統計数理（1993）
第41巻 第1号 33-46

計量経済モデルと見せかけの回帰

統計数理研究所　川　崎　能　典

(1993年3月　受付)

1. はじめに

図1はHendry (1980) から引用した回帰分析の一例で，英国の消費者物価指数の年次データ P_t を「ある時系列 C_t」で説明しようというものである．両変数のcross-plotに基づき，定数項，C_t, C_t^2 を説明変数に選び，更に予測能力の向上のために残差が1階の自己回帰過程に従うとして得られた回帰式が図中左上に報告されている．$R^2=0.998$，係数はいずれも有意，1976年以降の模擬予測も極めて良好である．

しかしながら，消費者物価の動きを説明するこの時系列 C_t の正体は，英国の年間降雨量を年を追うごとに累積していったデータにすぎないのである．降雨量の累積値と物価の関係を説明する経済理論は世の中にないであろうし，逆に気象学の立場からも説明はつくまい．この回帰分析はナンセンスなのである．「計量経済学――錬金術か科学か？(Econometrics――Alchemy or Science?)」と表題をつけたこの論文において，応用計量経済での回帰分析の実態はこの極端な例からそう遠くない，とHendryは批判する．

二つの一変量時系列（平均差引済み）を与え，一方を他方に回帰することを考えよう．定常過程の枠内で議論すれば，生成過程が統計的に独立であれば，回帰係数（OLS推定量）は標本数 $T\uparrow\infty$ のとき0に確率収束する．しかし，二つの時系列が独立に発生させたランダムウォーク

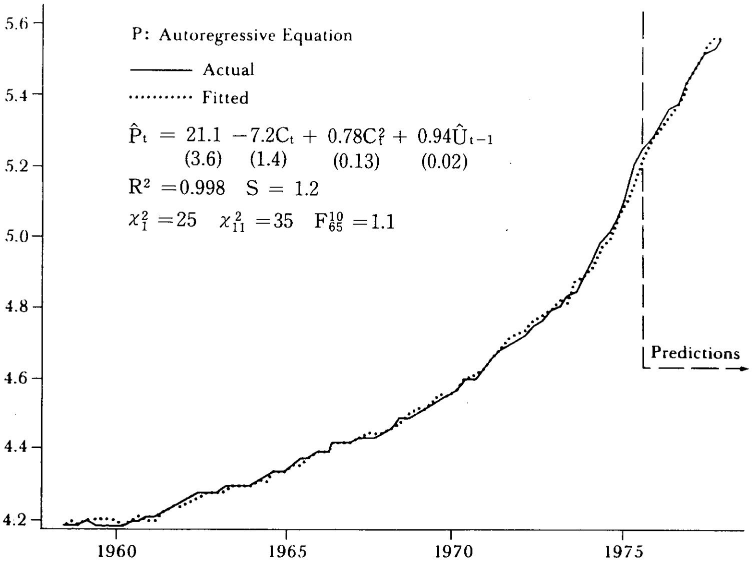

図1.　見せかけの回帰の一例（Hendry (1980) から引用）.

の場合にはOLS推定量は0には確率収束せず，極限において非退化の分布に従う．この事実を指して，spurious regression あるいは nonsense correlation などと呼ぶ．ここでは「見せかけの回帰」と一貫して表現する．

見せかけの回帰を扱った論文としては，古くはYule (1926) までさかのぼることができるが，経済学の分野でこの論点を再び強調したのはGranger and Newbold (1974) である．本稿では，まず次章でGranger and Newbold (1974) に示された数値実験とそこからのインプリケーションを振り返り，第3章でそれを解析的に裏付けたPhillips (1986) の業績を紹介する．Phillips (1986) での主要な道具はinvariance principle と continuous mapping theorem であるが，これらは単位根・共和分の検定統計量の極限分布を導く際にも共通のものである．我々が関心ある統計量の大部分はこの二つの道具のルーティーンワーク的な応用で極限分布が求められるが，その範囲をはみ出す統計量の扱いについても触れる．

こうしたナンセンスな回帰分析は古くから認識されていたにも拘らず，応用計量経済の分野で真剣に認識されるに至るには何度かの re-emphasize が必要であった．第4章では，計量経済学の教科書に掲げられる接近法の問題点を，見せかけの回帰と絡めながら検討する．見せかけの回帰が経済の実証分析に与えるマイナスのインパクトを考えれば，予備検定として単位根検定 (unit root tests) は重要な役割を果たしていると言えよう．第5章では，経済時系列のトレンドの記述という問題を中心に単位根検定について簡単に触れ，続いて便宜的なトレンドの処理が生み出す情報損失の危険性を回避する共和分 (cointegration) モデルについて解説する．

2. Granger-Newbold のシミュレーション

2.1 実証分析結果の傾向

国民所得分析のようなマクロ経済学のコアから，金融・財政・労働などの応用経済と呼ばれる分野にいたるまで，用いられる実証分析手法の大部分を線形回帰分析が占めていることは，広く認められる事実であろう．1970年代はじめのことだが，Granger-Newbold は各分野で「すぐれた実証分析」として経済学者に当時認知されていた実証分析結果，とりわけそこの報告されている回帰分析の結果を検討し，それらに共通して見かけられる傾向として以下の点を指摘した．

1) 報告されている回帰式の決定係数 (R^2) あるいは自由度修正済み決定係数 ($\bar{R}^2$) の値は極めて高く，ほとんど1に近い．2) 回帰係数の有意性を示す t 統計量は大きく，たいていの場合大きく2を越える．3) 一方，Durbin-Watson 統計量 (以下DW統計量) の値は極めて低く，0付近の値をとることが多い．彼らの指摘はおおむねこの3点であるといってよいだろう．

第3点に着目しよう．DW統計量とは，帰無仮説「回帰の誤差項が1次の自己回帰過程に従う」を検定するための統計量であり，回帰診断を構成する各種統計量の中でも最もポピュラーなものである．いま回帰の残差 (スカラーとしよう) を $\{e_t\}$, $t=1, \ldots, T$ と書くことにすればDW統計量は，

$$DW=\frac{\sum_{t=2}^{T}(e_t-e_{t-1})^2}{\sum_{t=1}^{T}e_t^2}$$

によって定義される．DW の分子で，T が大きいときに $\sum e_t^2$ と $\sum e_{t-1}^2$ を同一視すれば，

$$\hat{\rho}=\frac{\sum_{t=2}^{T}e_t e_{t-1}}{\sum_{t=1}^{T}e_t^2}$$

とするとき，$DW\simeq 2(1-\hat{\rho})$ であることが容易に確認できる．つまり，残差に系列相関がなけれ

ば DW は2に近い値をとるはずであり，DW が0に近いということは回帰モデルの残差系列が強い正の自己相関を持つことを示している．

残差に正の自己相関が生じている場合には，原因として回帰式に重要な変数が欠落している可能性が考えられるが，$\bar{R}^2$ がほとんど1に近いケースでは，そうした変数の欠落は考えにくい．一方経済時系列の動きはかなり緩慢であり，1次の自己相関は大きい．Granger-Newbold は，1)～3) の現象が，自己相関の強い時系列どうしをラグを考慮せずに回帰することで引き起こされる性質として導けるのではないかと考え，シミュレーションを行った．そこで得られた結論の多くは，後に Phillips (1986) によって解析的な裏付けを得た．

2.2 シミュレーションからの推測

Granger-Newbold は，X_t，Y_t という，各々標本数が50のランダムウォーク系列を，平均0分散1の正規乱数からそれぞれ独立に生成し，Y_t を定数項つきで X_t に回帰するという実験を100回繰り返した．

彼らの提出した回帰係数の t 統計量のヒストグラム（表1）によれば，X_t の回帰係数の有意性を伝統的な「t 比2の基準」に従って検定すると，100回中77回は「2つの変数は無関係」という帰無仮説を誤って棄却してしまう（表中の S は t 比）．t 比2の名目サイズは約5% であるから，over-rejection の方向に大きくバイアスがかかっていることになる．

また，この100回の実験での DW の平均は0.32で，2.1節での観測3) と整合的である．新たに独立なランダムウォークを説明変数に追加すると，平均値は若干上向くものの2.0にはほど遠く，依然として正の相関が残差に示される．一方 R^2 は100回の平均で0.26と極端に1に近いわけではないが，そもそも独立なプロセスで説明しようとしていることを考えれば，観測1) と矛盾するものではない．説明変数として新たに独立なランダムウォーク系列を追加していくにつれて，R^2 が0.7（この数字自体に深い意味はない）を越える頻度は増大し，見かけ上のあてはまりのよさは向上していく傾向にあることが報告されている．

計量経済モデルにおいては，概して予測性は重要視されず，回帰係数の有意性検定を通じてさまざまな経済学的仮説の当否を定性的に論ずることに重きが置かれている．だが，このシミュレーション結果の警告するところでは，経済時系列をレベル（あるいは対数レベル）のまま回帰分析に持ち込めば，モデルの解釈で重大な過ちを引き起こしかねないことになる．いかに R^2 で見てあてはまりがよく係数が有意であっても，DW の値が極端に低い回帰分析は見せかけの回帰である可能性が高いから，そこで主張されている結論には十分注意が必要になる．なお，第1章で挙げた Hendry (1980) の例では，残差にコクラン・オーカット法を適用しなければ $DW=0.1$ である．ラフな指標としては，決定係数より DW が低い回帰や，コクラン・オーカット法により DW が上方修正されている場合は注意が必要であるとされる．

表1．Granger and Newbold (1974) のシミュレーション結果．

S:	0–1	1–2	2–3	3–4	4–5	5–6	6–7	7–8
Frequency:	13	10	11	13	18	8	8	5
S:	8–9	9–10	10–11	11–12	12–13	13–14	14–15	15–16
Frequency:	3	3	1	5	0	1	0	1

3. Phillipsによる解析的裏付け

Granger-Newboldの研究は，経済時系列をレベル変数のまま扱って回帰式を特定化し推定することの問題点を指摘し，見せかけの回帰のいわば「傍証」を揃えたと言っていい．Phillips (1986) は見せかけの回帰のメカニズムを解析的に究明したが，そこに示された漸近理論はGranger-Newboldの実験結果の多くを説明する．

3.1 データ生成過程についての仮定

データ生成過程 (data generating process ; DGP) については，以後 $\{x_t\}$, $\{y_t\}$ という2つの一変量時系列が，各々独立なランダムウォーク過程にしたがって発生していると仮定する；

$$y_t = y_{t-1} + v_t, \qquad x_t = x_{t-1} + w_t, \qquad t = 1, 2, \ldots. \tag{3.1}$$

ここで v_t, w_t は平均0の独立同一分布でそれぞれ分散 σ_v^2, σ_w^2 を持つと仮定する．Phillips (1986) ではi.i.d.の仮定は緩められるが，それは確率論における結果 (McLeish (1975a), Herrndorf (1984)) を並行移入して行われる．要は，次節に述べるinvariance principleが成り立つ範囲で v_t, w_t に自己相関を許すのである．弱定常（かつshort-memoryの）ARMA過程で表現されるような相関構造を持つプロセスはMcLeishやHerrndorfの議論の範疇におさまるので，DGPは1階の階差で弱定常（すなわち $I(1)$ 過程，integrated of order 1）としても以下の結論に変わりはない．また，多変量でのinvariance principleの成立を示すのに v_t と w_t の共分散が0である必要はないが，「独立なプロセスである」という設定上見せかけの回帰では0である．プロセスの初期値 (x_0, y_0) については，1) 確率1で或る定数，2) 特定の既知の分布に従って発生，のいずれを仮定してもよいが漸近理論の導出には $x_0 = y_0 = 0$ としても全く支障はない．以上の枠組みのもとで最小二乗回帰

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{u}_t, \qquad t = 1, \ldots, T\,; \tag{3.2}$$

を行ったときの $\hat{\alpha}, \hat{\beta}, t_\alpha$ と t_β (α, β の t 比), R^2, DW などの極限分布，あるいは極限分布が存在するためのノーマライザーを求める際に有効な道具がinvariance principleである．

3.2 Invariance principle

本節では，確率論でよく知られた結果であるinvariance principleについて解説しておく．本節の内容は，畠中 (1991) の簡潔な記述ならびにBillingsley (1968) によった．以後 $\xrightarrow{\mathscr{D}}$ は法則収束 (convergence in distribution), $\Rightarrow$ は弱収束 (weak convergence), $W(\cdot)$ はウィーナー過程，$[M]$ は M を越えない最大の整数を表す．

$\{\xi_t\}$ がi.i.d.で $E(\xi_t)=0$, $E(\xi_t^2)=\sigma^2$ のとき，$S_T = \xi_1 + \cdots + \xi_T$ とすると，ξ_t の分布によらず，$T \uparrow \infty$ とともに $(\sqrt{T}\sigma)^{-1} S_T \xrightarrow{\mathscr{D}} N(0,1) (\equiv W(1))$ となる．これは中心極限定理 (CLT) であるが，invariance principleはその一般化である．いま

$$X_T(r) = \frac{1}{\sqrt{T}\sigma} S_{[Tr]}, \qquad 0 \le r \le 1 \tag{3.3}$$

とおけば，$X_T(r)$ は区間 $[0,1]$ で定義された確率過程である．このとき ξ_t の分布によらず，$T \uparrow \infty$ とともに $X_T(r) \xrightarrow{\mathscr{D}} W(r), 0 \le r \le 1$ が成り立つ．これは，$[0,1]$ 上のrandom elementの列 $\{X_T\}$ が誘導する確率測度（に対応する分布）がウィーナー測度（による分布）に弱収束するということであり，pathの分布を議論している．右辺の W はrandom elementでその分布がウィーナー測度により定まると考えてもよいし，W を測度と思ってもよい．後者なら X_T

の分布を例えば P_T として $P_T \Rightarrow W$ と書くべきところであるが，$X_T \xrightarrow{\mathscr{D}} W$ のような 'hybrid terminology' が実際によく用いられる（Billingsley (1968), p. 24 を参照）.

さて，この結果は CLT の拡張であり，functional CLT あるいは極限分布が ξ_t の分布によらないことを強調して invariance principle と呼ばれる（Billingsley (1968), Sections 10 and 16）. Donsker の定理のように連続な random element を考えてもよいが，(3.3) のように見本過程が右連続関数となるように random element を与える方が直接的である（我々のケースではイノベーション列の基準化された部分和）.

更に，random element $X_T(r)$ $(0 \leq r \leq 1)$ がある連続な写像で移った先の極限分布は，その同じ写像による $W(r)$ $(0 \leq r \leq 1)$ の像の分布に等しい（Continuous mapping theorem, Billingsley (1968), pp.29-31, Theorem 5.1 ならびに Corollary 1）.

単位根を持つ非定常過程の各種統計量の極限分布を求めるのに，invariance principle と continuous mapping theorem が有効な方法であることを最初に示したのは White (1958) であろう. Phillips (1986) は確率論で知られた結果から DGP に関する前提条件を緩めた上で，White のやり方を見せかけの回帰に応用したと言ってよいだろう.

3.3 推定量・回帰診断統計量の漸近分布

さて，$\hat{\beta}$ などの極限分布を扱う前に，基本的な統計量について，非退化の極限分布ならびにそれが存在するためのノーマライザーのオーダーを見ておこう. 例えば $\sum x_t$ については，

$$
\begin{aligned}
(3.4) \qquad T^{-3/2}\sum_{1}^{T} x_t &= \sigma_w T^{-1}\sum_{i=1}^{T}\left\{\frac{1}{\sqrt{T}\sigma_w}\sum_{1}^{i-1} w_t\right\} + T^{-3/2}\sum_{i=1}^{T}(w_i + x_0)\\
&= \sigma_w \sum_{1}^{T}\int_{(i-1)/T}^{i/T} X_T(t)dt + o_p(1)\\
&= \sigma_w \int_0^1 X_T(t)dt + o_p(1)\\
&\xrightarrow{\mathscr{D}} \sigma_w \int_0^1 W(t)dt \qquad \text{as } T\uparrow\infty
\end{aligned}
$$

のごとくである. $X_T(t)$ は $[0,1]$ 上の右連続階段関数であり区間 $((i-1)/T, i/T]$ 上では t に依存せず定数であること，$T^{-1}=\int_{(i-1)/T}^{i/T} dt$ などに注意されたい. 残余項が $o_p(1)$ であるのは，初期条件 i), ii) のいずれにおいても $T^{-1/2}x_0 = o_p(1)$ であること，大数の法則から $T\uparrow\infty$ のとき $T^{-1}\sum_1^T w_i \to 0$ a.s. が示されることによる（独立同一分布の仮定を外して議論するときは，McLeish (1975b) に示される大数の強法則を利用する）. 最終行での法則収束では invariance principle と continuous mapping theorem が適用される. 同様にして，

$$
(3.5) \qquad T^{-2}\sum_{1}^{T} x_t^2 \xrightarrow{\mathscr{D}} \sigma_w^2 \int_0^1 W(t)^2 dt \qquad \text{as } T\uparrow\infty
$$

を得る. 次式を導くには別の議論を経由しなければならないのだが，それについては次の 3.4 節で触れることとして結果を先に述べておくと

$$
(3.6) \qquad T^{-2}\sum_{1}^{T} x_t y_t \xrightarrow{\mathscr{D}} \sigma_w \sigma_v \int_0^1 V(t)W(t)\,dt \qquad \text{as } T\uparrow\infty
$$

が成立する. ここで $V(t)$ と $W(t)$ は独立なウィーナー過程である. 見せかけの回帰たるゆえんが (3.6) 式に表れているといえよう. また (3.4)～(3.6) から，定常過程では $\sum x_t$ は $O_p(T^{1/2})$, $\sum x_t^2$ は $O_p(T)$ であるが，データが $I(1)$ であれば $\sum x_t$ は $O_p(T^{3/2})$, $\sum x_t^2$ は $O_p(T^2)$ である.

最終的に関心のある推定量や統計量の極限分布については，例えば $\hat{\beta}$ であれば，

$$\hat{\beta}=\frac{\sum y_t(x_t-\bar{x})}{\sum(x_t-\bar{x})^2}=\frac{T^{-2}\sum y_t x_t-T^{-1}\bar{y}\bar{x}}{T^{-2}\sum(x_t-\bar{x})^2}$$

として，これに invariance principle と continuous mapping theorem を適用する．同様の議論の繰り返しで，見せかけの回帰について以下の結論にいたる．$T\uparrow\infty$ のとき，

1. $\hat{\beta}$ は，独立なウィーナー過程 $V(t)$, $W(t)$ $(0\le t\le 1)$ の汎関数で表される確率変数に法則収束する．即ち OLS 推定量は一致性を持たない；

$$\hat{\beta}\xrightarrow{\mathscr{D}}\frac{\sigma_v\left\{\int_0^1 V(t)W(t)dt-\int_0^1 V(t)dt\int_0^1 W(t)dt\right\}}{\sigma_w\left\{\int_0^1 W(t)^2dt-\left(\int_0^1 W(t)dt\right)^2\right\}}. \tag{3.7}$$

2. R^2 も非退化の極限分布を持つ．
3. $\hat{\alpha}$, t_α, t_β は，$T^{-1/2}\hat{\alpha}$, $T^{-1/2}t_\alpha$, $T^{-1/2}t_\beta$ とノーマライズして非退化の極限分布を持つ．
4. DW は 0 に確率収束する．

R^2, t_β などについては，その極限分布の表現が煩雑を極めるのでここでは記さない．詳細は Phillips (1986) を参照されたい．

3,4 は Granger-Newbold のシミュレーション結果と整合的である．3 は，t 比が $\sqrt{T}$ のオーダーで発散していることを示すから，見せかけの回帰のもとで t 比 2 の基準に従えば，標本数の増大に伴って回帰係数の P-value は小さくなっていく．さらにこの点について Granger-Newbold との関係で言及すれば，彼らは自分達のシミュレーションの結果から，5% 有意水準の critical value として 11.2 なる数値を提唱したが，これは標本数が 50 のときしか目安にならないことになる．

3.4 多変量での問題点

ここでは，invariance principle と continuous mapping theorem をルーティーンワーク的に適用しただけでは解決できない問題について述べる．見せかけの回帰に帰着させて言えば，$T\uparrow\infty$ のとき，

$$\begin{pmatrix} T^{-1}\sum x_{t-1}w_t & T^{-1}\sum x_{t-1}v_t \\ T^{-1}\sum y_{t-1}w_t & T^{-1}\sum y_{t-1}v_t \end{pmatrix}\xrightarrow{\mathscr{D}}\begin{pmatrix} \sigma_w^2\int_0^1 W(t)dW(t) & \sigma_v\sigma_w\int_0^1 W(t)dV(t) \\ \sigma_v\sigma_w\int_0^1 V(t)dW(t) & \sigma_v^2\int_0^1 V(t)dV(t) \end{pmatrix}$$

をいかにして言うかである．

3.1 節の設定を多変量にし，$\{x_t\}$ は n 変量ベクトル値時系列で，DGP は以下のようであるとしよう．

$$x_t=x_{t-1}+u_t, \qquad t=1,2,\ldots$$

ここで $\{u_t\}$ は i.i.d. $(0,\Sigma)$ とする．$X_T(r)=T^{-1/2}\sum_1^{[Tr]}u_j$ としてスカラーの場合と同様に random element $X_T(r)$ を構成すると，適当な条件の下で多変量でも invariance principle が成立する；

$$X_T(r)\xrightarrow{\mathscr{D}}W(r) \qquad \text{as } T\uparrow\infty.$$

ここで $W(r)$ は n 次元のウィーナー過程で，その分散共分散行列は Σ になる．しかしこのと

き，$T^{-1}\sum x_{t-1}u_t'$ は $X_T(r)$ の連続汎関数として表せないので，3.3 節と同じ議論をたどって

$$T^{-1}\sum x_{t-1}u_t' \xrightarrow{\mathscr{D}} \int_0^1 WdW' \qquad \text{as } T\uparrow\infty$$

と結論できない（例えば Phillips (1988a), p. 255 を見よ）．この法則収束の証明としては Stock (1987), Phillips (1988a) があるが，証明の過程が複雑である．しかし Chan and Wei (1988) の Theorem 4.2 を利用すれば議論は簡潔である．Chan and Wei によれば，$\mathscr{F}_t$ を（u_t, u_{t-1}, ...）により生成される σ-field とするとき，$\{u_t\}$ が $\mathscr{F}_{t-1}$ に関して 2 乗可積分マルチンゲール差になっていれば，上の statement は成立する．Phillips (1988b) では $\{u_t\}$ に invariance principle が成立する範囲で相関を許し，u_t をマルチンゲールで近似して Chan and Wei の定理を援用し，

$$T^{-1}\sum x_{t-1}u_t' \xrightarrow{\mathscr{D}} \int_0^1 WdW'+\Lambda \qquad \text{as } T\uparrow\infty$$

の成立を示している．ただしここで $\Lambda=\sum_{k=1}^{\infty}E(u_0u_k')$ である．

4. 教科書的な接近法と見せかけの回帰

Granger and Newbold (1974) の指摘を振り返れば，当時学術雑誌レベルで報告される実証分析においてすら DW 統計量のような基本的な回帰診断統計量が軽視されていたのはたいへん奇妙なことのように思われる．この原因は，いったいどういうところに求められるだろうか．

図 2 は，計量経済モデルを作成するときの極めて典型的な接近法である．この図は Spanos (1986) からの引用であるが，ちなみに Spanos (1986) の文脈では，この接近法はひな型というよりはむしろ，批判の対象としてのこれまでの伝統的な接近法というニュアンスで挙げられている．こうした教科書的接近法（textbook approach）が見せかけの回帰の観点から問題なのは，経済理論から導かれる関係式の多くがレベルの観点で記述されているため素直に「理論の計測」を試みれば統計的推測が困難になることである．

いま貨幣需要を例に textbook approach をなぞってみよう．ある経済理論の立場からは，貨幣需要（M^D）は国民所得（Y）・物価水準（P）・利子率（I）によって決まるとされる．こうした依存関係が何らかの関数形 f で表されるという程度の意味でシンボリックに

$$M^D=f(Y,P,I)$$

と書いたのが，図中の Theory に該当する．Theoretical model とは，これに数学的に明示的な表現を与えたもので，たとえば

$$M^D=AY^{\alpha_1}P^{\alpha_2}I^{\alpha_3}$$

あるいは

$$\ln M^D=\alpha_0+\alpha_1\ln Y+\alpha_2\ln P+\alpha_3\ln I$$

などが theoretical model に相当するものである．このような特定の関数形に基づいて定性的分析を行うのが狭い意味での経済理論と言ってよかろう．

更に M^D, Y, P, I という理論に登場する変数につき，これら理論値の測定値として，利用可能ないくつかのデータが選ばれる．観測可能なデータ，$\tilde{M}_t$, $\tilde{Y}_t$, $\tilde{P}_t$, $\tilde{I}_t$ についてそれぞれ測定誤差がないものと仮定して貨幣需要関数を改めて

$$\ln\tilde{M}_t=\alpha_0+\alpha_1\ln\tilde{Y}_t+\alpha_2\ln\tilde{P}_t+\alpha_3\ln\tilde{I}_t$$

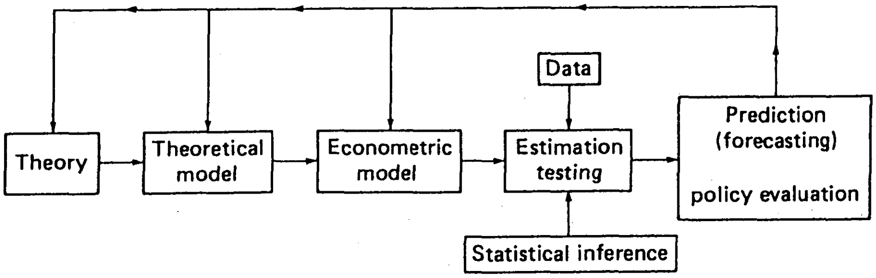

図2. 計量経済学の「教科書的アプローチ」(Spanos (1986) より引用).

と特定化し，後は方程式に誤差項を加えて出来上がるのが econometric model である．誤差分布に適当な仮定をおいてモデルを最小二乗推定し，回帰係数の有意性検定や各種診断統計量をたよりに経済理論の定性的・定量的分析を行う，これが statistical inference である．

計量経済学の方法論全般を議論するのが本稿の主旨ではないので，このフローチャートに対する批判は Spanos (1986) を参照して頂きたい．見せかけの回帰は，現象としてはこの図の statistical inference において生じる問題である．$\tilde{M}_t$, $\tilde{Y}_t$, $\tilde{P}_t$, $\tilde{I}_t$ といった現実のデータの自己相関構造を検討せずに，理論モデルから導かれる関係（多くはレベル変数で記述される）を盲目的に最小二乗推定すれば，推定は形式上は可能であり数値ははじき出されるものの，それに対し statistical inference は不能であることは3章に説明した．あるいは現在の知見をもって判断すれば，これらの回帰は見せかけの回帰であったのかもしれないし，5章に解説する共和分関係を推定していたのかもしれない．しかしいずれにしても推定したモデルからは何の推論も行えない．Granger and Newbold (1974) の言葉を借りれば，一部の経済学者にとっては経済理論から導かれる関係式は，あたかも「存在自体が真」であるかのように疑われることがなく，彼らは「レベル変数で盲目的に回帰分析を行うことで，とうてい受け入れがたい結論を生み出し」ていたわけである．

また，一見理論を信頼するかのようにレベルでの回帰に固執する姿勢の背後には，R^2 低下への不安がうかがえる．与えられたデータをレベルでなしに伸び率で見ることに当時の経済学者がいかに消極的であったかは，たとえば Plosser and Schwert (1978) に描かれている．一部の経済学者には，理論から導かれる式が R^2 の観点からサポートされることが重要であって，データから仮説を検証するというよりは，仮説が支持されるようなデータセット・計測期間を繰り返し探索する傾向にあり，この姿勢は「データ発掘 (data mining)」として批判の対象となった．

更に，理論への過度の信頼に基づくこのような姿勢を助長する背景として，経済学では実証から理論へのフィードバックが働きにくいことが挙げられよう．経済データは多くが指数などの加工系列である上サンプルサイズが不十分であることが多く，解釈に苦しむ実証結果が得られた場合でも，理論に疑いの目を向けるより「データがおかしい」と分析者が判断を下すことも少なくない．図2においてもデータはただ代入するだけであって，データの性質の検討やデータから何が推定可能かのチェックは欠落している．

また，見せかけの回帰に基づく予測が naïve forecast にほかならないことは明らかだろう．冒頭に挙げた Hendry (1980) の例に戻るが，残差に自己相関を入れなければ，R^2 が非常に高いにも拘らず予測は極めて悪い（図3参照）．

見せかけの回帰は，教科書的な計量経済分析にとっては非常にネガティブなものであり，改めて経済理論とデータの折り合いをつけることの難しさを感じさせる．経済時系列をレベルの

まま扱うことに伴う危険性（非定常変数を含む回帰の問題）について80年代を通して広く認識されていったことの背景には，予備検定としての単位根検定（unit root test）の理論的発展と分布表の提示に伴って「非定常か否か」の検定が最小二乗推定に基づいて簡単に行えるようになったこと，また同時期に従来の大型マクロ計量モデルに対する批判が浴びせられただけでなく，実証分析のありかたそのものにも疑問が投げかけられたことが考えられよう（Sims (1980), Hendry (1980), Leamer (1983), Pagan (1987) など）.

5. 単位根検定と共和分モデル

本章では，単位根検定と共和分モデルについて触れる．網羅的解説を目指すものではなく，前者については経済時系列のトレンドの記述という問題を中心に，後者については便宜的なトレンド処理がもたらす情報の損失を中心に解説する．

5.1 単位根検定

マクロの消費理論のように，ある変数が「ランダムウォークである」ことが理論モデルの解析の結果得られることもあるので，単位根検定の存在意義を見せかけの回帰だけに還元はできないが，実証分析に予備検定として重要な意味を持つことは異論のないところであろう．分析に用いる変数が「非定常であるとの仮説を棄却できない」のであれば，何らかの形でトレンドの除去を行ってから分析を開始するのでない限り，報告された回帰分析の信用性は著しく低いであろう．

ただ，当該経済変数が本当に確率的トレンドを持つのかと，個々の経済時系列のもつトレンドの性質がいかなるものであるかは，また別問題である．あるデータが，非確率的トレンドの周りに定常変動しているのと，確率的トレンドに支配されているのとでは，経済学的意味あいは大きく違う．イノベーションの線形結合で表現すれば明らかだが，前者においては過去に発

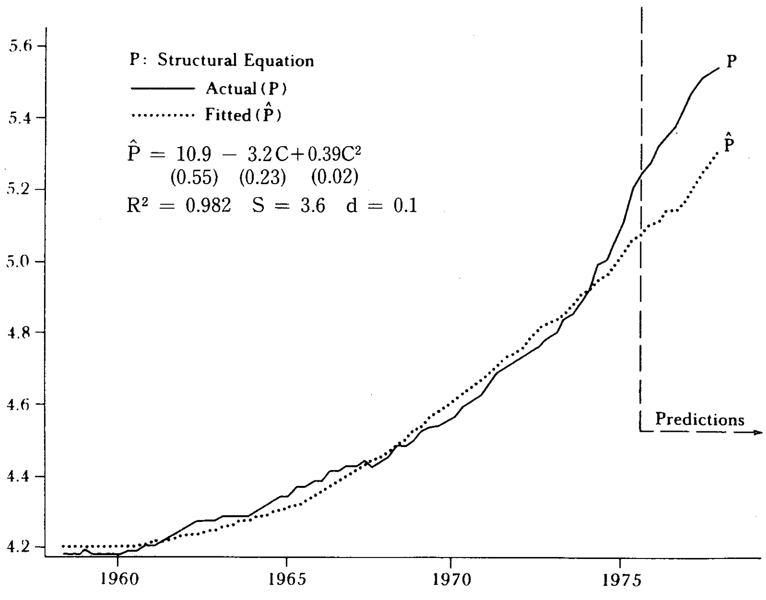

図 3.

生したショックの影響は時点が遠ざかるにしたがって0に収束するが，後者においてはショックの影響は恒久的である．一方，実際問題として頻繁に指摘されることだが，これまで提案されてきた単位根検定は非確率的トレンド周りに定常変動しているデータに対し，$I(1)$と誤認してしまうことがままある．

経済データのトレンド成分の記述についての論調を簡単にまとめれば，80年代は$I(1)$あるいは外生トレンド$+I(1)$とみなす向きが支配的で確率的トレンドの介在を積極的に認めてきたが，最近では必ずしもそうではない．前者についてよく引用されるのがNelson and Plosser (1982)であるが，そこでは大部分の経済時系列について「$I(1)$である」との仮説を棄却できないことが示されている．彼らの結論には勿論一定の留保が必要である．まず第一に，Dickey and Fuller (1979) をはじめ，これまで提案されてきた検定統計量のほとんどは，非定常が帰無仮説という点である．更に最近では，非確率的トレンドの特定化如何によっては，必ずしも$I(1)$と結論されないとの報告もある（構造変化を *a priori* に考慮し，それを非確率的トレンドのkinkやjumpとして与えるやりかたである）．制度的変化などの影響が考えられる場合には，特定の変節点を外生的ショックによるものとみなして構造変化を考慮するほうが，標本期間にわたって不変な線形トレンドを対立仮説に持ってくるやり方よりはフェアであろう．が，こうしたアプローチでは，事前に構造変化を与えるときの変化点の判断の仕方が問題である上，モデルの細分化を進めていけば，幾重にも検定を張り巡らす必要に駆られる点が問題である．

単位根検定のサーベイは少なからず存在するため本稿では検定の具体的な紹介はしないが，検定の手順がうまくまとめられているものとしてDolado et al. (1990) がある．またKunitomo (1992) では煩雑になりがちな検定統計量の分布の表現が重回帰の観点から整理されている．

5.2 共和分モデル

経済時系列のトレンド性質がどのような形で記述されることが適当か (trend versus random walk) については，現在のところ統一的な見解は存在しないし，明確な結論を得るのは難しいであろう．実証分析に対して見せかけの回帰が持つ危険性は深刻である一方，実際のデータ解析の場ではこれを避けるために何らかの方法でトレンドを除去することがまず要請される．経済理論（特に成長理論）はトレンドの描写にあたってデータ解析上有益な構造を与えてくれそうにもなく，現在のところトレンドの処理は便宜的なものにとどまっているといえよう．なかでも実際上最もポピュラーなトレンド除去の方法は階差 (differencing) であろう．データを対数変換した後に階差をとればそれは前期比伸び率であり，当該データが$I(1)$という前提が妥当なら，伸び率で議論する分には見せかけの回帰からは逃れることができる．

しかし一方で，このようなアプローチでは一般に情報の損失が起こっている．例として，次のような多変量ARからデータが発生しているとしよう．

$$x_t=\mu+\sum_{i=1}^{k}A_i x_{t-i}+\varepsilon_t \tag{5.1}$$

ここで，x_tは$p\times1$ベクトルで各成分は$I(1)$と仮定する．また，μは定数ベクトル ($p\times1$), A_iは$p\times p$の係数行列，誤差ベクトルε_tは独立に平均0分散Λの多変量正規分布に従うと仮定する．この多変量ARモデルを，左辺が階差系列Δx_tとなるように変形し，次の表現を得る；

$$\Delta x_t=\mu+\sum_{i=1}^{k-1}\Gamma_i\Delta x_{t-i}+\Pi x_{t-k}+\varepsilon_t \tag{5.2}$$

ただしここで，

(5.3) $$\Gamma_i = -\left(I_p - \sum_{j=1}^{i} A_j\right)$$

(5.4) $$\Pi = -\left(I_p - \sum_{j=1}^{k} A_j\right)$$

である．

この変形から明らかに見てとれるように，差分系列で議論しても情報の損失が生じないのは，Π が零行列の場合に限る．しかし，Π が零行列でないときに両辺が釣りあうためには，線形結合 Πx_{t-k} が定常となることが必要である．この場合を指して，複数の $I(1)$ 変数間に「共和分 (cointegration) が存在する」と定義するのである．

x_t の各成分が $I(1)$ という仮定のもとでは，Π は零行列かランクが落ちているかのいずれかであり，後者の場合には Π は，$p \times r$ $(r<p)$ 行列 α, β によって $\Pi = \alpha\beta'$ と表現される (Engle and Granger (1987), Johansen (1991))．このような分解は一意ではないが，β' の第1列をすべて1とするなど，恣意的な基準化を行えば α, β を定めることはできる．各成分が定常な $r \times 1$ ベクトル $\beta' x_{t-k}$ を作り出す行列 β の各列ベクトルは共和分ベクトル (cointegrating vectors) と，行列 α の各行ベクトルは調整ベクトル (loading vectors) と呼ばれる．共和分が存在すれば，行列 Π は必ずランク落ちしている．これは各成分の確率的トレンドの間に共線性が存在することの反映であり，その情報を追加した分説明能力の向上が期待されるであろう．

共和分ベクトルの推定もさることながら，多変量システムでは何通りの共和分が存在するか (行列 Π のランク) をデータから決めることが重要である．2変数であれば共和分が存在すれば1通りであるが，一般の p 変量では1個以上 $p-1$ 個まで考えられる．換言すれば，システムを支配している確率的トレンドは何本か，である．このようなやり方は Johansen (1988) に始まったが，Johansen の確立した共和分の推定･検定法について詳述することは本稿の目的外であるので，例えば川崎 (1992) を参照されたい．

一方このモデルの解釈であるが，(5.2) は経済学では「誤差修正モデル (Error Correction Model; ECM)」と呼ばれている．(5.2) で，定常変数を生み出す線形結合を経済システム内の長期的な均衡関係式とみなし，過去に生じた誤差 (均衡からの外れ具合) が均衡に向けて修正される動きが現在の短期的変動にも影響を与える，という解釈を施すのである．こうした均衡からのズレは不確実性や調整費用の存在に対応しているといえる．このように，1) 時系列モデルを起点としながらも簡明な経済学的解釈を与えること，2) 差分のみによる議論よりフィットの向上が期待されることから，近年実証分析で頻繁に利用されている．ただ，先述した Π の分解の一意性のなさから，均衡関係といっても恣意的なものといえよう．

ここまでは情報の損失という観点から共和分を眺めてきたが，最後に共和分が見せかけの回帰の重要な例外であることを確認しておこう．共和分とは複数の非定常変数の関係を指すから (5.1) のような多変量システムを考えるのが自然であるが，そもそもは単純に二つの時系列の差を作って，それを単一方程式の説明変数に加えるというようなやりかたが試みられていた．こうしたアドホックなやり方から脱するべく推測理論が登場するまでには約20年を要したが，Stock (1987) は二つの非定常時系列の線形結合を回帰で推定した場合の OLS 推定量の性質を論じた．形式的にはこれは見せかけの回帰と同じである．しかし，見せかけの回帰では OLS 推定量が一致性を持たなかったのに対し，二つの時系列に共和分が存在するケースでは OLS 推定量が真の値 (真の共和分ベクトル) に収束する速さは T^{-1} である．これは定常ケースでの OLS 推定量の収束スピード $T^{-1/2}$ よりも速く，super consistency と呼ばれている．Engle and Granger (1987) の2段階法はこの super consistency を利用し，誤差修正項を回帰の残差として求めてから誤差修正モデルに代入し再び OLS を実行するというものであり，Johansen の方

法と並んで計量経済分析ではよく用いられるやり方である.

謝 辞

1992年の夏学期に開講された一橋大学大学院における演習への参加を快諾くださった田中勝人氏からは，同セミナーで多くのことを教えていただきました．また，本稿の誤りならびに不適切な表現を指摘し，改稿にあたって参考になるコメントをしてくださった査読者の方に感謝いたします.

参 考 文 献

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.
Chan, N.H. and Wei, C.Z. (1988). Limiting distributions of least square estimators of unstable autoregressive processes, *Ann. Statist.*, **16**, 367-401.
Dickey, D.A. and Fuller, W. (1979). Distribution of estimators for autoregressive time series with a unit root, *J. Amer. Statist. Assoc.*, **74**, 427-431.
Dolado, J.J., Jenkinson, T. and Sosvilla-Rivero, S. (1990). Cointegration and unit roots, *Journal of Economic Surveys*, **4**, 249-273.
Engle, R. and Granger, C.W.J. (1987). Co-integration and error correction; representation, estimation and testing, *Econometrica*, **55**, 251-276.
Granger, C.W.J. and Newbold, P. (1974). Spurious regressions in econometrics, *J. Econometrics*, **2**, 111-120.
畠中道雄 (1991). 『計量経済学の方法』, 創文社, 東京.
Hendry, D. (1980). Econometrics —— alchemy or science?, *Economica*, **47**, 387-406.
Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables, *Ann. Probab.*, **12**, 141-153.
Johansen, S. (1988). Statistical analysis of cointegration vectors, *J. Econom. Dynamics Control*, **12**, 231-254.
Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, **59**, 1551-1580.
川崎能典 (1992). Johansenの共和分検定について, 金融研究, **11**, 99-120.
Kunitomo, N. (1992). Tests of unit root hypotheses in econometric models, Discussion Paper, 92-F-2, Faculty of Economics, University of Tokyo.
Leamer, E.E. (1983). Let's take out the con out of econometrics, *American Economic Review*, **73**, 31-44.
McLeish, D.L. (1975a). Invariance principles for dependent variables, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **32**, 165-178.
McLeish, D.L. (1975b). A maximal inequality and dependent strong laws, *Ann. Probab.*, **3**, 829-839.
Nelson, C.R. and Plosser, C.I. (1982). Trends and random walks in macroeconomic time series, *Journal of Monetary Economics*, **10**, 139-162.
Pagan, A. (1987). Three econometric methodologies: a critical appraisal, *Journal of Economic Surveys*, **1**, 3-24.
Phillips, P.C.B. (1986). Understanding spurious regressions in econometrics, *J. Econometrics*, **33**, 311-340.
Phillips, P.C.B. (1988a). Weak convergence to the matrix stochastic integral $\int_0^1 BdB'$, *J. Multivariate Anal.*, **24**, 252-264.
Phillips, P.C.B. (1988b). Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations, *Econometric Theory*, **4**, 528-533.
Plosser, C.I. and Schwert, G.W. (1978). Money, income and sunspots: measuring economic relationships and the effects of differencing, *Journal of Monetary Economics*, **4**, 637-660.
Sims, C.A. (1980). Macroeconomics and reality, *Econometrica*, **48**, 1-48.

Spanos, A. (1986). *Statistical Foundations of Econometric Modelling*, Cambridge University Press, Cambridge.

Stock, J.H. (1987). Asymptotic properties of least-squares estimators of cointegrating vectors, *Econometrica,* **55**, 1035-1057.

White, J.S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case, *Ann. Math. Statist.,* **29**, 1188-1197.

Yule, G.U. (1926). Why do we sometimes get nonsense-correlations between time series?――A study in sampling and the nature of time series, *J. Roy. Statist. Soc.,* **89**, 1-64.

Econometric Models and Spurious Regressions

Yoshinori Kawasaki

(The Institute of Statistical Mathematics)

Spurious regressions and other related topics in non-stationary time series analysis are explained. We briefly review Granger and Newbold's simulation study on spurious regressions and then the analytic results given by Peter Phillips. To avoid spurious regressions, unit roots tests play an important role as pre-tests, and cointegration approaches serve to recover the possible information loss caused by such a conventional detrending method as differencing. We also discuss some problems of usual 'textbook approach' in econometrics and also the difficulty of trend specification for macroeconomic time series.

Key words: Spurious regression, unit root test, cointegration, invariance principle.