日本統計学会誌
第 43 巻, 第 2 号, 2014 年 3 月
157 頁 ～ 183 頁

計測誤差と統計学

国友　直人*

Measurement Errors and Statistics

Naoto Kunitomo*

経済時系列において生じる計測誤差を巡る最近の話題をきっかけに，経済に関連する経済統計 (GDP 公表値の改訂と季節調整問題)，計量経済 (構造方程式の推定問題)，計量ファイナンス (拡散過程とマーケット・マイクロストラクチャ・ノイズ (market-microstructure-noise)，など応用統計分野における計測誤差の統計的処理の問題を振り返る．これまでに行われた計測誤差の統計学的分析から浮上する計測誤差と統計学についての一般的な論点を議論し，ビッグ・データ (big data) や近年のビジネス解析 (business analytics) における統計学の重要性の高まりが話題となっている現代日本における統計学・統計科学の今後の方向を展望する．

In this lecture we illustrate several measurement errors issues and their statistical analyses arisen in Government Statistics, Econometrics and Financial Econometrics. We argue that there are some common structures and methods in many statistical problems and it shall be beneficial for many statisticians to think the roles of measurement errors and their statistical analyses in the era of Big-Data.

キーワード: 四半期 GDP，季節調整，計測誤差，統計的関係と構造方程式，確率過程モデルとマーケット・マイクロストラクチャ・ノイズ，ラージ・データとビッグ・データ，計測誤差と統計学，統計学・統計科学の将来

1. はじめに

2013 年 3 月 8 日内閣府が公表した四半期別 GDP 速報 (2 次速報値) により 2012 年 10 月–12 月の四半期，実質，季節調整済 GDP は 1 次速報値 (年率換算成長率) マイナス 0.4 よりプラス 0.2 に改訂された．約 1 ヶ月前に 2012 年 10 月–12 月の実質 GDP 成長率はマイナス 0.4 と発表,「しかし実は成長率はプラス 0.2 であった」という意味である．「光よりも速い物理現象が観測された」というマスメディアによる報道がしばらく前に波紋を投げかけたが，後でよく調べてみると「観測機器に問題があり事実ではなかった」ようである．GDP など日本の政府が定期的に公表しているマクロ経済時系列は物理現象の観測結果とは異なり「経済時系列の統計学」の検討対象である．

* 東京大学経済学研究科：〒 113-0033 東京都文京区本郷 7-3-1 (E-mail: kunitomo@e.u-tokyo.ac.jp).

本稿ではこうした「経済時系列における計測誤差」をはじめとして，これまでに筆者が経験した経済に関連する政府統計 (official statistics)・経済統計 (economic statistics)，計量経済 (econometrics)，計量ファイナンス (financial econometrics)，などを中心にして応用統計分野における「計測誤差の統計的処理」の取り扱い，について振り返ってみる．そしてこれまで行われてきた幾つかの計測誤差が関わる統計学的分析を通じて，「計測誤差と統計学」についてのより一般的な論点を考察するとともに，2013 年時点の日本における今後の統計学・統計科学の方向を展望する．

2. 政府公表系列と計測誤差

最近では日本の主要なマクロ経済指標は 2008 年に発生したリーマン・ショックや 2011 年の東日本大震災などに大きな影響を受け，時系列データの公表値にもとづく景気判断や，経済の将来見通しを立てることが以前より困難となっている．例えばエコノミストが注目している国内総生産 (GDP) の政府公表系列では，2012 年 10〜12 月期の実質 GDP 成長率（季節調整済み年率換算）の 1 次速報値はマイナス 0.4%であったが，1 カ月後に公表の改定値（2 次速報値）は一転してプラスの 0.2%に上方修正された．さらに 2013 年 9 月 9 日に公表された GDP 成長率（季節調整済み年率換算）の改訂値は 2.6%から 3.8%に上方修正されている．こうした四半期 GDP の政府公表値は 1 次速報値が発表されてから約 1 カ月後に数値が改定され 2 次速報値となり，さらにかなり後に確報値が公表される．近年の経験では 1 次速報値，2 次速報値，確報値の間のギャップが無視できないことがしばしば生じている．そうした背景としては，政府統計を必要とする側からは早期に経済の状態を理解したいという要請があり，他方で GDP のような一国全体の経済状態についての詳細なデータに基づき正確な数値を得るにはかなりの手間・予算・時間がかかることから，こうした異なる方向からの要請についてのある妥協点として公表スケジュールが運営されていることが挙げられる．GDP はマクロ経済における消費や投資をはじめとする様々な経済活動の集計値であり，例えば 1 次速報の作成段階では法人企業統計（財務省）の最新値は未集計のため，GDP 集計値の推計では民間設備投資や在庫投資という変動が激しい数字を予測する必要がある．実際に 2012 年 10〜12 月の実質 GDP 公表値の一次速報では GDP 集計値とともに発表された民間設備投資はマイナス 2.6%であったが，1 カ月後の 2 次速報ではマイナス 1.5%と上方に修正されているのである．

現代的な統計学からはこうした政府が発表している時系列データをどう見たらよいだろうか？ここでの状況は，「ある時点での真の値」について観測誤差が存在するので，実際に観測される「観測値」から，統計的にどのように「真の値」を推定するかという問題と解釈できる．1 次速報と 2 次速報の相違は，それぞれの時点で利用可能な統計情報，基礎となるデータが同一ではないために発生する推定誤差が主たる原因である．ここで経済時系

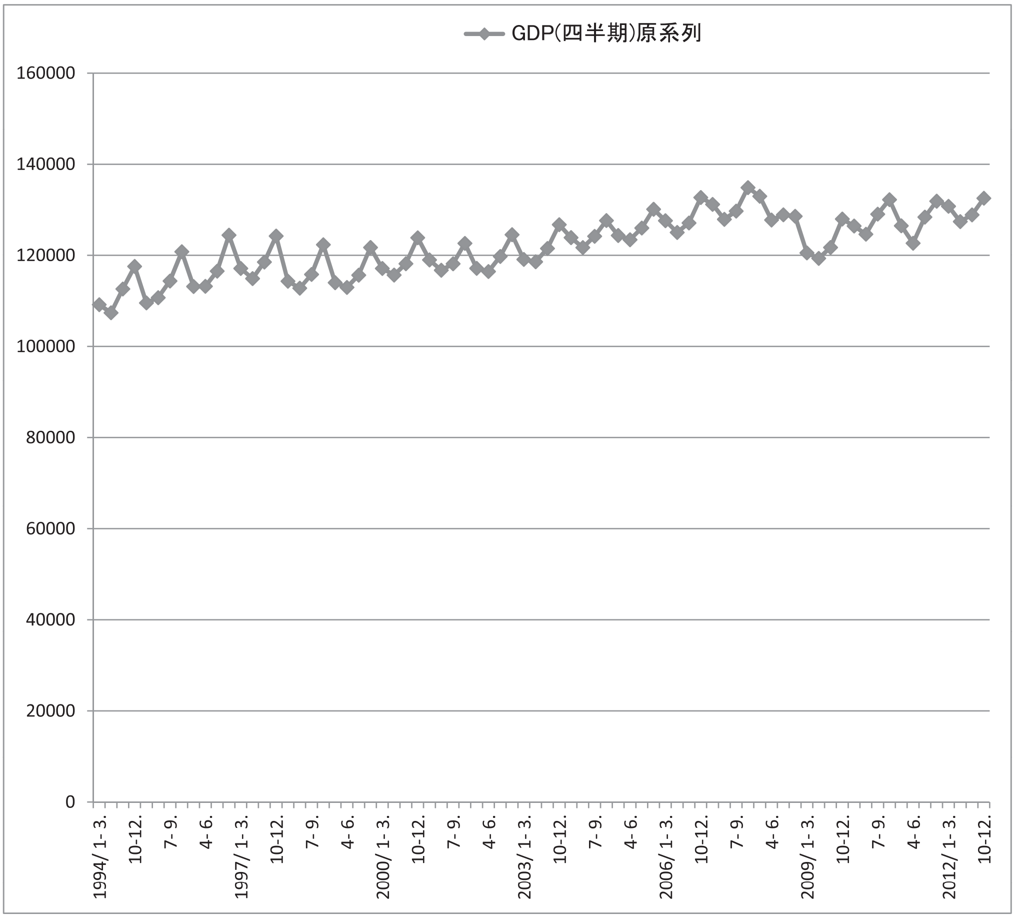

図 1　GDP 原系列.

列データは初等統計学で学ぶような，それぞれが「ある母集団から互いに独立に得られる標本 (sample)」と見なせる統計データとは異なり，時間の経過とともに時系列がとる経路，その経路上でとる数値の変化分が重要である[1)]．さらにマクロ経済の循環的変動や季節的変動まで考慮するとより複雑化し，ある時点での真の状態は「マクロ経済のトレンド・循環成分」,「季節変動成分」,「土曜・日曜など曜日数の変動成分」，などの構成要素の合成であり，実際の観測値には「不規則成分（あるいは計測誤差）」も加わると考えられる．これらの変動成分を統計学では「状態変数」(state variables) と呼ぶことが一般的である．観測される時系列データから状態変数の値を推定することは統計学の古典的な問題と比較するとかなり困難である．幸いにも近年の統計学，統計科学の進歩は目覚ましく，以前に比べると科学的に解決できることも少なくない．

GDP 速報では前期比から年率換算の成長率の動向に関心が集まる．図 1 に最近の四半期

1) 互いに独立な確率変数列の実現値とみなすことはできないので確率過程の実現値と解釈されるべきである．

GDP の原系列 (2013 年 3 月時点での公表されてる 1994 年〜2012 年の四半期原系列) を示しておくが，こうした数値より統計的な変換を行い，マスメディアを通じて内閣府が公表し，政府の経済の見通しや将来を見据える経済政策の基礎としているのである．図 1 から明らかなように原系列は 1 年間の中でもギクシャクと上下に変動している．そこで 3 カ月毎に集計される各基礎データから，ある種の変換を施して季節調整値の系列を作成し，その時系列より前期からの成長率をほぼ 4 倍した数値が公表されているのである．ここでエコノミストや政府関係者にとってのもっとも重要な情報は，1 年を周期とする季節的変動や，非常に短期的・偶然的に発生する不規則変動ではなく，経済のトレンドと循環変動である．そこで季節変動成分を取り除くために原系列データに対して，複雑な統計的操作である季節調整を行い，年率換算の変化率を推定している．例えば年末・年始における小売店の営業形態や夏期休業期間の時間的変化などを考えると容易に類推されるように，経済をとりまく環境は時間の経過と共に変化するので，結果としては変化する季節性なども処理する必要がある．ここで新聞などで報じられる年率換算値は 3 カ月間の瞬間的 GDP 増加率のほぼ 4 倍であって，実は「それまで 1 年間の実現値」でも「今後 1 年間の予測値」でもないことには注意が必要である[2)]．統計学から見ると，この年率換算値には直近の不規則成分・観測誤差（いわば「ノイズ」）の影響や状態変数の推定誤差の影響を非常に強く受けていると考えられる．

ここでは例として問題をかなり単純化して説明しよう．データ数を n，ある時点での観測系列 $Y_i\ (i=1,\ldots,n)$，真の状態変数 $\xi_i\ (i=1,\ldots,n)$，観測誤差 $U_i\ (i=1,\ldots,n)$ として

$$Y_i = \xi_i + U_i \quad (i=1,\ldots,n) \tag{2.1}$$

と表現する．真の状態はトレンド・循環成分 $TC_i\ (i=1,\ldots,n)$（さらに $TC_i = T_i + C_i$，トレンド成分 T_i と循環成分 C_i に分解することも良く行われる)，季節成分 $S_i\ (i=1,\ldots,N)$ すると，加法成分モデルは

$$\xi_i = TC_i + S_i \quad (i=1,\ldots,n) \tag{2.2}$$

となる．季節調整系列とは構成成分について加法モデルを仮定すると原系列 Y_i, 季節成分の推定値 $\hat{S}_i$ より $X_i = Y_i - \hat{S}_i\ (i=1,\ldots,n)$ により求められる[3)]．

さて経済時系列の統計学の展開は初等的な移動平均法（各時点より過去値と将来値のそれどれ数期分の平均を取る操作）から始まり，1950 年ごろより様々な統計的手法や統計的モデルが開発されてきた．季節調整法としては 60 年代に米国センサス局が開発した

---

2) 年率換算値を巡る問題点および改善可能性については国友・佐藤 (2010) を参照．

3) 内閣府が公表している季節調整済み系列は年率換算をかけ算で行うので年率成長率はほぼ $4X_i\ (i=1,\ldots,n)$ となる．

「X-11」，それを 90 年代に改良した「X-12-ARIMA」という手法が今でも重要な役割を果たしているが，実はいずれも統計学者からの評価は高くはない．多くの政府統計で用いられる X-11 や X-12-ARIMA などのプログラムのほかにも，統計数理研究所の北川源四郎氏が開発した「時系列成分分解法 (DECOMP)」なども知られている[4)]．DECOMP モデルではトレンド成分の d 次階差 $\Delta^d T_i = U_{Ti}$(平均 0, 分散 σ_T^2 の正規分布)，季節成分 ($s \geq 2$) $S_i + \cdots + S_{i-s+1} = U_{Si}$(平均 0, 分散 σ_S^2 の正規分布) が互いに独立な確率変数列により発生し，

$$\Delta^d T_i = U_{Ti} \sim N(0, \sigma_T^2) \ , \ \ S_i + \cdots + S_{i-s+1} = U_{Si} \sim N(0, \sigma_S^2) \ , \tag{2.3}$$

循環項 C_i は定常 AR モデルで表現できることが仮定されている[5)]．この DECOMP モデルは単純で明快な統計的構造を持つので，状態空間表現 (state space representation) を利用してフィルタリング (Kalman filtering を改良した square root filtering) により推定した季節調整値の解釈，その是非の判断も比較的容易である．他方，X-12-ARIMA は初等的な移動平均法をベースとしつつも，その内部では例えばデータの前処理・事前調整部分において Reg-ARIMA と呼ばれる統計モデルを利用するなど，古典的統計モデリングを多用している．古典的な移動平均法に基づく「X-12-ARIMA による原時系列データの変換」の詳細な中身を明快に理解することは難しいが[6)]，DECOMP モデルによる四半期データの季節成分推定において $d = 2, s = 4, C_i = 0, TC_i = T_i$ $(i = 1, \ldots, n)$ と設定すると，しばしば X-12-ARIMA による季節調整値と極めて類似した結果が得られることには重要な意味がある (国友 (2006, 2013b))．ここで季節階差操作 $\Delta_s = 1 - \mathcal{L}^s, \mathcal{L}^s Y_i = Y_{i-s}, \Delta Y_i = Y_i - Y_{i-s}$ より $\Delta\Delta_s = (1-\mathcal{L})(1-\mathcal{L}^s) = 1 - \mathcal{L} - \mathcal{L}^s + \mathcal{L}^{s+1}$ なので階差・季節階差を原系列に作用すると，統計的モデリングとしては時系列

$$X_i = \Delta\Delta_s Y_i = Y_i - Y_{i-1} - Y_{i-s} + Y_{i-s-1} \tag{2.4}$$

がある定常移動平均過程 (stationary moving average process) にしたがうことになる．このことから季節調整法として日本の政府統計において現在なおよく利用されている X-12-ARIMA は，トレンド・循環変動，季節変動，不規則変動について簡単な成分分解で得られる計算を，わざわざ複雑な計算過程として経験的に（長年の知恵とカンにより）実行して

---

4) DECOMP は Akaike (1980) が開発した BAYSIA (詳細は Akaike *et al.* (1985) を参照) を Kitagawa (1981) が square root filter を用いて改良したプログラムである．DECOMP の詳細は北川 (2005), HP 版 Web-DECOMP の利用については国友編 (2006) に収録されてい佐藤論文を参照されたい．

5) ラグ操作，階差操作は $\Delta = 1 - \mathcal{L}, \mathcal{L}Y_i = Y_{i-1}, \Delta Y_i = Y_i - Y_{i-1}$ により定められる．DECOMP モデルを仮定すれば原系列から成分分解が可能である．ここで他の時系列モデル，例えば移動平均 (moving average) 操作や Reg-ARIMA モデルを利用して成分分解を行っている X-12-ARIMA プログラムでもむろん可能である．

6) X-12-ARIMA の原マニュアルは国友編 (2004) により翻訳されているが，X-12-ARIMA アルゴリズムの概略は杉山他編 (2007) 収録の「移動平均と時系列成分モデル」などで説明されている．

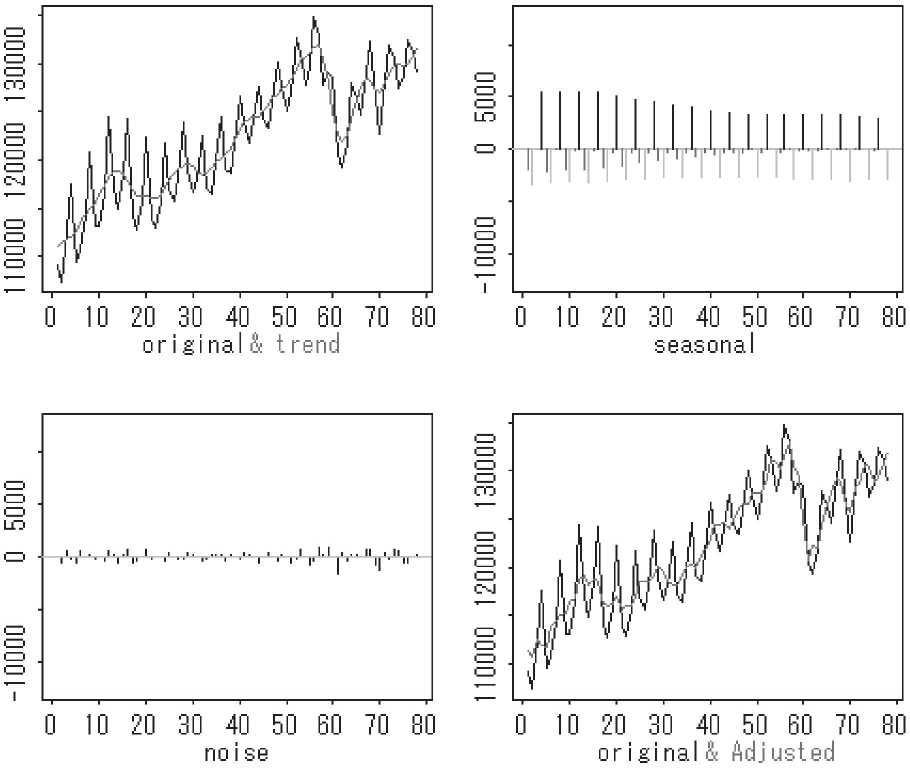

図 2 DECOMP による分析例．

いると解釈できるのである．ここで DECOMP を用いてその設定を X-12-ARIMA と整合的にして 2013 年 9 月に分析した結果を図 2 に示しておく．ここでは原系列に対数変換などを行っていないので原単位の大きさの影響が水準の変動幅が拡大 (図 1 を部分的に増幅している)，2008 年〜2009 年の落ち込みが大きいこと，季節性の推定値の振れ幅の変化に表れていること，などの観察事実が読み取れる．センサス局が X-11 法を開発した当時と比較すると，今日では簡単な操作により様々な統計的分析が容易に可能である例示である．

季節調整系列より推定され，マスコミなどの注目が集まる四半期 GDP の（年率換算の）前期比成長率の公表値は $R_i = (X_i - X_{i-1})/X_{i-1}$ $(i = 1, \ldots, n)$ の 4 倍（パーセンテージ表現ではその 100 倍）である．ここで $R_i = [(Y_i - Y_{i-1}) - (\hat{S}_i - \hat{S}_{i-1})]/[Y_{i-1} - \hat{S}_{i-1}]$ と分解できる．季節成分 $\hat{S}_{i-1}$ は Y_{i-1} に較べて小さいので分母の誤差を無視すると[7)]分子の $Y_i - Y_{i-1}$ 項，$\hat{S}_i - \hat{S}_{i-1}$ 項の相対的大きさ及び後者の推定誤差が重要となる．例えばデータ全体より 10 月〜12 月期のみ前期からの季節推定値の変化分についての標準偏差の大きさを計測してみると，この期の成長率はゼロ成長から乖離しているとは云えないことが分かる．すなわち本稿の冒頭で述べた速報値の改定幅は統計的には十分に小さく，改訂前の

7) 誤差評価には様々な課題があるのでここでは簡便な方法を利用した．

数値も改訂後の数値のいずれもがゼロ成長率から乖離しているとは云えないと解釈することが妥当であろう．

ここで言及した統計的時系列分析の観点より政府統計を分析した経済時系列に関するいくつかの新しい知見については国友 (2004, 2006)，国友・佐藤 (2010)，佐藤・国友 (2010)，高岡・国友 (2010)，国友・川崎 (2011) などが説明している．近年ではリーマンショックや東日本大震災など大きな経済変動の統計的モデリングなどを含め，さらなる今後の検討課題も少なくない．特に政府統計の実務にも幾つかの重要な論点があることを国友 (2013a, b) は説明しているが，ここでは統計学に関わる日本の政府統計の課題を指摘しておく．政府統計における季節調整値の利用においては「最適な季節調整」と「結果の頑健性（あるいは安定性)」という異なる要請がある．「良い季節調整」の基準が明確でないので，政府統計の担当者にとっては「正しい処理を定義することが容易でない」という意味で，対応が困難な問題であり「安定性」に偏りかねない．例えば役所における人事異動の結果，担当者が季節調整法の重要性を十分に理解できないまま，十分な引き継ぎが行われずに様々な問題が棚上げされる可能性などもありそうである[8)]．近年の日本のように成長率がゼロ付近にある場合には，成長率が高かった時代に比べると，時系列成分の計測方法の微妙な扱いにより，公表値が過去に遡って改定される可能性があり，統計学的に適切な処理をする必要性がさらに高まっている．国民への情報開示，判断材料の提供，政策の基礎となる政府統計の質の保証に関わることから，季節調整の妥当性，年率換算値の妥当性，計測誤差の評価などについて現代的な統計学の知識を基礎とした適切な処理を施すため，統計家を交えた努力が以前にも増して重要なのである．

3. 統計的関係・構造方程式と計測誤差

統計学や経済学では広い意味で解釈される計測誤差・観測誤差 (*measurement error*) の重要性について古くから考察されている．観測誤差モデルに関する統計的モデルの議論は Gauss による最小二乗法ほどは古くないが，少なくとも Adcock (1878) の研究まで溯ることができる．この問題は線形関数関係 (*linear functional relationship*) モデル，ある線形構造関係 (*linear structural relationship*) モデル, などの区別されることがあるが，通常は前者は固定母数型モデル，後者は変量型モデルを意味する．経済学の文脈ではノーベル賞経済学者 M. Friedman (1956) の消費関数に関する恒常所得仮説 (permanent income hypothesis) が著名であるが，統計的モデルに翻訳すると計測誤差モデルと同一である．統計的多変量解析分野において *linear functional relationship* (線形関数関係) モデルと *linear structural*

---

8) ここでの説明は政府統計に対する批判でないことを改めて言及しておく．日本の政府統計組織は例えば米国センサス局やヨーロッパ統計局などを始めとする組織と較べると，分権的組織形態，各セクションでの限られた人員，予算，時間的制約の中でかなりの成果を上げていることは確かである．

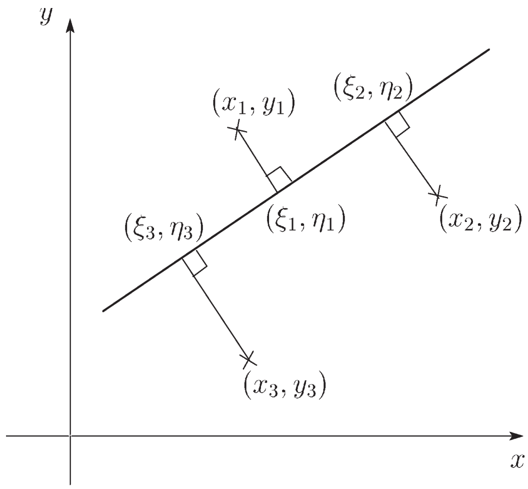

図 3 直交回帰 (二次元).

relationship (線形構造関係) モデルについては Anderson (1984) が包括的に議論している．本節ではこうした観測誤差の統計モデルを巡る古典的な議論とそこから派生する統計学上の問題を振り返ってみる．

Adcock (1878)

日本の大学で学ぶ統計学では線形回帰モデルや最小二乗法が定番の内容として定着している．二次元の変数 (X_i, Y_i) , $i = 1, \ldots, n$ についての観測値ベクトル (x_i, y_i) , $i = 1, \ldots, n$ が与えられたとき被説明変数 Y_i を説明変数 X_i の線形関係で説明するのが線形回帰モデルであるが，二つの変数の役割では説明する変数と説明される変数 (すなわち被説明変数) の区別がある．二つの変数が説明変数，被説明変数の区別がなくどちらにも測定誤差が存在するとき，確率変数 X の期待値 ξ と Y の期待値 η の間に線形関係 $\eta = \alpha + \beta\xi$ が存在し，それぞれに継続誤差が加わり観測ベクトル (x_i, y_i) が得られる状況から統計的線形関係の推定問題を Adcock (1878) は考察している．ここでは Adcock (1878) をやや現代風に解釈して，確率変数 X と Y が分散一定 σ^2 で互いに独立に二次元正規分布の標本として得られるとする．このとき尤度関数は

$$L(\alpha, \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_i - \xi_i)^2 + (y_i - \alpha - \beta\xi_i)^2]\right\} \tag{3.1}$$

で与えられる[9]．この尤度関数の最大化は損失関数

$$L_2 = \sum_{i=1}^{n}[(x_i - \xi_i)^2 + (y_i - \alpha - \beta\xi_i)^2] \tag{3.2}$$

[9] ここで ξ_i は母数と考えたので関数関係モデルを扱っている．ξ_i を確率変数とするのは構造関係モデルである．

の最小化問題であるので 2 次元平面上での点 (x_i, y_i) から点 (ξ_i, η_i) への距離の二乗和の最小化となり，直交回帰 (orthogonal regression) 問題とも呼ばれているが，視覚的には図 2 で示すような二次元観察データから「真の直線」の推定問題となる．母数 ξ_i について最適化すると損失関数は

$$L_3 = \frac{\sum_{i=1}^{n}[(y_i - \alpha - \beta x_i)^2]}{1 + \beta^2} \tag{3.3}$$

で与えられるので一般化最小二乗 (generalized least squares, GLS) 問題とも解釈できる．この問題を n 次元ユークリッド幾何的に考察すると興味深い解釈が得られるが，Adcock (1878) の導出から後に数多くの統計家が「互いに独立に解を発見」している．

さらに母数 α について (3.3) を最小化すると

$$L_4 = \frac{(1, -\beta) \sum_{i=1}^{n} \begin{pmatrix} y_i - \bar{y} \\ x_i - \bar{x} \end{pmatrix} (y_i - \bar{y}, x_i - \bar{x}) \begin{pmatrix} 1 \\ -\beta \end{pmatrix}}{(1, -\beta) \begin{pmatrix} 1 \\ -\beta \end{pmatrix}} \tag{3.4}$$

の最小化問題となる．したがって標本二次積率行列を

$$\mathbf{S}_n = \frac{1}{n-1} \sum_{i=1}^{n} \begin{pmatrix} y_i - \bar{y} \\ x_i - \bar{x} \end{pmatrix} (y_i - \bar{y}, x_i - \bar{x}) \tag{3.5}$$

とすると，ベクトル $\mathbf{b} \neq \mathbf{0}$ についての連立方程式

$$[\mathbf{S}_n - \lambda \mathbf{I}_2]\mathbf{b} = \mathbf{0} \tag{3.6}$$

を満足する解として得られる最小固有値 λ_1，および最小固有値に対応する固有ベクトル $\mathbf{b}_{ML} = (1, -b_{2.ML})'$ が求める問題の解に一致する．

Friedman (1956)

統計学の業績もある経済学者である Milton Friedman は米国における個人の膨大な消費データとマクロ消費動向データを整合的に説明できる恒常消費仮説により消費-所得に関する計量経済モデルを用いて分析した．観測される所得 Y, 観測される消費 C そのものという変数の関係ではなく，消費者にとって重要な所得-消費の関係は恒常所得 Y_p, 恒常消費 C_p について比例関係

$$C_p = kY_p \ \ (k \text{ は一定の定数}) \tag{3.7}$$

が成立すると考えた．観測される所得 Y には真の所得 (恒常所得) に偶然的な一時的所得 Y_t が加わり，観察される消費 C には真の消費 (恒常消費) に偶然的な一時消費 C_t が加えられていると見ると，個人の合理的な消費行動と集計されたマクロ時系列の変動を説明でき

ることを Friedman は主張したのである．ここで一時的所得を所得の観測ノイズ，一時的消費を消費の観測ノイズ，消費者にとっての所得-消費の関係は

$$Y = Y_p + Y_t \ ,\ C = C_p + C_t \ ,\ C_p = kY_p \tag{3.8}$$

と表現できる．すなわち経済学者に間で良く知られている Friedman の恒常所得仮説は統計学における線形関数関係モデルと同一であることが分かる．なお Friedman は実証分析として独自な統計的分析を展開している．

Anderson (1984)

統計学が関係する様々な分野においてしばしば異なる表現や名称の下に議論されている線形関数関係モデルおよび線形構造関係モデルについて統計的多変量解析 (statistical multivariate analysis) の観点から Anderson (1984) は包括的な議論を展開している．i $(i = 1, \ldots, n)$ について互いに独立な p 次元の観測可能な確率ベクトル $\mathbf{Y}_i$ に対して計測誤差ベクトル $\mathbf{U}_i$，$\mathbf{Y}_i$ の期待値ベクトル $\boldsymbol{\xi}_i$ として

$$\mathbf{Y}_i = \boldsymbol{\xi}_i + \mathbf{U}_i \ \ (i = 1, \ldots, n) \tag{3.9}$$

と表そう．ここで誤差項 $\mathbf{U}_i$ の期待値ベクトル $\mathbf{E}[\mathbf{U}_i] = \mathbf{0}$，分散共分散行列 $\mathbf{E}[\mathbf{U}_i\mathbf{U}_i^{'}] = \boldsymbol{\Sigma}_u$ と表す．真の p 次元ベクトル $\boldsymbol{\xi}_i$ が m 次元 $(1 \leq m < p)$ 線形空間上にあるときには $p \times q$ 行列 $\mathbf{B}$ $(q = p - m)$，定ベクトル $\mathbf{c}(q \times 1)$ により

$$\mathbf{B}^{'}\boldsymbol{\xi}_i = \mathbf{c} \ \ (i = 1, \ldots, n) \tag{3.10}$$

と表現できる．このとき Adcock モデルは $p = 2, q = 1, \mathbf{Y}_i = (Y_i, X_i), \mathbf{B} = (1, -\beta)^{'}, c = \alpha$ となる場合である．ここで $\mathbf{U}_i$ が互いに独立に $N_p(\mathbf{0}, \boldsymbol{\Sigma}_u)$ にしたがう確率変数，$\boldsymbol{\Sigma}_u$ が (正定符合行列の) 未知母数とすると一般には母数 $(\mathbf{B}, \boldsymbol{\Sigma}_u)$ を識別可能 (*identifiable*) ではないことに注意しよう[10]．このことより一般的状況で統計的推測を行うには何らかの条件を仮定する必要があることが分かる．例えば Adcock モデル (3.1) では $\boldsymbol{\Sigma}_u = \sigma^2\mathbf{I}_2$ としている．

一般に観測ベクトルから母数 $\mathbf{B}$ を推定する問題は線形関数関係モデルの一般化であるが，前の議論と同様に正規分布を仮定した尤度関数から導かれる最尤推定量 (maximum likelihood estimator) はベクトル $\mathbf{b} \neq \mathbf{0}$ についての連立方程式

$$[\mathbf{S}_n - \lambda\mathbf{I}_p]\mathbf{b} = \mathbf{0} \ , \mathbf{S}_n = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^{'} \tag{3.11}$$

---

[10] 母数ベクトル $\boldsymbol{\theta}$ が識別可能とは，観測データ数が仮に無限であれば観測データの分布から母数を一意的に定めることが可能となることを意味する．

を満足する解より小さい方から q 個の固有値に対応する固有ベクトル $\mathbf{b}_i$ $(i=1,\ldots,q)$ からなる行列 $\hat{\mathbf{B}}$ で与えられる．すなわちこの問題では標本から求められる固有値問題の解が重要な役割を演じていることが分かる．例えば推定される母数行列 $\hat{\mathbf{B}}$ が線形関係を満足することから，Anderson (1951) により導かれた縮小階数回帰 (reduced rank regression) 問題の解としても理解できるのである．このように統計的多変量解析における線形関数関係の分析の一般化から主成分分析 (principal components) 問題，因子分析 (factor analysis) 問題，同時方程式 (simultaneous equations) 問題，などにおける線型モデルの議論を統一的に理解できることを Anderson (1976, 1984) は示している．例えば行列 $\mathbf{B}$ の各列に直交する行列 $\mathbf{B}^{\perp}$ $(p\times m)$ とすると，(3.9) 式の左から正則行列 $\mathbf{\Gamma}=(\mathbf{B},\mathbf{B}^{\perp})'$ を掛けて整理すると

$$
\begin{aligned}
\mathbf{Y}_i &= \begin{pmatrix}\mathbf{B}'\\ \mathbf{B}^{\perp'}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{c}\\ \mathbf{0}\end{pmatrix}+\begin{pmatrix}\mathbf{B}'\\ \mathbf{B}^{\perp'}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{O}\\ \mathbf{I}_m\end{pmatrix}\mathbf{B}^{\perp'}\boldsymbol{\xi}_i+\mathbf{U}_i \qquad (3.12)\\
&= \boldsymbol{\mu}+\mathbf{\Lambda}\boldsymbol{\nu}_i+\mathbf{U}_i \qquad (i=1,\ldots,n)
\end{aligned}
$$

と表現できる．ここで $\mathbf{\Lambda}$ は $p\times m$ 行列 (階数は $0\le m\le p$)，$\boldsymbol{\nu}_i=\mathbf{B}^{\perp'}\boldsymbol{\xi}_i$ は $m\times 1$ ベクトル，$\boldsymbol{\mu}$ は定ベクトルである．ここで特に $q=p-1$ とすると $m=1$ となる 1 因子モデル (1-factor model) に対応するが，多因子モデルも同様に導けるが，(3.12) における分析での最も重要な特長は一般に関数関係が存在すると**係数行列の階数が退化する** (reduced rank) ことである．なお同論文によれば 1984 年の時点においても，ここでとりあげた固有値問題としての統計的分析は時間を超え，様々な形で統計学に関係する「主要な学術誌に独立に発表」されている．

なお統計理論家にとって興味深い変数誤差 (errors-in-variables) モデルでは偶然母数 (incidental parameter) $\boldsymbol{\xi}_i$ $(i=1,\ldots,n)$ は観測数 n に依存しているので統計学の標準的な理論的結果をそのまま適応することは妥当でないことが多いことに注意しておこう．例えば Neyman and Scott (1948) はこうした論点を最初に指摘したが，統計学の理論では標本数とともに「無限に母数が存在する場合」の問題である．ここでの統計モデルにおける有限標本における理論的結果としては，既に説明した推定方式がある損失関数の下で許容的 (admissible)[11] となることが Anderson *et al.* (1985) により示されていることである．また漸近的効率性に関する先駆的な研究としては，$p=2$ の場合について竹内 (1974) の 4 章には様々な理論的詳しい結果[12]が報告されているが，その後 Bickel and Ritov (1987) の研究がある．

---

11) このことは直交回帰推定を一様に改善することは可能ではないことを意味する．

12) この書物は日本語で書かれているので残念ながらその後，文献ではあまり引用されていないようである．

構造方程式・多操作変数・弱操作変数・パネルモデル

次に Anderson and Rubin (1949) などにより開発された[13]古典的な構造方程式 (*structural equation*) モデルの統計的分析に関する比較的最近の話題に言及しておこう．Anderson (1984) が明確化したように，構造方程式モデルは変数誤差 (errors-in-variables) モデル，観測誤差 (measurement errors) モデルを別の用語，別の記号で表現した統計モデルである．経済分析ではしばしば多くの変数が観測されるが，経済の議論では多くの中で鍵となる変数間の関係，すなわち構造方程式の分析が重要となる．経済全体で決定される分析上で鍵となる変数を内生変数，補助的な情報を提供する変数を外生変数，あるいは操作変数と呼んでいる．計量経済学 (econometrics) の歴史では経済学では基本的な考え方である市場により決定される価格と (取引) 数量が需要曲線と供給曲線という二つの構造方程式から決まるときの推定問題が 1940 年頃に統計家により考察されたことが重要な転機であったが，統計的モデリングにおける推定問題と識別問題が議論され続けている．

ここで観測される $p \times 1$ 内生変数ベクトル (endogenous variables) $\mathbf{Y}_i$ の真の値 $\boldsymbol{\xi}_i$ が外生変数 (exogenous variables, あるいは操作変数 instrumental variables) ベクトル $\mathbf{Z}_i$ で説明され，互いに独立な確率変数列の計測誤差を $\mathbf{V}_i$ とすると，

$$\mathbf{Y}_i = \boldsymbol{\xi}_i(\mathbf{Z}_i) + \mathbf{V}_i \ \ (i = 1, \ldots, n) \tag{3.13}$$

としよう．さらに内生変数ベクトルを $(Y_{1i}, \mathbf{Y}_{2i}^{'})\,'$ に分割し，内生変数ベクトル $\mathbf{Y}_{2i}$ に対する $G_2 \times 1$ 係数ベクトル $\boldsymbol{\beta}_2$, 外生変数ベクトル $\mathbf{Z}_i = (\mathbf{Z}_{1i}^{'}, \mathbf{Z}_{2i}^{'})^{'}$ を分割し，$K_1 \times 1$ 係数ベクトル $\boldsymbol{\gamma}_1$, 誤差項 u_i は $N(0, \sigma^2)$ にしたがう確率変数とするとき，線形構造方程式モデル

$$Y_{1i} = \boldsymbol{\beta}_2^{'}\mathbf{Y}_{2i} + \boldsymbol{\gamma}_1^{'}\mathbf{Z}_{1i} + u_i \ \ (i = 1, \ldots, n) \tag{3.14}$$

の推定問題が古典的な課題である．ここで誤差項 u_i の期待値 0, 分散 σ^2 としておくと，$G \times 1$ $(p = G_2 + 1)$ ベクトルの内生変数 $\mathbf{Y}_i = (Y_{1i}, \mathbf{Y}_{2i}^{'})^{'}$, $K \times 1$ $(K = K_1 + K_2)$ ベクトルの外生変数 $\mathbf{Z}_i = (\mathbf{Z}_{1i}^{'}, \mathbf{Z}_{2i}^{'})^{'}$ に関する誘導型 (reduced form) 表現を

$$\mathbf{Y}_i = \boldsymbol{\Pi}^{'}\mathbf{Z}_i + \mathbf{V}_i \tag{3.15}$$

とする．誤差項 $\mathbf{V}_i$ の期待値はゼロ・ベクトル，共分散行列 $\boldsymbol{\Omega}$，$\boldsymbol{\Pi}^{'}$ は母数行列であって $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2^{'})^{'}$ とすると線形制約

$$\boldsymbol{\beta}^{'}\boldsymbol{\Pi}^{'} = (\boldsymbol{\gamma}_1^{'}, \mathbf{0}^{'}) \tag{3.16}$$

---

[13] 正確にはシカゴに存在していたコールズ委員会に属する研究者などが中心となり開発されたが，関係者の中で T. C. Koopmans, T. Haavelmo, L. R. Klein などはノーベル経済学賞を受賞している．

が導かれるので，**係数行列の階数が退化**している．したがって構造方程式モデルの未知母数を識別するには一般には係数行列について階数条件の考察が必要となる．計量経済学では $K \times 1$ ベクトルの外生変数 $\mathbf{Z}_i$ を操作変数 (instrumental variables) と呼んでいる．古典的な価格-数量の市場均衡モデルは $p = 2, \mathbf{Y}_i = (P_i, Q_i)'$($P$ と Q は価格と数量) 及び外生変数による表現される．

ここで誤差項の (多次元) 正規性を仮定すると十分統計量は

$$\hat{\mathbf{\Pi}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \ , \ \mathbf{C} = \mathbf{Y}'(\mathbf{I}_G - \mathbf{P}_Z)\mathbf{Y} \tag{3.17}$$

となる．ただし $n \times p$ 行列 $\mathbf{Y} = (\mathbf{y}_i')$ ($\mathbf{y}_i$ は $\mathbf{Y}_i$ の観測ベクトル), $n \times K$ 行列 $\mathbf{Z} = (\mathbf{z}_i')$($\mathbf{z}_i$ は $\mathbf{Z}_i$ 観測ベクトル), $n \times n$ 行列 $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ は階数 $K(< n)$ の射影行列である．こうした設定の下で Anderson and Rubin (1949, 1950) が導いた係数行列の階数制約の下で得られる制限情報最尤推定量 (limited information maximum likelihood estimator, 通称 LIML) はベクトル $\mathbf{b} \neq \mathbf{0}$ についての連立方程式

$$[\mathbf{G} - \lambda\mathbf{C}]\mathbf{b} = \mathbf{0} \tag{3.18}$$

を満足する解より最小固有値 λ_1 に対応する固有ベクトル $\mathbf{b}_{LI} = (1, -\mathbf{b}_{2.LI}')'$ となる．ここで $p \times p$ 二次標本行列

$$\mathbf{G} = \mathbf{Y}'(\mathbf{P}_Z - \mathbf{P}_{Z_1})\mathbf{Y} \tag{3.19}$$

としたが，最尤推定量や尤度比統計量はこれら二次積率行列 $\mathbf{G}, \mathbf{C}$ の関数であるから自然な推定量としては二つの二次形式の関数 $\mathbf{b} = \phi(\mathbf{G}, \mathbf{C})$ が考えられる．(3.18) において特に $\lambda = 0$ に設定すると二段階最小二乗 (the two-stage least squares) 推定量が得られる．

こうした線形構造方程式モデルは経済学の計量分析では 1950 年代から様々な形で利用されてきているが，1980 年頃には T. W. Anderson，竹内啓，佐和隆光，森棟公夫，佃良彦（および末席として筆者）などにより標本分布論が展開されたことが興味深い．この頃の研究成果，例えば Takeuchi and Morimune (1985) による一般的結果などについては森棟 (1985) が説明している．有限標本での統計的方法の最適性については明確な結果を得ることは一般には簡単ではないが[14]，仮説検定についての理論的結果としては尤度比検定の許容性に関する Anderson (2011) の結果が知られている．さらに必ずしも正規性などを仮定しない十分統計量のみに依存するとは限らない．

よりノンパラメトリックな方向への議論として一般化積率法 (GMM)[15]や Owen (2001)

---

[14] 既に引用した Anderson *et al.* (1985) と同一の結果が Chamberlain (2007, *Econometrica*) により「独立に導かれている」ことも興味深い．

[15] 数理統計学では Godambe (1960) の研究に端を発する推定方程式 (estimating equation) 論に対応するが，計量経済学ではマクロ経済分析における経済時系列分析上での必要性から導入された背景から Hansen (1982) の GMM(the generalized method of moment) と呼ばれている．

による経験尤度法 (empirical likelihood method, 略して EL) の適用などが理論的研究では顕著な動向となっている．GMM 推定では構造方程式は操作変数 $\mathbf{z}_i$ と構造方程式誤差 u_i 直交条件 $\mathbf{E}[\mathbf{z}_i u_i] = 0$ を意味することに注目する．一般には非線形構造を含むが特に分散均一の場合に制約してこの条件を標本で置き換えると二段階最小二乗法に対応すると解釈できる．他方，EL 推定では Qin and Lawless (1994) は nonparametric 尤度 $\Pi_{i=1}^n p_i$ を制約条件 ($\sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i u_i = 0$, ただし u_i は推定方程式の誤差，あるいは (18) の構造方程式を表す) の下で最大化することを提唱した．この推定方法は LIML 推定に対応すると解釈できるが，GMM 推定と EL 推定については関連する様々な論点がある．こうした計量経済学における近年の議論の一端は例えば Hayashi (2000), Kunitomo and Matsushita (2009), 国友 (2011), Kunitomo (2012) などにより知ることができる．

さらに 2000 年頃よりミクロ計量分析の応用上の観点より説明変数の数 K が多い場合の多操作変数 (*many instruments*) 問題, 説明変数・操作変数の識別力が弱い場合の弱操作変数 (*weak instruments*) 問題に特に関心が持たれるようになっている．前者は操作変数の数が多い場合を想定し，操作変数の数 K を標本数に依存させて K_n，極限として $K_n/n \to c$ $(0 \le c < 1, c$ は一定値) と言う状況を想定する漸近論による分析が可能である．ここで古典的な統計的多変量解析の枠組みでは, 誤差項の正規性の仮定の下で標本二次行列 $\mathbf{G}$ は p 次元非心ウシャート (non-central Wishart) 分布 $\mathcal{W}_p(K_{2n}, \mathbf{\Theta}_n, \mathbf{\Omega})$ にしたがう．ここで $K_{2n} = K_n - K_1$, 非心度は標本数 n, 操作変数の数 K_n に依存し $\mathbf{\Pi}'\mathbf{Z}'(\mathbf{P}_Z - \mathbf{P}_{Z_1})\mathbf{Z\Pi}$ となるので，この量を $\mathbf{\Theta}$ とおくと階数が退化している行列となる．（なお $\mathbf{\Omega}$ は誤差項の分散共分散行列である．）他方，$\mathbf{C}$ は誤差項の正規性の下では p 次元ウシャート分布 $\mathcal{W}_p(n - K_n, \mathbf{\Omega})$, にしたがい $\mathbf{G}$ とは独立に分布する事実が重要である．非心度は標本数 n, 操作変数の数 K_n に依存することを明示的に $\mathbf{\Theta}_n$(階数が退化する行列) と表現すると，通常の状況では標本数が増加するにつれて増大すると仮定することが自然である．多操作変数問題では変数の数も増加するので非心度 $\mathbf{\Theta}_n$ も同時に大きくなる．他方，弱操作変数問題とは非心度 $\mathbf{\Theta}_n$ が標本数の増加と共にあまり増加しない場合，つまりデータの増加に伴い情報量があまり増加しない状況，と定式化できるので推定方法は鍵となる n に依存する量，$n, K_n, n - K_n, K_{2n}$ $(K_n = K_1 + K_{2n}), \mathbf{\Theta}_n$ の挙動に依存して決まることがわかる．多操作変数問題においては，$0 < c < 1$ のときにも誤差項に正規分布を仮定しなくても一定の積率が存在するなど正則条件の下で Kunitomo (1980) が初期の推定量の性質の研究などを行っているがここでは Anderson *et al.* (2010) が得た漸近的結果に言及しておく．

ここで $\mathbf{z}_i^{(n)}, i = 1, 2, \ldots, n,$ を $K_n \times 1$ ベクトル $(K_n = K_1 + K_{2n}, n > 2)$, $\mathbf{v}_i, i = 1, 2, \ldots, n,$ を $(1 + G_2) \times 1$ 確率ベクトルで $\mathbf{z}_1^{(n)}, \ldots, \mathbf{z}_n^{(n)}$ と独立，$\mathcal{E}(\mathbf{v}_i) = \mathbf{0}$, $\mathcal{E}(\mathbf{v}_i \mathbf{v}_i') = \mathbf{\Omega}$ (a.s.), LIML 推定量 $\mathbf{b}_{LI} = (1, -\boldsymbol{\beta}_{2.LI}')'$ とする．

(i) $c=0$ のとき

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI}-\boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Psi}^*)\ , \tag{3.20}$$

ただし $\boldsymbol{\Psi}^*=\sigma^2\boldsymbol{\Phi}_{22.1}^{-1}$, $\sigma^2=\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}$ という表現を得る．

(ii) $0<c<1$ のとき

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI}-\boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Psi}^{**})\ , \tag{3.21}$$

ただし

$$\boldsymbol{\Psi}^{**}=\sigma^2\boldsymbol{\Phi}_{22.1}^{-1}+\boldsymbol{\Phi}_{22.1}^{-1}\left\{c_*\left[\boldsymbol{\Omega}\sigma^2-\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\right]_{22}+\left[(\boldsymbol{\Xi}_{3.2}+\boldsymbol{\Xi}_{3.2}^{'})+\eta\boldsymbol{\Gamma}_{44.2}\right]\right\}\boldsymbol{\Phi}_{22.1}^{-1}$$

と表現される．（母数行列 $\boldsymbol{\Phi}_{22.1},\boldsymbol{\Xi}_{3.2},\boldsymbol{\Gamma}_{44.2}\ \eta$ の定義は省略する．）ここで $c_*=c/(1-c)$ であるが，特に $G_2=1$ のとき $[\boldsymbol{\Omega}\sigma^2-\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}]_{22}=\omega_{11}\omega_{22}-\omega_{12}^2=|\boldsymbol{\Omega}|$ となる．

さらに最近の研究として Anderson *et al.* (2011), Kunitomo (2012) は制限情報最尤法 (limited information maximum likelihood method) がある種の最適性を持つことを示している．

関連する問題としては近年ではパネル・データの計量経済分析が極めて盛んになっていることにも言及しておく．近年ではパネルデータの整備とともに個人（あるいは企業や国・地域など）$i=1,\ldots,N$ に関する時間 $t=1,\ldots,T$ におけるデータ $\mathbf{Y}_{it},\mathbf{Z}_{it}$ が利用可能であるとき，個別効果 $\eta_i,\boldsymbol{\pi}_i$ $(i=1,\ldots,N)$ を含む線形構造方程式

$$Y_{1it}=\boldsymbol{\beta}_2^{'}\mathbf{Y}_{2it}+\boldsymbol{\gamma}_1^{'}\mathbf{Z}_{1it}+\eta_i+u_{it}\ \ (i=1,\ldots,N;t=1,\ldots,T) \tag{3.22}$$

の推定問題が浮上する．ここで内生変数ベクトル $\mathbf{Y}_{it}=(Y_{1it},\mathbf{Y}_{2it}^{'})^{'}$, 外生変数ベクトル $\mathbf{Z}_{it}$ についての誘導型を

$$\mathbf{Y}_{it}=\boldsymbol{\Pi}^{'}\mathbf{Z}_{it}+\boldsymbol{\pi}_i+\mathbf{V}_{it} \tag{3.23}$$

とすると，(3.23) に左から $(1,-\boldsymbol{\beta}_2^{'})$ を乗じると個別効果は $\eta_i=(1,-\boldsymbol{\beta}_2^{'})\boldsymbol{\pi}_i$ $(i=1,\ldots,N)$ と表現される．こうした定式化では多くの実際の状況に対応するように個人の数 N が多ければ推定すべき母数の数 $\eta_i,\boldsymbol{\pi}_i$ $(i=1,\ldots,N)$ も多くなるので統計学では偶然母数 (incidental parameter) が存在する問題と理解できる．ここでの統計的分析ではクロスセクション次元 N と時系列次元 T の設定により様々な可能性が生じる．観測値ベクトル $(\mathbf{Y}_{it},\mathbf{Z}_{it})$ が互いに独立な確率変数列とすることが妥当ならば，時間変数 t について差分をとった変数 $\Delta\mathbf{Y}_{it}=\mathbf{Y}_{it}-\mathbf{Y}_{i,t-1}$, $\Delta\mathbf{Z}_{it}=\mathbf{Z}_{it}-\mathbf{Z}_{i,t-1}$ についての関係は母数 $\boldsymbol{\pi}_i$ と独立になるので階差操作 (difference) が有力な手段となる．しかしながら，実際の分析ではパネル構造方程式は説明変数 $\mathbf{z}_{it}$ にラグ付き内生変数 $\mathbf{y}_{it-s}$ $(s\geq 1)$ を含むなどと動学化される．例えば説明変数 $\mathbf{Z}_{it}=\mathbf{Y}_{i,t-1}$ という動学パネルデータの場合には階差操作はあまり良くないことが知られている．こうした動学パネル構造方程式 (*Dynamic Panel Structural Equation*) 問題

において Akashi and Kunitomo (2010, 2012) はデータに対する前向き (後ろ向き) フィルターを利用したパネル制限情報最尤推定法 (PLIML) を提唱し，クロスセクションの観測数 N, 時系列の観測数 T がともに大きい場合にも操作変数を有効に利用する意味で最適性があることを示している．動学パネルモデルでは観測できない個別効果 (一種の観測誤差と見なせる) の処理が重要であり，時間についての階差操作の他にも

$$\mathbf{y}_{it}^{(f)} = c_t \left[\mathbf{y}_{it} - \frac{1}{T-t}(\mathbf{y}_{it+1} + \cdots + \mathbf{y}_{iT}) \right] \tag{3.24}$$

で定められる前向きフィルター (forward-filter, $c_t^2 = (T-t)/(T-t+1), t = 1, ..., T-1, T \geq 2$) を利用する方法，あるいは操作変数に対して

$$\mathbf{z}_{it-1}^{(b)} = b_t \left[\mathbf{z}_{it-1} - \frac{1}{t}(\mathbf{z}_{it-2} + \cdots + \mathbf{z}_{i0} + \mathbf{z}_{i(-1)}) \right] , \tag{3.25}$$

で定められる後ろ向きファイルター (backward-filter, $b_t^2 = t/(t+1), t = 1, ..., T-1$) などを利用することが有力な分析手段となっている．すなわちパネルデータにおける統計分析の問題では原データの変換によって変数誤差を巡る数理的構造を考慮することが重要となる．

計測誤差と変数誤差モデルの拡張

ここまでやや駆け足ではあるが線形構造方程式モデルに限定した最近の展開を説明した．近年の統計分析では非線形構造方程式モデルや非線形パネル方程式モデルの利用も盛んに行われている．パネル・データ問題ではクロスセクション次元での不均一性，時系列次元での定常・非定常性，構造方程式，などを考慮すると様々な可能性が広がり，例えば計量経済分野においては様々な方向で活発に議論されている．

さて限られた範囲であるがより統計学分野全般に視野を広げてみると，米国統計学会会長を経験した著者による書物 Fuller (1987) は生物統計学など様々な応用上に重要な論点を議論している[16)]．その後，生物統計学における応用例を念頭にした Carroll *et al.* (2006) が多くの非線形問題を議論しているが，生物統計学への応用では繰り返し観測が可能なことが現実に少なくないようである．こうした状況では誤差分布や誤差共分散 $\mathbf{\Sigma}_u$ は識別可能 (identifiable) であるので推定問題が意味を持つ．この状況における線型モデルの場合には (3.9) 式と (3.10) 式の代わりに

$$\mathbf{Y}_{ij} = \boldsymbol{\xi}_i + \boldsymbol{\mu} + \mathbf{U}_{ij} \quad (i = 1, \ldots, n; j = 1, \ldots, k) \tag{3.26}$$

において，真のベクトルが m 次元 $(m < p)$ 線形空間上にあり $\boldsymbol{\xi}_i^{'}\mathbf{B} = \mathbf{c}$ と表現されるが，(3.10) を拡張した非線形関数関係モデルが応用上での重要な展開につながっている．この

16) 著者は統計学の応用分野，とりわけ生物統計学には詳しくないので不完全で部分的な言及となる．

場合には例えば誤差項の分散共分散行列 $\mathbf{\Sigma}_u$ は

$$\hat{\mathbf{\Sigma}_u} = \frac{1}{k-1}\sum_{j=1}^{k}(\mathbf{Y}_{ij}-\bar{\mathbf{Y}}_{i\cdot})(\mathbf{Y}_{ij}-\bar{\mathbf{Y}}_{i\cdot})' \tag{3.27}$$

により推定可能である．

特に誤差項の確率分布の識別に最低限必要な $k=2$ で $\boldsymbol{\xi}_i$ が未知の確率分布 $F_{\boldsymbol{\xi}}$ にしたがうとき $\mathbf{Y}_{ij}$ の確率分布から $F_{\boldsymbol{\xi}}$ が識別可能となることが注目しよう．この方向の研究では Kotlarski の補題と呼ばれている結果，すなわち観測誤差と真の変量について i.i.d. を仮定すると (Y_{i1}, Y_{i2}) の同時特性関数 $\psi(u_1, u_2)$ により ξ_i の特性関数 $\phi_\xi(t)$ が

$$\phi_\xi(t) = \exp\left[\int_0^t \frac{\frac{\partial\psi(0,u_2)}{\partial u_1}}{\psi(0,u_2)}du_2\right] \tag{3.28}$$

により与えられることが重要である．この関係を利用した Tong and Vuong (1998) による研究を一つの契機として，構造方程式のノンパラメトリック推定が新たな展開を見せている．例えば Schennach (2004) による非線形関係の研究と経済分析への応用などを挙げることができる．

共和分関係と多次元時系列

経済時系列の統計的分析において構造方程式モデルに関係する近年にもつながる話題としては共和分 (co-integration) 問題の展開が挙げられる[17]．マクロ時系列の変動の観察から多くの場合には統計的時系列分析で標準的とされている定常時系列と見なすことはできないが，階差操作を行うことで定常時系列分析に還元することが実用的とされてきた．この場合には原系列は $I(1)$，階差系列は $I(0)$ と呼ばれる．他方，複数の重要なマクロ時系列の間にはかなり単純な関係が観察されている．この二つの要素を結びつける共和分関係とは Engle and Granger (1987) が導入した非定常系列における線形関数関係であるが，Johansen (1991) により最尤法による統計的推測方法が検討されている．p 次元内生変数 $\mathbf{Y}_i$ に対して説明変数ベクトル $\mathbf{Z}_i$ としてラグ付き変数 $\mathbf{Y}_{i-j}$ $(j=1,\ldots,q)$，係数行列 $\mathbf{\Pi}_j$ $(j=1,\ldots,1), \mathbf{\Pi}^*$ とすると

$$\mathbf{Y}_i = \sum_{j=1}^{q}\mathbf{\Pi}_j'\mathbf{Y}_{i-j} + \mathbf{\Pi}^{*'}\mathbf{Z}_i^* + \mathbf{V}_i\ , i=1,\ldots,n \tag{3.29}$$

となり，誘導型表現 (3.29) は多次元自己回帰型の離散時系列モデルとなる．ここで通常は $\mathbf{V}_i$ は i について互いに独立な確率変数列，$\mathbf{Z}_i^*$ は定数や季節ダミーなどを含む純粋な外生変数を想定する．さらに階差操作 Δ を順次に施すと適当に定義された係数行列 $\mathbf{A}_j$ を利用

[17] 日本では山本拓教授などの貢献がある．

すると

$$\Delta \boldsymbol{Y}_i = \boldsymbol{A}_1^{'} \boldsymbol{Y}_{i-1} + \sum_{j=2}^{q} \boldsymbol{A}_j^{'} \Delta \boldsymbol{Y}_{i-(j-1)} + \boldsymbol{\Pi}^{*'} \mathbf{Z}_i^{*} + \boldsymbol{V}_i \tag{3.30}$$

と表現できる．ここで時間の経過にともなう確率差分方程式を満足する内生変数ベクトル $\mathbf{Y}_i$ の動的挙動は固有方程式

$$\left| \lambda^q \mathbf{I}_p - \sum_{j=1}^{q} \lambda^{q-j} \boldsymbol{\Pi}_j \right| = 0 \tag{3.31}$$

の根の位置により決定される．簡単化して $\boldsymbol{\Pi}^* = \mathbf{O}$ とすると，もし (3.31) のすべての根の絶対値が 1 より小さければ $\mathbf{Y}_i$ は定常となる．他方，もし $\mathbf{A}_1 = \mathbf{O}$ であれば $\Delta \mathbf{Y}_i$ が定常となりうる．ここで C. Granger が指摘した重要な論点は**行列 $\mathbf{A}_1$ の階数がゼロでなく退化**していれば定常性と非定常性の混合状態を表現することが可能となることである[18]．ここで内生変数 $\mathbf{y}_i$ の線形和が定常構造を持つことと，$\mathbf{X}_i = \boldsymbol{\beta}^{'} \mathbf{Y}_i$ が定常過程となるような $p \times q$ 行列 $\boldsymbol{\beta}(p > q \geq 1)$ が存在する条件は表現 (3.30) における行列 $\mathbf{A}_1$ の階数が退化する条件を意味する．したがって数理的には既に議論した構造方程式の制約条件と類似の構造を持っているので，その統計学的な議論の基本構造は Anderson (1951, 1984) で展開された (標本固有値問題の）研究の非定常過程への拡張であることが分かる．すなわち既に説明した Anderson (1984) が与えた視点より標本から求められる固有ベクトルと固有値が鍵となることが理解できよう．$\hat{\boldsymbol{B}}$ を係数行列の最尤推定量 (より正確には階数制約の下での拡張された LIML 推定量) とするとき，国友 (2011) が説明しているように n に依存する行列 $\mathbf{D}_n$ がとれて

$$\mathbf{D}_n(\hat{\boldsymbol{B}} - \boldsymbol{B}) \xrightarrow{w} \mathbf{M}(1)^{-1} \int_0^1 \mathbf{m}(t) d\mathbf{W}^{'}(t) \mathbf{C} \tag{3.32}$$

という弱収束 (weak convergence) の結果を得る．ここで $\mathbf{W}(t)$ はブラウン運動であり確率積分の形式で表現されるが，確率行列 $\mathbf{M}(1)$, ベクトル $\mathbf{m}(\mathrm{t})$, 行列 $\mathbf{C}$ の定義は外生変数，係数行列に依存するが省略する．

なお，ここでの議論を非線形時系列の扱いに拡張する試みが行われているが，なお大方の統計家が合意できる統計的モデリングまでには至っていない．また 2 節で議論したように実際に観察される経済時系列はトレンド成分，循環成分，季節成分，不規則成分などの合成と考えられる．多次元時系列の場合にはこれら構成成分の統計モデルは複雑になる．他方，経済学者が政府統計で求められる季節調整済系列を分析することも少なくないが，統計家と経済学者を同時に満足する統計的モデリングはなお未解決な問題として残されていると思われる．

[18] Engle and Granger (1987) によるが例えば国友 (2011)8 章が説明している．

4. 確率過程モデルと計測誤差

近年では経済学と密接に関連するファイナンスと呼ばれる分野において連続時間の確率過程モデルの利用が盛んである[19]．こうした動向は学問的問題の発展のみならず，ファイナンス実務における確率過程の応用に関わる数理的議論が関係することが広く認識されるようになったことが重要な要因となっている．こうした中では当然の研究動向として，実際に観察される金融時系列データから連続確率過程モデルの妥当性を検証しようとする統計的分析も盛んになっている．近年ではほとんどすべての金融取引データを入手することも可能となり，高頻度金融データの分析も注目されるようになっている．

より精密な観測データが得られると利用する確率過程モデルをより正確に知ることが可能となることが確率過程に関する統計学での一時期までの常識であったが，金融データでは実は必ずしもそうとは限らないことが明らかになっている．こうした論点を理解することは統計学の応用として金融実務的にも重要と思われる．ここでは確率過程モデルと計測誤差について Kunitomo and Sato (2008, 2011)，Kunitomo and Misaki (2013) などで取りあげた問題と結果を紹介しよう．ファイナンス計量分析では連続時間の p 次元確率過程 $\boldsymbol{\xi}(t)$ $(0 \leq t \leq T)$ が連続時間のマルチンゲール過程

$$\boldsymbol{\xi}(t) = \boldsymbol{\xi}(0) + \int_0^t \mathbf{C}_\xi(s) d\mathbf{W}(s) \tag{4.1}$$

にしたがう場合の統計分析が統計的金融リスク管理への応用上の問題としても重要である．ここで $\mathbf{C}_\xi(s)$ はボラティィリティ関数 (あるいは行列値をとる関数) であり確率的 (stochastic) な場合も含むものとする．また $\mathbf{W}(s)$ は q 次元ブラウン運動，積分は Ito 式の確率積分を意味する．連続時間のマルチンゲール過程は一種の数学的に理想化された確率過程モデルであり，高頻度金融データが利用可能になるにつれて，現実に観察される高頻度金融データと整合的でないことが明らかになっている．ここで実際に市場で取引される (対数) 価格ベクトル $\mathbf{Y}(t_i^n)$ が離散時間 $(0 = t_0^n < t_1^n < \cdots < t_n^n = T$, ここでは簡単に為に $T = 1$ に固定する) に観測され

$$\mathbf{Y}(t_i^n) = \boldsymbol{\xi}(t_i^n) + \mathbf{U}(t_i^n) \ \ (i = 1, \ldots, n) \tag{4.2}$$

とする．ここで $\mathbf{U}(t_i^n)$ は期待値 0, 分散共分散 (一定)$\boldsymbol{\Sigma}_u$ で互いに独立な確率変数とする．この計測誤差はファイナンス分野では金融市場における取引により発生するマーケット・マイクロストラクチャ・ノイズと呼ばれ，直接には観測できない確率変数列であるので，計測誤差・観測誤差と見なすことができる．ここで観測誤差がなければ (すなわち $\mathbf{U}(t_i^n) = \mathbf{0} \ \ (i = 1, \ldots, n)$ であれば)，数学的な連続時間の確率過程モデルでは時間間隔

[19] 関連する文献は多岐に及ぶが例えば統計数理 (2009, 2011) の特集号がある．日本では吉田朋広教授のグループの研究が顕著である．

が小さくなるにつれてモデル誤差は小さくなるので離散時間における実現ボラティリティ (realized volatility) は

$$RV_n(0,1)=\sum_{i=1}^{n}(\mathbf{Y}(t_i^n)-\mathbf{Y}(t_{i-1}^n))(\mathbf{Y}^{'}(t_i^n)-\mathbf{Y}^{'}(t_{i-1}^n)) \tag{4.3}$$

で定められるが，観測幅が短くなる (単位時間間隔内で多くの観測値が得られる) とき (すなわち $n\to\infty$ かつ $T=1$ は固定) 連続時間の確率過程における二次変分量 (quadratic variation)

$$\langle\boldsymbol{\xi},\boldsymbol{\xi}\rangle(1)=\int_0^1\mathbf{C}_\xi(s)\mathbf{C}_\xi^{'}(s)ds \tag{4.4}$$

に収束することが確率過程論では知られている[20)]．しかしながら実際の金融市場では経済活動の性質から「連続時間の確率過程における標準的議論が成立しない」という事実が重要である．例えば図 4 は大阪証券取引所 (OSE) におけるある日の日経平均 225-先物の 1 秒，10 秒おきに観察される価格変動系列を示しておいた．連続時間のある時点での価格とはその時点までの直近の時点での注文と需要と供給が一致した取引価格を意味することは，市場経済における取引の仕組みを理解すると当然の事であり，連続時間の確率過程論で想定しているある確率過程の実現系列と見なせるか否か十分に検討すべきである．

さて特に $p=q=1$ とすると統計的な推定問題は市場での価格変動性を和分ボラティリティ (Integrated Volatility)

$$IV=\int_0^1\sigma_\xi^2(s)ds \tag{4.5}$$

を観測データから推定することである (ここでは $\sigma_\xi^2=\mathbf{C}_\xi(s)\mathbf{C}_\xi^{'}(s)$ を意味する)．観測データは高頻度金融データであるから $T=1$ と設定し，固定された区間 $[0,1]$ 上の細分数 n が十分に大きいときにボラティリティを推定するのが課題である．この問題については例えば Barndorff-Nielsen *et al.* (2008) の研究を始めとして幾つかの統計的方法が提唱されているが，ここでは Kunitomo and Sato (2008, 2011) が提唱している Separating Information Maximum Likelihood (SIML) method を取りあげてみる．$n\times p$ の観測行列 $\mathbf{Y}=(Y(t_i^n)^{'})$ に対して変換

$$\mathbf{Y}_n^*=\sqrt{n}\mathbf{P}_n\mathbf{C}_n^{-1}\left(\mathbf{Y}-\bar{\mathbf{Y}}_0\right) \tag{4.6}$$

を考える．ただし初期条件 $\bar{\mathbf{Y}}_0=\mathbf{1}_n\otimes\mathbf{y}_0^{'}$,

$$\mathbf{P}_n=(p_{jk})=\left(\sqrt{\frac{2}{n+\frac{1}{2}}}\cos\left[\frac{2\pi}{2n+1}\left(j-\frac{1}{2}\right)\left(k-\frac{1}{2}\right)\right]\right)\ , \tag{4.7}$$

20) 連続時間の確率解析では基本的な量であり，一般には確率変数となる二次変分についての詳細は例えば長井 (1999) を参照されたい．実際には金融市場での取引ではある価格の需要サイドの注文に対して供給サイドの注文が一致したときにのみ取引が成立する．通常では最小の注文の大きさと価格の変動の最小単位が存在している．

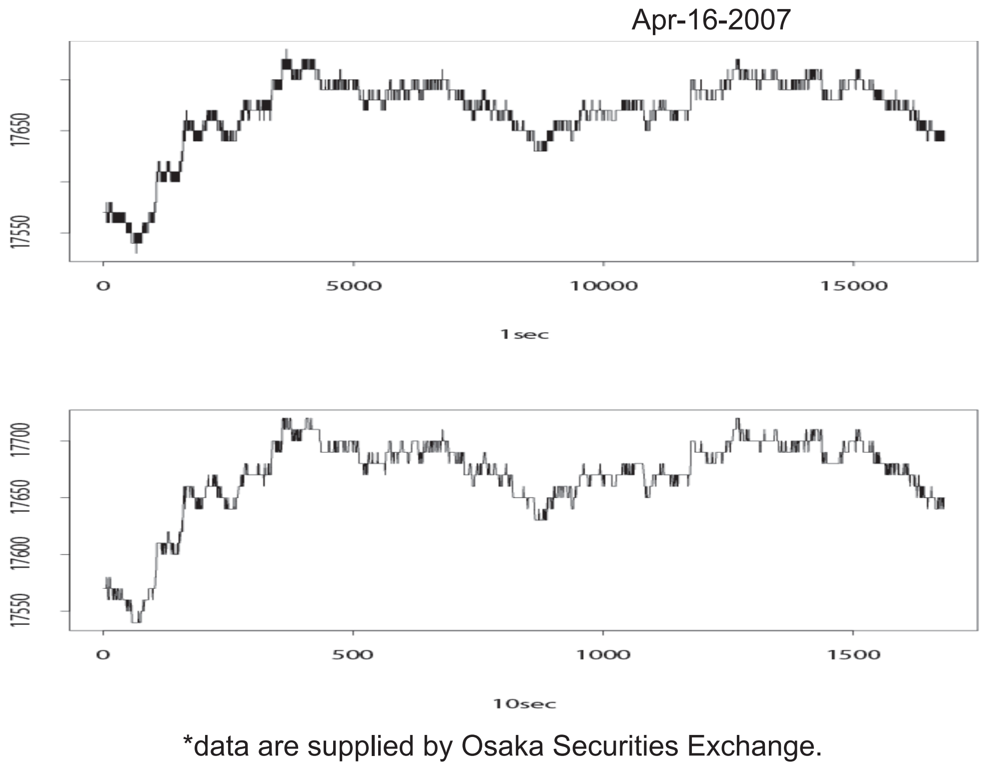

図 4　高頻度金融データの例 (日経 225 先物).

$$
\mathbf{C}_n = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 1 & 1 & 0 & \ldots & 0 \\ 1 & 1 & 1 & \ldots & 0 \\ 1 & \ldots & 1 & 1 & 0 \\ 1 & \ldots & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ -1 & 1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \ldots & 0 \\ 0 & \ldots & -1 & 1 & 0 \\ 0 & \ldots & 0 & -1 & 1 \end{pmatrix}^{-1} \tag{4.8}
$$

である．このとき Kunitomo and Sato (2008, 2011) は $\mathbf{Y}^* = (\mathbf{y}_k^{*'})$ の情報を利用した分散共分散行列 (Integrated Variance-Covariance Matrix)

$$
\mathbf{\Sigma}_\xi = \int_0^1 \mathbf{\Sigma}_\xi(s) ds \tag{4.9}
$$

の推定量 ($\mathbf{\Sigma}_\xi(s) = \mathbf{C}_\xi(s)\mathbf{C}_\xi^{'}(s)$) として分離情報最尤推定量 (separating information maximum likelihood estimator, 通称 SIML) を提案している．さらに $\mathbf{Y}_n^* = (\mathbf{y}_k^{*'})$ より

$$
\hat{\Sigma}_\xi = \frac{1}{m_n} \sum_{k=1}^{m_n} \mathbf{y}_k^* \mathbf{y}_k^{*'} , \tag{4.10}
$$

ノイズの分散共分散推定量を

$$\hat{\Sigma}_u = \frac{1}{l_n} \sum_{k=n-l_n+1}^{n} a_{k,n}^{-1} \mathbf{y}_k^* \mathbf{y}_k^{*'} , \tag{4.11}$$

とすることを提唱している．この定式化では $m_n = n^{\alpha}$ $(0 < \alpha < 1/2)$, $l_n = n^{\beta}$ $(0 < \beta < 1)$ としたが，$n \to \infty$ の時の漸近的性質として一致性や漸近正規性（あるいはより一般的には混合正規性や安定収束 (stable convergence)）などの条件を調べている．なお

$$a_{k,n} = 4n \sin^2 \left[\frac{\pi}{2} \left(\frac{2k-1}{2n+1} \right) \right] \ (k = 1, \ldots, n) \, . \tag{4.12}$$

である．

ここで多次元観測値に対する真の確率過程における線形関係を推定する方法として，ベクトル $\mathbf{b} \neq \mathbf{0}$ についての連立方程式

$$\left[\hat{\Sigma}_{\xi} - \lambda \, \hat{\Sigma}_u \right] \mathbf{b} = \mathbf{0} \tag{4.13}$$

を満足する解から得られる最小固有値 λ_1 および最小固有値に対応する固有ベクトル $\mathbf{b}_{SI}$ を Kunitomo (2013) が考察している．この方法は一定区間の中の観測数 n が大きい場合には漸近的に正当化できるが，有限標本でも良い性質を持つことが期待できよう．さらにここで導入した SIML 法については Sato and Kunitomo (2011) では離散観測される $Y(t_i^n)$ に対して真の状態 $X(t)$ との非線形調整モデル

$$Y(t_i^n) = h(\boldsymbol{\xi}(t_i^n), Y(t_{i-1}^n), U(t_i^n)) \tag{4.14}$$

を考察している．$h(\cdot, \ldots, \cdot)$ は非線形関数であるが，例えば金融市場で観察される価格は，実際には取引所の定めるティックサイズの整数倍でしか変動しない離散観測値であることに鑑みれば，このような非線形変換による定式化により実態をより正確に反映できるという意味で整合的となる．また実際には金融取引は事前に決まった時刻に定期的に起きるのではなく需要側と供給側からの注文がランダムに一致して成立すると考えるべきであり，その場合の SIML 推定の性質について Kunitomo and Misaki (2013)，Misaki and Kunitomo (2013) が調べている．SIML 法について今後，高頻度金融データによる様々な応用が期待できる．

5. 計測誤差と統計学

本稿では経済統計，計量経済，計量ファイナンスという経済に関連する統計科学での 3 つの応用分野における計測誤差の扱いから生じる重要な統計的問題について，筆者の経験を元に幾つかの事例の議論を提示した．ここで取りあげたそれぞれの話題についてはいずれも関連する研究は既に膨大な数にのぼるが，著者の研究・教育活動の守備範囲の問題も

あり，個々の項目についての論点を網羅的に議論することせず，かなり主観的な選択の中で説明したことに改めて言及しておくとともに，前節までの考察から得られる結論および統計学の展望を述べる．

本稿での主要な第一の論点として幾つかの具体的例を取りあげて説明したように，一見すると全く別の分野における重要な課題についての検討の中から，統計学という観点から見直してみると，全く無関係に思えるある分野の研究課題が別の分野の統計的分析としばしば類似し，深く関連しているということである．すなわちある問題の統計的解法が実は他の問題の解法のヒントとなりうることが少なくないことの重要性が理解できるであろう．こうした論点は本稿で取りあげた 3 つの以外の研究分野，例えば生物統計 (biometrics) や心理統計学 (psychometrics) においても重要であり，「計測誤差」あるいは「観測誤差」を考慮することが鍵となることが少なくないと想像するが，残念ながら筆者には十分な経験がないので，ここで例示することは差し控えたい．

第二の論点は統計学の今日までの発展を振り返ると，統計学は古くから計測誤差，観測誤差を明示的に導入することにより飛躍的に発展してきた，と云えることであろう．その意味では現代的な統計学や統計科学では計測誤差の重要性を強調しておくことは全くのムダな作業では無いと思われる．例えばかなり短い期間であったが本稿を準備する作業を通じて幾つかの未解決の統計的課題があることに気が付かされた．例えば政府統計では新しい季節調整モデルの開発や年率 GDP 推計の作成上の問題点を指摘することは容易ではあるが，完全な解決までの道のりは短くなさそうである．既に 3 節で例示したように近年になり展開されている統計的モデル分析，中でも統計的時系列解析，統計的多変量解析，ノンパラメトリック解析，計算統計学，などの成果を利用することにより解決できる課題も少なくないと思われる．また統計的関係や構造方程式についてでは弱操作変数問題や多操作変数問題は完全な解決にはほど遠く，例えば統計的多変量解析における高次元問題との関わりは現在の興味深い話題の一つである．計量ファイナンスの問題では数理ファイナンスにおける連続時間確率過程と計量ファイナンスにおける離散時間観測問題の関連して様々な確率論や統計学上の課題が浮かび上がってくる．

他方，数学的に興味深い確率モデルに応用があるはずと信じて探すことに意味があるだろうか．統計学の発展を振り返るとき，様々な応用分野ではそれぞれの事情より特定の研究課題が登場してくるので，純粋な数理モデルを機械的に当てはめることにより有益な成果を挙げた例は少ないことに気がつく．むしろそれぞれの応用分野において生じる課題から問題の重要性を解釈し，理解する中で既存の確率モデル・統計モデルを修正していく中で良い研究成果が挙がることが多いように思われる．今後とも「観察できる現実と理論の健全な対話」こそが統計学における重要で生産的研究へのアプローチと考えられる．

最後になるが，最近になり日本のマスメディアでも「ビッグ・データ (big data)」や「ビ

ジネスにおける最強の武器としての統計学」などの話題がしばしば登場するようになっている．例えば経済・経営分野におけるマーケティングなどでのビッグ・データの応用は一昔まででは考えられないレベルのデータが容易に処理できるようになったことなどが背景にあると考えられる[21)]．他方，例えば政府統計・経済統計分野においても近年ではミクロ・データ (micro data) の活用が大きな研究テーマとなっているが，政府統計データの二次利用として個票データは古典的な統計学が想定していたデータ量から見るとかなり大きい．つまりラージ・データ (large data) の処理が既にこうした分野では重要となっているのである[22)]．こうした統計学・統計科学をとりまく統計データの状況は多くの統計家にとっては特に奇異ではないのであるが，経済・経営・生物などをはじめ多くの応用分野で扱われる統計データの大きさは一昔前の数理統計学の教科書が想定していた状況と比較するとかなり大きい場合が少なくない．

こうした中でも統計学側からこれまで報告されているビッグ・データの分析例を除いてみると，比較的標準的な既存の統計的方法に依存していることが少なくない．標準的教科書に記述されている説明を超えて，例えば消費者に関する膨大なデータ，あるいはDNA解析で得られる膨大な情報を利用してその中から「隠れた構造」や「統計的関係」を見いだすには安定した構造とランダムな誤差，計測誤差を区別することが重要であり，変数誤差 (errors-in-variables) モデルの役割も重要と考えられる．こうした目的のためには密度関数や条件付期待値・条件付き分散のノンパラメトリック推定やセミパラメトリック推定など統計学・統計科学における近年の研究動向と整合的であり，応用可能性が高まると考えられる．ラージ・データやビッグ・データの登場は応用分野のデータ分析やデータ解析ビジネスに携わる必ずしも統計家ではない人々と統計学者との交流の必要性を社会において認識する重要な機会となっていると考える．

ここでより具体的な例を挙げてみると，観察可能な変数間の関係の分析手段としてはこれまで応用上ではある種の定番であった被説明変数の期待値を説明する線形回帰 (linear regression) 分析に加えて，被説明変数の分布・分位点を説明する分位点回帰 (quantile regression) 分析がより適切な統計的方法となる状況が多くなる可能性が高い[23)]．分位点回帰とは説明変数ベクトル $\mathbf{z}_i$ を条件とした被説明変数 Y の $\tau-$ 分位点 $(0 < \tau < 1)$ を $Q_\tau(y_i|\mathbf{z}_i)$

---

21) 例えば 2012 年〜2013 年にかけて日本で発刊された幾つかのビジネス系の雑誌が特集を組んでいる．ビッグデータの定義は必ずしも明確に理解されているわけではないが，ここでは従来のデータベースの管理やデスクトップ計算機の処理能力を超えたデータと理解しておく．

22) ラージ・データは本稿での造語である．古典的統計学の理論が t 検定など small sample theory として登場したことに鑑み，データの大きさをスモール・データ small data, ラージ・データ large data, ビッグ・データ big data と分けてると，それぞれに対応して小標本理論，大標本理論，ビッグ・データ理論が構想できるかもしれない．

23) 分位点回帰問題の詳細な解説は Koenker (2005)，日本語では加藤他 (2009) などが参考となろう．後者では線形計画法との関連や損害保険の請求額の大きさによる要因分解の応用例を説明している．

とすると

$$Q_\tau(y_i|\mathbf{z}_i) = \boldsymbol{\beta}(\tau)'\mathbf{z}_i \quad (i = 1, \ldots, n) \tag{5.1}$$

とする統計モデルである．ここで係数ベクトル $\boldsymbol{\beta}(\tau)'$ が分位点の水準に依存することが特徴であり，推定量を求める統計計算では最小二乗法ではなく分位点に依存する損失関数

$$\rho_\tau(u) = \tau u \ (u \geq 0), \rho_\tau(u) = (\tau - 1)u \ (u < 0) \tag{5.2}$$

に基づき解を求めるが，この解を線形計画法の解としても表現することが可能である．例えば分位点回帰問題の経済における典型的な応用例としては賃金や所得を説明する要因が所得水準により異なる場合の分析であり，こうした状況で変数誤差が存在する場合の統計的方法についても例えば Chernozhukov and Hansen (2006) などの研究がデータへの応用と共に活発化しつつある．

さて本稿では限られた分野における幾つかの具体例を通じて，統計学・統計科学では古くから今日に至るまで統計データの分析と分析結果の活用の方法を開発することで発展していることを繰り返し指摘した．ビッグ・データは新たな形式の統計データの提示という意味では統計学・統計科学へ新たな検討課題を投げかけていると捉えることも可能であろう．ここで与えられたデータの中に「安定した統計的関係」が見いだせてこそ，データを分析する者が通常は期待するであろうデータを発生させている現象の理解と共に，まだ実現していない現象，あるいは人々の将来の行動の予測可能性がはじめて生じると考えられることをあらためて指摘しておきたい．大量のデータが利用可能となった現代において統計データの分析者が想定している観測データに含まれる真の関係と (広い意味での) 計測誤差・観測誤差の区別，計測誤差の適切な統計的処理の重要性が高まっていると考えられるのである．本稿で議論した計測誤差を巡る幾つかの問題の考察を通じて，こうした現代社会において，これまで以上に統計学・統計科学の有用性を発揮できることを読者諸氏が確信できれば幸いである．

謝辞

原稿に対する雑誌の編集者とレフェリーからの詳細かつ建設的コメントに感謝する．また元の草稿に対する赤司健太郎氏と加藤賢悟氏からのコメントにも感謝する．

参　考　文　献

Adcock, R. J. (1878). A problem in least squares, *Analyst*, **5**, 53–54.

Akaike, H. (1980). Seasonal adjustment by a Baysian modeling, *J. Time Ser. Anal.*, **1**, 1–13.

Akaike, H., Ozaki, T., Ishiguro, Ogata, Y., Kitagawa, G., Tamura, Y. H., Arahata, E., Katsura, K. and Tamura, Y. (1985). *TIMSAC-84: Part-I*, Computer Science Monograph No. 22, The Institute of Statistical Mathematics (統計数理研究所).

Akashi, K. and Kunitomo, N. (2010). The limited information maximum likelihood approach to dynamic panel structural equations, Discussion paper CIRJE-F-708, Graduate school of Economics, University of Tokyo.

(http://www.cirje.e.u-tokyo.ac.jp/research/ この HP アドレスは以下の文献でも同様), forthcoming in *Ann. Inst. Stat. Math.*

Akashi, K. and Kunitomo, N. (2012). Some properties of the LIML estimator in a dynamic panel structural equation, *Journal of Econometrics*, **166**, 167–183.

Anderson, T. W. (1951). Estimating linear restrictionson regression coefficients for multivariate normal distributions, *Ann. Math. Stat.*, **22**, 327–351.

Anderson, T. W. (1976). Estimation of linear functional relationships: approximate distributions and connections to simultaneous equations in econometrics, *J. R. Stat. Soc. B*, **38**, 1–36.

Anderson, T. W. (1984). Estimating linear statistical relationships, *Ann. Stat.*, **12**, 1–45.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd Edition, John-Wiley.

Anderson, T. W. (2011). Optimal significance tests in simultaneous equation models, Unpublished Manuscript.

Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations, *Ann. Math. Stat.*, **20**, 46–63.

Anderson, T. W. and Rubin, H. (1950). The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equation, *Ann. Math. Stat.*, **21**, 570–582.

Anderson, T. W., Stein, C. and Zaman, A. (1985). Best invariant estimation of a direction parameter, *Ann. Stat.*, **13**, 526–533.

Anderson, T. W., Kunitomo, N. and Matsushita, Y. (2010). On the asymptotic optimality of the LIML estimator with possibly many instruments, *Journal of Econometrics*, **157**, 191–204.

Anderson, T. W., Kunitomo, N. and Matsushita, Y. (2011). On finite sample properties of alternative estimators of coefficients in a structural equation with many instruments, *Journal of Econometrics*, **165**, 58–69.

Barndorff-Nielsen, O., Hansen, P., Lunde, A. and Shephard, N. (2008). Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise, *Econometrica*, **76**(6), 1481–1536.

Bickel, P. J. and Ritov, Y. (1987). Efficient estimation in the errorsin variables model, *Ann. Stat.*, **15**(2), 513–540.

Carroll, R., Ruppert, D., Stefanski, L. and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models*, 2nd Edition, Chapman & Hall.

Chamberlain, G. (2007). Decision theory applied to an instrumental variables model, *Econometrica*, **75**(3), 609–652. （訂正記事：Chamberlain, G. (2009). Comment on “Decision theory applied to an instrumental variables model”, *Econometrica*, **76**(6), 1565.）

Chernozhukov, V. and Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models, *Econometrica*, **132**, 491–525.

Engle, R. and Granger, C. (1987). Co-integration and error correction: Representation, estimation and testing, *Econometrica*, **55**, 251–276.

Friedman, M. (1956). *A Theory of Consumption Function*, Princeton University Press.

Fuller, W. (1987). *Measurement Error Models*, John-Wiley.

Godambe, V. P. (1960). An optimum properties of regular maximum likelihood estimation, *Ann. Math. Stat.*, **31**, 1208–1211.

Hansen, L. (1982). Large sample properties of generalized method of moments estimation, *Econometrica*, **50**, 1029–1054.

Hayashi, F. (2000). *Econometrics*, Princeton University Press.

Johansen, S. (1991). Estimation and hypothesis testing of cointegrated vectors in Gaussian vector autoregressive models, *Econometrica*, **59**, 1551–1580.

加藤賢悟，国友直人，増田智己 (2009).「Lasso 分位点回帰の理論と損害保険への応用」『日本統計学会誌』**38**(2), 121–148.

Kitagawa, G. (1981). A nonstationary time series model and its fitting by a recursive filter, *J. Time Ser. Anal.*, **2**(2), 103–116.

北川源四郎 (2005).『時系列解析入門』岩波書店.

Koenker, R. (2005). *Quantile Regression*, Cambridge University Press.

Kunitomo, N. (1980). Asymptotic expansions of distributions of estimators in a linear functional relationship and simultaneous equations, *J. Am. Stat. Assoc.*, **75**, 693–700.

国友直人（編）(2004).『解説 X-12-ARIMA(2002)』CIRJE-R-1(研究報告), 東京大学日本経済国際共同研究センター (CIRJE, http://www.cirje.e.u-tokyo.ac.jp/research/R1ab.html).

国友直人（編）(2006). 『季節調整法 X-12-ARIMA と日本の官庁統計』東京大学日本経済国際共同研究センター, 研究報告 CIRJE-R-5.

国友直人 (2011).『構造方程式モデルと計量経済学』朝倉書店.

Kunitomo, N. (2012). Improving the LIML estimation with many instruments and persistent heteroscedasticity, *Ann. Inst. Stat. Math.*, **64**, 881–910.

Kunitomo, N. (2013). On estimating structural relationships of diffusion processes, in preparation.

国友直人 (2013a).「政府統計と季節調整」『統計』6 月号, 17–21.

国友直人 (2013b).「GDP 統計に改善の余地」『日本経済新聞』, 2013 年 7 月 4 日.

国友直人，川崎能典 (2011). 「ベンチマーク問題と経済時系列：GDP 速報と GDP 確報を巡って」『経済学論集 (東京大学経済学部)』**2**(19), 77-1.

Kunitomo, N. and Matsushita, Y. (2009). Asymptotic expansions and higher order properties of semi-parametric estimatorsin a linear simultaneous equations, *J. Multivar. Anal.*, **100**, 1727–1751.

Kunitomo, N. and Misaki, H. (2013). The SIML estimation of integrated covariance and hedging coefficient under micro-market noise and random sampling, Discussion Paper CIRJE-F-893, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/).

Kunitomo, N. and Sato, S. (2008). Separating information maximum likelihood estimation of realized volatility and covariance with micro-market noise, CIRJE DP F-581, University of Tokyo, (http://www.e.u-tokyo.ac.jp/cirje/research/), forthcoming in *North American Journal of Economics and Finance*, Elsevier.

国友直人，佐藤整尚 (2010).「GDP 速報の推定法の改善について」『経済学論集 (東京大学経済学部)』**2**(21), 76-3.

Kunitomo, N. and Sato, S. (2011). The SIML estimation of realized volatility of Nikkei-225 futures and hedging coefficient with micro-market noise, *Math. Comput. Simul.*, **81**, 1272–1289, North-Holland.

Misaki, H. and Kunitomo, N. (2013). On robust properties of the SIML estimation of volatility under micro-market noise and random sampling, Discussion Paper CIRJE-F-892, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/).

森棟公夫 (1985). 『経済モデルの推定と検定』共立出版.

長井英生 (1999). 『確率微分方程式』共立出版.

Neyman, J. and Scott, E. L. (1948). Consistent estimates base on partially consistent observations, *Econometrica*, **16**, 1–32.

Owen, A. (2001). *Empirical Likelihood*, Chapman & Hall.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations, *Ann. Stat.*, **22**, 300–325.

佐藤整尚，国友直人 (2010). 「景気判断と平滑化問題：GDP 公表値を巡って」『経済学論集 (東京大学経済学部)』**2**(21), 72–87.

Sato, S. and Kunitomo, N. (2011). A robust estimation of integrated volatility under micro-market adjustments and round-off errors, CIRJE Discussion Paper, University of Tokyo.

Schennach, S. (2004). Estimation of nonlinear models with measurement error, *Econometrica*, **72**, 33–75.

杉山，藤越，杉浦，国友（共編）(2007). 『統計データ科学活用辞典』（共編）朝倉書店.

高岡慎，国友直人 (2010). 「最近のマクロ経済変動と季節調整 (貿易統計を題材に)」『経済学論集 (東京大学経済学部)』**76**(1), 56–74.

竹内啓 (1974). 「統計的推定の漸近理論」教育出版.

Takeuchi, T. and Morimune, K. (1985). The third-order efficiency of the extended maximum-likelihood estimatorsin a simultaneous equation system, *Econometrica*, 53-1.

統計数理 (2009).『確率過程の統計解析』統計数理研究所.

統計数理 (2011).『金融リスクの統計解析』統計数理研究所.

Tong, L. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicator, *J. Multivar. Anal.*, **65**, 139–165.