CIRJE-J-248

計測誤差と統計学

東京大学大学院経済学研究科
国友直人

2013 年 9 月

Measurement Errors and Statistics

Abstract

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 ara of Big-Data.

計測誤差と統計学*

国友直人†

2013 年 9 月

要約

経済時系列において生じる計測誤差についての最近の話題をきっかけに、経済に関連する経済統計、計量経済、計量ファイナンス、など統計科学、応用統計の研究分野における「計測誤差の統計的処理」の問題を振り返る。これまでに提案されている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、「Big-Data（ビッグデータ）」や「統計学の重要性」が話題となっている時代において今後の統計学、統計科学の方向性を展望する。

鍵言葉

四半期別 GDP 速報, 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イクロ・マーケットノイズ, ビッグ・データ (Big-Data), 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)-0.4 より 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-0.4 と発表、「しかし実は成長率は 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。

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･経済統計 (economic statistics)、計量経済 (econometrics)、計量ファイナンス (financial econometrics)、などの統計科学・応用統計の研究分野にお

*日本統計学会・会長講演案 (2013-9-12). 元の原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感謝する。

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

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今後の統計学の方向性を展望する。

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 年のリーマン・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 は１次速報値から約１カ月後に数値が改定されて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ある。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、これまで二つの要請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を挙げることができる。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、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 の公表値では１次速報で民間設備投資がマイナス 2.6 %となったが、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推定誤差により生じる。ここで経済時系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、その経路上での変化分そのものが重要である[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。

[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.。

< 図 1(GDP 原系列) を挿入 >

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８年間の四半期GDP の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て経済政策の基礎としているのである。図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カ月間の各基礎データからある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。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、観測する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のである。ここで年始の小売店の営業形態や夏期休業期間の時間的変化など、経済をとりまく環境変化とともに変化する季節性も処理する必要もある。ところが、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。

ここで取りあげた例を単純化して説明しよう。ある時点での観測系列 Y_i $(i = 1, \cdots, n)$、真の状態変数 ξ_i $(i = 1, \cdots, n)$、観測誤差 U_i $(i = 1, \cdots, n)$ として

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

と表現する。真の状態はトレンド・循環成分 TC_i $(i = 1, \cdots, n)$（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 $(i = 1, \cdots, N)$ すると、加法成分モデルは

$$\xi_i = TC_i + S_i \ \ (i = 1, \cdots, n) \tag{2}$$

となる。

さて経済時系列の統計学の展開は初等的な移動平均法（過去と将来の数期分の平均を取る）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「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[3]。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、

$$\Delta^d T_i = U_{Ti} \sim N(0, \sigma_T^2) \ \ , \ \ S_i + \cdots + S_{i-s+1} = U_{Si} \sim N(0, \sigma_S^2) \tag{3}$$

と書けることが仮定される。(ここで $\Delta = 1 - \mathcal{L}, \mathcal{L}Y_i = Y_{i-1}, \Delta Y_i = Y_i - Y_{i-1}$ である。) 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

[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]例えば北川 (2005) を参照。

representation) を利用した最適フィルタリング (Kalman フィルターリング) 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。一方、X-12-ARIMA は初等的な移動平均法をベースとしつつも、その内部で様々な統計的モデルを多用している。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、DECOMP モデルによる四半期データの季節成分推定において $d=2, s=4, C_i=0$ $(i=1,\cdots,n)$ 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))。季節階差操作 $\Delta_s = 1-\mathcal{L}^s, \mathcal{L}^s Y_i = Y_{i-s}, \Delta_s Y_i = Y_i - Y_{i-s}$ として、階差・季節階差を原系列に作用すると時系列

$$X_i = \Delta\Delta_s Y_i = Y_i - Y_{i-1} - Y_{i-s} + Y_{i-s-1} \tag{4}$$

がある定常移動平均過程 (stationary moving average process) にしたがうことがわかる。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）実行していると解釈できるのである。

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･高岡 (2010), 国友･佐藤 (2010,2011), 国友・川崎 (2010) 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[4]。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、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 Errors*) の重要性について古くから考察されている。観測誤差 (Measurement Error) モデルに関する統計的モデルの議論は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

[4] 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。日本の政府統計の担当者は限られた人員、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。

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、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経済学者 M. Friedman (1958) の消費関数に関する恒常所得仮設 (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 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。

＜図 2(直交回帰) を挿入＞

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) , $i = 1, \cdots, n$ についての観測値ベクトル (x_i, y_i) , $i = 1, \cdots, n$ 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) が得られる状況を考えよう。このときの推定問題を Adcock (1978) は考察している。ここでは Adcock (1978) をやや現代風に解釈すると、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

$$L(\alpha, \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^n} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} [(x_i - \xi_i)^2 + (y_i - \alpha - \beta\xi_i)^2]\} \tag{5}$$

で与えられる[5]。この尤度関数の最大化は損失関数

$$L_2 = \sum_{i=1}^{n} [(x_i - \xi_i)^2 + (y_i - \alpha - \beta\xi_i)^2] \tag{6}$$

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るので直交回帰 (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「真の直線」の推定問題である。母数 ξ_i について最適化すると損失関数は

$$L_3 = \frac{\sum_{i=1}^{n} [(y_i - \alpha - \beta x_i)^2]}{1 + \beta^2} \tag{7}$$

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 (1978) 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。

ここで母数 α について (7) を最小化すると

$$L_4 = \frac{(1, -\beta) \sum_{i=1}^{n} \begin{pmatrix} y_i - \bar{y} \\ x_i - \bar{x} \end{pmatrix} (y_i - \bar{y}, x_i - \bar{x}) \begin{pmatrix} 1 \\ -\beta \end{pmatrix}}{(1, -\beta) \begin{pmatrix} 1 \\ -\beta \end{pmatrix}} \tag{8}$$

[5] ここで ξ_i は母数と考えたので関数関係 (functional relationship) 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 (structural relationship)) モデルである。

の最小化問題となる。したがって標本二次積率行列を

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}\begin{pmatrix} y_i - \bar{y} \\ x_i - \bar{x} \end{pmatrix}(y_i - \bar{y}, x_i - \bar{x}) \tag{9}$$

とすると、固有方程式

$$[\mathbf{S} - \lambda\mathbf{I}_2]\mathbf{b} = \mathbf{0} \tag{10}$$

を満足する最小固有値 λ_1 および最小固有値に対応する固有ベクトル $\mathbf{b} = (1, -b_2)'$ を求める問題に一致する。

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 (permanent income hypothesis) により消費-所得に関する計量経済モデルを用いて説明した。観測される所得 Y, 観測される消費 C そのものではなく、消費者にとっての所得-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

$$C_p = kY_p \ \ (k \text{ は一定の定数}) \tag{11}$$

が成立すると考える。観測される所得には一時的所得 Y_t, 観察される消費 C には一時的消費 C_t が加えられていると見ると、個人の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 Firedman 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とみると

$$Y = Y_p + Y_t\ ,\ C = C_p + C_t\ ,\ C_p = kY_p \tag{12}$$

となるので、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そのものであることが分かる。なお Friedman は実証分析では (しばしば一昔前の経済学者がそうであるように) 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置いてしばしば別々に議論されていた線形関数関係 (linear functional relationship) モデルおよび線形構造関係 (linear structural relationship) モデルについて、統計的多変量解析の観点から Anderson (1984) は包括的な議論を展開している。i $(i = 1, \cdots, n)$ 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 $\boldsymbol{\xi}_i$ とすると

$$\mathbf{Y}_i = \boldsymbol{\xi}_i + \mathbf{U}_i \ \ (i = 1, \cdots, n) \tag{13}$$

と表そう。（ここで誤差の期待値ベクトル $\mathbf{E}[\mathbf{U}_i] = \mathbf{0}$, 分散共分散行列 $\mathbf{E}[\mathbf{U}_i\mathbf{U}_i'] = \mathbf{\Sigma}_u$ としておく。）ここで真のベクトルが m 次元 $(1 \le m < p)$ 線形空間上にあるとすると $p \times q$ 行列 $\mathbf{B}$ $(q = p - m)$、定ベクトル $\mathbf{c}(q \times 1)$ により

$$\mathbf{B}'\boldsymbol{\xi}_i = \mathbf{c} \ \ (i = 1, \cdots, n) \tag{14}$$

と表現される。Adcock のモデルは $p = 2, q = 1, \mathbf{Y}_i = (Y_i, X_i), \mathbf{B} = (1, -\beta)', c = \alpha$ となる場合である。ここで $\mathbf{U}_i$ が互いに独立に $N_p(\mathbf{0}, \mathbf{\Sigma}_u)$ にしたがう確率変数、$\mathbf{\Sigma}_u$ が (正定

符合行列の) 未知母数とすると一般には母数 $(\mathbf{B}, \mathbf{\Sigma}_u)$ を識別可能 (identifiable) ではないことに注意する[6]。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 (5) では $\mathbf{\Sigma}_u = \sigma^2 \mathbf{I}_2$ 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、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) は固有方程式

$$(15) \qquad [\mathbf{S} - \lambda \mathbf{I}_p]\mathbf{b} = \mathbf{0}\ , \mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})'$$

を満足する小さい方から q 個の固有値に対応する固有ベクトル $\mathbf{b}_i$ $(i = 1, \cdots, q)$ からなる行列 $\hat{\mathbf{B}}$ で与えられる。したがってここでは標本から計算される固有値問題の解が重要な役割を演じている。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ランク次元 (reduced rank regression) 問題の解としても得られる。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、(13) 式の左から正則行列 $\mathbf{\Gamma} = (\mathbf{B}, \mathbf{B}^{\perp})'$ を掛けて再び整理すると

$$(16) \qquad \mathbf{Y}_i = \boldsymbol{\mu} + \mathbf{\Lambda}\boldsymbol{\nu}_i + \mathbf{U}_i \ \ (i = 1, \cdots, n)$$

と表現できる。ここで $\mathbf{\Lambda}$ は $p \times m$ 行列 (階数は $m \leq p$)、$\boldsymbol{\nu}_i = \mathbf{B}^{\perp\prime}\boldsymbol{\xi}_i$ 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。なお同論文によれば 1984 年の時点でも、ここでとりあげた固有値問題としての分析は時間を超え、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) モデルに関する有限標本における理論的結果としては、ここで導かれた推定方式はある損失関数の下で許容的 (admissible) [7]となることが Anderson, Stein and Zaman (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理論的結果としては、$p = 2$ の場合について竹内 (1974) の 4 章には様々な詳しい結果[8]が報告されているが、その後 Bickel and Ritov (1987) の研究がある。

構造方程式・多操作変数・弱操作変数・パネルモデル

次に Andesrson and Rubin (1949) などにより開発された[9] 構造方程式 (*structural equation*) モデルの統計的分析におけ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

[6] ここで母数ベクトル $\boldsymbol{\theta}$ が識別可能 (*Identifiable*) 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。

[7] 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。

[8] 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、あまり引用されていないようである。

[9] 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. Havelmo, L.R.Klein　などはノーベル経済学賞を受賞している。

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である。以来、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、

$$\mathbf{Y}_i = \boldsymbol{\xi}_i(\mathbf{Z}_i) + \mathbf{V}_i \quad (i = 1, \cdots, n) \tag{17}$$

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 $\boldsymbol{\gamma}_1$, $N(0, \sigma^2)$ にしたがう誤差項 u_i とするとき、線形構造方程式モデル

$$Y_{1i} = \boldsymbol{\beta}_2' \mathbf{Y}_{2i} + \boldsymbol{\gamma}_1' \mathbf{Z}_{1i} + u_i \quad (i = 1, \cdots, n) \tag{18}$$

の推定問題を考えよう。(誤差項 u_i の期待値 0, 分散 σ^2 とする。) $G \times 1$ ($p = G_2 + 1$) 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

$$\mathbf{Y}_i = \mathbf{\Pi}' \mathbf{Z}_i + \mathbf{V}_i \tag{19}$$

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

$$\boldsymbol{\beta}' \mathbf{\Pi}' = (\boldsymbol{\gamma}_1', \mathbf{0}') \tag{20}$$

が導かれるので、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では $p = 2, \mathbf{Y}_i = (P_i, Q_i)'$($P$ と Q は価格と数量) である。
ここでさらに誤差項に (多次元) 正規性を仮定すると十分統計量は

$$\hat{\mathbf{\Pi}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \ , \ \ \mathbf{C} = \mathbf{Y}'(\mathbf{I}_G - \mathbf{P}_Z)\mathbf{Y} \tag{21}$$

となる。(ここで $n \times p$ 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。) Anderson and Rubin (1949) 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) は固有方程式

$$[\mathbf{G} - \lambda \mathbf{C}]\mathbf{b} = \mathbf{0} \tag{22}$$

を満足する最小固有値 λ_1 に対応する固有ベクトル $\mathbf{b} = (1, -\mathbf{b}_2')'$ である。ここで $p \times p$ 標本行列

$$\mathbf{G} = \mathbf{Y}'(\mathbf{P}_Z - \mathbf{P}_{Z_1})\mathbf{Y} \tag{23}$$

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。特に $\lambda=0$ に設定すると二段階最小二乗推定量が得られる。

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（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-Morimune(1985) の結果などについて森棟 (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。こうした中で既に引用した 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」ことも興味深いが、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) [10]や Owen (2001) による経験尤度法 (empirical likelihood method) の応用など Non-parametric な統計的方法の応用が著しいが、計量経済学 (Econometrics) における近年の議論の一端は Hayashi (2000) や国友 (2011) で知ることができる。

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問題 *many instruments*、説明変数・操作変数の識別力が弱い場合の弱操作変数問題 *weak instruments* 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ものである。後者は誤差項の正規性を仮定すると統計量 $\mathbf{G}$ がしたがう非心ウイッシャート分布の非心度が十分に大きくない場合を意味することが多い。

なお多操作変数問題については、$0<c<1$ 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。最近の研究としては Anderson, Kunitomo and Matsushita (2010, 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であ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）$i=1,\cdots,N$ に関する時間 $t=1,\cdots,T$ におけるデータ $\mathbf{Y}_{it},\mathbf{Z}_{it}$ が利用可能であるとしよう。このとき個別効果 $\eta_i,\boldsymbol{\pi}_i$ $(i=1,\cdots,N)$ を含む線形構造方程式

$$Y_{1it}=\boldsymbol{\beta}^{'}\mathbf{Y}_{2it}+\boldsymbol{\gamma}_1^{'}\mathbf{Z}_{1it}+\eta_i+u_{it}\ \ (i=1,\cdots,N;t=1,\cdots,T) \tag{24}$$

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

$$\mathbf{Y}_{it}=\mathbf{\Pi}^{'}\mathbf{Z}_{it}+\boldsymbol{\pi}_i+\mathbf{V}_{it} \tag{25}$$

とする。この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 $\eta_i,\boldsymbol{\pi}_i$ $(i=1,\cdots,N)$ も多くなるので古典的な統計的問題ではない攪乱母

[10]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 Hansen (1982) の GMM と呼ばれている。

数 (Incidenral Parameters) が存在する統計的問題となる。ここでの統計的分析ではクロスセッション次元 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) は有力な手段となる。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。こうした動的パネル構造方程式問題 *Dynamic Panel Structural Equation* 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、ある種の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

$$\mathbf{y}_{it}^{(f)} = c_t\Big[\mathbf{y}_{it} - \frac{1}{T-t}(\mathbf{y}_{it+1} + \cdots + \mathbf{y}_{iT})\Big] \tag{26}$$

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

$$\mathbf{z}_{it-1}^{(b)} = b_t\Big[\mathbf{z}_{it-1} - \frac{1}{t}(\mathbf{z}_{it-2} + \cdots + \mathbf{z}_{i0} + \mathbf{z}_{i(-1)})\Big]\ , \tag{27}$$

で定められる後ろ向きファイルター (backward-filter, ただし $b_t^2 = t/(t+1), t = 1, ..., T-1)$ などを利用することが有力な分析手段となっている。

計測誤差と変数誤差モデルの拡張

本節ではこれ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展開を説明したが、近年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るようになっている。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、非構造、などを考慮すると様々な可能性があり、計量経済分野においても活発に議論されている。

本稿の筆者により限られた範囲であるが視野をより統計学分野全般に広げると、米国統計学会長を経験した著者による書物 Fuller (1987) は様々な応用上に重要な論点を議論している[11]。その後、生物統計学における応用例を念頭にした Carroll, Ruppert, Stefanski and Crainiceanu (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的となる場合が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 $\boldsymbol{\Sigma}_u$ は識別可能 (identifiable)、推定可能である。このとき線型モデルの場合には (13) 式と (14) 式の代わりに

$$\mathbf{Y}_{ij} = \boldsymbol{\xi}_i + \boldsymbol{\mu} + \mathbf{U}_{ij} \quad (i = 1, \cdots, n; j = 1, \cdots, k) \tag{28}$$

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、(14) 式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。この場合には例えば誤差項の分散共分散行列 $\boldsymbol{\Sigma}_u$ は

$$\hat{\boldsymbol{\Sigma}}_u = \frac{1}{k-1}\sum_{j=1}^{k}(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i\cdot})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i\cdot})' \tag{29}$$

[11] 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部分的な言及となる。

により推定可能となる。

さらに計量経済学の問題として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 $F_{\boldsymbol{\xi}}$ にしばがう場合、$\mathbf{Y}_{ij}$ の確率分布から $F_{\boldsymbol{\xi}}$ が識別可能となることが注目されている。この方向では *Kotlarski* の補題による結果では、観測誤差と真の変量について i.i.d. の仮定の下で (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 $\phi_\xi(t)$ は

$$\phi_\xi(t) = \exp\left[\int_0^t \frac{\frac{\partial \psi(0,u_2)}{\partial u_1}}{\psi(0,u_2)} du_2\right] \tag{30}$$

により与えられることが重要である。Tong and Vuong (1998) による研究を一つの契機として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) による非線形推定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

1980 年代-1990 年代における構造方程式モデルに関係して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[12]。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、Jonhansen (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 $\mathbf{Y}_{i-j}\ (j=1,\cdots,q)$、係数行列 $\boldsymbol{\Pi}_j\ (j=1,\cdots,1), \boldsymbol{\Pi}^*$ とすると

$$\mathbf{Y}_i = \sum_{j=1}^{q} \boldsymbol{\Pi}_j^{'} \mathbf{Y}_{i-j} + \boldsymbol{\Pi}^{*'} \mathbf{Z}_i^* + \mathbf{V}_i\ , i=1,\cdots,n \tag{31}$$

となり、誘導型 (31) は多次元自己回帰型の離散時系列モデルとなる。（ここで $\mathbf{V}_i$ 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。）さらに階差操作を何回か施し適当に定義された係数行列 $\mathbf{B}_j$ を利用すると

$$\Delta \boldsymbol{Y}_i = \boldsymbol{B}_1^{'} \boldsymbol{Y}_{i-1} + \sum_{j=2}^{q} \boldsymbol{B}_j^{'} \Delta \boldsymbol{Y}_{i-(j-1)} + \boldsymbol{\Pi}^{*'} \mathbf{Z}_i^* + \boldsymbol{V}_i \tag{32}$$

と表現できる。ここで時間の経過にともなう内生変数ベクトルの挙動は固有方程式

$$|\lambda^q \mathbf{I}_p - \sum_{j=1}^{q} \lambda^{q-j} \boldsymbol{\Pi}_j| = 0 \tag{33}$$

の根の位置により決定されるが、特に内生変数ベクトル間の関係は行列 $\mathbf{B}_1$(および $\boldsymbol{\Pi}^{*'}$) の階数により定常性と非定常性の混合状態がある得ることが重要である[13]。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、

$$\mathbf{X}_i = \mathbf{B}^{'} \mathbf{Y}_i \tag{34}$$

が定常過程となるような $p \times q$ 行列 $\mathbf{B} (p > q \geq 1)$ が存在する条件は表現 (32) における行列 $\mathbf{B}_1$ の階数が退化する条件となるので、数学的には既に議論した構造方程式の制約条件と同一である。したがってその統計学的な議論の基本構造を振り返って見ると Anderson

[12] 日本では山本拓教授などの貢献がある。

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

(1951), Anderson (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。したがって既に説明した 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。

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。

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[14]。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確率過程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。

< 図 3(高頻度金融データ・日経 225-Futures) を挿入 >

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重要と思われるが、ここでは確率過程モデルと計測誤差について Kunitomo and Sato (2008, 2012), 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

$$\boldsymbol{\xi}(t) = \boldsymbol{\xi}(0) + \int_0^t \boldsymbol{\sigma}(s) d\mathbf{B}(s) \tag{35}$$

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 $\boldsymbol{\sigma}(s)$ はボラティィリティ関数 (あるいは行列)、$\mathbf{B}(s)$ は q 次元ブラウン運動、積分は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、高頻度ファイナンス・データが利用可能になるにしたがって現実に観察される高頻度金融データとは矛盾することが知られるようになっている。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

$$\mathbf{Y}(t_i^n) = \boldsymbol{\xi}(t_i^n) + \mathbf{U}(t_i^n) \quad (i = 1, \cdots, n) \tag{36}$$

[14]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。

とする。ここで $\mathbf{U}(t_i^n)$ は期待値 0, 分散共分散 (一定)$\mathbf{\Sigma}_u$ 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マイクロ・マーケット・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 $\mathbf{U}(t_i^n) = \mathbf{0} \ \ (i = 1, \cdots, n)$ であれば)、数学的な連続時間モデルでは時間間隔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二乗変動和（実現ボラティィリティ (realized volatility) と呼ばれている) は

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

は $n \rightarrow \infty (T = 1$ に固定) のとき連続時間の量

$$
(38) \qquad < \boldsymbol{\xi}, \boldsymbol{\xi} >_1 = \int_0^1 \boldsymbol{\sigma}(s) \boldsymbol{\sigma}^{'}(s) ds
$$

に収束することが確率過程論では知られている[15]。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モデルがそのままでは成立しない」という事実が重要である。例えば図 3 はある日の日経平均 225-Futures 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、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は十分に検討すべき問題なのである。
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ィリティ (Integrated Volatility)

$$
(39) \qquad IV = \int_0^1 \sigma^2(s) ds
$$

を観測データから推定することである。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の推定が問題である。この問題については Barndorff-Nielsen, O., P. Hansen, A. Lunde and N. Shephard (2008) の研究など幾つかの統計的方法が提唱されているが、個々では 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

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

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

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

$$
(42) \qquad \mathbf{C}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & 0 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \cdots & 0 \\ 0 & \cdots & -1 & 1 & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}^{-1}
$$

[15] 例えば長井 (1999) を参照。

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)

$$\mathbf{\Sigma}_\xi = \int_0^1 \sigma^2(s)ds \tag{43}$$

の分離情報最尤推定量 (separating information maximum likelihood estimator, 通称 SIML) を $\mathbf{Y}_n^* = (\mathbf{y}_k^{*'})$ より

$$\hat{\Sigma}_\xi = \frac{1}{m_n}\sum_{k=1}^{m_n} \mathbf{y}_k^* \mathbf{y}_k^{*'} \ , \tag{44}$$

ノイズの分散共分散推定量を

$$\hat{\Sigma}_u = \frac{1}{m_n}\sum_{k=n-l_n+1}^{n} a_{k,n}^{-1} \mathbf{y}_k^* \mathbf{y}_k^{*'} \ , \tag{45}$$

とすることを提唱し、$n \to \infty$ の時の漸近的性質を調べている。ここで

$$a_{k,n} = 4n \sin^2\left[\frac{\pi}{2}\left(\frac{2k-1}{2n+1}\right)\right] \ (k = 1, \cdots, n) \ . \tag{46}$$

である。さらに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、固有方程式

$$\left[\hat{\Sigma}_\xi - \lambda \, \hat{\Sigma}_u\right] \mathbf{b} = \mathbf{0} \tag{47}$$

を満足する最小固有値 λ_1 および最小固有値に対応する固有ベクトル $\mathbf{b}$ 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。

ここで導入した SIMS 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

$$Y(t_i^n) = h(\boldsymbol{\xi}(t_i^n), Y(t_{i-1}^n), U(t_i^n))) \tag{48}$$

を考察している。実際に金融市場で観察される価格の変動幅はティックサイズと呼ばれる離散間隔であり、こうした定式化は整合的となる。また実際には取引はランダムに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 Misaki and Kunitomo (2013), Kunitomo and Misaki (2013) が調べているが、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よう。

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における３つの研究分野において、計測誤差の扱いから生じる重要な統計的分析について、筆者の経験を元に幾つかの事例の議論を行った。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、講演の性格や著者の能力の問題もあり、本稿では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、主観的な選択であることに改めて言及しておく。

本稿の主要な第一の論点は、一見すると全く別の研究分野における重要な課題について検討した中からの議論を、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に対する分析が実はしばしば類似する別の分野の統計的分析と関連してということである。ある問題の統計的解法が実は他の問題の解法に関わっている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、例えば生物統計や教育統計においても重要であり、「計測誤差」あるいは「観測誤差」を考慮することが重要であると想像するが、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。

第二に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。かなり短い時間であるが本稿を準備する作業を通じて幾つかの未解決の問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用意ではあるが、完全な解決までの道のりは短くなさそうである。しかしながら２節で議論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のである。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統計的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、あるいは応用があると信じて探すことに意味があるだろうか？筆者はこのような研究アプローチには比較的否定的である。様々な研究分野では様々な事情より研究課題が登場してくるのであり、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あまりないと思われる。むしろ研究課題にとり重要な課題を解釈し、理解する中で既存の確率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。「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」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 (Big-Data) 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[16]。消費者に関する膨大なデータ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、変数誤差モデル (errors-in-variables) の役割も重要と考えられる。こうした状況で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がより重要になると考えられる。さらに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) 分析よりも、むしろ被説明変数の分布・分位点を説明す

[16]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。

る分位点回帰 (quantile regression) 分析がより適切な統計的方法となる可能性が高い[17]。分位点回帰とは被説明変数 Y の期待値ではなく説明変数ベクトル $\mathbf{z}_i$ を条件とした Y の $\tau-$ 分位点 $(0<\tau<1)$ を $Q_\tau(y_i|\mathbf{z}_i)$ とすると

$$Q_\tau(y_i|\mathbf{z}_i)=\boldsymbol{\beta}(\tau)^{'}\mathbf{z}_i\ \ (i=1,\cdots,n) \tag{49}$$

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

$$\rho_\tau(u)=\tau u\ (u\geq 0),\rho_\tau(u)=(\tau-1)u\ (u<0) \tag{50}$$

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。

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。大量のデータが利用可能となった現代においてこそ統計データの分析者が想定している真の関係と (広い意味での) 計測誤差・観測誤差を含む観測データの区別、計測誤差の問題の統計的処理の重要性が高まっているのである。本稿の考察を通じて、こうした現代においては、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が確信できれば幸いである。

参考文献

[1] Adcock, R. J. (1878), "A Problem in Least Squares," *Analyst*, 5, 53-54.

[2] Akashi, K. and N. Kunitomo (2012), "Some Properties of the LIML Estimator in a Dynamic Panel Structural Equation," *Journal of Econometrics*, Vol.166, 167-183.

[3] Akashi, K. and N. Kunitomo (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).

[4] Anderson, T.W. (1951), "Estimating Linear RestrictionsonRegression Coefficients forMultivariate NormalDistributions," *Annals of Mathematical Statistics*, 22, 327-351.

[17] 分位点回帰問題の詳細な解説は Koenker (2006)、日本語では加藤・国友・増田 (2008) 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。

[5] Anderson, T.W. (1976), "Estimation of Linear Functional Relationships : Approximate Distributions and Connections to Simultaneous Equations in Econometrics," *Journal of the Royal Statistical Society*, B, Vol. 38, 1-36.

[6] Anderson, T.W. (1984), "Estimating Linear Statistical Relationships," *Annals of Statistics*, Vol. 12, 1-45.

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

[8] Anderson, T.W. (2011), "Optimal Significance Tests in Simultaneous Equation Models," Unpublished Manuscript.

[9] Anderson, T.W., N. Kunitomo and Y. Matsushita (2010), "On the Asymptotic Optimality of the LIML Estimator with Possibly Many Instruments," *Journal of Econometrics*, Vol. 157, 191-204.

[10] Anderson, T.W., N. Kunitomo, and Y. Matsushita (2011), "On Finite Sample Properties of Alternative Estimators of Coefficients in a Structural Equation with Many Instruments," *Journal of Econometrics*, Vol. 165, 58-69.

[11] Anderson, T. W. and Rubin, Herman (1949), "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Mathematical Statistics,* **20**, 46–63.

[12] Anderson, T.W. and H. Rubin (1950), "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equation," *Annals of Mathematical Statistics*, Vol. 21, 570-582.

[13] Anderson, T.W., C. Stein and A. Zaman (1982), "Best Invariant Estimation of a Direction Parameter," *Annals of Statistics*, Vol. 13, 526-533.

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

[15] Bickel,P.J. and Y. Ritov (1987), "Efficient estimation in the errorsin variables model," *The Annals of Statistics*, Vol.15-2, 513-540.

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

[17] Chernozhkov, V. and C. Hansen (2006), "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132, 491-525.

[18] Engle, R. and C. Granger (1987), "Co-integration and Error correction : Representation, Estimation and Testing," *Econometrica*, 55,251-76.

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

[20] Fuller, W. (1987),"Measurement Error Models," *John-Wiley.*

[21] Godambe, V. P. (1960), "An optimum properties of regular maximum likelihood estimation," *Annals of Mathematical Statistics*, 31, 1208-1211.

[22] Johansen, S. (1991), "Estimation and hypothesis testing of cointegrated vectors in Gaussian vector autoregressive models," *Econometrica*, 59, 1551-80.

[23] Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimation," *Econometrica*, 50, 1029-1054.

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

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

[26] Kunitomo, N. (1980) "Asymptotic Expansions of Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of American Statistical Association*, 75, 693-700.

[27] Kunitomo, N. (2012), "Improving the LIML estimation with many instruments and persistent heteroscedasticity," *Annals of Institute of Statistical Mathematics*, 64, 881-910.

[28] Kunitomo, N. (2013), "On Estimating Structural Relationships of Diffusion Processes," In preparation.

[29] Kunitomo, N. and S. Sato (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.

[30] Kunitomo, N. and S. Sato (2011), "The SIML Estimation of Realized Volatility of Nikkei-225 Futures and Hedging Coefficient with Micro-Market Noise," *Mathematics and Computers in Simulation*, 81, 1272-1289, North-Holland.

[31] Kunitomo, N. and Misaki (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/).

[32] 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/).

[33] Owen, A. (2001),"Empirical Likelihood," *Chapman* & *Hall*.

[34] Sato, S. and N. Kunitomo (2011) "A Robust Estimation of Integrated Volatility under Micro-Market Adjustments and Round-off Errors," CIRJE Discussion Paper, University of Tokyo.

[35] Schennach, S. (2004), "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 72, 33-75.

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

[37] Tong, Li and Vuong Q. (1998), "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicator," *Journal of Multivariate Analysis*, 65, 139-165.

[38] 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.

[39] 北川源四郎 (2005)「時系列解析入門」, 岩波書店。

[40] 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)。

[41] 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.

[42] 国友直人 (2011),「構造方程式モデルと計量経済学」, 朝倉書店。

[43] 国友直人 (2013a),「政府統計と季節調整」, 統計, 6 月号, 17-21。

[44] 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。

[45] 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。

[46] 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。

[47] 統計数理 (2009)「確率過程の統計解析」, 統計数理研究所。

[48] 統計数理 (2011)「金融リスクの統計解析」, 統計数理研究所。

[49] 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。

[50] 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。

[51] 竹内啓 (1974) 「統計的推定の漸近理論」, 教育出版。

[52] 森棟公夫 (1985) 「経済モデルの推定と検定」, 共立出版。

[53] 長井英生 (1999) 「確率微分方程式」, 共立出版。

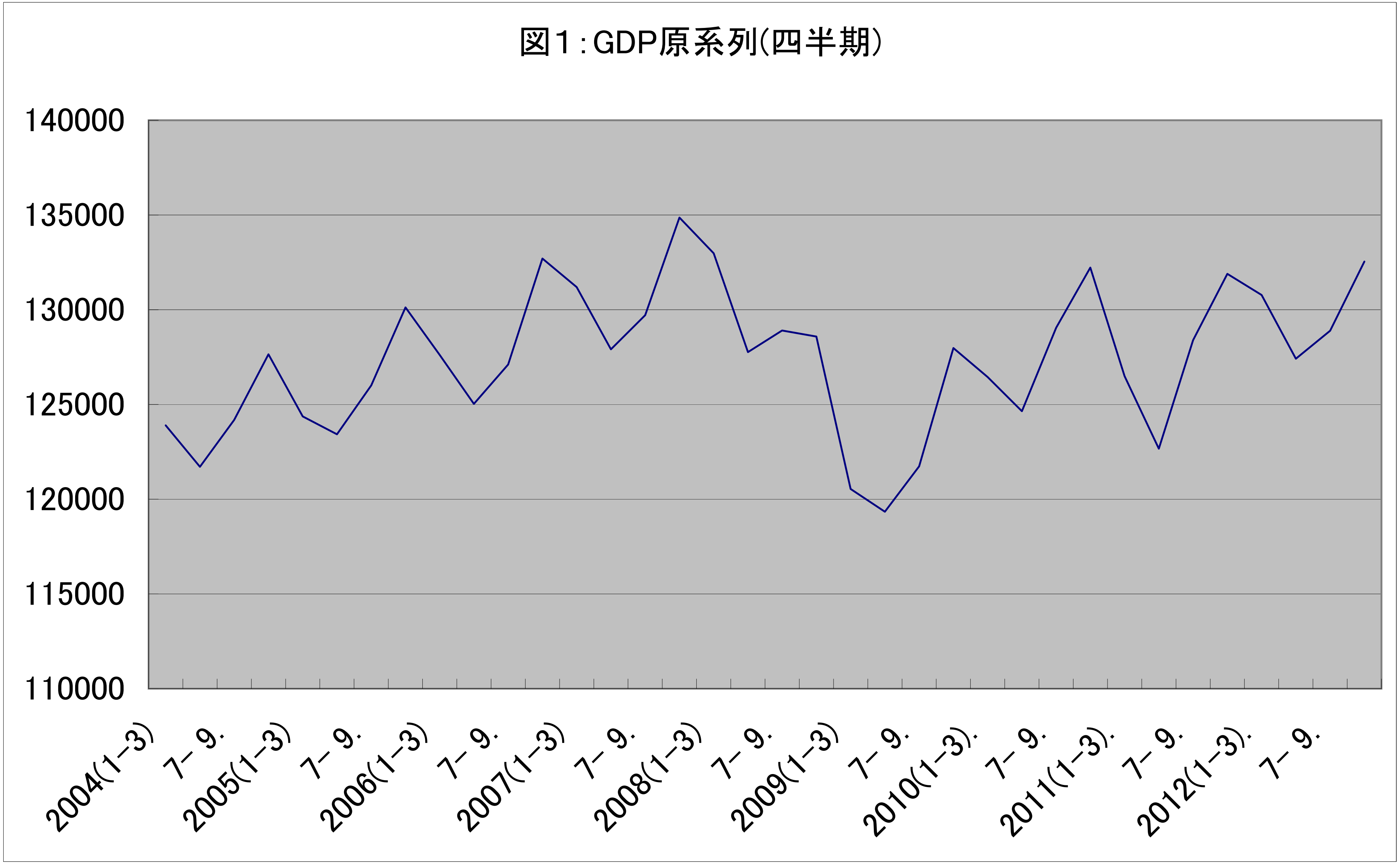


図1：GDP原系列(四半期)
140000
135000
130000
125000
120000
115000
110000
2004(1-3)
7-9.
2005(1-3)
7-9.
2006(1-3)
7-9.
2007(1-3)
7-9.
2008(1-3)
7-9.
2009(1-3)
7-9.
2010(1-3).
7-9.
2011(1-3).
7-9.
2012(1-3).
7-9.

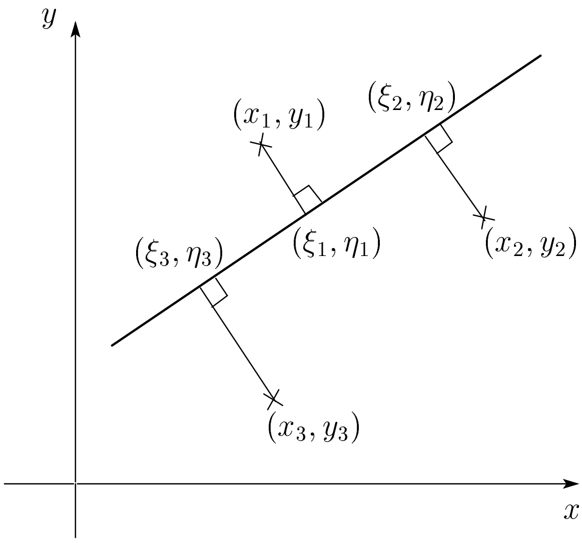
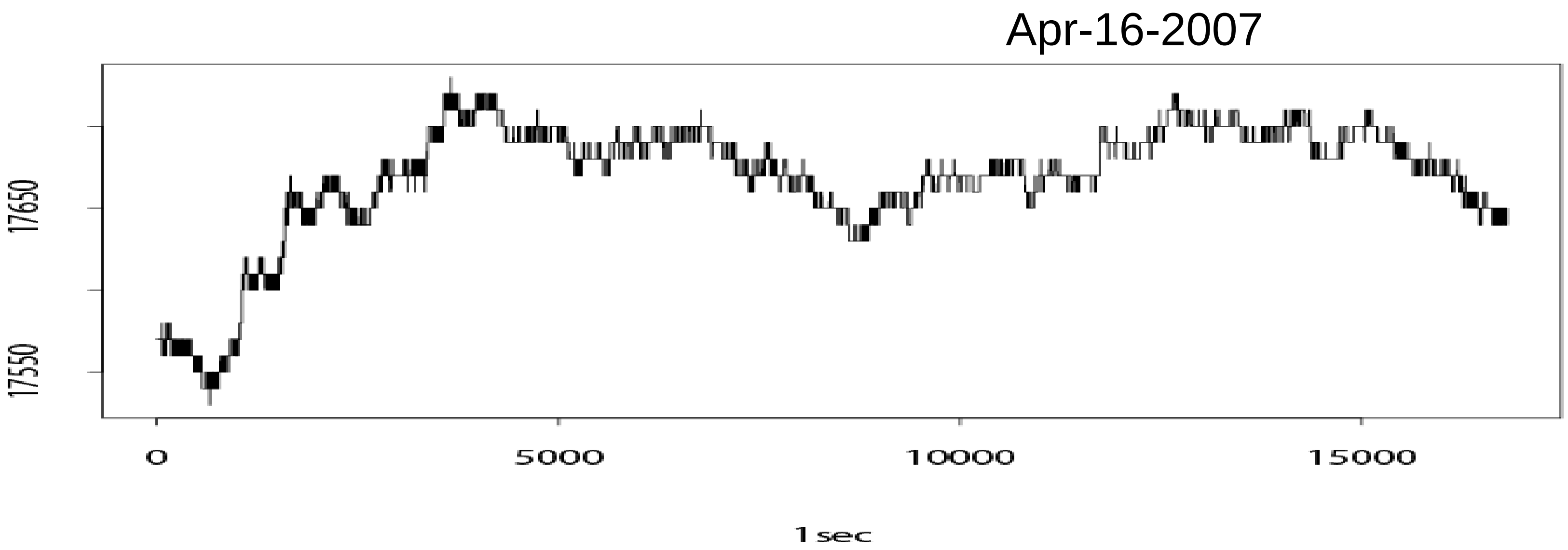
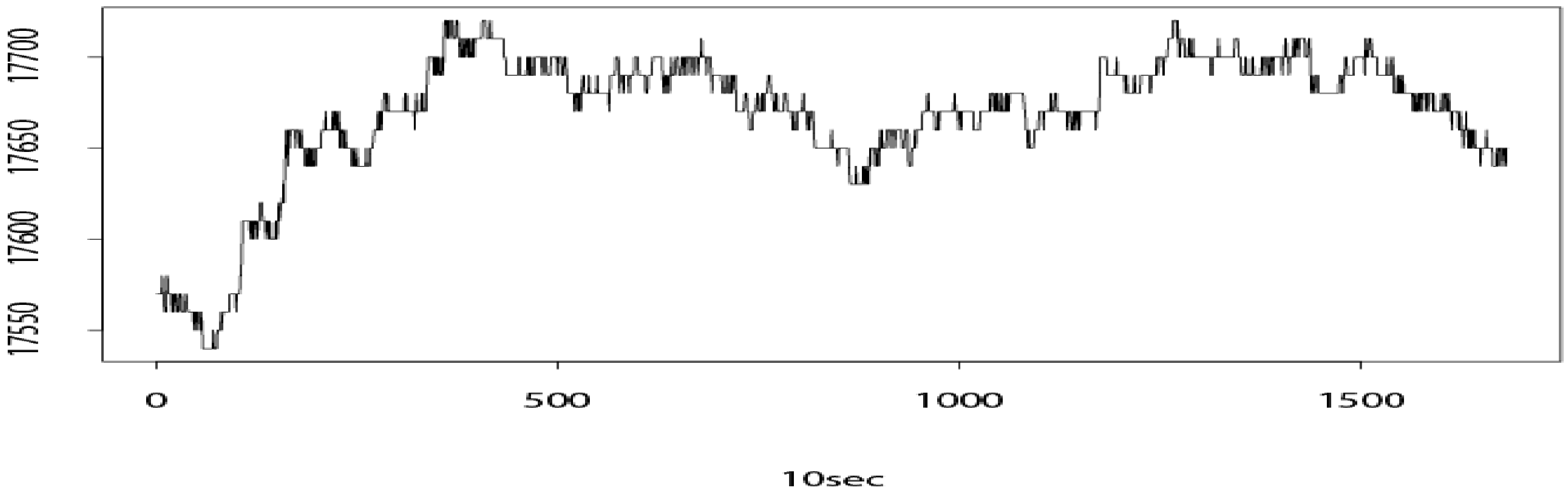


図 1: 直交回帰の例

図3: Nikkei-225-Futures Prices





*data are supplied by Osaka Securities Exchange.