



慶應義塾大学ビジネス・スクール

回帰分析シリーズ 5

— 回帰分析の制約 —

今までの統計ノートでは、いくつかの仮定の下での回帰分析を紹介してきたが、現実にはこれらの仮定をすべて満たすことはまずあり得ない。したがって回帰分析を用いて現実の世界をモデル化しようとする、これらの仮定のいくつかを緩めたり修正しなければ、誤った結果を導き出すことも十分に考えられる。このノートでは、修士論文などで陥りやすい回帰分析の限界や制約についてまとめている。それらは、誤った回帰モデルの構築についてのスペック・エラー、説明変数間の相関関係についての多重共線性、誤差項の分散不均一性、そして誤差項の自己相関という4つのトピックである。

1：スペック・エラー

これまで、回帰モデルには必要な変数がすべて含まれており、また不必要な変数が一切除外されていることを仮定してきた。しかし現実には回帰モデルを構築する場合、このような仮定が成立しない方がはるかに多い。

たとえば真のモデルが $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ であるとしよう。ここで $E(\mathbf{u}) = \mathbf{0}$ 、また $E(\mathbf{u}\mathbf{u}') = \sigma^2\mathbf{I}$ と仮定すると、 $\boldsymbol{\beta}$ の不偏推定値 \mathbf{b} は、 $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ によって求めることが可能であった。しかし、もし分析者が誤ったモデル $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ を構築してしまったならば、 \mathbf{b} は $(\mathbf{X}\cdot'\mathbf{X}\cdot)^{-1}\mathbf{X}\cdot'\mathbf{y}$ によって推定されることになる。ここで $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ を代入すると、 \mathbf{b} は、

$$\begin{aligned}\mathbf{b}\cdot &= (\mathbf{X}\cdot'\mathbf{X}\cdot)^{-1}\mathbf{X}\cdot'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}\cdot'\mathbf{X}\cdot)^{-1}\mathbf{X}\cdot'\mathbf{u} \\ &= (\mathbf{X}\cdot'\mathbf{X}\cdot)^{-1}\mathbf{X}\cdot'\mathbf{X}\boldsymbol{\beta} \quad \therefore E(\mathbf{u}) = \mathbf{0}\end{aligned}\tag{1}$$

となる。これは誤ったモデルによって推定された回帰係数が、もはや真のパラメータではないことを示すものである。このように行列 \mathbf{X} の失敗のことをスペック・エラーという。このようなスペック・エラーには、(1) 必要な変数をモデルから除外してしまったケースと、(2) 不必要な変数を組み込んだケースとが考えられる。ここでは、その2つのケースを解説することにより、スペック・エラーが推定にどのような影響をもたらすか示す。

1.1 : 必要な変数の除外

ここで真の説明変数の行列を \mathbf{X} 、必要な変数の除外した行列を \mathbf{X}_\bullet とし、それぞれ次のような変数を含んでいるものとする。

$$\text{(真)} \quad \mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2 \ \cdots \ \mathbf{X}_r; \mathbf{X}_2 \ \cdots \ \mathbf{X}_r] = [\mathbf{x}_1 \ \mathbf{x}_2]$$

$$\text{(誤)} \quad \mathbf{X}_\bullet = [\mathbf{X}_1 \ \mathbf{X}_2 \ \cdots \ \mathbf{X}_r] = \mathbf{x}_1$$

したがって回帰モデルはそれぞれ次のように表すことができる。

$$\text{(真)} \quad \mathbf{y} = \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \mathbf{u} \quad (2)$$

$$\text{(誤)} \quad \mathbf{y} = \mathbf{x}_1\beta_1 + \mathbf{u} \quad (3)$$

ここで \mathbf{X} は r を含む k の説明変数から構成され、 \mathbf{X}_\bullet は r だけを含んでいるものとする。誤ったモデルの β の推定値は、

$$\mathbf{b}_1 = (\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{y} \quad (4)$$

である。ここで (4) 式の \mathbf{y} に (2) 式を代入すると

$$\begin{aligned} \mathbf{b}_1 &= (\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'(\mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \mathbf{u}) \\ &= \beta_1 + (\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{x}_2\beta_2 + (\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{u} \\ &= \beta_1 + (\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{x}_2\beta_2 \quad \therefore E(\mathbf{x}_1'\mathbf{u}) = 0 \end{aligned} \quad (5)$$

と定義できる。ここで $(\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{x}_2$ は、 x_1 対する x_2 の回帰係数である。したがって誤った回帰モデルが真の説明変数をすべて含んでいたとしても、不偏推定値とはなり得ないのである。このような不必要な変数のモデルへの組み込みは、分散にも影響を与える。「回帰分析シリーズ 3 - 行列 -」において、真の行列 \mathbf{X}_1 を \mathbf{Y} について回帰させると、その残差 \mathbf{e} は次のようになることを述べた。

$$\begin{aligned} \mathbf{e} &= \mathbf{y} - \mathbf{x}_1\mathbf{b} \\ &= \mathbf{y} - \mathbf{x}_1(\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{y} \\ &= [\mathbf{I} - \mathbf{x}_1(\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1']\mathbf{y} \\ &= \mathbf{M}\mathbf{y} \end{aligned} \quad (6)$$

そして \mathbf{M} は巾等行列であることが分かっているので、残差平方和は、

$$\text{RSS} = \mathbf{e}'\mathbf{e} = \mathbf{y}'\mathbf{M}\mathbf{y} \quad (7)$$

である。この (7) 式に (2) 式を代入すると、

$$\begin{aligned}
\mathbf{y}'\mathbf{M}_1\mathbf{y} &= (\mathbf{x}_1\boldsymbol{\beta}_1 + \mathbf{x}_2\boldsymbol{\beta}_2 + \mathbf{u})'\mathbf{M}_1(\mathbf{x}_1\boldsymbol{\beta}_1 + \mathbf{x}_2\boldsymbol{\beta}_2 + \mathbf{u}) \\
&= (\mathbf{x}_2\boldsymbol{\beta}_2 + \mathbf{u})'\mathbf{M}_1(\mathbf{x}_2\boldsymbol{\beta}_2 + \mathbf{u}) \quad \therefore \mathbf{M}_1\mathbf{x}_1 = [\mathbf{I} - \mathbf{x}_1(\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1']\mathbf{x}_1 = \mathbf{0} \\
&= \mathbf{u}'\mathbf{M}_1\mathbf{u} + \boldsymbol{\beta}_2'\mathbf{x}_2'\mathbf{M}_1\mathbf{x}_2\boldsymbol{\beta}_2 + 2\boldsymbol{\beta}_2'\mathbf{x}_2'\mathbf{M}_1\mathbf{u}
\end{aligned} \tag{8}$$

ここで(8)式の期待値をとると、

$$\begin{aligned}
E(RSS) &= E(\mathbf{u}'\mathbf{M}_1\mathbf{u}) + \boldsymbol{\beta}_2'\mathbf{x}_2'\mathbf{M}_1\mathbf{x}_2\boldsymbol{\beta}_2 \\
&= \sigma^2(n-r) + \boldsymbol{\beta}_2'\mathbf{x}_2'\mathbf{M}_1\mathbf{x}_2\boldsymbol{\beta}_2
\end{aligned} \tag{9}$$

または、

$$\begin{aligned}
E\left(\frac{RSS}{n-r}\right) &= E(\mathbf{u}'\mathbf{M}_1\mathbf{u}) + \boldsymbol{\beta}_2'\mathbf{x}_2'\mathbf{M}_1\mathbf{x}_2\boldsymbol{\beta}_2 \\
&= \sigma^2 + \frac{1}{n-r} \boldsymbol{\beta}_2'\mathbf{x}_2'\mathbf{M}_1\mathbf{x}_2\boldsymbol{\beta}_2 > \sigma^2
\end{aligned} \tag{10}$$

となり、分散 σ^2 を最小化しない。

1.2 : 不必要な変数の採用

次に不必要な変数をモデルに含んでしまうという逆のケースを想定してみよう。この場合には、真のモデルと誤ったモデルを次のように定義する。

$$(真) \quad \mathbf{y} = \mathbf{x}_1\boldsymbol{\beta}_1 + \mathbf{u}$$

$$(誤) \quad \mathbf{y} = \mathbf{x}_1\boldsymbol{\beta}_1 + \mathbf{x}_2\boldsymbol{\beta}_2 + \mathbf{u}$$

誤ったモデルの $\boldsymbol{\beta}$ を最小二乗法で推定すると、

$$\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \mathbf{x}_1 \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_k \boldsymbol{\beta}_1 \\ \mathbf{0} \end{bmatrix} \tag{11}$$

これは推定値 b_1 と b_2 がともに不偏推定値をとることを示すものである。次に分散については、

$$\begin{aligned}
\mathbf{y}'\mathbf{M}\mathbf{y} &= (\mathbf{x}_1\boldsymbol{\beta}_1 + \mathbf{u})'\mathbf{M}(\mathbf{x}_1\boldsymbol{\beta}_1 + \mathbf{u}) \\
&= \mathbf{u}'\mathbf{M}\mathbf{u} + \boldsymbol{\beta}_1'\mathbf{x}_1'\mathbf{M}\mathbf{x}_1\boldsymbol{\beta}_1 + 2\boldsymbol{\beta}_1'\mathbf{x}_1'\mathbf{M}\mathbf{u} \\
&= \mathbf{u}'\mathbf{M}\mathbf{u} \quad \therefore \mathbf{M}\mathbf{x}_1 = \mathbf{0}
\end{aligned} \tag{12}$$

ここで期待値をとると、

$$\begin{aligned}
E(RSS) &= \sigma^2 \text{tr}(\mathbf{u}'\mathbf{M}\mathbf{u}) \\
&= \sigma^2 \text{tr}(\mathbf{M}) \\
&= (n-k) \sigma^2
\end{aligned} \tag{13}$$

これらの結果から、必要な変数を除外するよりも不必要な変数を採用する方が好ましく思われるかも知れない。しかしこれらの議論は、 x_1 と x_2 との相関が存在しないことを仮定している。もし真のモデルと不必要な変数を含んだモデルとに相関が存在するならば、(13)の分散は真のモデルの分散よりも大きくなる。たとえば、

$$\text{(真)} \quad y = \beta_1 x_1 + u \quad \text{(誤)} \quad y = \beta_1 x_1 + \beta_2 x_2 + u$$

とすると、不必要な変数を含んだモデルの β_1 の分散は、

$$\text{Var}(b_1) = \frac{\sigma^2}{\sum x_2^2 (1 - r_{12}^2)}$$

となり、これは真のモデルの分散

$$\text{Var}(b_1) = \frac{\sigma^2}{\sum x_1^2}$$

よりも大きくなる。したがってスペック・エラーのどちらが好ましいかという問題について、一概に解答を出すことはできないのである。

2：多重共線性

多重共線性 (multicollinearity) とは、重回帰モデルにおいて説明変数間に高い相関が存在することによって、推定値の精度が著しく損なわれる問題に言及するものである¹⁾。たとえば次のような2つの説明変数をもつ回帰モデルを例にして、多重共線性の影響を考えてみよう。

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u \quad (14)$$

(14) 式の偏差型は、

$$y = \beta_2 x_2 + \beta_3 x_3 + u \quad \therefore \bar{u} = 0 \quad (15)$$

である。ここで推定値 b_1 と b_2 の分散は、「回帰分析シリーズ3 - 重回帰分析 -」の(50)式によって与えられる。

¹⁾ 多重共線性は説明変数間の高い相関というよりも、その共分散と標準偏差が近似している場合に問題となると指摘されることもある (マダラ著「計量経済分析の方法」1992年, マグロウヒル)。つまり相関係数は共分散を標準偏差で割ったものとして計算されるため、相関の高さは共分散と標準偏差が近似した結果ということを強調する。

$$r_{23} = \frac{\text{COV}(x_2, x_3)}{\sqrt{\text{Var}(x_2)} \sqrt{\text{Var}(x_3)}}$$

$$\text{Var}(b_2) = \frac{\sigma^2 \sum x_3^2}{\sum x_2^2 \sum x_3^2 - (\sum x_2 \sum x_3)^2} = \frac{\sigma^2}{(1 - r_{23}^2) \sum x_2^2}$$

および、

$$\text{Var}(b_3) = \frac{\sigma^2 \sum x_2^2}{\sum x_2^2 \sum x_3^2 - (\sum x_2 \sum x_3)^2} = \frac{\sigma^2}{(1 - r_{23}^2) \sum x_3^2}$$

ここで r_{23}^2 は説明変数 X_2 と X_3 との重相関係数である。上式が意味するのは、 r_{23}^2 が高くなればなるほど b_1 と b_2 の分散は大きくなるということである。さらに、 b_1 と b_2 の推定値も r_{23}^2 に敏感に反応することにより精度が著しく低下するのである。

このような多重共線性の影響を例題を用いて考察することにしよう。表1は1977年から1994年までの新日本製鐵の財務指標であるが、分析者が新日本製鐵の自己資本利益率は当座比率、売上高経常利益率、総資産利益率、資産回転率という4つの指標に依存するという仮説を構築し、この仮説を重回帰モデルによって検証するものとしよう。

(表1) 新日本製鐵の会計指標

	当座比率	売上高 経常利益率	総資産利益率	資産回転率	自己資本利益率
n	(X2)	(X3)	(X4)	(X5)	(Y)
1	60.41	-0.85	-0.94	0.66	-3.56
2	60.40	1.22	0.14	0.70	0.53
3	60.09	3.81	2.47	0.78	9.22
4	72.55	6.17	2.01	0.78	7.66
5	81.56	7.86	3.06	0.81	13.02
6	54.16	6.72	2.03	0.77	10.22
7	55.55	2.90	0.98	0.66	5.36
8	56.18	-0.58	-0.38	0.63	-2.17
9	48.27	1.34	1.02	0.75	5.94
10	48.74	3.17	1.14	0.78	6.84
11	50.52	0.22	0.08	0.73	0.49
12	55.18	0.45	-0.17	0.75	-0.96
13	61.89	3.68	1.02	0.88	5.44
14	59.27	5.62	2.10	0.92	13.09
15	58.17	6.41	3.17	0.85	21.61
16	58.01	3.40	1.33	0.71	10.21
17	54.70	0.31	0.46	0.67	3.70
18	60.17	2.36	0.86	0.75	6.72

この回帰モデルを統計処理した結果は表2に記載している。もちろんこのような例は極端なものである。しかし分析者がよほど注意してモデルを構築しないと、同じような誤りを行うことは今までの修士論文発表などから容易に指摘できる。

まずここで気づくことは、各変数の t 値が小さいにもかかわらず決定係数が極めて高いことである。またサンプル数を18から13に減らして再度回帰分析をおこなった結果、各回帰係数の値や符号が変わったり、 t 値が大きくなっていることが示されている。このように多重共線性をもつ回帰モデルは、サンプルによって敏感に反応することが指摘でき、したがって導き出された数値に対する信頼性が極めて小さくなる。このように多重共線性の存在は回帰分析に対して好ましくない影響を与える。

多重共線性に関連して、その存在が推定値に対してどのような影響を与えるかについて、分散拡大因子 (variance inflation factor: VIF) や条件数 (condition number)、サイルの共線性テスト、行列分解などといった尺度が提唱されている。また多重共線性を取り除く方法としても、リッジ回帰や変数型の変更、主成分回帰、VIF の大きな変数の除去などが提唱されているが、これらについてはこのノートでは解説しない。ただ多重共線性が発生しているときには、次のような傾向が見受けられることを指摘しておく。

- (1) 標準誤差が大きい。
- (2) t 値が低いのに決定係数が高い。
- (3) サンプル数が少ないのに決定係数が高い。
- (4) サンプル数を少し変えただけで各変数の t 値が大きく変動する。

(表2) 多重共線性をもつ回帰分析の結果

サンプル数 = 18

DEP VAR: Y		N: 18			
MULTIPLE R: 0.949					
SQUARED MULTIPLE R: 0.901					
ADJUSTED SQUARED MULTIPLE R: 0.871					
STANDARD ERROR OF ESTIMATE: 0.966					
VARIABLE	COEFFICIENT	STD ERROR	STD COEF	T	P(2 TAIL)
CONSTANT	-6.048	3.531	0.000	-1.713	0.110
X2	0.062	0.037	0.184	1.696	0.114
X3	0.017	0.124	0.039	0.134	0.895
X4	1.670	0.708	0.725	2.360	0.035
X5	4.538	4.150	0.132	1.094	0.294
ANALYSIS OF VARIANCE					
SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	110.614	4	27.653	29.612	.201779E-05
RESIDUAL	12.140	13	0.934		

サンプル数=13

DEP VAR: Y N: 13					
MULTIPLE R: 0.978					
SQUARED MULTIPLE R: 0.956					
ADJUSTED SQUARED MULTIPLE R: 0.871					
STANDARD ERROR OF ESTIMATE: 0.618					
VARIABLE	COEFFICIENT	STD ERROR	STD COEF	T	P (2 TAIL)
CONSTANT	-5.899	2.755	0.000	-2.141	0.065
X2	0.111	0.048	0.196	2.334	0.048
X3	-0.570	0.162	-1.490	-3.523	0.008
X4	5.761	1.703	2.340	5.367	.67E-03
X5	0.321	2.913	0.011	0.110	0.915
ANALYSIS OF VARIANCE					
SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	66.696	4	16.674	43.698	.177013E-04
RESIDUAL	3.053	8	0.382		

相関係数

PEARSON CORRELATION MATRIX					
	X2	X3	X4	X5	Y
X2	1.000				
X3	0.544	1.000			
X4	0.435	0.931	1.000		
X5	0.261	0.699	0.679	1.000	
Y	0.271	0.860	0.940	0.680	1.000
NUMBER OF OBSERVATIONS:		18			

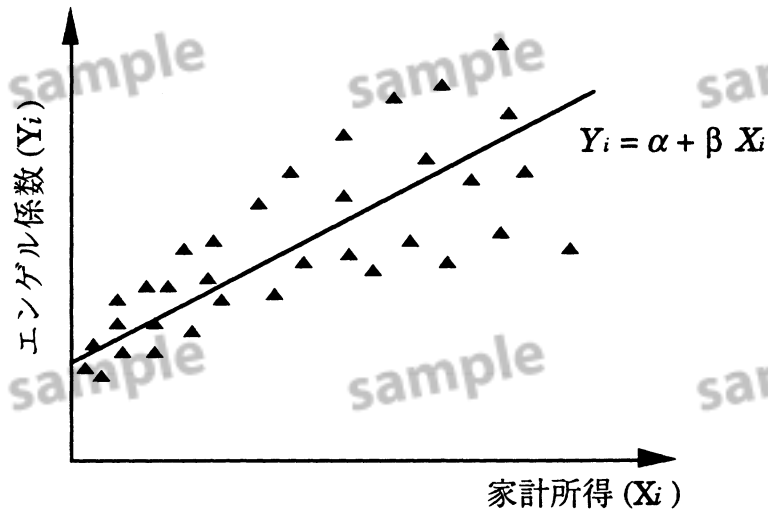
3：分散の不均一性

回帰分析の制約の一つとして分散の不均一性 (heteroscedasticity) がある。既に最小二乗法を用いて推定線を導き出すには、いくつかの仮定が必要であることは述べた。その仮定の一つは、(16) 式のように誤差項 u の分散に均一性を求めている。

$$E(\mathbf{uu}') = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} \quad (16)$$

しかし実際にデータを収集して回帰分析を行うとき、特にクロス・セクションのデータを取り扱うときには、分散が均一ではない状況が多く見られる。たとえばエンゲル係数を家計所得によって説明しようとする場合、所得が高くなるにつれてエンゲル係数のバラツキが大きくなることが指摘されている。このような状況を図1に示しているが、そのときの分散は(17)式のような構造になる。

(図1) エンゲル係数と家計所得の関係



$$E(\mathbf{uu}') = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix} = \Phi = \sigma^2 \Omega \quad (17)$$

このような分散が不均一な状況でOLSを適用すれば、回帰係数の推定値は不偏であるものの、精度に関連する分散はもはや最良ではなくなる。この章では、(1) 分散が不均一な状況でOLSを適用した場合に想定される影響、(2) その検出方法、(3) 解決方法の3点について解説する。

まずここでのモデルを次のように定義する。

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (18)$$

この(18)式にOLSを適用すると推定値 \mathbf{b} は、

$$\mathbf{b} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \quad (19)$$

である。さらに誤差項の平均値がゼロで、説明変数 \mathbf{X} と \mathbf{u} の相関がないものと仮定すると、(19)式の期待値は(20)式ようになる。つまり分散の不均一性の下でOLSを適用しても、その推定値は不偏であることが分かる。

$$E(\mathbf{b}) = \boldsymbol{\beta} \quad \therefore E(\mathbf{u}) = \mathbf{0}, E(\mathbf{X}'\mathbf{u}) = \mathbf{0} \quad (20)$$

次に推定値 \mathbf{b} の分散を表せば、

$$\begin{aligned} \text{Var}(\mathbf{b}) &= E[(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})'] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{uu}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \quad (21)$$

となる。これは、もはや分散が均一の仮定において求められた $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ によって、 \mathbf{b} の分散を測定することができないことを示すものである。したがって分散が不均一な場合には OLS を適用することはできない。このような不都合を解消しようとする一つの手法が、「一般化最小二乗法 (generalized least squares: GLS)」と呼ばれるものである。

いま (18) 式のモデルの両辺に $(n \times n)$ の行列 \mathbf{T} を掛けると、

$$\mathbf{T}\mathbf{y} = (\mathbf{T}\mathbf{X})\boldsymbol{\beta} + \mathbf{T}\mathbf{u} \quad (22)$$

そしてこの $\mathbf{T}\mathbf{u}$ の分散を求めると、

$$E(\mathbf{T}\mathbf{u}\mathbf{u}'\mathbf{T}') = \sigma^2 \mathbf{T}\boldsymbol{\Omega}\mathbf{T}' \quad (23)$$

もし $\mathbf{T}\boldsymbol{\Omega}\mathbf{T}' = \mathbf{I}$ となるような \mathbf{T} を見つけることができれば、OLS による推定が可能になる。ここで (22) 式に OLS を適用すれば、

$$\mathbf{b}\cdot = (\mathbf{X}'\mathbf{T}'\mathbf{T}\mathbf{X})^{-1}\mathbf{X}'\mathbf{T}'\mathbf{T}\mathbf{y} \quad (24)$$

さらに $\mathbf{T}\boldsymbol{\Omega}\mathbf{T}' = \mathbf{I}$ により $\mathbf{T}'\mathbf{T} = \boldsymbol{\Omega}^{-1}$ なので、これを (24) 式に代入すると、

$$\mathbf{b}\cdot = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} \quad (25)$$

となり、また分散共分散行列は、

$$\text{Var}(\mathbf{b}\cdot) = \sigma^2(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} \quad (26)$$

である。(26) 式を OLS による分散 (21) 式と比較すると、

$$\frac{\text{Var}(\mathbf{b}\cdot)}{\text{Var}(\mathbf{b})} = \frac{(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}}{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}} = \frac{(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})}{(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X})}$$

となり、これは 1 より小さいことが分かっている。

もし $\boldsymbol{\Omega}$ がすでに知られており、分散にどのような定数を掛けるかが分かっているものとする、 $\boldsymbol{\Omega}$ は (27) 式のように示すことができる (もちろん σ^2 は依然として未知である)。

$$\boldsymbol{\Omega} = \begin{pmatrix} \frac{1}{n_1} & 0 & \dots & 0 \\ 0 & \frac{1}{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{n_n} \end{pmatrix} \quad (27)$$

したがってその逆行列は、

$$\Omega^{-1} = \begin{pmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_n \end{pmatrix} \quad (28)$$

となる。また $\mathbf{T}'\mathbf{T} = \Omega^{-1}$ により変換行列 \mathbf{T} は次のようになる。

$$\mathbf{T} = \begin{pmatrix} \sqrt{n_1} & 0 & \cdots & 0 \\ 0 & \sqrt{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{n_n} \end{pmatrix} \quad (29)$$

3.1 : 不均一分散の検出

では分散の不均一性はどのように検出されるのだろうか。ここでは Goldfeld=Quandt の検定、Breusch=Pagan の検定、Glesjer の検定という3つの代表的な検定法方を紹介する。

▼ Goldfeld=Quandt の検定

もしサンプル数が少なく、また分散の不均一性を引き起こしている変数が1つの場合、Goldfeld=Quandt の検定方法を使用することができる。この検定方法の基本的な仮定は、サンプルは分散が異なる2つのグループに区分でき、それぞれのグループ内の分散は等しくなるというものである。そこでの帰無仮説と対立仮説は次のようになる。

$$H_0: \sigma_1^2 = \sigma_2^2 \quad H_1: \sigma_1^2 < \sigma_2^2$$

具体的な手続きは以下のように行われる。

(1) サンプルを説明変数 X_i の大きさによって並べ替える (すべてのサンプル数を T とする)。

(2) 並べ替えられたサンプルの中央に位置するいくつかのサンプルを取り除く (取り除かれたサンプル数を c とする)。

(3) 2つに区分されたサンプルを用いて、それぞれのグループごとに回帰を行う (サンプル数はそれぞれ $(T-c)/2$ である)。

(4) 回帰によって測定された残差平方和を X_i の小さい方からそれぞれ RSS_1 と RSS_2 とし、統計量 λ を求める。

$$\lambda = \frac{RSS_2}{RSS_1} \sim F((T-c-2k)/2, (T-c-2k)/2)$$

(5) この統計量 λ は、自由度がそれぞれ $(T-c-2k)/2$ の F 分布にしたがうことが分かっている。したがって、もし λ が $F_{0.95}$ よりも大きければ、2つのグループの分散が等しいという帰無仮説を 5% 水準で棄却できる。

一般に、Goldfeld=Quandt の検定方法の信頼性は取り除かれるサンプル数 c に依存するといわれている。もし c が多ければ、分散の違いをより明確に認識することができる反面、自由度を犠牲にする。これまでの経験から、 $T=30$ のとき $c=4$ 、 $T=60$ のとき $c=10$ が合理的であることが分かっている。

▼ Breusch=Pagan の検定

Breusch=Pagan の検定は、分散が次のような関数で表すことができると仮定する。

$$\sigma_i^2 = h(\mathbf{z}_i' \boldsymbol{\alpha}) = h(\alpha_1 + \mathbf{z}_i' \boldsymbol{\alpha}^*)$$

$h(\cdot)$ はある特定されていない関数型である。また \mathbf{z}_i' は分散の不均一性を引き起こす観察可能な変数のベクトルであり、 $\boldsymbol{\alpha}'$ は未知の係数のベクトルである。一般に \mathbf{z}_i' には、いくつかの説明変数 X が用いられる。

$$\mathbf{z}_i' = (1 \quad \mathbf{z}_i^*) = (1, z_{i2}, z_{i3}, \dots, z_{is})$$

$$\boldsymbol{\alpha}' = (\alpha_1 \quad \boldsymbol{\alpha}^*) = (\alpha_1, \alpha_2, \dots, \alpha_s)$$

ここで $\mathbf{z}_i = \mathbf{1}$ になるように設定すれば、帰無仮説は次のようになる。

$$H_0: \boldsymbol{\alpha}^* = \mathbf{0} \quad \text{もしくは} \quad H_0: \alpha_2 = \alpha_3 = \dots = \alpha_s = 0$$

つまり帰無仮説はすべての i について $\sigma_i^2 = h(\alpha_1)$ 、つまり一定値であることを示している。Breusch=Pagan の検定は、次のような手続きによって行われる。

- (1) まず OLS を行うことによって推定誤差 \mathbf{e} のベクトルを求める。
- (2) 推定誤差 \mathbf{e} を用いて分散の推定値を計算する。

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n}$$

- (3) 関数 $h(\cdot)$ を特定し、

$$\frac{e_i^2}{\hat{\sigma}^2} = \mathbf{z}_i' \boldsymbol{\alpha} + v_i$$

を回帰することにより説明された平方和 (ESS) を計算する。

(4) (3) で計算された ESS を 2 で割った数 ψ は、自由度 $(p-1)$ のカイ二乗分布にしたがう。したがって、もし $\psi > \chi_{0.95}^2(p-1)$ ならば分散が均一であるという帰無仮説が 5% 水準で棄却されることになる。しかし Breusch=Pagan の検定の最大の問題は、関数 $h(\cdot)$ を分析者自らが特定することが必要とされるため、関数の客観性が究めて乏しくなるという点にある。

▼ Glesjer の検定

Glesjer の検定は、特定の 1 つの説明変数 Z が分散の不均一性を引き起こしているときに利用できる。その手続きは、

(1) $Y_i = \alpha + \beta Z_i$ を回帰させ、推定誤差 e_i を計算する。

(2) (1) で計算された e_i の絶対値、もしくは e_i^2 を被説明変数として OLS を行う。

$$|e_i| = \delta_0 + \delta_1 Z_i^h + v_i$$

h は 1, 0.5, -1 などの特定の数値であるため関数は非線形になる。またここでの帰無仮説は傾き δ_1 がゼロというものであり、もしこの仮説が棄却されれば分散の不均一性の存在が指摘されることになる。

3.2 : 分散不均一性の下での推定

もし分散の不均一性が検出された場合、どのように解決すればよいのかを検討する。すでに Ω が知られている場合には、回帰モデルを Ω で割ることによって OLS をそのまま適用すればよい。しかし現実にはこのようなケースは究めて希であり、一般的には真の分散は未知である。そのような限界を克服する方法として、たとえば 2 段階 Aitken 推定値 (2SAE) がある。この手法は基本的に Glesjer の検定とよく似ている。すでに (16) 式によって Ω が既知であるときの一般化最小二乗法について述べてきたが、現実には Ω は未知であり、その場合には (16) 式はもはや使用できない。しかしもし Ω がその推定値 $\hat{\Omega}$ で置き換えるならば、その推定値 $\hat{\mathbf{b}}$ は 2 段階 Aitken 推定値と呼ばれる。

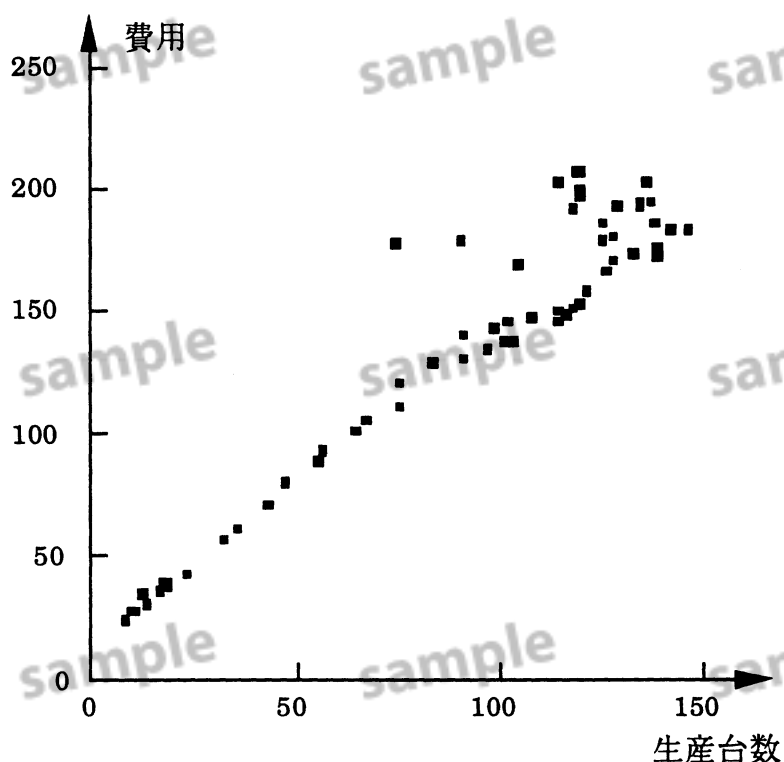
ここでは例題を用いて、データの対数化についてだけ述べることにする。表 3 は 1961 年から 1991 年までに日産自動車が生産した車両台数と、費用（売上原価と販売費及び一般管理費の合計したもの）を示したものである。この表から、生産台数の増加に伴って費用が増加しているだけでなく、費用のバラツキも大きくなっていることが理解できる。図 2 は生産台数を横軸に、費用を縦軸にとり、データをプロットしたものである。

(表3) 日産自動車の生産台数と費用の推移

年月	費用 (百億円)	生産台数 (万台)	年月	費用 (百億円)	生産台数 (万台)	年月	費用 (百億円)	生産台数 (万台)
61.3	23.1	8.0	71.9	130.5	91.5	81.9	176.1	138.4
61.9	27.1	9.7	72.3	140.6	91.3	82.3	180.5	128.1
62.3	27.5	10.6	72.9	143.2	99.0	82.9	170.6	128.1
62.9	34.4	11.8	73.3	146.0	102.3	83.3	185.5	137.9
63.3	31.1	13.0	73.9	134.1	97.3	83.9	193.5	136.6
63.9	36.3	16.0	74.3	129.5	84.0	84.3	194.3	134.2
64.3	38.9	17.4	74.9	137.5	101.1	84.9	192.7	134.7
64.9	38.2	17.8	75.3	137.8	103.3	85.3	202.9	135.8
65.3	36.9	17.4	75.9	146.9	107.9	85.9	193.0	128.5
65.9	37.4	17.8	76.3	148.1	116.1	86.3	185.8	125.2
66.3	42.7	22.6	76.9	145.6	114.1	86.9	177.6	74.3
66.9	56.8	32.3	77.3	149.4	114.2	87.3	179.4	125.2
67.3	61.5	35.3	77.9	157.4	121.2	87.9	178.1	90.7
67.9	71.1	42.9	78.3	152.6	119.4	88.3	169.6	104.8
68.3	80.1	47.3	78.9	151.0	118.0	88.9	197.2	119.2
68.9	88.9	55.3	79.3	166.5	126.0	89.3	191.8	117.8
69.3	92.4	56.5	79.9	172.7	133.1	89.9	206.3	119.4
69.9	100.7	64.4	80.3	183.6	145.8	90.3	199.2	119.1
70.3	104.7	66.9	80.9	171.6	138.7	90.9	206.1	118.9
70.9	110.7	75.2	81.3	182.7	142.1	91.3	201.7	114.1
71.3	119.9	75.1						

最初にGoldfeld=QuandtとGlesjerの検定を用いて、このモデルの分散が果たして不均一かどうかを調査してみた。まずGoldfeld=Quandtの検定では、データを説明変数の大きいものから並び替え、その中心に位置するデータを取り除くことから始まる。ここでは例題のデータ数61(T)から13のデータを取り除き(c)、残りを2つのグループに分けた。次にそれぞれのグループごとに回帰を行なった結果、残差平方和は $RSS_1 = 173.370$ と $RSS_2 = 3,480.566$ であった。 RSS_2 を RSS_1 で割った統計量 λ は自由度が(23, 23)のF分布にしたがう。例題の統計量 λ は20.075であり、これは $F_{0.95} = 2.01$ よりも大きい。したがって、この2つのグループの分散が均一であるという帰無仮説は棄却された。

(図2) 生産台数と費用の散布図



またGlesjerの検定は、まずすべてのデータについてOLSを適用し、その推定誤差 e_i の絶対値を求める。次に $|e_i|$ を被説明変数としたOLSによって回帰係数 $\hat{\delta}_1$ を推定する。ここでは h に 1, -1, 0.5 という3つの数値を組み入れたが、すべての h について回帰係数 $\hat{\delta}_1$ が有意であることが導き出されている。これはGoldfeld-Quandtの検定と同様に、データの分散が均一であるという帰無仮説が棄却されたことを意味する。

	係数 ($\hat{\delta}_1$)	標準誤差	t 値	確率
$h = 1$	0.105	0.033	3.128	0.003
$h = 0.5$	1.624	0.516	3.146	0.003
$h = -1$	-139.827	56.460	-2.477	0.016

このように分散の不均一性が指摘されたデータに対して、対数化を行うことによって結果がどのように変化したかを分析してみる。ここでは両対数関数 (double-log function または log-log function) を採用したが、導き出される回帰係数は通常の場合の係数とは意味が異なってくる。

$$Y_i = \alpha + \beta X_i + u_i \quad (30)$$

$$\beta = \frac{dY_i}{dX_i} \quad (31)$$

(30) 式のような通常の回帰式での回帰係数 β は、説明変数が「1単位」変化したときの被説明変数の変化を測定する。しかし両対数変換を行った回帰係数は、説明変数が「1%」変化したときの被説明変数の変化率を測定するものである。たとえば (32) 式のような弾力性回帰モデルを構築し、両辺を対数に変換すると (33) 式が得られる。

$$Y_i = \exp(\alpha) X_i^\beta \quad (32)$$

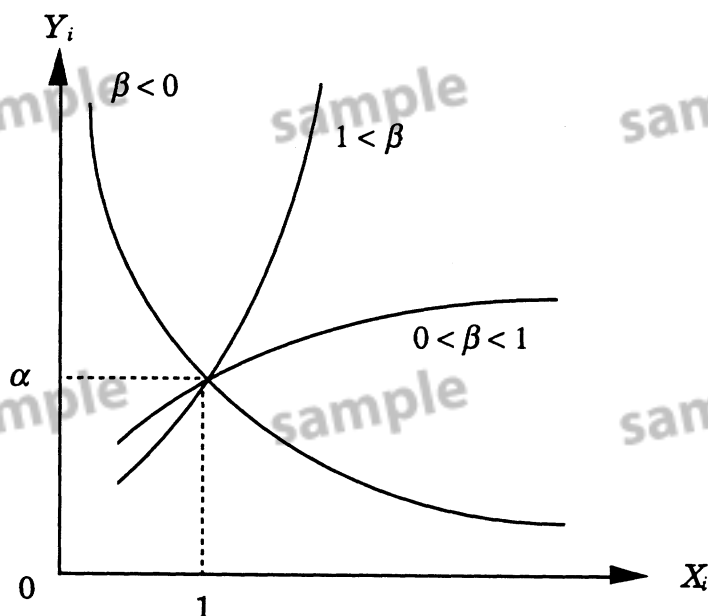
$$\ln(Y_i) = \alpha + \beta \ln(X_i) + u_i \quad (33)$$

したがって回帰係数 β は、

$$\beta = \frac{d[\ln(Y_i)]}{d[\ln(X_i)]} = \frac{dY_i}{dX_i} \cdot \frac{X_i}{Y_i} \quad (34)$$

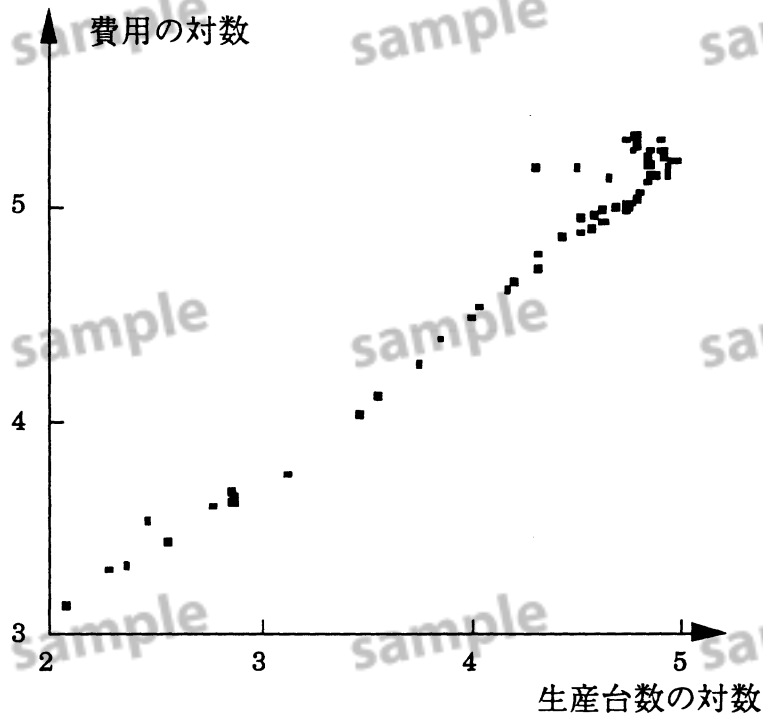
もし β が 0 より大きければ増加関数となるが、図 3 のように $1 < \beta$ の場合と $0 < \beta < 1$ との場合とでは二階の導関数の符号が異なる。また $0 > \beta$ の場合には常に減少関数となる。

(図 3) 両対数関数



例題のデータについて両対数変換を行い散布図を作成してみた (図 4)。この図から分かるように、説明変数が大きくなることによって発生するバラツキが、図 2 と比較して小さくなっている。

(図4) 両対数変換による散布図



このことを確認するために Goldfeld=Quandt の検定を再び行った。Goldfeld=Quandt の検定では、残差平方和がそれぞれ $RSS_1=0.073$ と $RSS_2=0.100$ となり統計量 λ は 1.370 であった。この水準は $F_{0.95} = 2.01$ よりも小さく、したがって分散の均一性が棄却されていない。ここで対数化を行っていないモデルと両対数モデルとの β の分散を比較をしてみると、両対数モデルでの分散は約 11% に減少していることが理解できる。

$$Var(b_{non-log}) = \frac{\sum e^2}{(n-2) \sum x^2} = \frac{17,486.442}{(59)(118,424.490)} = 2.503 - E3$$

$$Var(b_{log}) = \frac{0.686}{(59)(42.022)} = 2.767 - E4$$

4 : 自己 (系列) 相関

分散の均一性と同様に、OLS を用いるための仮定の一つに誤差項が相関しないというものがあった。しかし特に時系列データを取り扱う場合、誤差項の相関が発生することが多く見受けられる。このような相関を「自己相関 (autocorrelation)」、あるいは系列相関と呼ばれる。この自己相関とは、 t 期の誤差項 u_t が過去や将来の誤差項 ($u_{t \pm k}$) と相関をもつことをいい次のような式で表される。

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_k u_{t-k} + \varepsilon_t \quad |\rho| < 1$$

このような自己相関が発生する理由はいくつかある。たとえば重要な説明変数が除外されたケースである。つまり回帰モデルの誤差項には、除外された変数間の相関が集約されるためである。また過去の出来事が現在や将来に影響する場合にも自己相関が発生する。たとえば企業の設備投資や新製品開発の効果は、通常長期にわたって認識されるからである。

自己相関が存在するモデルにOLSを適用すると、好ましくない推定や予測が行われることになる。まず次のような単回帰モデルを想定し、その影響について議論しよう。

$$\begin{aligned} y_t &= \beta x_t + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \end{aligned} \tag{35}$$

ここで小文字の x, y, u は偏差型を表している。また ε は、

$$E(\varepsilon) = 0, \quad E(\varepsilon'\varepsilon) = \sigma_\varepsilon^2 \mathbf{I}$$

である。(35)式にOLSを適用すると、(21)式と同じ

$$\text{Var}(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

が得られる。 u は1階の自己相関をしているので、

$$\mathbf{\Omega} = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix} \tag{36}$$

したがって、

$$\mathbf{\Omega}^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix} \tag{37}$$

であることが知られている。(21)式に(36)式を代入すると、

$$\text{Var}(b) = \frac{\sigma_u^2}{\sum_{t=1}^n x_t^2} \left(1 + 2\rho \frac{\sum_{t=1}^{n-1} x_t x_{t+1}}{\sum_{t=1}^n x_t^2} + 2\rho^2 \frac{\sum_{t=1}^{n-2} x_t x_{t+2}}{\sum_{t=1}^n x_t^2} + \dots + 2\rho^{n-1} \frac{\sum_{t=1}^{n-2} x_1 x_n}{\sum_{t=1}^n x_t^2} \right) \tag{38}$$

逆に、もし Ω が既知であるとする (26) 式が適用できる。

$$\text{Var}(b_*) = \frac{\sigma_u^2}{\sum_{t=1}^{n-1} x_t^2} \left(\frac{1 - \rho^2}{1 + \rho^2 - 2\rho \sum_{t=1}^{n-1} x_t x_{t+1} / \sum_{t=1}^{n-1} x_t^2} \right) \quad (39)$$

ここで $\sum_{t=1}^{n-1} x_t x_{t+1} / \sum_{t=1}^{n-1} x_t^2 = \lambda$ とおき、(39) 式を (38) 式で割れば、

$$\begin{aligned} \frac{\text{Var}(b_*)}{\text{Var}(b)} &= \frac{(1 - \rho^2)}{(1 + \rho^2 - 2\rho\lambda)(1 + 2\rho\lambda + 2\rho^2\lambda^2 + \dots)} \\ &= \frac{(1 - \rho^2)(1 - \rho\lambda)}{(1 + \rho^2 - 2\rho\lambda)(1 + \rho\lambda)} \end{aligned} \quad (40)$$

となる。(40) 式は、 $\rho = 0$ 、または $\rho = \lambda$ のときに (38) 式と (39) 式とが等しくなることを示している。また ρ が大きくなるにしたがって (40) 式の数値は小さくなる。つまり自己相関しているモデルに OLS を適用すると、真の分散よりも過小推定されるのである。その結果として、導き出される決定係数や推定される回帰係数の t 値が過大に評価されるという弊害をもたらす。したがって、モデルに自己相関が発生しているかどうかの検定は究めて重要な作業である。

4.1 : 自己相関の検定

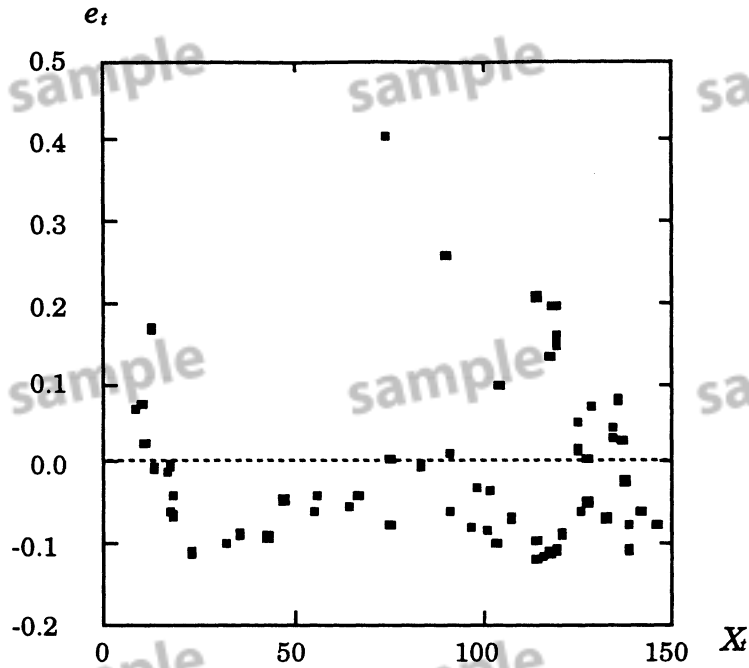
自己相関が発生しているのかを検定するについては多くの検定方法が提唱されている。ここではその中でも特に頻繁に用いられる Durbin-Watson の検定を紹介する。彼らの手法は一階の自己相関に関する検定方法で、OLS による推定誤差 e を用いて次のような統計量が与えられる。

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (41)$$

推定誤差の平均はゼロであることから、残差は水平軸の上下にプロットされる。Durbin-Watson 検定の重要な仮定は、もし正の自己相関が発生しているならば、残差の違い $(e_t - e_{t-1})$ は残差自体 (e_t) よりも徐々に小さくなる傾向があり、逆にもし負の自己相関が発生しているならば、連続した残差は逆の符号となるというものである。したがって負の自己相関の場合には、残差の違い $(e_t - e_{t-1})$ は残差自体 (e_t) よりも徐々に大きくなるというものである。

たとえば、表3の例題を OLS によって推定した誤差と説明変数との散布図を図5に示している。この図からも理解できるように、残差は連続して水平軸の回りに散らばらずに、 e_t と e_{t-1} とが同じ座標内で接近している。このような傾向が、正の自己相関が発生している例である。

(図4) 残差の散布図



(41) 式の d 統計量を展開すると、

$$d = \frac{\sum_{t=2}^n e_t^2 + \sum_{t=2}^n e_{t-1}^2 - 2\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}$$

$$= 1 + \frac{\sum_{t=2}^n e_{t-1}^2}{\sum_{t=1}^n e_t^2} - 2 \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \quad (42)$$

ここでサンプル数が大きくなると暫時的に $\sum_{t=2}^n e_{t-1}^2 \approx \sum_{t=1}^n e_t^2$ となるので、(42) 式は次のように簡略化できる。

$$d \approx 2(1 - r) \quad (43)$$

ここで $r = \sum_{t=2}^n e_{t-1}^2 / \sum_{t=1}^n e_t^2$ であり、 e_{t-1} の e_t に対する回帰係数である。また (43) 式の r は $-1 < r < 1$ によって d の範囲は 0 から 4 であり、 d が 2 のとき自己相関は発生していないと考えてもよい。さらに (43) 式はサンプル数が大きいときに適用できると述べたが、サンプル数が限られているときには d の期待値は次の近似式によって与えられる。

$$E(d) \approx 2 + \frac{2(k-1)}{n-k} \quad (44)$$

d 統計量による自己相関についての検定は次のようになる。ただし du と dL は、それぞれの有意水準の下での上限と下限であり、ここでの帰無仮説は自己相関がない ($\rho \neq 0$) というものである。

- ・ $d < dL$ ならば自己相関がないという帰無仮説は棄却される
- ・ $d > du$ ならば帰無仮説は棄却されない
- ・ $dL < d < du$ ならば結論を保留する

d 統計量は、ほとんどの統計パッケージにおいて表示される。たとえば例題の d 統計量は 0.670 という数値が導き出されたが、サンプル数 61、説明変数の数 1 のときの 99% 水準は $du = 1.383$ である。したがって例題のデータに自己相関が発生していないという帰無仮説は棄却されることになる。

4.2：自己相関のもとでの推定

これまでに自己相関が発生しているモデルを OLS によって推定することによる弊害を述べてきた。このような弊害を取り除くために、Cochrane-Orcutt 法や Prais-Winsten 法、最尤法などいくつかの方法が提唱されている。

▼ Cochrane = Orcutt 法

次のような単純回帰を想定すると、

$$Y_t = \alpha + \beta X_t + u_t, \quad t = 1, 2, \dots, n \quad (45)$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad (46)$$

(46) 式を用いて (45) 式を変形すると、

$$Y_t - \rho Y_{t-1} = \alpha(1 - \rho) + \beta(X_t - \rho X_{t-1}) + \varepsilon_t, \quad t = 2, 3, \dots, n \quad (47)$$

となり、ここで $Y_t - \rho Y_{t-1} = Y_t^*$ 、 $X_t - \rho X_{t-1} = X_t^*$ 、 $\alpha(1 - \rho) = \alpha^*$ 、 $\beta = \beta^*$ おくと、

$$Y_t^* = \alpha^* + \beta^* X_t^* + \varepsilon_t, \quad t = 2, 3, \dots, n \quad (48)$$

になり、この変換を Cochrane-Orcutt 変換と呼ぶ。具体的な手続きは、

- (1) (45) 式に OLS を適用し ρ を推定する。ここで推定誤差を e_t とすれば、

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2} \quad (49)$$

(2) ρ の推定値を $Y_t - \rho Y_{t-1} = Y_t^*$ と $X_t - \rho X_{t-1} = X_t^*$ に代入し、(48) 式に再び OLS を適用する。ただしサンプル数は $n-1$ である。このようなプロセスを繰り返すことにより、 ρ を最小化する「反復 Cochrane-Orcutt 法」と呼ばれる方法もある。

Cochrane=Orcutt 法は、変換の段階で自由度を1つ減らし、最初の期のデータを推定に使用せず、さらに推定の精度が悪いことから現在ではほとんど使用されることはない。Cochrane=Orcutt 法に似た手法でありながら、Cochrane=Orcutt 法よりも優れているのがPrais=Winsten 法である。

▼ Prais = Winsten 法

Prais=Winsten 法とCochrane=Orcutt 法との大きな違いは、Prais=Winsten 法は ρ の推定にDurbin=Watson 比を用いることと、さらに最初の期のデータもパタメータの推定に利用する点にある。Prais=Winsten 法の手続きは、

(1) ρ の推定に次の式を用いる。

$$\hat{\rho} = \frac{n^2 \left(1 - \frac{d}{2}\right) + k^2}{n^2 - k^2} \quad (50)$$

ここで d はDurbin=Watson 比であり、サンプル数の増大によって (50) 式は漸次的に $1 - d/2$ に近づく。

(2) (37) 式の Ω^{-1} を用いて、次のような変換行列 \mathbf{T} を求める ($\mathbf{T}'\Omega\mathbf{T} = \mathbf{I}$)。

$$\mathbf{T} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix} \quad (51)$$

(51) 式を (45) 式に代入すると、 $\mathbf{T}\mathbf{y} = (\mathbf{T}\mathbf{X})\boldsymbol{\beta} + \mathbf{T}\mathbf{u}$ より、

$$\begin{bmatrix} \sqrt{1-\rho^2} \cdot Y_1 \\ Y_2 - \rho Y_1 \\ \vdots \\ Y_n - \rho Y_{n-1} \end{bmatrix} = \begin{bmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2} \cdot X_1 \\ 1-\rho & X_2 - \rho X_1 \\ \vdots & \vdots \\ 1-\rho & X_n - \rho X_{n-1} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \sqrt{1-\rho^2} \cdot u_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (52)$$

(3) (50) 式によって推定された ρ を (52) 式に代入し、定数項なしのOLSを行う。

▼ 最尤法

Prais=Winsten 法と同じ程度に自己相関をもつデータに対して有効なのが最尤法

(maximum likelihood) と呼ばれるものである。最尤法についての詳しい解説は別の機会に譲るとして、ここでは基本的な考え方にだけとどめることにする。

いま X_1, X_2, \dots, X_n というデータが与えられたとする。もし母集団が離散的分布で各データが独立であるならば、母集団から X_1, X_2, \dots, X_n が得られる確率は、

$$f(X_1, X_2, \dots, X_n; \theta) = \prod_{i=1}^n f(X_i; \theta)$$

と書くことができる。ここで $f(X_i; \theta)$ は母集団 θ のもとで X_i が得られる確率を示す記号である。いま母集団が θ_1 と θ_2 の2つの場合を想定したとき、もし

$$f(X_1, X_2, \dots, X_n; \theta_1) > f(X_1, X_2, \dots, X_n; \theta_2)$$

ならば θ_1 を推定値として選ぶ。 $f(X_1, X_2, \dots, X_n; \theta)$ は確率密度であり、これを θ の関数とすると尤度関数 (L) と呼ばれる。そしてもし母集団の推定値 $\hat{\theta}$ がすべての θ について成立するとき、この $\hat{\theta}$ を最尤推定値という。 X_1, X_2, \dots, X_n は固定されているので、最尤推定値は

$$\frac{\partial L}{\partial \theta} = 0$$

によって求めることができる。また尤度関数 (L) は対数に置き換えられることもある。最尤法は次のような定理のもとに成り立っている。

もしランダムな変数 X が密度関数 $f(x)$ をもち、変数 Z が X と1対1の関係の関数であるならば、 Z の密度関数は $f(z) = |dx/dz| f(x)$ である。

ここで $|dx/dz|$ は、 x を z について偏微分した行列の行列式 (determinant) の絶対値を示している。ここで、ベクトル $\hat{\theta}$ を k 個の未知のパラメータ θ の推定値とし、Cramer = Rao 定理を用いて情報行列 (information matrix) を求めると、

$$\mathbf{R}(\theta) = -E \left(\frac{\partial^2 \ln L}{\partial \theta \partial \theta'} \right) = -E \begin{bmatrix} \frac{\partial^2 \ln L}{\partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_2} & \dots & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_k} \\ \frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_2^2} & \dots & \frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln L}{\partial \theta_k \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_k \partial \theta_2} & \dots & \frac{\partial^2 \ln L}{\partial \theta_k^2} \end{bmatrix}$$

となる。(45)式と(46)式を考えてみると、

$$\varepsilon_t = u_t - \rho u_{t-1} = (Y_t - \alpha - \beta X_t) - \rho (Y_{t-1} - \alpha - \beta X_{t-1}) \quad (t = 2, 3, \dots, n)$$

ここで u が正規分布にしたがうことを仮定し、尤度関数を対数化すると、

$$\begin{aligned} L &= \ln \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^{n-1} \exp \left[-\sum \frac{(\varepsilon_t^2)}{2\sigma^2} \right] \\ &= -\frac{n-1}{2} \ln(2\pi) - \frac{n-1}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum \varepsilon_t^2 \quad (t = 2, 3, \dots, n) \end{aligned}$$

となる(ただしデータ数は1つ減少する)。ここで未知数の α , β , ρ , σ_ε^2 について L を偏微分すると、

$$\frac{\partial \ln L}{\partial \alpha} = \frac{(1-\rho)}{\sigma_\varepsilon^2} \sum \varepsilon_t$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{\sigma_\varepsilon^2} \sum (X_t - \rho X_{t-1}) \varepsilon_t$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{1}{\sigma_\varepsilon^2} \sum u_{t-1} \varepsilon_t$$

$$\frac{\partial \ln L}{\partial \sigma_\varepsilon^2} = -\frac{n-1}{2\sigma_\varepsilon^2} + \frac{1}{2\sigma_\varepsilon^4} \sum \varepsilon_t^2$$

さらにこれらの2階の偏微分は、

$$\frac{\partial^2 \ln L}{\partial \alpha^2} = -\frac{(n-1)(1-\rho)^2}{\sigma_\varepsilon^2}$$

$$\frac{\partial^2 \ln L}{\partial \beta^2} = -\frac{1}{\sigma_\varepsilon^2} \sum (X_t - \rho X_{t-1})^2$$

$$\frac{\partial^2 \ln L}{\partial \rho^2} = -\frac{1}{\sigma_\varepsilon^2} \sum u_{t-1}^2$$

$$\frac{\partial^2 \ln L}{\partial (\sigma_\varepsilon^2)^2} = -\frac{n-1}{2\sigma_\varepsilon^4} - \frac{1}{\sigma_\varepsilon^6} \sum \varepsilon_t^2$$

次に交差偏微分は、

$$\frac{\partial^2 \ln L}{\partial \alpha \partial \beta} = -\frac{1-\rho}{\sigma_\varepsilon^2} \sum (X_t - \rho X_{t-1})$$

$$\frac{\partial^2 \ln L}{\partial \alpha \partial \rho} = -\frac{1}{\sigma_\varepsilon^2} [\Sigma \varepsilon_t + (1-\rho) \Sigma u_{t-1}]$$

$$\frac{\partial^2 \ln L}{\partial \alpha \partial \sigma_\varepsilon^2} = -\frac{1-\rho}{\sigma_\varepsilon^2} \Sigma \varepsilon_t$$

$$\frac{\partial^2 \ln L}{\partial \beta \partial \rho} = -\frac{1}{\sigma_\varepsilon^2} [\Sigma \varepsilon_t X_{t-1} + \Sigma \varepsilon_{t-1} (X_t - \rho X_{t-1})]$$

$$\frac{\partial^2 \ln L}{\partial \beta \partial \sigma_\varepsilon^2} = -\frac{1}{\sigma_\varepsilon^4} \Sigma \varepsilon_t (X_t - \rho X_{t-1})$$

$$\frac{\partial^2 \ln L}{\partial \rho \partial \sigma_\varepsilon^2} = -\frac{1}{\sigma_\varepsilon^4} \Sigma \varepsilon_t u_{t-1}$$

これらのマイナスの期待値をとると、

$$\mathbf{R} = -E \begin{bmatrix} \frac{\partial^2 \ln L}{\partial \alpha^2} & \frac{\partial^2 \ln L}{\partial \alpha \partial \beta} & \frac{\partial^2 \ln L}{\partial \alpha \partial \rho} & \frac{\partial^2 \ln L}{\partial \alpha \partial \sigma_\varepsilon^2} \\ \frac{\partial^2 \ln L}{\partial \beta \partial \alpha} & \frac{\partial^2 \ln L}{\partial \beta^2} & \frac{\partial^2 \ln L}{\partial \beta \partial \rho} & \frac{\partial^2 \ln L}{\partial \beta \partial \sigma_\varepsilon^2} \\ \frac{\partial^2 \ln L}{\partial \rho \partial \alpha} & \frac{\partial^2 \ln L}{\partial \rho \partial \beta} & \frac{\partial^2 \ln L}{\partial \rho^2} & \frac{\partial^2 \ln L}{\partial \rho \partial \sigma_\varepsilon^2} \\ \frac{\partial^2 \ln L}{\partial \sigma_\varepsilon^2 \partial \alpha} & \frac{\partial^2 \ln L}{\partial \sigma_\varepsilon^2 \partial \beta} & \frac{\partial^2 \ln L}{\partial \sigma_\varepsilon^2 \partial \rho} & \frac{\partial^2 \ln L}{\partial (\sigma_\varepsilon^2)^2} \end{bmatrix}$$

$$= \frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} (n-1)(1-\rho)^2 & (1-\rho) \Sigma (X_t - \rho X_{t-1}) & 0 & 0 \\ (1-\rho) \Sigma (X_t - \rho X_{t-1}) & \Sigma (X_t - \rho X_{t-1})^2 & 0 & 0 \\ 0 & 0 & (n-1) \sigma_u^2 & 0 \\ 0 & 0 & 0 & \frac{n-1}{2\sigma_\varepsilon^2} \end{bmatrix}$$

となる。さらに行列 \mathbf{R} の点線内に注目するとこれは $\mathbf{X}' \cdot \mathbf{X} \cdot$ に等しい。ここで $\mathbf{X} \cdot$ は、

$$\mathbf{X} \cdot = \begin{bmatrix} 1-\rho & X_2 - \rho X_1 \\ 1-\rho & X_3 - \rho X_2 \\ \vdots & \vdots \\ 1-\rho & X_n - \rho X_{n-1} \end{bmatrix}$$

である。さらに \mathbf{R}^{-1} は分散行列であることから、 α, β の分散は次のように求められ

る¹⁾。

$$\text{Var} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \sigma_{\varepsilon}^2 (\mathbf{X}' \cdot \mathbf{X})^{-1}$$

表4に、表3の例題を用いた統計パッケージの結果を掲載している。ここで使用した統計パッケージは、TSP (Time Series Processor) バージョン4.3である。TSPは自己相関のもとでの推定について、Cochrane=Orcutt法、最尤法、Hildreth=Lu法、グリッド法の4つのオプションをもっている。表4の結果から、自己相関のもとでのOLSの推定値とCochrane=Orcutt法、最尤法との違いを確認するとよい。ただしこれらの推定方法は、自己相関が1階という仮定のもとでの統計手法であることを認識しておかなければならない。

(表4) TSPによる推定結果

最尤法

FIRST-ORDER SERIAL CORRELATION OF THE ERROR MAXIMUM LIKELIHOOD ITERATIVE TECHNIQUE			
Mean of dependent variable			= 3.76612
Std. dev. of dependent var.			= 7.65564
Sum of squared residuals			= 3368.68
Variance of residuals			= 57.0963
Std. error of regression			= 7.55621
R-squared			= .148427
Adjusted R-squared			= .133993
Durbin-Watson statistic			= 2.57358
Rho (autocorrelation coef.)			= .993821
Standard error of rho			= .815323E-02
t-statistic for rho			= 121.893
F-statistic (zero slopes)			= 2.58953
Log of likelihood function			= -211.101
Variable	Estimated Coef.	Standard Error	t-statistic
Constant	95.6418	62.7341	1.52456
UNIT	.304273	.080391	3.78491

¹⁾ \mathbf{R}^{-1} が分散行列となるのは、次のCramer = Rao の定理による。

$$\text{Var}(\tilde{\theta}) \geq \frac{-1}{E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right)} = \frac{-1}{\mathbf{R}}$$

コクラン・オーカット法

**FIRST-ORDER SERIAL CORRELATION OF THE ERROR
COCHRANE-ORCUTT ITERATIVE TECHNIQUE**

Mean of dependent variable	= 7.68091
Std. dev. of dependent var.	= 7.64628
Sum of squared residuals	= 2878.72
Variance of residuals	= 49.6331
Std. error of regression	= 7.04507
R-squared	= .165462
Adjusted R-squared	= .151073
Durbin-Watson statistic	= 2.78244
Rho (autocorrelation coef.)	= .964119
Standard error of rho	= .034272
t-statistic for rho	= 28.1314
F-statistic (zero slopes)	= 11.4995
Log of likelihood function	= -201.259

Variable	Estimated Coef.	Standard Error	t-statistic
Constant	177.967	27.4929	6.47322
UNIT	.262421	.077385	3.39109

OLS

METHOD OF ESTIMATION = ORDINARY LEAST SQUARES

Mean of dependent variable	= 132.286
Std. dev. of dependent var.	= 59.0312
Sum of squared residuals	= 17478.8
Variance of residuals	= 296.250
Std. error of regression	= 17.2119
R-squared	= .916402
Adjusted R-squared	= .914985
Durbin-Watson statistic	= .669803
F-statistic (zero slopes)	= 646.759
Schwartz Bayes. Info. Crit.	= 5.79265
Log of likelihood function	= -259.120

Variable	Estimated Coef.	Standard Error	t-statistic
Constant	19.4622	4.95358	3.92892
UNIT	1.27198	.050016	25.4314

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

sample

不 許 複 製

慶應義塾大学ビジネス・スクール

Contents Works Inc.