関の存在が示唆されることになる．

具体的には，ダービン＝ワトソン検定は次のように行う．

ダービン＝ワトソン検定の手続き

(i) 有意水準 γ を指定する．

(ii) (12) のダービン=ワトソン統計量 DW を計算する．

(iii) 有意水準 γ のもとで，説明変数の数 k と標本サイズ n に対応した 2 つの有意点 d_L と d_U を付表 6 から求めて，次のように判断する．

$DW > d_U$　　のとき H_0 を受容

$DW < d_L$　　のとき H_0 を棄却

$d_L \le DW \le d_U$　　のとき結論保留

結論保留の場合がある理由

ダービン＝ワトソン統計量の帰無仮説のもとでの分布は，データ数と説明変数の個数だけでなく，説明変数のとる値そのものに依存して変化する．この点で，t 統計量や F 統計量と異なる．そのようなすべての説明変数ごとに分布表を作成することは不可能であるから，ダービン＝ワトソン検定は，データ数と使われる説明変数の個数ごとに，統計量の下限と上限の分布を求め，これらの分布に基づいて受容，棄却の判断をする検定である．

具体的には，統計量の値が下限の分布の $100\gamma\%$ 点未満になると，真の分布においても $100\gamma\%$ 点未満になる．他方，統計量の値が上限の分布の $100\gamma\%$ 点を越えると，真の分布においても $100\gamma\%$ 点を越える．このような場合には，それぞれ受容と棄却の判断が正当化される．しかし，統計量の値が中間にあるときは，真の分布においてどこに位置するかが不確定であり，受容，棄却のいずれの判断もできないことになる．このような理由から，結論を保留せざるを得ない場合が生じる．

対立仮説が $H_1\!:\!\rho < 0$ の場合

経済データでは例外的であるが，対立仮説が $H_1\!:\!\rho < 0$ の場合には，検定統計量として $4 - DW$ を考えることにより，上述の手続きに従って行うことができる．

〔**例題 6.3**〕**付録** 4 の家計消費支出（y）と家計可処分所得（x）の実質値データに対して単回帰モデルをあてはめて，系列相関の検定を行え．

（解）単回帰モデルの推定結果は

$$y = -17.04 + 0.911x \qquad R^2 = 0.992, \bar{R}^2 = 0.991$$
$$(-4.51)\ (50.8) \qquad \hat{\sigma} = 4.09, DW = 0.345$$

である．決定係数や t 値は申し分ないが，DW の値は小さい．実際，**付表 6** から，$n = 24$，$k = 1$ の場合，有意水準 1% における d_L は 1.04 であるから系列相関がないという仮説は棄却される．**図 6-6** には残差がプロットされている．残差にパターンが見られるので，系列相関の存在はこの図からも読みとることができる．

系列相関がある場合の推定

ダービン＝ワトソン検定により系列相関が示唆される場合には，それを考慮した推定を考えねばならない．そのための一つの方法は，(11) を変形して

$$\begin{aligned} y_t - \rho y_{t-1} &= \alpha + \beta x_t + u_t - \rho(\alpha + \beta x_{t-1} + u_{t-1}) \\ &= \alpha(1-\rho) + \beta(x_t - \rho x_{t-1}) + \varepsilon_t \end{aligned} \qquad (14)$$

⇒

$$y_t^* = \alpha^* + \beta^* x_t^* + \varepsilon_t \qquad (15)$$

ただし，$y_t^* = y_t - \rho y_{t-1}$，$x_t^* = x_t - \rho x_{t-1}$，$\alpha^* = \alpha(1-\rho)$，$\beta^* = \beta$ である．一般には ρ は未知なので，その場合には (13) で定義した推定量 $\hat{\rho}$ でおきかえればよい．系列相関がある場合に，このようにして推定量を求める方法を**コクラン＝オーカット法**という．

〔**例題 6.4**〕例題 6.3 と同一のデータに対して，コクラン＝オーカット法により α と β の推定値を求めよ．また，変換後のモデルにおける残差をプロットせよ．

（解）パラメータ ρ の推定値として (13) から 0.760 が得られるので，これを (14) の ρ に代入して，変換後のモデル (15) は

$$y^* = -5.24 + 0.925x^* \qquad R^2 = 0.955, \bar{R}^2 = 0.953$$
$$(-2.11)\ (21.1) \qquad \hat{\sigma} = 2.23, DW = 1.467$$

のように推定される．ダービン＝ワトソン統計値が前よりもかなり大きく，系列相関は非常に弱いものとなっている．実際，$n = 23$，$k = 1$ の場合，有意水準 1% における d_U は 1.19，有意水準 5% における d_U は 1.44 であるから，系列相関がないという帰無仮説は受容される．なお，t 値の観点からは例題 6.3 で考えた系列相関を無視したモデルの方がもっともらしく見える．しかし，その場合にはすでに述べたように，t 値が本来の意味の t 値ではなくなるので有意性検定も正当性がなくなることに注意されたい．パラメータの推定値を変換することにより，もとのモデルは

$$y_i = -\frac{5.24}{1-0.76} + 0.925x_i = -21.83 + 0.925x_i \qquad \rho \text{ の推定値} = 0.760$$

と推定されることになる．**図 6-7** は実際のデータと推定された 2 つの回帰直線（実線：最小 2 乗法，破線：コクラン＝オーカット法）を図示したものである．また，**図**

6-8 はコクラン＝オーカット変換後のモデルの残差プロットである．**図 6-6** と異なり，特別のパターンが見られないことがわかる．

3　ラグ付き変数

時系列データの回帰分析では，系列相関を考慮すると同時に，説明変数として被説明変数の**ラグ**（遅れ）を追加する場合がある．例えば，

$$y_t = \beta_0 + \delta y_{t-1} + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + u_t \qquad u_t = \rho u_{t-1} + \varepsilon_t \tag{16}$$

を考えよう．ここで，1 期前のラグ付き変数 y_{t-1} の係数 δ については $|\delta|<1$ を仮定する．AR(1) に従う誤差項のパラメータ ρ についても $|\rho|<1$ とする．これらの条件は非定常な場合を排除するためのものである．

実際の分析では，まず ρ が 0 であるかどうかの検定を行う必要がある．そのためにはダービン＝ワトソン検定が考えられる．しかし，被説明変数のラグを含む (16) のようなモデルでは，ダービン＝ワトソン統計量の分布は，$\rho>0$ ならば右方向へ，$\rho<0$ ならば左方向へシフトすることが知られている．その結果，例えば ρ が正であっても検定統計値は 2 に近い値をとる傾向があり，系列相関の存在を見過ごす危険が増大することがわかっている．

上記の理由から，説明変数の中に被説明変数のラグが含まれる (16) のような回帰モデルにおいては，系列相関のための検定統計量としては，

$$h = \hat{\rho}\sqrt{\frac{n}{1-n\hat{V}(\hat{\delta})}} \tag{17}$$

を使うのが適切であるとされている．ここで，$\hat{\delta}$ は LSE，$\hat{\rho}$ は最小 2 乗残差に基づく ρ の推定量であり，$\hat{V}(\hat{\delta})$ は $\hat{\delta}$ の分散の推定量である．検定統計量 (17) はダービンの h **統計量**と呼ばれる．h 統計量に基づく検定は大標本検定であり，$\rho=0$ のもとで漸近的に標準正規分布に従うことが知られている．h が大きいときには正の系列相関が示唆される．

〔**例題 6.5**〕**付表 4** の年次マクロ・データにおいて，例題 5.1 と同様の輸入関数をコクラン=オーカット法で推定せよ．また，輸入額の 1 期前のラグを説明変数として追加したモデルも推定し，両者を比較せよ．

(解) 輸入額を y，GDE を x，交易条件指数を z として，モデルは例題 5.1 のモデルからは DW = 0.657 となるから，系列相関がないという仮説は棄却される．実際，**付表 6** より有意水準 1% のもとで，$n=24$，$k=2$ のとき $d_L=0.96$ である．そこで，系列相関を考慮するとパラメータ ρ の推定値として 0.645 が得られる．これより，変換後のモデルは

$$y^* = -0.836 + 1.053x_1^* + 0.0979x_2^* \qquad R^2 = 0.679, \bar{R}^2 = 0.647$$
$$(-2.22)\ \ (5.90)\qquad (0.61) \qquad\qquad \hat{\sigma} = 0.069, DW = 1.528$$

と推定される．ただし $y_t^* = \log y_t - 0.645\times\log y_{t-1}$ であり，x_{1t}^*, x_{2t}^* も同様に定義される．変換後のモデルでは DW の値が改良されたことがわかる．また，t 値の観点から

GDE の説明力は有意である．しかし，交易条件は有意でないことがわかる．

次に，ラグを導入したモデルは

$$y = -1.302 + 0.658 y_{-1} + 0.444 x_1 + 0.248 x_2 \qquad R^2 = 0.953, \overline{R}^2 = 0.946, F = 122.6$$
$$(-2.79)\ (4.87)\qquad (3.02)\qquad (2.89)\qquad s = 0.064, DW = 1.811, h = 0.333$$

のように推定される．変数はすべて原データの対数値である．$DW = 1.811$ と 2 に近い値であるが，説明変数として被説明変数のラグが入っているから統計値としては不適切である．しかし，h 統計値は 0.333 なので，系列相関がないという帰無仮説は受容される．このモデルでは，コクラン・オーカット変換によるモデルとは異なり各変数が有意となっている．交易条件も輸入変動の説明要因とする立場からは，このモデルの方が説明力が高く，より適切であると思われる．**図 6-9** には，例題 5.1 で推定したラグ導入前のモデルと，ここで考えたラグ導入後のモデルの残差をプロットしてある．前者にはパターンが見られるが，後者についてはパターンがかなり消えていることが読みとれる．

４　分布ラグ・モデル

本節ではラグ付き変数のバリエーションとして，説明変数の中に通常の説明変数のラグが入ったモデルを考えよう．例として，x_t を当期の説明変数とするとき，

$$\begin{aligned} y_t &= \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_m x_{t-m} + v_t \\ &= \alpha + \beta(L) x_t + v_t \end{aligned} \tag{18}$$

と表されるモデルを考えよう．ここで，v_t は誤差項であるが，必ずしも独立ではないものとする．また，L は**ラグ・オペレータ**と呼ばれ，

$$Lx_t = x_{t-1}, \quad L^2 x_t = L(Lx_t) = Lx_{t-1} = x_{t-2}, \ \cdot\cdot\cdot, \quad L^j x_t = x_{t-j}$$

のように，時点を戻す作用がある．ただし，定数に作用させても不変である．そして，

$$\beta(L) = \beta_0 + \beta_1 L + \cdots + \beta_m L^m$$

は，L の多項式である．ラグ・オペレータやその多項式を使う第一の理由は，(18) の 2 つの表現からわかるように表現を簡略にすることにある．第二の理由は，あとで示すように，種々の計算をする場合に非常に便利である，ということである．

ところで，(18) のモデルに含まれるパラメータ β_h（$h = 0,1,\cdots,m$）は h 期前の説明変数の限界性向を表しており，これらの値の大小により，当期の被説明変数に対して，いずれの期の影響が大きいかを相対的に知ることができる．このようなモデルを**分布ラグ・モデル**という．この場合，β_0 は即時的な限界性向を表すので**衝撃乗数**と呼ばれる．また，係数の部分和 $S_i = \beta_0 + \beta_1 + \cdots + \beta_i$ は**中間乗数**，総和 S_m は**全乗数**と呼ばれ，ともに限界性向に時間的経過を加味した概念である．

平均ラグ

分布ラグ・モデル (18) において，$\beta_0, \beta_1, \cdots, \beta_m$ がすべて正であるものとしよう．このとき，説明変数が被説明変数に影響を与えるまでの平均的なタイム・ラグを求めることができる．現実との対応では，ある経済政策の効果が現れるまでに要する平均的な期

間と解釈できる．この値を**平均ラグ**といい，(18) のモデルに対しては，

$$平均ラグ=\sum_{h=0}^{m}\frac{\beta_h}{\beta_0+\beta_1+\cdots+\beta_m}h=\sum_{h=1}^{m}\frac{\beta_h}{S_m}h \tag{19}$$

で定義される．ここで，ラグの値に関する次の確率分布を考えれば，平均ラグは，この確率分布の平均にほかならないことが了解されよう．

ラグ	0	1	・・・	h	・・・	m
確率	β_0/S_m	β_1/S_m	・・・	β_h/S_m	・・・	β_m/S_m

メディアン・ラグ

同様にして，**メディアン・ラグ**が定義される．それは，上の確率分布のメディアンであり，説明変数の影響の半分が現れるまでに要する期間としての意味合いをもつ．

アーモン・ラグ

(18) のような分布ラグ・モデルにおいては，ラグの長さ m が大きくなるとともに，推定すべきパラメータの数が増え，しかも多重共線の可能性も高まる．この危険を回避する一つの方法は，$\beta_0,\beta_1,\cdots,\beta_m$ がとる値に制約を課して，これらがある曲線，例えば 3 次関数

$$f(h)=\gamma_0+\gamma_1 h+\gamma_2 h^2+\gamma_3 h^3 \qquad (3<m)$$

上にのっていると想定して推定することである．こうすることによって，元のパラメータ $\beta_0,\beta_1,\cdots,\beta_m$ は $\gamma_0,\gamma_1,\gamma_2,\gamma_3$ で表され，(18) 式は

$$\begin{aligned} y_t=\alpha&+\gamma_0(x_t+x_{t-1}+\cdots+x_{t-m})+\gamma_1(x_{t-1}+2x_{t-2}+\cdots+mx_{t-m})\\ &+\gamma_2(x_{t-1}+4x_{t-2}+\cdots+m^2x_{t-m})+\gamma_3(x_{t-1}+8x_{t-2}+\cdots+m^3x_{t-m})+v_t \end{aligned} \tag{20}$$

と変形される．推定すべきパラメータの個数が減少し，同時に多重共線の危険を回避することが可能となることが了解されよう．このように，簡単な多項式の関係が想定されたラグの係数のことを，発案者にちなんで**アーモン・ラグ**と呼ぶ．アーモン・ラグの構造をもつモデル (20) に含まれるパラメータの LSE は，誤差項 v_t が独立，同一分布に従うならば BLUE となる．

分布ラグ・モデルは，前節で考えたモデル，すなわち，被説明変数のラグが通常の説明変数とともに説明変数として入ったモデルと密接な関係をもつことが示される．例として，

$$y_t=\alpha+\delta y_{t-1}+\beta x_t+\varepsilon_t, \qquad |\delta|<1 \tag{21}$$

を考えよう．ここで，ε_t は独立，同一分布に従う誤差項である．このモデルは，$\beta=0$ ならば平均が 0 でない AR モデルであるが，それに説明変数が加わった形をしていることから，**ARX モデル**と呼ばれることもある．ARX の X は Exogenous（外生的）の x を取り出したものである．

さて，ARX モデル (21) を y に関して解いてみよう．そのためには，ラグ・オペレータを使うのが便利である．まず，(21) は，

$$(1-\delta L)y_t=\alpha+\beta x_t+\varepsilon_t$$

と表すことができる．この表現を y に関して解くことができ，

$$y_t = \frac{\alpha}{1-\delta L} + \frac{\beta}{1-\delta L} x_t + \frac{1}{1-\delta L}\varepsilon_t$$
$$= \alpha^* + \sum_{j=0}^{\infty} \gamma^*_j x_{t-j} + v_t \qquad (22)$$

となる．ここで，$\alpha^* = \alpha/(1-\delta)$，$\gamma^*_j = \beta\delta^j$ であり，v_t は AR(1) モデル

$$v_t = \delta v_{t-1} + \varepsilon_t$$

に従っている．

したがって，通常の説明変数 x_t 以外に，被説明変数 y_t の 1 期前のラグを含むモデル (21) は，x_t の無限のラグをもつ分布ラグ・モデル (22) の形にも表現されることがわかった．この場合のラグの係数 $\gamma^*_j = \beta\delta^j$ は，ラグの長さとともに幾何級数的に減衰する．このようなラグ構造を**コイック・ラグ**という．しかしながら，現象を同程度に記述できるモデルならば，より単純の方がよいというモデル選択の基準からは，当然ながら，(22) よりも (21) の方がよりよいモデルである．

以上のような理由から，統計的推測においては，分布ラグ・モデルよりも，ARX モデルのように被説明変数のラグを含む時系列モデルの方がよく使われている．ただし，(22) の表現も乗数や平均ラグなどの概念を考える際には必要である．第 8 章のトピックである同時方程式モデルにおいては，分布ラグ・モデル (22) は ARX モデル (21) の**最終形**と呼ばれている．

5　期待を含むモデル

企業や家計の行動理論においては，予想価格や期待インフレ率などのように，将来のある時点で実現する値に関する予想や期待を変数として含む形でモデルが構成される場合がある．以下では，そのようなモデルの代表的な 3 つの例について述べることにする．

適応的期待（Adaptive Expectations）**モデル**

単回帰モデル

$$y_t = \alpha + \beta x^*_t + \varepsilon_t \qquad (23)$$

を考えよう．ここで，誤差項 ε_t は独立，同一分布に従うものとする．そして，y_t はある財の時点 t における供給量，x^*_t は時点 t において形成されるその財の予想価格であり，予想が

$$x^*_t = \lambda x^*_{t-1} + (1-\lambda)x_t = \frac{1-\lambda}{1-\lambda L}x_t, \qquad 0<\lambda<1 \qquad (24)$$

に従って形成されるものとする．x_t は現実の価格である．このとき，(24) を (23) に代入すれば，結局の所，

$$y_t = \alpha(1-\lambda) + \lambda y_{t-1} + \beta(1-\lambda)x_t + \varepsilon_t - \lambda\varepsilon_{t-1} \qquad (25)$$

を得る．このモデルを**適応的期待モデル**と呼ぶ．

モデル (25) は，前節で考えた ARX モデルのバリエーションである．ARX モデルにおいては誤差項が独立であったが，

$$v_t = \varepsilon_t - \lambda\varepsilon_{t-1} \tag{26}$$

はもはや独立ではない．(26) も第 9 章で議論する時系列モデルの 1 つであり，次数 1 の**移動平均モデル**，略して MA(1) モデルと呼ばれる．MA は Moving Average の略である．

以上のことから，モデル (25) は ARMAX モデルと呼ばれる．すなわち，適応的期待モデルとは，ARMAX モデルにほかならない．推定の観点からは，(26) で定義された誤差項 v_t が明らかに y_{t-1} と相関をもつので，LSE は一致性をもたない．この場合の推定は，MLE を使うか，あるいは次章で説明する操作変数推定量を使うのが望ましい．

部分調整（Partial Adjustment）**モデル**

時点 t におけるある企業の実際の製品在庫量を y_t，望ましい在庫量を y_t^* とする．ここで，在庫調整は

$$y_t - y_{t-1} = \delta(y_t^* - y_{t-1}) + \varepsilon_t, \qquad 0 < \delta < 1 \tag{27}$$

に従って行われるものとする．誤差項 ε_t は独立，同一分布に従う確率変数である．y_t^* が観測可能ならば，これは ARX モデルと同様なので，統計的推測の観点からは特に問題はない．しかし，y_t^* は観測不可能であり，時点 t の売上高 x_t に依存して，

$$y_t^* = \alpha + \beta x_t \tag{28}$$

により決められているものとする．以上より，(28) を (27) に代入することにより，

$$y_t = \alpha\delta + (1-\delta)y_{t-1} + \delta\beta x_t + \varepsilon_t \tag{29}$$

が得られる．モデル (29) を**部分調整モデル**と呼ぶ．部分調整モデルは ARX モデルにほかならない．適応的期待モデル（＝ARMAX モデル）の誤差項が独立でなく，y_{t-1} と相関しているのに対して，部分調整モデル（＝ARX モデル）の誤差項は独立である．そして，y_{t-1} とは無相関となるので，LSE も望ましい性質をもつことになる．

合理的期待（Rational Expectations）**モデル**

ここで導入される期待は，確率論的な意味での予測量である．今，ある財の時点 t における価格を p_t，時点 $t-1$ において p_t を予測するために利用可能な情報の集まりを I_{t-1} とする．このとき，p_t の最適な予測量は，

$$p_t^e = E\left(p_t | I_{t-1}\right)$$

で与えられる．この期待値を p_t の**合理的期待**という．合理的期待とは，利用可能な情報を与えたときの条件付き期待値にほかならず，合理的期待を含むモデルを**合理的期待モデル**という．

合理的期待に関しては，次の性質が成り立つ．

合理的期待の性質
（i）予測誤差 $v_t^e = p_t - p_t^e$ は無相関
（ii）合理的期待 p_t^e と予測誤差 v_t^e は無相関
（iii）$V(p_t) \geq V(p_t^e)$

しかし，理論上はともかく，合理的期待を計算することは一般に困難である．したがって，合理的期待モデルの推定も困難をともなう．実行可能な 1 つの方法は，合理的期待を実現値でおきかえて LSE を求めるものであるが，この方法は好ましくない．このことを示すために，次の例題を考えよう．

〔**例題 6.6**〕実質貨幣需要に関するモデルが，

$$y_t = \alpha + \beta(p^e_{t+1} - p_t) + \varepsilon_t \tag{30}$$

で表されるものとする．ここで，y_t は実質貨幣需要の対数，p_t は物価水準の対数，p^e_{t+1} は時点 t において計算される p_{t+1} の合理的期待である．また，ε_t は独立，同一分布に従う誤差項であり，p_t および p^e_{t+1} と独立である．このモデルにおいて，p^e_{t+1} の代りに p_{t+1} を使って求められる α と β の LSE の性質を議論せよ．

（解）推定しようとするモデルは，予測誤差を v^e_{t+1} として，$p^e_{t+1} = p_{t+1} - v^e_{t+1}$ を (30) に代入することにより，

$$y_t = \alpha + \beta(p_{t+1} - p_t) + \varepsilon_t - \beta v^e_{t+1}$$

となる．このとき，α と β の LSE は一致性を失う．なぜなら，p_{t+1} と v^e_{t+1} は相関しているので，説明変数 $p_{t+1} - p_t$ と誤差項 $\varepsilon_t - \beta v^e_{t+1}$ も無相関ではないからである．一致性を保証する推定としては，次章で説明する操作変数推定法がある．