

基礎的な Regression の理論の仮定と対応

東京大学 大学院経済学研究科

河野 愛一郎

平成 20 年 10 月 23 日

本項では、基礎的な計量経済学の理論について説明する。

(1) OLS (Ordinary Least Square)

1. Assumption in Finite-Sample Theory

以下、 $y_i, x_{ik}, \varepsilon_i (i=1,2,\dots,n; k=1,2,\dots,K)$ をそれぞれ、regressand、regressor、error term とする (K は regressor の数、 n は sample の数である)。

(1) (linearity) $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_K x_{iK} + \varepsilon_i$

(2) (strict exogeneity) $E(\varepsilon_i | x_1, x_2, \dots, x_K) = 0$

(3) (no multicollinearity) $\{x_1, x_2, \dots, x_K\}$ is linearly independent.

(4) (homoskedasticity) $E(\varepsilon_i^2 | x_1, x_2, \dots, x_K) = \sigma^2$

(5) (no correlation between observations)

$$E(\varepsilon_i \varepsilon_j | x_1, x_2, \dots, x_K) = 0 (i, j = 1, 2, \dots, n; i \neq j)$$

(6) (normality of the error term) $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n | x_1, x_2, \dots, x_K\} \sim N(0, \sigma^2)$

以上の仮定に基づいて、 $\{\beta_1, \beta_2, \dots, \beta_K\}$ に相当するとされる OLS estimator が求まり、その estimator に対する Test (検定) が行われる。

2. Assumption に対する Test

(2) (strict exogeneity) $E(\varepsilon_i | x_1, x_2, \dots, x_K) = 0$

$E(\varepsilon_i | x_1, x_2, \dots, x_K) = 0$ を null hypothesis とする t-Test を行う。これが accept される、つまり、p-Value が大であることが好ましい。

(3) (no multicollinearity) $\{x_1, x_2, \dots, x_K\}$ is linearly independent.

ひとつの regressor を他の regressors で regression してみて、R-squared (以下、 R^2 とする) が高ければ、regression 同士で相関が発生 (multicollinearity) していることになる。この場合、estimator が正しい値を取らなかつたり、有意になりにくくなつたりするので問題である。一般には、

$$VIF \equiv \frac{1}{1 - R^2}$$

で定義される VIF (Variance Inflation Factor) が 10 を越えたら multicollinearity が発生していると考えられる。この VIF は、regressor それぞれにおいて存在するので、ひとつの regressors の set につき、その regressor の数の分だけ求めなければならない。

(4) (homoskedasticity) $E(\varepsilon_i^2 | x_1, x_2, \dots, x_K) = \sigma^2$

White Test によって判断される。以下、この方法を述べる。
まず、 ε_i^2 を他の regressor とその全ての second moment ($x_i x_j (i, j = 1, 2, \dots, n)$) で regress する (つまり、このときの regressor の総数は $K+1$ の H_2 となる)。その regression における R-squared を R^2 とすると、White test では、homoskedasticity が violate されればされるほど、 nR^2 は、degree of freedom (自由度) $K+1$ の χ^2 -distribution に従うとされる。つまり、null hypothesis が accept される方が良く、そのためには p-Value が大でなければならない。

ちなみに、 ε_i^2 は実際には観測不可能であるが、White Test では Large-Sample Theory に基づいて、OLS estimate の際に観測される deviation (残差) を使用している。そのため、(1) (linearity) (2) (strict exogeneity) (6) (normality of the error term)が violate されているかどうかに関わらず、Finite-Sample Theory における t-Test や F-Test を実施する。

(5) (no correlation between observations)

$$E(\varepsilon_i \varepsilon_j | x_1, x_2, \dots, x_K) = 0 \quad (i, j = 1, 2, \dots, n; i \neq j)$$

本来は Durbin-Watson Test を実施する。しかし、Time Series Data ではなく、時期を固定した Cross section Data を用いた分析では、通常、この Assumption が violate されるとは考えないので、Test を行わない。

(6) (normality of the error term) $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_K | x_1, x_2, \dots, x_K\} \sim N(0, \sigma^2)$

normality (正規性) の Test には、Shapiro-Wilk test を実施する。これは、normality に従うことを null hypothesis とした Test である。つまり、null hypothesis が accept される方が良く、そのためには p-Value が大でなければならない。

3. Test によって Assumption が violate されている場合

(1) (linearity) (2) (strict exogeneity) (6) (normality of the error term)が violate されている場合は、Finite-Sample Theory における t-Test や F-Test が有効ではなくなる。そこで、後述する Large-Sample Theory を使って Test を行う。しかし、この時点では OLS estimator そのものが無効になったわけではない。有効かどうかは Large-Sample Theory における Assumption を満たしているかどうかによる。

(3) (no multicollinearity) が violate されている場合は、multicollinearity を引き起こしている regressor を取り除くことで、再度、regression と Test を実施する。

(4) (homoskedasticity) が violate されている場合には、OLS estimator そのものが無効となる。その際の解決方法のひとつとしては、GLS (Generalized Least Square) estimator を使用する。その上で、t-Test や F-Test を行わなければならない。

(5) (no correlation between observations) が violate されている場合には、Large-Sample Theory を使用し、その Assumption も violate されているのであれば Time Series Analysis を実施する。

(2) GLS (Generalized Least Square) estimator

GLS では、(1) OLS in Finite-Sample Theory における Assumption のうち、(1) (linearity) (2) (strict exogeneity) (3) (no multicollinearity) が成立していれば、estimator は有効 (正確には unbiased で efficient) となる。Time Series Data を扱わない場合は、(5) (no correlation between observations) を前提にした Eicker-White の方法で GLS estimator を求めればよい。

ちなみに、 ϵ_i^2 は実際には観測不可能であるが、Eicker-White の方法では、一部、Large-Sample Theory に基づいて estimate を行っている。このため、White test の箇所でも述べたが、(1) (linearity) (2) (strict exogeneity) (6) (normality of the error term) が violate されているかどうかに関わらず、Finite-Sample Theory における t-Test や F-Test を OLS と同様、GLS に対しても実施すればよい。

(3) 線形重回帰分析で2つの改善方法

○まずは Finite-Sample Theory における Ordinary Least Squares = OLS (通常最小二乗法) で regression や Test を行う。

→しかし、OLS を行うには、いくつかの Assumption を満たしていることが前提

→regression の結果、もし Assumption を満たしていないなら、Large-Sample Theory を採用したり、別の推定法 (GLS や GMM など) を使う必要がある。

○決定係数 (R-squared) と、regressors の t-value を検証

・・・もし当てはまりが悪ければ、改善 (指標の求め方がミスかも) または、外して再度回帰分析を行う。

→繰り返して改善する。

・ 分析のアルゴリズム

