

2009年度応用マクロ経済学講義ノート: モデル推計

阿部修人
一橋大学

平成 21 年 12 月 11 日

概要

最小二乗法、最尤法、非線形最小二乗法による推計を Matlab でおこない比較する。

1 本章であつかう問題

経済変数 Y と X があり (例えば、 Y が消費、 X に金利と国民所得、 $X = (X_1, X_2)$)、以下の関係が成立しているとする。

$$Y = X\beta + e$$

ここで、 β は一定の値をとる未知のパラメータであり、 e は誤差項である。上の式そのものが正しいものであることを我々は知っており、かつ、 Y と X の値も知っているとは仮定する。しかしながら、 β と e の値については、我々は知らないとは仮定する。

ここで、追加的に直行条件、すなわち、「 e は X と相関をもたない」と仮定を課す。例えば、 $\beta = (\beta_1, \beta_2)'$ と、二つの変数があるとし、 $\beta_1 > 0$ としよう。このとき、 X_1 が大きいほど、 Y は大きくなるが、もしも誤差項 e が X_1 と負の相関をもつとすると、 X_1 が大きいほど、 e が小さくなり、その分 Y も小さくなってしまう。このとき、 e と X_1 の相関を事前に知らない限り、 X_1 と Y の間の関係を知ることができなくなる。

本章での議論は、線形モデルに限定しているが、OLS を除き、非線形モデルでもほぼ同様の議論が可能である。ただし、非線形推計では、漸近分布を用いることになる。また、最尤法でも、出てくる推計量の分布は漸近的なものである。

1.1 OLS の復習

Ordinary Least Squares(OLS) は、

$$\text{Min}_{\beta} (Y - X\beta)'(Y - X\beta)$$

をみたす β を求める手法である。

最小化の一階条件は

$$X'(Y - X\hat{\beta}) = 0$$

であるが、これは、 $(Y - X\hat{\beta})$ と X' の内積がゼロ、すなわち、ベクトル $(Y - X\hat{\beta})$ とベクトル X' が直行していることを意味する。上記の条件は正規方程式と呼ばれる。

正規方程式を整理すると

$$\begin{aligned} X'Y &= X'X\hat{\beta} \\ \hat{\beta} &= (X'X)^{-1} X'Y \end{aligned}$$

したがって、

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1} X'(X\beta + e) \\ &= \beta + (X'X)^{-1} X'e \end{aligned}$$

ここで、 e の期待値がゼロで、分散が σ^2 なら、

$$\begin{aligned} E(\hat{\beta}) &= \beta, \\ \text{Var}(\hat{\beta}) &= (X'X)^{-1} (X'X) (X'X)^{-1} \text{Var}(e) \\ &= (X'X)^{-1} \text{Var}(e) \end{aligned}$$

実際には、

$$\varepsilon = Y - X\hat{\beta}$$

と定義すると、

$$\widehat{\text{Var}}(e) = \text{Var}(\varepsilon)/(n - k)$$

ただし、 $n - k$ は自由度である。

そして、

$$\text{Var}(\hat{\beta}) = (X'X)^{-1} \text{Var}(\varepsilon)/(n - k)$$

となる。

ちなみに

$$\begin{aligned}\lim_{n \rightarrow \infty} \hat{\beta} &= \lim_{n \rightarrow \infty} (X'X)^{-1} X'(X\beta + e) \\ &= \beta + \lim_{n \rightarrow \infty} (X'X)^{-1} X'e \\ &= \beta + \lim_{n \rightarrow \infty} \left(\frac{X'X}{n} \right)^{-1} \frac{X'e}{n}\end{aligned}$$

となる。通常は、 $\lim_{n \rightarrow \infty} \left(\frac{X'X}{n} \right)$ は、ある正値定符号行列に収束し、 $\frac{X'e}{n}$ は、中心極限定理を適用できるように、 $\frac{X'e}{\sqrt{n}}$ が、正規分布に従うという仮定をおく。すると、OLS の推計量は一致性をもつことを示すことができる。そして、Greene の教科書では、これらを保証する条件は十分に緩いものだと議論されている。しかしながら、説明変数の一つが非定常、例えば Random Walk に従うと仮定すると (GDP や消費等の、多くのマクロ変数は非定常であると考えられる)、OLS の推計量は一致性を持たなくなる。

X の一つの変数が w_t であり、Random Walk に従っていると仮定すると、 $X'X$ の中で、w に対応するものは $\sum w_t^2$ となる。 w_t を、下記のような iid のショック項により生成される Random Walk だとする。

$$w_t = w_{t-1} + \xi_t, w_0 = 0, var(\xi_t) = 1$$

したがって

$$w_t = \sum_{s=1}^t \xi_s$$

この関係を用いると、

$$\sum_{t=1}^n w_t^2 = \sum_{t=1}^n \left(\sum_{r=1}^t \sum_{s=1}^t \xi_r \xi_s \right)$$

ところで、iid の仮定より

$$E(\xi_r \xi_s) = 0 \quad \text{if } r \neq s$$

したがって、r と s が一致するところのみが残る。

$$\sum_{t=1}^n \sum_{r=1}^t E(\xi_r^2) = \sum_{t=1}^n \sum_{r=1}^t 1 = \sum_{t=1}^n t = \frac{1}{2}n(n+1)$$

この右辺は、n の二乗の order になっている。これは、 $X'X$ を n で割っても、収束しないことを意味している。

現実的には、我々は無限の観察数を持っていない。たとえ説明変数が非定常過程にしたがっていても、 $\frac{X'X}{n}$ の値を計算することは可能であり、なんらかの推計量を得ることができる。OLS を実践する時には、なによりもまず、

推計の大前提である、(1) 誤差項が説明変数と直行していることの保証 (他の説明変数と相関があるような重要な変数が抜け落ちていないか、サンプル抽出の際に、被説明変数と説明変数の間に、モデルが想定している以外の関係が出てくるような手法をとっていないか、等など)を確認する必要があるが、それと同じように、(2) 説明変数が定常であるかどうか、にも注意を払う必要がある。これは、以下の ML や NLS でも同様である。

説明変数が 3 個の場合、Matlab では OLS は下記のように書くことができる。

```

%%%%%%%%%%%%%%
betahat_ols = (inv(X'*X))*X'*Y;
res = Y-X*betahat_ols;
sigmasq = res'*res/(n-3);% 説明変数が三個だから、自由度は n-3,n は観測数で事前に宣言
sigmahat = inv((X'*X))*sigmasq;
std_ols=[sqrt(sigmahat(1,1)),sqrt(sigmahat(2,2)),sqrt(sigmahat(3,3))];
%%%%%%%%%%%%%%

```

1.2 最尤法の復習

Maximum Likelihood (ML), 最尤法は、誤差項の分布に特定の仮定をおいて推計する手法である。OLS が線形モデルに限定されているのに対し、0 と 1 しかとらないような PROBIT や TOBIT、もしくは、もっと複雑なモデルを推計するのに適している一方、分布の仮定に全面的に依存するという欠点もある。誤差項が、互いに独立な、非常に多くの確率変数のみで構成されていると考えられる場合は、誤差項が正規分布に従うと考えることは、中心極限定理により、それほど非現実的ではない。しかし、分布が正規分布に従うかどうかは、残差の分布形状などをみて、慎重に検証する必要がある。

一方、最尤法は、OLS の仮定に「加えて」、誤差項の分布にも仮定を課すため、その仮定が正しい限りは、非常に効率的な推計が可能になる。ただし、OLS は、Gauss Markov の定理より、誤差項が綺麗な分布をしていれば、BLUE になっており、ML を行う必要はない。モデルが線形であり、誤差項が綺麗 (iid など) な性質を持っている限り、ML を行う必要ないことになる。マクロ経済学では、モデルは線形とは限らず、かつ、誤差項にも様々な構造が入ることが多いため、ML、もしくは、後述する非線形最小二乗法 (NLS、あるいは GMM) が用いられることは非常に多い。

経済モデルが

$$Y = g(X; \beta) + e$$

で表されており、かつ、 e が、iid で、なおかつ、各 e_i が分布 F 、密度関数 f

に従っているとすると、対数尤度 (log likelihood) は下記のように定義される。

$$\ln L = \sum_{i=1}^n \ln f(Y_i - g(X_i; \beta))$$

ただし、 n は観察数である。

例えば、 $Y = X\beta + e$ で、 F が平均値ゼロ、分散が σ^2 の正規分布であるとすると、対数尤度は

$$\ln L = \sum_{i=1}^n \left[\ln \sigma^2 + \ln(2\pi) + \frac{(Y_i - X_i\beta)^2}{\sigma^2} \right]$$

である。ML とは、この対数尤度を最大化させるような β と σ^2 の値を求める手法である。

ML で得られたパラメータ、 β の漸近分布は下記のようにして求めることができる。

対数尤度をパラメータ β に関して微分すると

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^n \frac{\partial \ln(Y_i - g(X_i; \beta))}{\partial \beta} \equiv \sum_{i=1}^n \Delta_i(\beta)$$

最適化されていれば、

$$\sum_{i=1}^n \Delta_i(\hat{\beta}) = 0.$$

ここで

$$\Delta = \sum_{i=1}^n \Delta_i$$

とし、最適点でテイラー展開すると

$$\Delta(\beta) = \Delta(\hat{\beta}) + \frac{\partial \ln \Delta(\hat{\beta})}{\partial \beta} (\beta - \hat{\beta}) = 0$$

ここで、

$$\Delta_0 = \frac{\partial \ln L(\hat{\beta})}{\partial \beta}$$

とすると、

ところで、対数尤度のヘシアン、すなわち

$$\frac{\partial \ln \Delta(\hat{\beta})}{\partial \beta} = \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = \sum_{i=1}^n \sum_{i=1}^n \frac{\partial^2 \ln(Y_i - g(X_i; \hat{\beta}))}{\partial \beta \partial \beta'} \equiv \sum_{i=1}^n H_i(\hat{\beta}) \equiv H_0$$

となり、

$$\Delta(\beta) = \Delta(\hat{\beta}) + H_0(\beta - \hat{\beta})$$

したがって、

$$\begin{aligned}(\beta - \hat{\beta}) &= (-H_0)^{-1} \Delta(\hat{\beta}) \\ \sqrt{n}(\beta - \hat{\beta}) &= (-H_0)^{-1} \sqrt{n} \Delta(\hat{\beta})\end{aligned}$$

ところで

$$\sqrt{n} \Delta(\hat{\beta}) \Rightarrow N\left(0, -E_0\left(\frac{1}{n} H_0\right)\right)$$

となることが知られている。これを用いると

$$\sqrt{n}(\beta - \hat{\beta}) \Rightarrow N\left(0, \left(-E_0\left(\frac{1}{n} H_0\right)\right)^{-1}\right)$$

問題は、 $\left(-E_0\left(\frac{1}{n} H_0\right)\right)^{-1}$ をどう求めるか、である。

$$\left(-E_0\left(\frac{1}{n} H_0\right)\right)^{-1} I(\hat{\beta})^{-1} \equiv \left(-E_0 \frac{\partial^2 \ln L}{\partial \beta \partial \beta'}\right)^{-1} = \left(E_0 \left(\frac{\partial \ln L}{\partial \beta'}\right) \left(\frac{\partial \ln L}{\partial \beta'}\right)'\right)^{-1}$$

だから、最適点における対数尤度のヘシアンの逆行列、あるいはグラディエントの外積の逆行列を求めればよいことがわかる。

1.3 Non Linear Least Squares

Non Linear Least Squares (NLS), 非線形最小二乗法は、モデルの残差二乗をパラメータに関する非線形関数とみなし、それを最小化させる手法である。残差二乗は二次形式ととして定義することが可能であり、様々な Weight 行列を考えることができる。NLS は、ML のように誤差項の分布に対して仮定を置く必要がないため、ミクロからマクロまで、非常に多くのケースで利用されている。ただし、非線形の最適化問題 (あるいは一階条件の非線形連立方程式の解法) を含むため、推計量を得ることは OLS と比べて遥かに難しく、得られた推計量の分布も漸近的性質に依存することになる。

一般的に

$$y = h(x, \beta) + e$$

で、誤差項が iid であるような過程を考える。このとき、NLS は、

$$\text{Min} (y - h(x, \beta))' (y - h(x, \beta))$$

を解くことで得られる。もしも、方程式が複数ある場合 (例えばマクロの資源制約とオイラー方程式)、方程式の数が s 本であれば、 $s \times s$ の正値定符号行列 Ω を用いて、

$$\text{Min} (y - h(x, \beta))' \Omega (y - h(x, \beta))$$

のような Weighted Least Squares を最小化させることで、 β を推計することができる。 Ω は正値定符号であれば、どのような行列でも構わないが、もっとも効率的な Weight Matrix が存在することが知られており、それは、GMM と呼ばれる手法と一致する。しかしながら、効率的であるはずの GMM (Hansen による Weight 行列の採用) は、必ずしも一般的に行われているわけではない。最適な Weight 行列を、我々は通常知らないで、データから推計せねばならないが、Weight 行列の成分、特に非対角成分の推計に関しては誤差が大きいため、small sample においては、 β の推計結果の誤差も非常に大きくなると可能性が指摘がされており、Weight 行列は使わない、あるいは単に、観察数をコントロールするための Weight 行列の使用が推奨されている (Altonji and Segal (1996))。とはいえ、マクロ経済学では、GMM は極めて Popular な推計手法であり、特にマクロの集計量を用いたオイラー方程式の推計に関しては、GMM Weight を用いることが伝統的に確立された手法となっている。ここでは、GMM は漸近的には効率的であるが、small sample の場合は、GMM 以外の推計量を試す価値は大きいとだけ指摘しておく。

さて、NLS

$$\text{Min} (y - h(x, \beta))' (y - h(x, \beta))$$

の解を $\hat{\beta}$ とする。このとき、一階条件より

$$(y - h(x, \hat{\beta})) \frac{\partial h}{\partial \beta} = 0$$

この一階条件を β に関して、テイラー近似を行うことで、 $\hat{\beta}$ の漸近分布を知ることができる。詳細は Amemiya(1985) 等の専門書に任せるが、結論としては、NLS の推計量 $\hat{\beta}$ は漸近的に下記の分布に従う。

$$\hat{\beta} \sim N \left[\beta, \frac{\sigma_2}{n} (Q_0)^{-1} \right],$$

$$Q_0 = p \lim \frac{1}{n} \frac{\partial h'}{\partial \beta} \frac{\partial h}{\partial \beta}$$

実践するときには

$$\hat{\beta} \sim N \left[\beta, \hat{\sigma}_2 \left(\frac{\partial h'}{\partial \beta} \frac{\partial h}{\partial \beta} \right)^{-1} \right]$$

と計算できる。 $\frac{\partial h'}{\partial \beta} \frac{\partial h}{\partial \beta}$ は、グラディエントの外積、すなわちヘシアン行列の近似となる。無論、この推計が可能であるためには、 $\frac{\partial h'}{\partial \beta} \frac{\partial h}{\partial \beta}$ が正則でなければならない。

1.4 Bootstrap

ML や NLS では、標準誤差は漸近分布から作成される。OLS と異なり、観察数が有限の場合は、NLS 推計量は正規分布に従うとは限らず、情報行列から求められる標準誤差も真の標準誤差とは異なるものになっている。特に、非線形性が強ければ強いほど、漸近分布と小標本の間の乖離は著しくなる傾向にあり、必ずしも十分な観察数を確保できない場合は深刻な問題となる。さらに、操作変数を用いる GMM 等を行う際には、漸近分布に全面的に依存することになり、有限標本の場合の標準誤差の推計量は著しく不正確になる可能性がある。

近年 (といっても 20 年以上の期間が過ぎているが)、漸近分布を用いず、経験分布から推計量の標準誤差を推計する手法が広く利用されるようになってきている。Bootstrap という、変わった名前の手法は、コンピューターの計算力が発達している近年では、極めて一般的に利用されており、今後は、線形モデル以外では Bootstrap による標準誤差推計が基本となる可能性もあるほどである。

詳細は Davidson and MacKinnon (2004) 等の計量経済学の教科書を参照してもらいたい。アイディアは極めてシンプルである。観察数が N の標本があるとすると、その中から、「重複を許して」、何度もリサンプリングを行うのである。例えば、50 回リサンプリングを行い、推計を行うと、50 個の推計量を得ることができる。得られた推計量の分布の標準偏差は、求めたい推計量の標準誤差になっているとみなすのである。Bootstrap は、極めて簡単な手続きで、中央値から、複雑な構造方程式の解まで、多くの標準誤差を求めることが可能となる。この手法には様々な Variation があるので、Efron の古典や近年の教科書を参照してもらいたい。ノイズを新たに加えたり、回帰の残差を用いる等、非常に多くの種類の Bootstrap が考案されている。

1.5 MATLAB による OLS, ML, NLS の比較

optimization toolbox にある `fminunc` と、前の講義ノートで紹介した数値 Gradient を計算するプログラムを用い、ここで紹介された三つの推計手法を比較してみよう。

推計するモデルは

$$Y = 1.5 + 2x_1 + 3x_2 + e$$

のように、線形であると仮定するが、これは OLS を ML や NLS と比較す

るためであって、ML や NLS の場合は、様々な非線形モデルを試してもらいたい。

まずメインプログラムである。

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% 2009年度応用マクロ経済学講義用
%
% OLS と ML、NLS の比較
% 線形モデル
%  $Y = 1.5 + 2x_1 + 3x_2 + e$ 
% の推定量 OLS と ML および NLS で比較
% ML を解くときの標準誤差の推計方法の比較
%
% 阿部修人
% 2009年11月13日
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
clear all
global Y X;
%
% まず、fake のデータ作成
%
n=50;
%
% 真のパラメター
%
beta=[1.5 2 3];%  $Y = 1.5 + 2x_1 + 3x_2 + e$ 
X=ones(n,3);
for i=1:n
X(i,2) = 2+0.1*randn(1);
X(i,3) = 3+1*randn(1);
end
Y=ones(i,1);
% 被説明変数の作成
for i =1:n
Y(i)=X(i,:)*beta'+0.1*randn(1);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% OLS による推計
%
```

```

betahat_ols = (inv(X'*X))*X'*Y;
res = Y-X*betahat_ols;
sigmasq = res'*res/(n-3);
sigmahat = inv((X'*X))*sigmasq;
std_ols=[sqrt(sigmahat(1,1)),sqrt(sigmahat(2,2)),sqrt(sigmahat(3,3))];
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ML による推計
param = ones(1,4);
options = optimset('MaxfunEvals', 20000, 'TolFun', 0.000000001, 'Max-
Iter', 100000, 'Display', 'iter');
options.HessUpdate=('bfgs');
[xopt,fval1]=fminunc('like1test', param);
betahat_ml=xopt;
global ind
%
% 標準誤差求める (1) グラディエントの外積
for ind =1:n
G1(ind,:)=numgrad1('mltgradfun',xopt);
end
COV2=inv(G1'*G1);
std_grad=[sqrt(COV2(1,1)),sqrt(COV2(2,2)),sqrt(COV2(3,3))];
%
%
% 標準誤差求める (2) Finite Differenc Hessian
H1=hessianvar('like1test',xopt);
cra2=inv(H1);
sdt_finhessian=[sqrt(cra2(1,1)) sqrt(cra2(2,2)) sqrt(cra2(3,3)) ];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% NLS による推計
param = ones(1,3);
options = optimset('MaxfunEvals', 20000, 'TolFun', 0.000000001, 'Max-
Iter', 100000, 'Display', 'iter');
options.HessUpdate=('bfgs');
[xop_nls,fval2]=fminunc('nlstest1', param);
for ind =1:n
G2(ind,:)=numgrad1('nlsgradfun',xop_nls);
end
for ind =1:n

```

```

res_nls(ind,:)=feval('nlsgradfun',xop_nls);
end
sigma_nls=(res_nls'*res_nls)/n;
COV_NLS=inv(G2'*G2)*sigma_nls;
std_grad_nls=[sqrt(COV_NLS(1,1)),sqrt(COV_NLS(2,2)),sqrt(COV_NLS(3,3))];
%
betahat_nls=xop_nls;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%bootstrap%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
data =[Y X];
m=500;
bootco_nls=zeros(m,3);
param=[1 1 1];
for i=1:m
for b=1:n
Xboot = ceil(n*rand(n,1));
data1(b,:)=data(Xboot(b,:),:);
end
Y=data1(:,1);
X=data1(:,2:4);
[xop_nls,fval2]=fminunc('nlstest1', param);
bootco_nls(i,:)=xop_nls;
end
std_bootstrap_nls=sqrt(var( bootco_nls));
% 出力
betahat_ols=betahat_ols';
betahat_ols
betahat_ml
betahat_nls
std_ols
std_grad
sdt_finhessian
std_grad_nls
std_bootstrap_nls

```

	betahat_ols N=50			betahat N=500		
OLS	1.5246	1.994	2.9915	1.4705	2.0087	3.0022
ML	1.5246	1.994	2.9915	1.4705	2.0087	3.0022
NLS	1.5247	1.994	2.9915	1.4705	2.0087	3.0022
	std			std		
OLS	0.2899	0.1447	0.0131	0.0813	0.0402	0.0041
ML(grad)	0.3127	0.1533	0.0133	0.0901	0.0441	0.0041
ML(finite hessian)	0.1039	0.0706	0.0112	0.025	0.0178	0.0036
NLS(grad)	0.2811	0.1403	0.0127	0.081	0.0401	0.004

推計結果

ここでは、モデルが線形であるので、OLS と ML および NLS のパラメータ推計量は完全に一致するはずである。しかしながら、標準誤差は、ML と NLS は漸近分布、OLS は Finite Sample の分布を出しているため、多少ずれる。

Table より、観察値数に依存せず、Finite Hessian による標準誤差は他の手法に比べて著しく異なっており、信頼性が低いことが分かる。