イミテーションデータによるもっともらしいデータ分析
── SNA データを非現実的な過程から生成，分析する ──

中敷領　孝能

要　　旨

研究者が彼らのテーマをいくつかのモデルを使って研究するとき，彼らは－意識的か無意識的かを問わず－しばしば彼らのモデルを多かれ少なかれ制約されたモデル空間の中で作成しなければならない。そして構築されたモデルがそれらのテーマを十分に説明するとき，モデルは正しいものとみなされ，そして研究者の結論もまた正当なものとみなされる。しかし，われわれはこれらの結論ははじめのモデル空間が制約されていたので導き出されたのかもしれないということを考える必要がある。

本稿で，私はこの自明の問題を SNA データを分析することで示す。簡単な消費関数は統計学や計量経済学でポピュラーである。私はこの関数を再度取り扱う。現実のデータ構造から導き出されるいくつかの統計量から，私は GDP や消費のデータを発生させる。しかし，これらのデータは非現実的な方法で発生させられる。

データがどこから来ようが，研究者はそれらのデータを分析することができる。そして，よりもっともらしい結果はわれわれにより結果の信頼性をより確信させる。

私は本稿を次のように論述する。生成されたデータの確率的な特性について説明した後，単純な回帰モデルからはじめ，和分や共和分を含む時系列分析に進む。本稿の，あまり取り上げないがひとつの焦点は現代日本経済分析における価格の役割である。

1. 統計的推論の正当性

多くの場合，経済学者は経済分析にあたって特定の正統 (オーソドックス) とされる手順を踏むことを求められる。そして，その手順を遵守すれば，その結論の正当性は保障される。この，いわば「常識」を本稿では基本にかえって改めて考えてみることにする。

最初に以下のような，ほとんど現在ではむしろ誰も使わないかもしれない以下のような式を考えてみよう。

$$C=\alpha+\beta Y$$

ここで Y は GDP なり国民所得をあらわし，C は民間最終消費支出ないしその仲間たちを表している。名目値であるか実質値であるかはさしあたり関係ないのだが，日本経済の現状に照らせば，実質値のほうが本稿の論述にはより適することになる。

さてこの方程式は，実際には特定の C と Y のペアについて，3 点以上で成立することはない，つまり，直線には乗ってこないことを理由として，次のように書き換えられる。

$$C=\alpha+\beta Y+u$$

そして，現実の Y，C はまずは次のように説明される。α と β が定数であることはもちろん，Y もばらつきはあるものの定数 (だから $\alpha+\beta Y$ までは決まっている)，そして，u は，あたかも「0 を平均として，特定の標準偏差を持つ数が書いてある球」要するに平均 0 の iid の確率変数をそれらがしまってある箱から取り出すようにして取り出され，その結果，C の値が決定されるのだと。

しかし，通常考えられているようなスタンダードな問題，つまり今度取り出す玉が前 (やもっと前) に引いた玉の値に影響を受けるだとか，箱から一回引くごとに，玉に書いてある数字の絶対値が膨らんでくるとか，引く人 (Y の特定の値) によって玉の値が影響を受けるとかいったことのほかに，そもそもこの式には別の問題がある。

それは，Y は定数であるにもかかわらず，C は確率変数だという想定である。

当然ながら，Y は C とそのほかの変数の合計であることは経済学部の初等教育で学ぶような内容である。したがって，Y は確率変数 C と何らかの和であるわけだ。ところが，上の式では Y は「定数」になるとされている。C が確率変数であるにもかかわらず，C となにかの和であるところの Y が確率変数ではないということを信じることには，かなりの困難が伴う。

実際のところは，Y が確率変数であってもさほど問題がないというところに落ち着かせるのが教科書の一般的なパターンである。Y と撹乱項 u との間に相関がなければ，β の推定には漸近的に問題がないからである。

無論このことは正しい。しかし，これらのいわば問題解消的なアプローチ，つまり正統とされる統計処理手順によって確認されることが，元のデータが一意的なデータ発生プロセスに基づくものであるということを保障するものではないという，当然のことを改めて本稿で示すこ

とにする。

あらかじめ断っておくが，本稿の内容はいわゆる見せかけの回帰とはあまり関係がない。見せかけの回帰は，実際に関係のない変数の間にあたかも関数関係が存在するかのように見える現象であったのに対し，本稿の取り上げる内容は，異なるデータ発生メカニズムが，同一の統計的推論を導くということを確認することにあるからである。より単純にいえば，本稿で取り上げるデータの間には明白な関連がある。

叙述は次のように行う。まずもっともらしいデータの発生方法を提示し，その確率的特徴を一定程度示す。次に一般的な回帰分析で作られたデータによってどのような結果が得られるかを示し，最後に時系列分析で同様のことを行う。

本稿であまり明示することはないが，ひとつ問題意識とされていることは物価による名目値と実質値の間の関係である。とりわけ回帰分析で計算結果が何を意味しているのかという部分でいくつかの示唆をしておいた[1)]。

2. 偽の DGP

2.1 フェイク DGP の導入

本稿では先にもあげたように典型的ではあるが，しかしおそらくあまり使われない消費関数 $C=\alpha+\beta Y$ などを取り上げることにしよう。

ただし，その「DGP (Data Generating Process)」は現実をほとんど全く反映しないものとする。具体的には，次のようなプロセスによって作成する。

まず，実質暦年の SNA データから民間最終消費支出（C）の平均および標準偏差，そして $GDP(Y)$ の値から民間最終消費支出を引いたもの（So）[2)] の平均および標準偏差を計算する（$So=Y-C$）。そして，平均と標準偏差を C と So とにあわせた一様分布に従う乱数を発生させる。GDP のデータは C と So の合計から得られる。

コンピュータの一様乱数は通常 0 から 1 の間で発生させるから，この乱数を X として，次のように C および So を作成する。C と So の作成に使用する X は乱数の異なる実現値である。

---

1) 本稿では，大体において教科書的な手法を取り上げる。したがっておよそ枯れた手法が多いが，個々の手法については参照文献をあげる。計算には R を使用するが，とりわけ間瀬茂『R プログラミングマニュアル』(2007) 数理工学社を参考にした。

2) 本稿の内容にはなんら関係ないし，どのように表記してもよいだろうが，「So」は「そのほか」の意のほかにドイツ語の “Sonstige” からとった。

$C=\mathrm{Mean}(C)+(X-0.5)(\mathrm{SD}(C)\sqrt{12})$

$So=\mathrm{Mean}(So)+(X-0.5)(\mathrm{SD}(So)\sqrt{12})$

Mean および SD はそれぞれ平均および (母) 標準偏差を計算するものとするが，$1/\sqrt{12}$ は，0 から 1 の一様乱数の標準偏差である。

現実の SNA の実質暦年データからは 30 年間 (1980 年から 2009 年) で，C の平均を 253.7839 兆円，標準偏差を 45.0869 兆円，So の平均を 195.7067 兆円，標準偏差を 37.8154 兆円と計算することができる。

なぜ正規乱数ではなく一様乱数を選択したかといえば，そのほうが現実の Y (Y は一様にはならないが) や C のデータをまだより正しく反映していた時期があったからである。これが，実質値ではなく名目値を分析の対象とし，1990 年 (あたり) 以降に限るなら，一様分布というのはリアリティに欠けるだろう。しかし，1980 年代までのように一定程度成長する情況を含む場合，一様乱数は説得力を持つだろう[3)]。なお，時間的な構造の導入は後に説明する。

$Y=C+So$ であるが，Y の分布を考えることにしよう。まず平均は $\mathrm{Mean}(C)+\mathrm{Mean}(So)$ になる。また，分散は C，So が独立であることを考えればそれぞれの分散の和になるので，これから標準偏差を求めることができる。上記データから計算すると Y の標準偏差は 58.8458 兆円になる。この間の現実の GDP のデータの標準偏差は 82.5514 兆円だから，かなり標準偏差は小さいといえるだろう。

これはあくまで平均と分散に限られているので，厳密な分布を求めることにしよう。分布を考える場合，確定部分 $\mathrm{Mean}(C)$ および $\mathrm{Mean}(So)$ はさしたる重要性を持たない。そこで，

---

3)　図 2.11 で実際の実質 GDP と消費の散布図をあげた。その横軸ないし縦軸への射影が GDP と消費それぞれの分布をあらわすが，それらは中央にデータが集中し，中心から外れるにしたがい頻度は小さくなるが，ある程度離れたところでもデータがあるといったような正規分布の形状とは明白に異なり，「直観的」には一様分布のほうがまだ「まし」と思えるだろう。この直感の正当性を示す。本稿では Y (GDP) については一様分布とはしていないが (もっとも相関を強く入れれば一様分布に近くなる)，Y，C ともに見てみることにする。

与えられた期限での Y の歪度，尖度はそれぞれ −0.789，−0.700，これから計算される正規性検定の指標 (ボウマン・シェントン，ジャック・ベラ，正規性の帰無仮説のもとで自由度 2 の χ^2 分布) は 20.23，C についてはそれぞれ −0.975，−0.634，21.26 で双方とも正規性を棄却するには十分な値である。

しかしこれだけでは一様分布のもっともらしさを説明したことにならない。そこで一様分布を帰無仮説としコルモゴロフ・スミルノフ統計量 (2 つ) から求められる有意確率を計算すると，Y について 0.799，0.564，C について 0.535，0.956 であり，これらは通常使用される 10％，5％といった確率より十分大きく，一様分布にしたがうという仮説は棄却できない。

なお，ほかの部分でも一部述べたが，一様分布を支持する原因は中・低度の成長期には成長により，1995 年ぐらいからはデフレーターによるものと考えてよいと思っている。

$(X-0.5)(\mathrm{SD}(C)/\sqrt{12})$ および $(X-0.5)(\mathrm{SD}(So)/\sqrt{12})$ を考えることにするが，0.5 を引いているのは計算の便宜上であって，$0.5\mathrm{SD}(C)/\sqrt{12}$ などの部分も定数であるから，実際には確率変数 X に定数がかかっている部分のみ考慮すればよい。

今，一般に，X_a，X_b を 0 から 1 までの値をとる一様乱数とし，$0<b\leqq a$ として $Z=aX_a+bX_b$ を考えることにする[4)]。

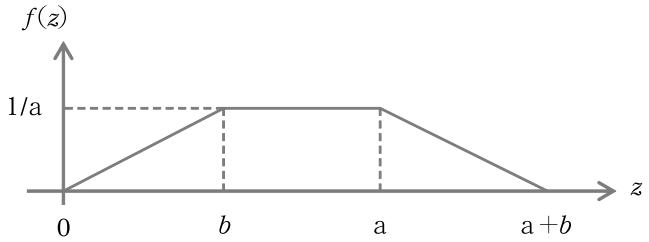
Z は，最小値 0，最大値 $a+b$ をとる確率変数である。

その確率密度関数は次のようになる。

$$\begin{aligned} f(z)&=(1/ab)z && (0\leqq z<b)\\ f(z)&=1/a && (b\leqq z\leqq a) \\ f(z)&=(a+b-z)/ab && (a<z\leqq a+b) \end{aligned} \qquad (2.1.1)$$

図で表すと次のようになる。

図 2.1



累積分布関数は $f(z)$ を積分すると得られるが，それよりも密度関数のグラフにあらわされるシンメントリーな台形の面積を考えたほうが効率的だろう。$F(z)$ を累積分布関数として次のように書くことができる。

$$\begin{aligned} F(z)&=z^2/(2ab) && (0\leqq z<b)\\ F(z)&=b/(2a)+(z-b)/a && (b\leqq z\leqq a) \\ F(z)&=1-(a+b-z)^2/(2ab) && (a<z\leqq a+b) \end{aligned} \qquad (2.1.2)$$

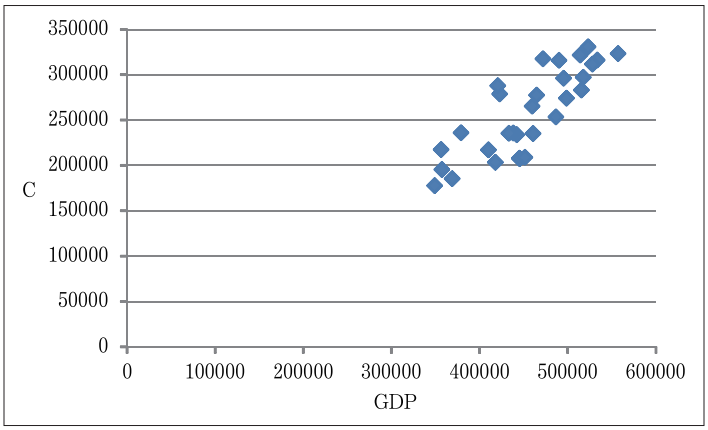
$a=1$ とするといわば標準的な場合を考えることができる。このとき，b の値は小さいほうの一様変数の「倍率」と見なすことができる。$b=0$ なら z は標準的な一様確率密度にしたが

4)　本稿の文脈では $a=\mathrm{SE}(C)$，$b=\mathrm{SE}(So)$ となる。

う確率変数と見なすことができるし (ただし，密度関数の 3 式のうち $b<z<a$ の部分しか意味を持たない)，$b=1$ なら，z は標準的な一様確率密度にしたがう確率変数の 2 つの和になる。このときの密度関数のグラフが二等辺三角形状になることはよく知られている。したがって，$a=1$ のときの b の値，あるいは a と b の比が，確率密度関数の形状を決定する基本的な要因になる。a の絶対値はいわば拡大ファクターである。

「30 年分」のデータを適当に発生させたものの一例をその散布図で示す。なお，本稿の計算は基本的に R (version 2.11.1) で行い，追試を可能にするため set.seed (0) で乱数の初期化を行ったものであることを記しておく。

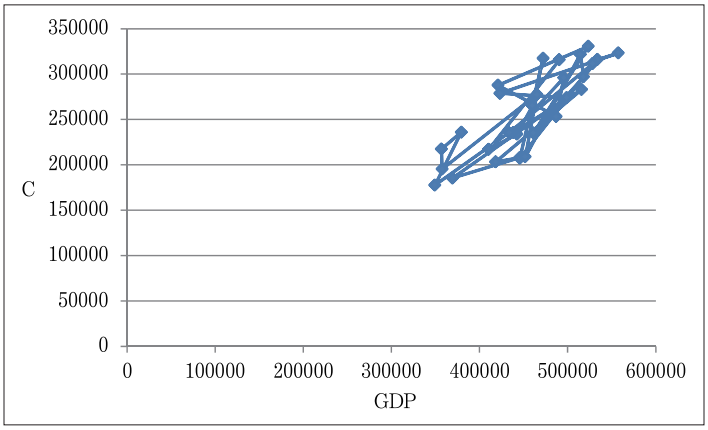
図 2.2



図を見れば，実際の実質 *GDP* および実質民間最終消費支出のデータに比べれば圧倒的にばらつきが大きいことが即座にわかるが，見る人がデータを読むことに慣れていなければ，違和感を持たない可能性も一概には否定できない。

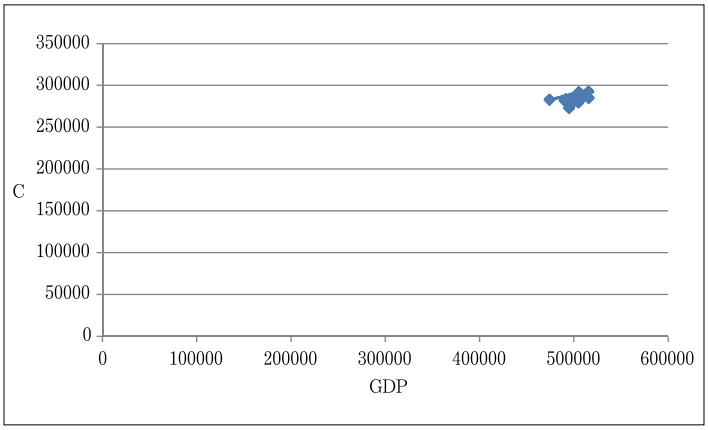
しかし，データを線分でつながないタイプの散布図で描いたので，ある程度違和感が軽減されているといえる。この散布図を，データを線分でつなぐタイプに変更してみよう。

図 2.3



これであれば，かなり違和感がもたれるに違いない。もっとも，近年の名目データの情況が続けば，名目データのものとして提示されれば，そのうちに違和感が軽減される可能性なしとしない。念のため，1995−2009 の名目の GDP と民間最終消費支出の散布図を掲載しておく[5)]。

図 2.4



2.2　順序統計量に変換する

以上に見たように，データそのものを俯瞰して見る限りは全くオリジナル (SNA データ) と

5)　改めて確認するまでもないだろうが，図中右上のしみのような部分が名目 *GDP-C* の 14 年分のデータである。そして，やや左にある端点は，1995 年ではなく 2009 年のものである。

かけ離れているとまではいえない部分もあるにせよ，時間構造を考えた場合，限りなく疑わしいデータでしかない。そこで，ここでは極めて簡単かつ乱暴に時間構造を導入する。その方法は，Y の昇順で C とともに並べ替えてしまうというものである。

このとき，並べ替えられた Y は順序統計量となるが，順序統計量の性質からもはや Y の各要素はそれぞれ独立ではない。Y は C と So から作ったが，Y は C とその他の和なのだから，Y が大きいまたは小さい場合，C も大きいまたは小さい可能性が高いだろう。したがって，C もまた独立にならない。

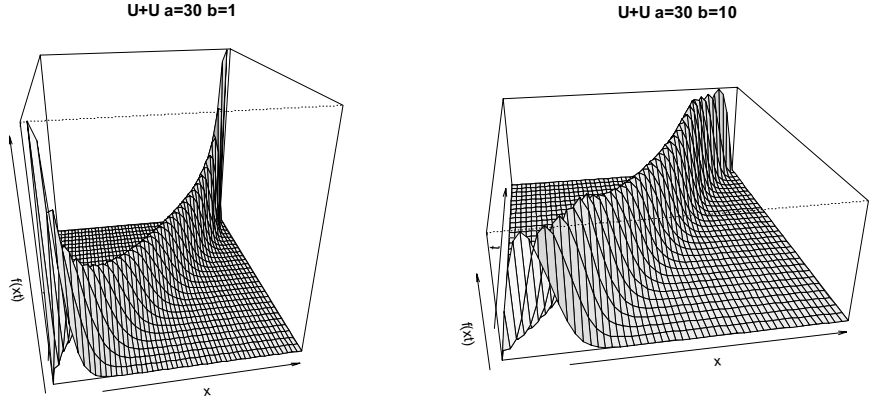
C の分布を検討することはかなりの困難が予想されるが，並べ替えに使用した Y についてはよく知られた式を使用することによって各 Y_t の分布を検討することが一定程度可能である (ここでの例では $t=1\ldots n$)。

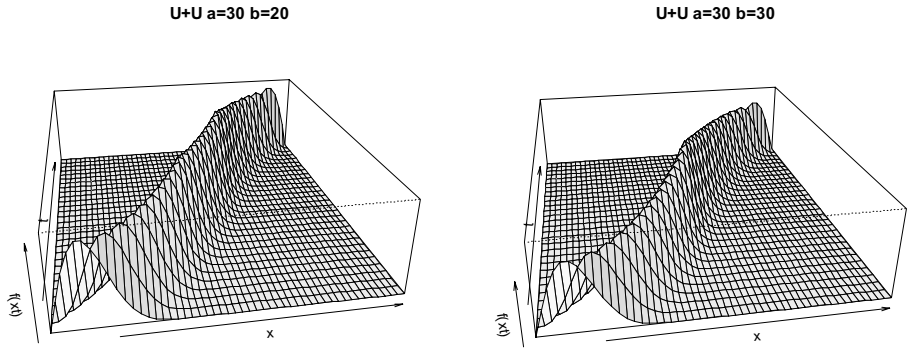
Y_t の周辺分布は (2.1.1) 式および (2.1.2) 式から次の $\tilde{f}$ に定数 $n!/((t-1)!(n-t)!)$ をかけて求めることができる (ここでは煩雑さを避けるため $x=Y_t$ とおく)。

$$\begin{aligned}
&\tilde{f}(z)=x/(ab)\cdot(x^2/(2ab))^{t-1}\cdot(1-x^2/(2ab))^{n-t} && (0\leqq x<b)\\
&\tilde{f}(z)=(1/a\cdot(2x-b)/(2a))^{t-1}\cdot(1-(2x-b)/(2a))^{n-t} && (b\leqq z\leqq a)\\
&\tilde{f}(z)=(a+b-x)/(ab)\cdot(1-(a+b-x)^2/(2ab))^{t-1}\cdot((a+b-x)^2/(2ab))^{n-t} && (a<z\leqq a+b)
\end{aligned} \tag{2.1.1}$$

本稿では $n=30$ であるから，b を変化させるとして $a=30$ とおいて b をそれぞれ 1, 10, 20, 30 と変化させて ($a=1$ と基準化するとそれぞれ $b=1/30$，$1/3$，$2/3$，1) $t=1\ldots30$ の順序統計量 Y_t の周辺分布を求めることにする。ただし，その密度関数は単なる多項式の積分となるので明示的に求めることは可能であるが，実際にはかなり煩雑である。したがって，ここでは密度関数を図示しておこう。

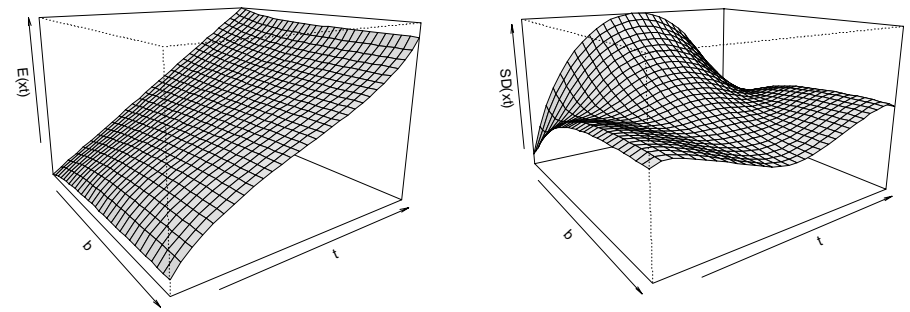
図 2.5





(2.2.1) 式から数値的に平均および標準偏差を求めることができる。これもまた図示しておく。

図 2.6



以上の検討からは，a と b の比率に応じてもとの密度関数が相当程度変わるので，順序統計量の分布も相当程度変化するといえるといえる。一般的に言えば，b の比率が小さければ大きい，または小さい順序統計量の分布はより大きい，または小さい値に集中し，b の比率が高まるにつれ，いわば平準化してくることになる[6)]。上の $b=30$ では，2 つの同じ一様分布の和になるから中心部分での実現値が多く，それはグラフ上では稜線の傾き (高さではなく，xt 平面上の) が中心付近で大きくなることで示される。また，端の部分では「圧縮」を受けるので分布がゆがむ。

先に述べたように，順序統計量は一般に独立でないから，その加減乗除も単純には考えられ

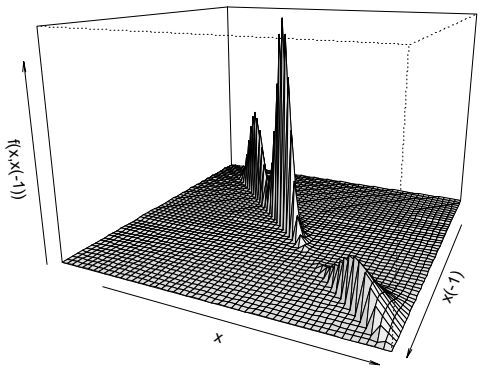
6) 計算では，b の比率が 60％あたりで最も順序統計量の分布が平準化する

ないことが予想される。経済系の統計分析では階差変数をとることがよくあるから，時間に関してひとつ異なる順序統計量の同時分布を検討しておくことが有用だろう。

同時密度関数自体の明示は場合分けが煩雑なので Appendix A にまわす。

本稿の偽 DGP では a と b の比率 45.0869 と 37.8154 の比 (先に見たように民間最終消費支出の標準偏差及びそのほかの標準偏差)，すなわちほぼ 6 : 5 あるいは 1 : 0.839 をまず考え，そのもとで Y_{30} および Y_{29} の同時密度関数，Y_{15} および Y_{14}，Y_5 および Y_4 の同時密度関数をひとつのグラフにあらわしてみよう[7]。ひとつのグラフにあらわすのは，a と b の比がこれぐらいであると，順序統計量の同時分布のそれぞれはほぼ十分に「分離」しており，理論上は不正確なグラフになるが実際上の影響はほぼなく，比較上のメリットがあるからである。

図 2.7



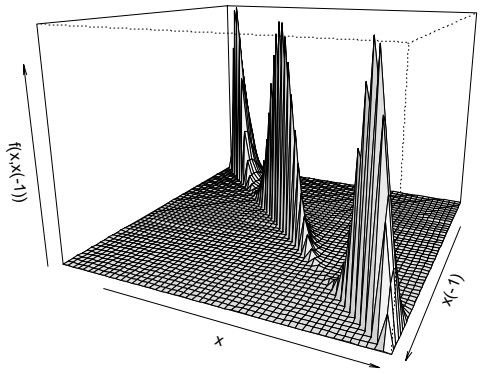
$a : b = 30 : 25$ という高い b の比率においては，かなり順序統計量の中央のもの ($n/2$ 前後) は集中することがわかる。一方で中央から離れるにしたがって 1 つ異なる順序統計量の間での同時密度関数も広い分布をとるようになる。したがって，Δy を考えたときにも，それらが異なる t に関して同一分布になるとは言えない。

一方，参考として $a : b = 30 : 5$ という低い b の比率のケースを図示しておく。ただし，こちらの図では先の図における Y_5 および Y_4 の同時密度関数のかわりに Y_2 および Y_1 の同時密度関数が描かれている。なぜなら，Y_5 および Y_4 では，Y_{15} および Y_{14} の密度関数の影響をいまだにある程度受けるからである。

---

7)　図の手前から Y_{30} と Y_{29}，Y_{15} と Y_{14}，Y_5 と Y_4 の同時密度関数である。

図 2.8



この低い b，つまり Y の分布がさほど一様分布と変わらない場合では，さほど 1 つ異なる順序統計量の間での同時密度関数が t によって変わらないと評価してよいであろう。

任意の Δy どうしでは負の相関がある。これは，ある Δy が大きければ，その他の Δy の和のとりうる範囲が狭くなることから理解できるであろう。また，相関係数は，およそ 1/n のオーダーで減少する。したがって，$n=30$ であればごく弱い負の相関があるということになるが，これは一般の経済データと著しくその性質を異にする点である。

y と Δy の相関について述べておこう。$2<i\leqq t\leqq n$ においては y_t と Δy_i の間には正の相関がある。$t<i\leqq n$ においては相関の値は負になる。y_t と Δy_t の間の相関係数は，t が大きくなるにつれて小さくなる。$b=0$，つまり Y が一様分布の場合と $b=1$ という 2 つの一様分布に従う変数の和という両極端の表を $n=5$ と 6 の場合について掲げておく。これらの表は y のとりうる値を 1/20 ごとに区切って各値を計算したもので，やや精度に欠けているが大まかな傾向をつかむには充分であろう。計算方法については Appendix B にまわす。

表 2.1

$b=0$，$n=5$

	Δy_2	Δy_3	Δy_4	Δy_5
Δy_2	1.000	−0.207	−0.198	−0.200
Δy_3	−0.207	1.000	−0.206	−0.198
Δy_4	−0.198	−0.206	1.000	−0.207
Δy_5	−0.200	−0.198	−0.207	1.000

	Δy_2	Δy_3	Δy_4	Δy_5
y_2	0.638	−0.320	−0.312	−0.315
y_3	0.446	0.449	−0.449	−0.446
y_4	0.315	0.312	0.320	−0.638
y_5	0.199	0.197	0.197	0.199

$b=1$，$n=5$

	Δy_2	Δy_3	Δy_4	Δy_5
Δy_2	1.000	−0.174	−0.091	−0.069
Δy_3	−0.174	1.000	−0.118	−0.091
Δy_4	−0.091	−0.118	1.000	−0.174
Δy_5	−0.069	−0.091	−0.174	1.000

	Δy_2	Δy_3	Δy_4	Δy_5
y_2	0.386	−0.437	−0.242	−0.182
y_3	0.263	0.347	−0.347	−0.263
y_4	0.182	0.242	0.437	−0.386
y_5	0.112	0.150	0.262	0.495

$b=0$，$n=6$

	Δy_2	Δy_3	Δy_4	Δy_5	Δy_6
Δy_2	1.000	−0.177	−0.165	−0.165	−0.167
Δy_3	−0.177	1.000	−0.175	−0.163	−0.165
Δy_4	−0.165	−0.175	1.000	−0.175	−0.165
Δy_5	−0.165	−0.163	−0.175	1.000	−0.177
Δy_6	−0.167	−0.165	−0.165	−0.177	1.000

	Δy_2	Δy_3	Δy_4	Δy_5	Δy_6
y_2	0.653	−0.263	−0.254	−0.254	−0.257
y_3	0.470	0.473	−0.357	−0.348	−0.352
y_4	0.352	0.348	0.357	−0.473	−0.470
y_5	0.257	0.254	0.254	0.263	−0.653
y_6	0.166	0.164	0.164	0.164	0.166

$b=1$，$n=6$

	Δy_2	Δy_3	Δy_4	Δy_5	Δy_6
Δy_2	1.000	−0.185	−0.097	−0.061	−0.053
Δy_3	−0.185	1.000	−0.112	−0.065	−0.061
Δy_4	−0.097	−0.112	1.000	−0.112	−0.097
Δy_5	−0.061	−0.065	−0.112	1.000	−0.185
Δy_6	−0.053	−0.061	−0.097	−0.185	1.000

	Δy_2	Δy_3	Δy_4	Δy_5	Δy_6
y_2	0.388	−0.450	−0.258	−0.162	−0.141
y_3	0.269	0.304	−0.362	−0.224	−0.198
y_4	0.198	0.224	0.362	−0.304	−0.269
y_5	0.141	0.162	0.258	0.450	−0.388
y_6	0.089	0.102	0.162	0.270	0.482

以下に Y_t どうしの相関を掲げておくが，一様確率変数からでも，一様確率変数からの和でもそれほど大きな差はない。ここではの一様確率変数からの和について $n=6$ の場合のものを掲げておく。全て正であることと，時間の間隔が開くほど相関係数が小さくなるさまを見ることができる。

表 2.2

	y_1	y_2	y_3	y_4	y_5	y_6
y_1	1.000	0.621	0.450	0.333	0.237	0.149
y_2	0.621	1.000	0.714	0.527	0.375	0.237
y_3	0.450	0.714	1.000	0.738	0.527	0.333
y_4	0.333	0.527	0.738	1.000	0.714	0.450
y_5	0.237	0.375	0.527	0.714	1.000	0.621
y_6	0.149	0.237	0.333	0.450	0.621	1.000

2.3　Y に順序を導入した場合の C の分布

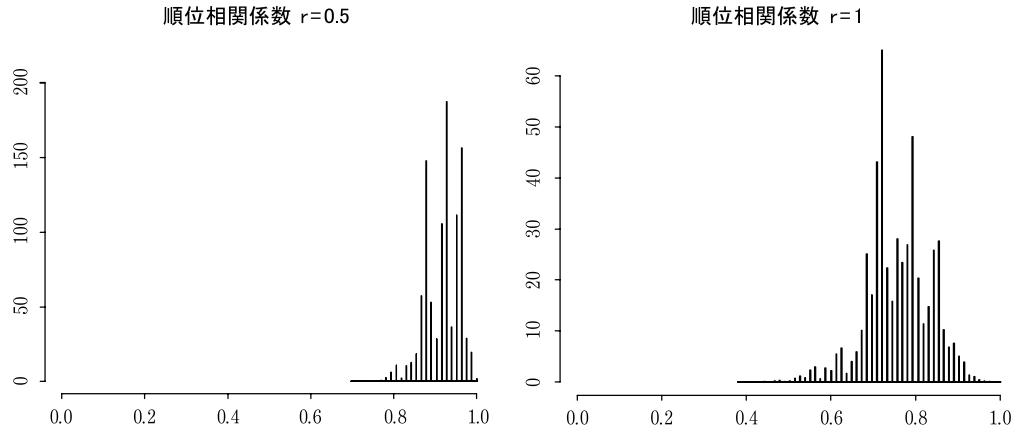
Y の大きさで導入した t によって作られる C_t は一般に単なる順序統計量ではない。

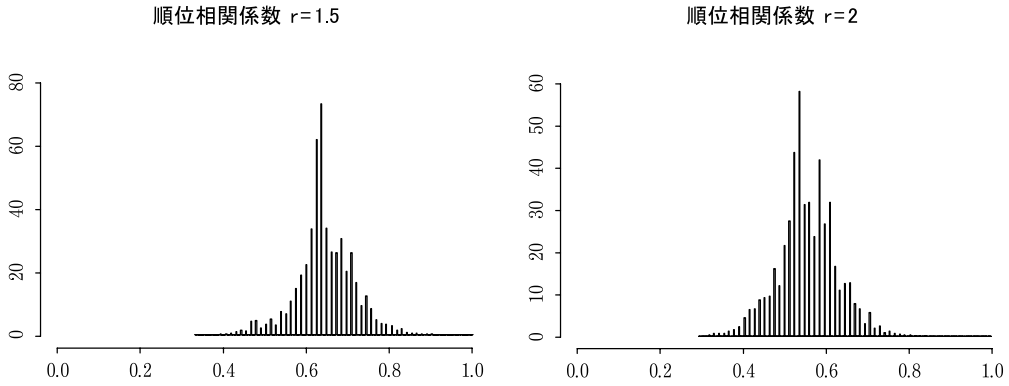
ただし，$b=0$ であれば，当然ながら $Y=C$ となり，Y も C も順序は一致し，かつ一様確率変数となる。しかし，$C=Y-So$ であることを考慮すると b の値が大きくなるにしたがって，C の順序が Y の順序から「撹乱」されてくることが予想される。そして，a と b があらかじめわかっていれば，荒っぽくではあるが C の順序がどのように Y の順序と異なるかはある程度の見当をつけることができる。

方法としては，a 側の一様乱数の (n 個の) 順序統計量の (順序 t の) 期待値が $at/(n+1)$ であり，b 側の一様乱数の同様の期待値が $bt/(n+1)$ になることを使う。まず，順序だてて並べられた $Xa_t(t=1\ldots n)$ を作成する。次に，$Xb_t(t=1\ldots n)$ をそれに加え，Xa_t+Xb_t の大きさを全体の中で比較する。ここで，Xb_t はどの Xa_t と加えられるかについては総当たりで計算する。考え方としては 2.2 に類似する。

どのようにもとの Xa_t の順序と Xa_t+Xb_t の順序が異なるかは，それ自体として検討してもよいが，たとえば Spearman の順位相関係数を用いることもできる。以下では定数 r を導入して，$aX\mathrm{a}+rbXb$ の順序と並べられた Xa との間の順序相関係数の例を示す。データの大きさは 10 とし，ここでの C と Y の例に従って $a=6$，$b=5$ とする。$r=1$ のとき，本稿での例にほぼ同じになるが，r を変えることで a と b との比を変更することができるので，異なる r の値について示しておく。

図 2.9





ここでの例に即していえば ($r=1$)，相関係数がやや低いことからある程度 C の順位は「乱されている」といえる。

2.4　C と So に相関を導入する

前項までは，C と So の間に相関を想定しなかった。

ある意味では，これが Y の低分散をもたらしていたといえる。現実のデータから計算される C と So の相関係数は 0.983 であり，無相関であることや独立であるといったことはまずいえない値である。

先に進む前に，0.983 といったような極端な相関係数が生じる理由を考察しておこう。

さまざまな考え方がありうるが，本稿で考えているようなデータについては 2 つの条件が満たされるだけで，特に特定のデータ発生プロセスを考えなくてもかなり高い相関係数が生じることを留意しておこう。一つ目の条件は，全体の中でその一部分の比率が安定していること。もうひとつは，一定程度，全体の大きさが (あるいは，一つ目の条件が満たされれば一部分でもよいが) 変動することである。

このことは，$Y=C+So$ の間に登場するどの変数（Y，C，So）の間でも当てはまる。まず，比率が安定していることはそれぞれのデータが正比例のライン近くに乗ってくることを潜在的に意味し，そして変数の一定程度の変動が実際にデータが直線の近くにばらつくということを保障する。つまり，高い相関係数をもたらすことになる。もし，第一の条件を満たしていたとしても，第二の条件を満たさない場合，計算上データの散らばりが小さいので相関係数は高い値にならない。

実際のところ，1980 年から 2009 年までの実質民間最終消費支出の割合は最大で 58.8％，最小で 54.7％，すなわち 30 個のデータで最大でも 4.1％しか変動していない。そして，この間

に 2 倍近くに額が膨らんでいる。そうしてみれば高い相関係数がもたらされることはある意味当然である。

その他の実質 GDP の需要側構成各項目について，対 GDP 比率の変化をまとめておく。ただし，在庫品にかかわるものは省略する。

表 2.3

	民間最終消費支出	民間住宅	民間企業設備	政府最終消費支出	公的固定資本形成	純輸出	輸出	輸入
最大比率	58.84％	6.58％	18.09％	18.73％	8.94％	4.87％	16.05％	11.18％
最少比率	54.67％	2.50％	11.80％	14.02％	3.35％	−0.15％	7.13％	5.57％
最大−最少	4.16％	4.08％	6.28％	4.71％	5.59％	5.02％	8.92％	5.60％
平均	56.55％	4.67％	14.51％	16.00％	6.65％	1.72％	9.88％	8.16％
標準偏差	0.99％	1.09％	1.56％	1.29％	1.57％	1.21％	2.56％	1.81％

輸出の変動が大きい。これは，この間傾向的に輸出の比率が大きくなっているためである。ただし純輸出レベルではやや変動が小さくなっている。それに次いで変動が大きいのは民間企業設備の 6.3％となっている。ただし，全体には劇的な変動があるとまではいえないだろう。

事実上ほとんど名目的には「成長」のなくなった 1995 年以降のデータについて実際のデータから計算される C と So の相関係数を計算してみよう。名目の場合には，比率の変化は最大で 4.5％，相関係数は−0.028，実質の場合には比率の変化は最大で 2.8％，しかし相関係数は全く異なり 0.872 である。この極端な相違は，データの変動のもたらしたものといってよいだろう[8)]。先の図に見たように，名目値はあまり変化していないにもかかわらず，実質値は最大と最小の間で 15％ほど変化しているからである[9)]。したがって仮に安定した関係が存在したとしても，名目値を使用する限り相関係数や t 値によるある種の統計的推論では関係が検出されないことには注意が必要である。

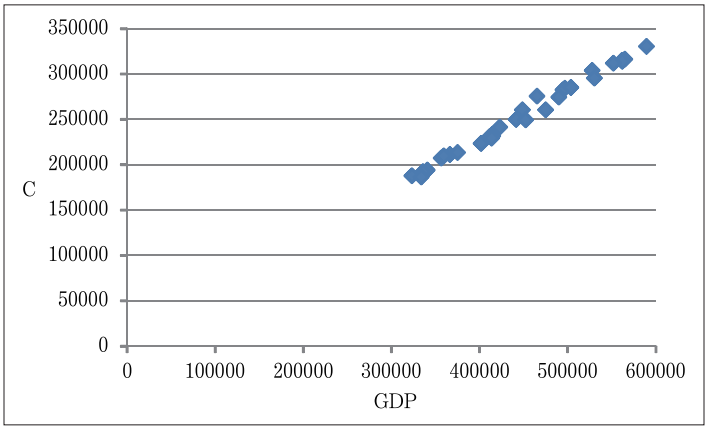
本稿での 2 つの 0 から 1 までの値をとる確率変数に一定の相関を導入するにあたっては次の方法をとった。まず相関係数が現実の値，すなわち 0.983 である 2 変量正規分布 (平均 0，分散 1) に従う乱数を発生させ，おのおのの確率変数の実現値と正規分布の累積密度関数の逆関

8) GDP デフレータによって実質 GDP が「インフレート」されているのではないかという疑念が生じるかもしれないが，名目 GDP と実質 GDP は別個に項目別に項目ごとの物価上昇率から計算されており，これらの値から GDP デフレータが総合されて計算されることになっている。
9) よって実質値を分析に使うということには，実際的な必要性があるともいえる。
10) 正規乱数発生には R の MASS パッケージの mvrnorm を使用した。

数を使用して 0 から 1 の一様分布に従う 2 変数を生成する[10]。高い相関で C や So を作成すると，その和は一様変数に近づくことも予想されるだろう。

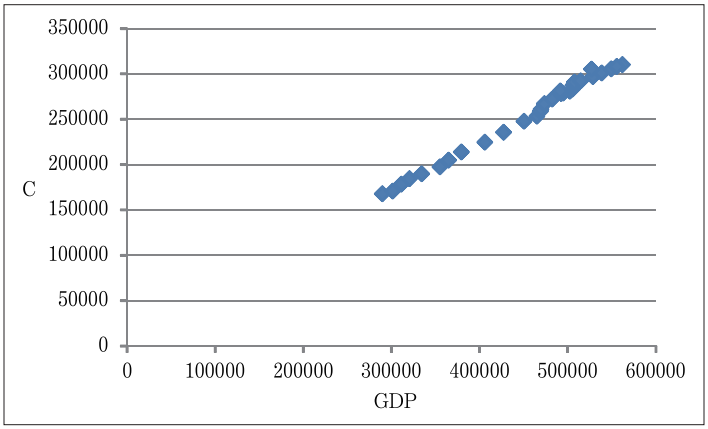
相関を導入した場合のデータの例をあげておく。かなりリアルなデータに見えるだろうが，無論これは α, β, $GDP(Y)$ を前もって決めておいて，しかる後に適当に u の生成条件を設定して C を作成したものではない。

図 2.10



参考として，1980−2009 の実際の実質 GDP-C の散布図を掲げておく[11]。

図 2.11



11)　ある程度慣れた人なら，左右どちらにデータが固まっているかを見ることで，どちらが本物か見抜くことができるだろう。

3. シンプルな回帰分析

3.1 Y と C

これからは，上の方法で作ったデータを使用して，いくつかの分析を行ってみよう。

最初に取り上げるのはシンプルな消費関数 $C=\alpha+\beta Y\ (+u)$ である。

まず，実際のデータ (1980−2009，実質暦年) でシンプルな回帰分析を行うと次のようになる。括弧内は t 値である。

$C=9158.0+0.544\,Y$

(2.31)　(62.65)

S.E. 3927.9　　$\overline{R}^2$ 0.993　　*D.W.* 0.83

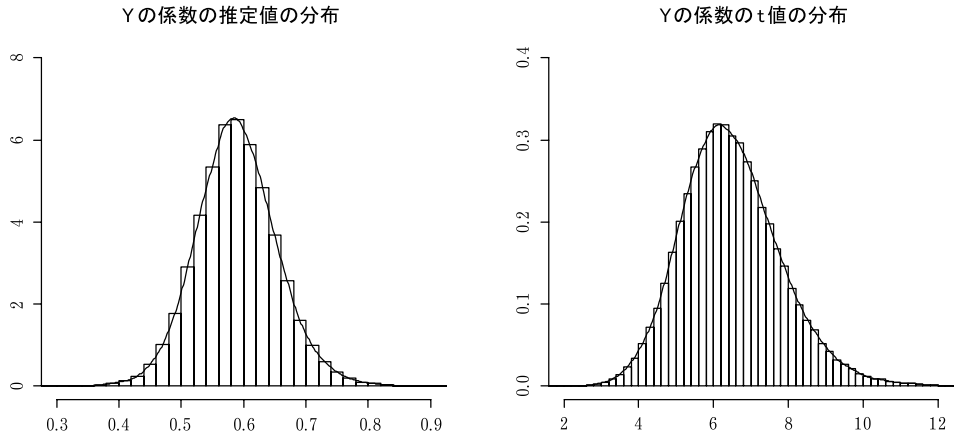
よく知られているように t 値や $\overline{R}^2$ の値は高いものの，DW 比は低い。

次に，上のもっともシンプルな方法 (時間構造なし，相関なし) で 100000 個の (30 個の) データを発生させ，統計量の分布を見てみよう。β の推定量の分布と t 値の分布，あるいはその他の統計量の結果は次のようになる。図において折線はカーネル推定値である。

表 3.1

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$ の t 値	$\hat{\beta}$ の t 値	$\overline{R}^2$	F	*D.W.*	*S.E.*
平　均	−10463.1	0.588	−0.24	6.48	0.572	43.71	2.00	29064.3
標準偏差	29363.9	0.064	0.69	1.31	0.099	18.03	0.35	3299.7

図 3.1



これを見る限り，作られたデータでも $C=\alpha+\beta Y$ が推定できるような気がするかもしれないし，t 値などからも，特に問題は感じられないであろう。

3.2　Y と I

同様に，そういう関数が実際に役に立つかどうかはともかく，$I=\alpha+\beta Y\ (+u)$ という「投資関数」を推定してみよう。I や Y の作り方は 2.1 に従う。I は民間企業設備とする。

前項同様に実際のデータでシンプルな回帰分析を行うと次のようになる。

$I=13655.5+0.177\,Y$

$(-2.31)\quad(13.22)$

$S.E.\ 6053.6\qquad \overline{R}^2\ 0.857\qquad D.W.\ 0.43$

C のときと同様の傾向といえるだろう。

再び前項と同じように 100000 個のデータを発生させ，統計量の分布を同様に見る。データを発生させる式は次のようになる。単位は 10 億円である。$(X-0.5)$ の係数を $\sqrt{12}$ で割ると，それぞれの標準偏差が得られる。データを発生させる式は次のようになる。

$I=65880.0+54505.6\,(X-0.5)$

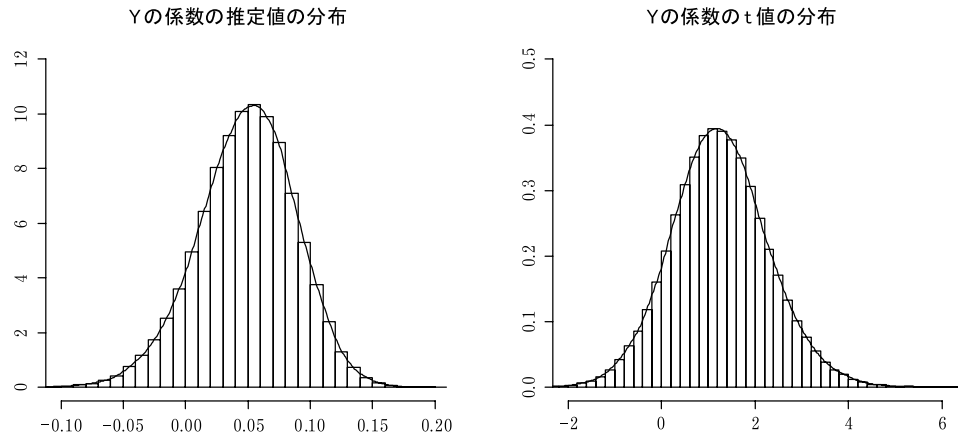
$So=383610+236236.2\,(X-0.5)$

前項と同様に処理しよう。

表 3.2

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$ の t 値	$\hat{\beta}$ の t 値	$\overline{R}^2$	F	$D.W.$	$S.E.$
平　均	43979.9	0.049	2.36	1.25	0.045	2.65	2.00	15298.9
標準偏差	17854.8	0.039	0.92	1.04	0.084	3.23	0.36	1447.2

図 3.2



明らかに感じられるであろうことは，この回帰式は一般に支持できないということである。

まず，t 値が低いことが目に見えて指摘されるだろう。t 値が 1.8 を超える割合は 28.7％，2 を超える比率は 22.5％になる。しかし，実際には，そのデータが拠ってきたところは別として，この回帰式は t 値の意味では本来支持されるべきなのである。

t 値が低いことは，次の理由から説明される。ひとつには，I と Y の (真の) 相関係数が低いこと，もうひとつはデータが少ないことである。前者は要するに So の標準偏差のほうが I のそれよりかなり大きい，すなわち 4 倍にもなるという，つまりシグナルーノイズ比の問題がある。Y と I の間の相関係数は，$a/\sqrt{(a^2+b^2)}$ と計算される (a と b は SD(I) および SD(So) に比例する)。また，よく知られているように単回帰の場合には t 値と標本相関係数の間には次の関係がある。

$$t=\sqrt{\frac{n-2}{1-r^2}}\,r \qquad (3.2.1)$$

I について a，b，n の値を実際に入れて計算すると真の相関係数は 0.225，真の相関係数を上式に代入して得られる t の値は 1.22 となる。同じことを C で行うと真の相関係数と，それを前提した t 値はそれぞれ 0.766 と 6.31 になる。これらは結果と符合する。I について，以上の計算で t の値を 2 にするためには $n=77$ 程度 (つまり，約 2.5 倍のデータ) が必要である。

本稿では最初に，基本的な分析を考え直すとした。t 値が低いことは必ずしも特定の回帰式に正統性がないことを意味しないし，一方で，標本の大きさを増やせば，被説明変数と関係がないでもないが，さほど重要性の高くない変数でもその t 値を 0 から十分に離すことができることには留意すべきである。このことは，とりわけ，回帰式を使ったなんらかの命題の「証明」

には，いくつかの場合大きな問題を根本的に抱えざるを得ないことをも意味している。

次にβの推定値を考えてみよう。

$I=\alpha+\beta Y$という式をたてているからには，Yが1単位増加したときにIがどれほど増えるかとか，αが無視できるならば，IとYの間にどのような比例関係，あるいは割合が想定されるかといったことを知りたいわけである。言い換えると，$\Delta I/\Delta Y$またはI/Yを知りたいということがあるわけだ。そして，それはIではなくCのときはある程度解明されたような結果になった。

一般に，被説明変数がy_t，説明変数の標本平均からの偏差がx^*_tであるとき，βの最小二乗推定量を$\Sigma(yx^*_t)/\Sigma({x^*_t}^2)$と書くことができる。現在の場合，$x^*$も$y$も確率変数であり，nが一定程度あればその中心をおよそ$a^2/(a^2+b^2)$と計算することができる。つまり，βの推定値はIとSoの標準偏差の比からおよそ決まってくる。ここでは$a^2/(a^2+b^2)$の値は0.051と計算される。CとSoで同じことを行うと0.587となる。これらは推定結果とほぼ符合する。しかし，前項の最小二乗推定が実際のデータからの結果と似ているように見えるのは偶然の一致だったのだろうか。

2.4において，GDPの需要面での各要素のGDPに対する割合の安定性について検討した。いわば「平均」について見た。ここでは，その標準偏差について比較してみてみる。

表 3.3

(平均および標準偏差の単位：10億円)

	GDP	民間最終消費支出	民間住宅	民間企業設備	政府最終消費支出	公的固定資本形成	純輸出	輸出	輸入
平均	449,490.5	253,783.9	20,280.8	65,880.0	72,492.5	29,058.8	8,027.2	46,032.6	38,005.4
対GDP列比率(A)	100.0％	56.5％	4.5％	14.7％	16.1％	6.5％	1.8％	10.2％	8.5％
標準偏差	82,551.4	45,086.9	3,404.7	15,734.4	17,369.0	6,308.4	6,778.4	19,250.0	14,312.0
対GDP列比率(B)	100.0％	54.6％	4.1％	19.1％	21.0％	7.6％	8.2％	23.3％	17.3％

とりわけ輸出入関連など個々の項目ではかなり異なる値になっているものもあるものの，個々の需要項目の平均と標準偏差のGDPのそれらに対する比率(AおよびB)を比較した場合，いくつかの項目においてはさほど差がない(なお，ここで「平均」としているのは結局のところ全年を通じたGDPに対する割合である)。仮に各項目がGDPにほぼ比例していると仮定すると，その標準偏差も同様になるからこのことはとりわけ奇妙なことではない。

そこで，GDP全体に占める各(ある)項目の割合をkとし，標準偏差もkに比例するもの

と (思い切って) 見なしてしまおう。そうすると，β の推定値の中心を $k^2/(k^2+(1-k)^2)$ と概算することができる。これに I の場合の実際の k の値 $(I/Y)=0.147$ から計算すると 0.029，C の場合の $(C/Y)=0.565$ から計算すると 0.627 となる。また，先に述べた β の中心と真の相関係数の関係から，真の相関係数は β の中心の平方根 (ただし，正負は場合による) であることがわかる。したがってさらに (3.2.1) 式から β の t 値を計算することができ，投資関数では 0.91，消費関数では 6.03 となる。これらはよく結果を説明している。したがって，現実のデータの構造から，以上の C，I の関数の推定および t 値はかなり説明され，3.1 および 3.2 で見た C と I の分析の違いは偶然とはいえない。

3.3 順序だてられた Y と C

本項では，順序だてられたデータについてある程度検討しよう。時間構造の入れ方は 2.2 に従う。

時間を入れた場合の $C=\alpha+\beta Y$ を推定してみることにする。α や β の推定値などは全く変わらないので，DW 統計量のみの変化だけがありうるが，実際のところさほど変化はない。時間を入れない場合，DW の平均が 2.07，標準偏差が 0.40 であったのに対し，時間を導入しても平均が 2.00，標準偏差が 0.35 とさほど変化はない。

次に『平成 21 年度年次経済財政報告』の付注 2 にあるような消費関数を推定しよう[12)]。そこでは，実質民間消費を，その 1 期前の変数と，実質可処分所得の今期の値に回帰させている。本稿では「実質可処分所得」の概念はないので，Y で代用する。実際のデータによる回帰式は次のようになる。なお，年次経済財政報告の式では定数項はない。

$$C=11636.1+0.557C_{(-1)}+0.230Y$$
$$(5.77)\quad(10.17)\qquad(7.30)$$
$$S.E.\ 1798.0\qquad \overline{R}^2\ 0.998\qquad D.W.\ 1.79$$

再び作られたデータで検討してみよう。

12) 2010 年 11 月現在では内閣府のサイトからダウンロード可能である。
http://www5.cao.go.jp/j-j/wp/wp-je09/09p00000.html

表 3.4

	$\hat{\alpha}$	$\hat{\beta}_{C(-1)}$	$\hat{\beta}_Y$	$\hat{\alpha}$ のｔ値	$\hat{\beta}_{C(-1)}$ のｔ値	$\hat{\beta}_Y$ のｔ値	$\overline{R}^2$	F	$D.W.$	$S.E.$
平　均	−13506.9	−0.041	0.617	−0.27	−0.23	4.09	0.54	19.68	2.00	29380.7
標準偏差	36044.3	0.201	0.137	0.75	1.11	0.96	0.11	8.69	0.11	3427.8

$C_{(-1)}$ についてはひどく魅力のない数字になっている。また，$C_{(-1)}$ と Y の双方のｔ値が 1.8 を超えている場合は全体の 3.1％しかない (境界値を 1.8 ではなく 2 にすると 1.9％)。それに，実際のデータでは $C_{(-1)}$ の説明力のほうが大きくなっているのに対し，作られたデータでは Y の説明力が高くなっている。

この結果から見れば，$C_{(-1)}$ は C に対して影響力を持っていないように見えるが，2.3 で，C_t の分布は C の順序統計量に一定の変更を与えたものであることを見た。したがって，C と $C_{(-1)}$ の間に相関関係のようなものが検出されないということはないであろう。実際，$C=\alpha+\beta C_{(-1)}$ という式を作られたデータから推定すると以下のようになる。

表 3.5

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$ のｔ値	$\hat{\beta}$ のｔ値	$\overline{R}^2$	F	$D.W.$	$S.E.$
平　均	121876.4	0.533	2.93	3.49	0.279	13.95	2.32	37042.4
標準偏差	37089.0	0.146	0.60	1.34	0.152	10.59	0.20	4281.4

要するに，このデータの作り方では $C_{(-1)}$ より Y のほうが説明力が大きいため，先の $C=\alpha+\beta_1 C_{(-1)}+\beta_2 Y$ という式では $C_{(-1)}$ が「埋没」してしまったのである。なお，$C=\alpha+\beta C_{(-1)}$（$+u$）という式は定数項のある AR (1) 式であるとも見なすことができる。定数項はほぼ正で有意であるといってよい。

3.4　C と So に相関を導入した場合

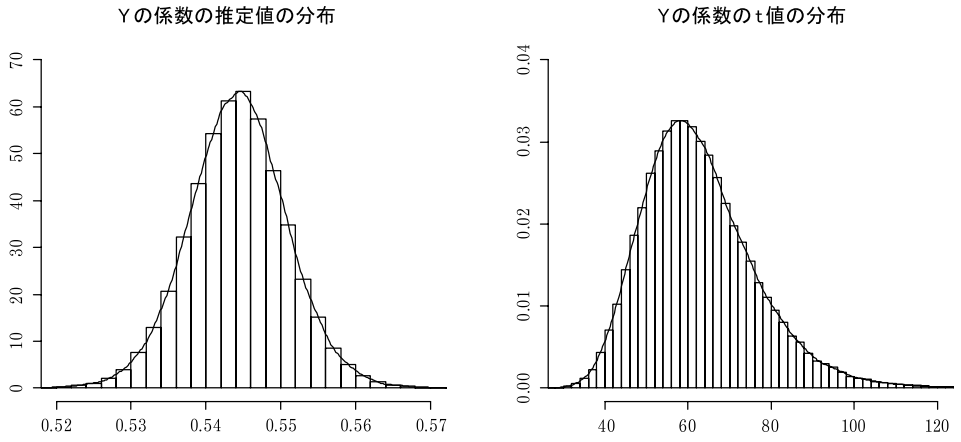
2.4 で説明した方法によって相関関係を導入しよう。もちろん，このとき，C や I，So，Y は一様変数ではないが，一定の幅に収まる変数にはなっている。

最初に，$C=\alpha+\beta Y$ を検討する。実際のデータからは C と So の相関係数は 0.983 と計算された。同じように 100000 回，回帰を行ってみよう。

表 3.6

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$ のｔ値	$\hat{\beta}$ のｔ値	$\overline{R}^2$	F	$D.W.$	$S.E.$
平　均	9133.0	0.544	2.31	62.75	0.992	4119.04	2.00	3962.8
標準偏差	3053.8	0.007	0.90	13.48	0.003	1870.05	0.35	639.3

図 3.3



DW 比を除いてほとんど 3.1 に記述した現実のデータからの結果と見まがうばかりである。

次に，$I=\alpha+\beta Y$ を検討しよう。実際のデータからは I と So の相関係数は 0.893 と計算される。集約表のみ掲載する。

表 3.7

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$ のｔ値	$\hat{\beta}$ のｔ値	$\overline{R}^2$	F	$D.W.$	$S.E.$
平　均	−13184.8	0.176	−2.20	13.23	0.842	184.11	2.00	6063.1
標準偏差	4857.1	0.011	0.99	3.00	0.059	87.83	0.35	920.2

これもまた，3.2 の現実のデータからの結果とほぼ変わらない。3.2 で見た，作られたデータからの「投資関数」が受け入れられないように感じられたのに対し，こちらは正反対の結果である。

やや一般的に論じよう。3.2 における記号 a，b と C や I と So の間の相関係数 ρ を使う。C (や I) と Y の相関係数は，$(a+\rho b)/\sqrt{(a^2+2\rho ab+b^2)}$ で計算できる。標本相関係数をｔ値に

変換する式 (3.2.1) は既述である。また，β の中心は $(a^2+\rho ab)/(a^2+2\rho ab+b^2)$ で計算できる。これに a，b，ρ の値を代入すると，消費関数においては t 値はおよそ 62.6，β の推定値はおよそ 0.544，投資関数においては t 値は 10.5，β の推定値はおよそ 0.177 を中心に分布することになる。

結局のところ，通常のデータで $C=\alpha+\beta Y$ などを分析した場合，β の推定値が Y における C の割合などに近ければもっともらしいのである。ここでの設定では Mean(C)/(Mean(C)＋Mean(So)) に近ければもっともらしくなる。ところが，3.2 に見たように，Mean(C) と Mean(So) などの比は，GDP 構成項目の各比率が比較的安定しているもとでは SD(C) と SD(So) の比とさほど変わらない。そして，先の $(a^2+\rho ab)/(a^2+2\rho ab+b^2)$ は ρ が大きくなるほど，$a/(a+b)$，したがっておよそ Mean(C)/(Mean(C)＋Mean(So)) に近づくといえる。

さらに，3.2 で見たような，GDP に占める各需要項目の比率 k を導入して一段簡単化しよう。k と標本相関係数の 2 つから，消費関数における β の推定値の中心はおよそ 0.557，その t 値は 64.38，投資関数においては 0.146 と 12.68 と計算することができる。この結果はこの節の結果はもちろん実際のデータと比較してもかなり異なるというほどのものではないであろう。

言い換えるなら，このタイプの実際のデータにおける回帰分析の β に関する結果はほぼ特定の需要項目の比率とそれ以外の部分との相関係数のわずか 2 つで説明される。

次に Y の大きさで並べた $C=\alpha+\beta Y$ を考えよう。常に乱数の種を一定にしているから，DW だけが先の DW とは変わる。しかし，ここでも 3.3 で見たように大きな変化はない。DW の平均は 2.10，標準偏差は 0.39 となる。

最後に C を一期前の C と当期の Y で説明するモデルを調べてみよう。

表 3.8

	$\hat{\alpha}$	$\hat{\beta}_{C(-1)}$	$\hat{\beta}_Y$	$\hat{\alpha}$ の t 値	$\hat{\beta}_{C(-1)}$ の t 値	$\hat{\beta}_Y$ の t 値	$\overline{R}^2$	F	$D.W.$	$S.E.$
平　均	9258.4	−0.023	0.557	2.11	−0.19	8.49	0.99	1826.45	2.06	4014.3
標準偏差	3704.8	0.130	0.071	0.92	1.08	1.94	0.00	864.83	0.31	660.5

3.3 で見たように，これだけ見れば $C_{(-1)}$ はやはり魅力の薄い変数となっている。

4. 時系列分析

4.1 ADF 検定

本節では時系列的な側面について分析してみよう。偽のデータの作られ方は一般的な時系列分析で取り上げられるものとかなり異なっていたが，結果はどうなるであろうか。

まず，Y (GDP) については並べ替えを行ったので，常に増加する数列となる。そうであればトレンドまわりの定常時系列または非定常時系列と判断される可能性がある。まずこの事項を考える。

最初に，現実のデータに ADF 検定を行った結果を記す[13]。上が検定統計量の値，下が判定結果である。表中「＊＊＊」とあるのは 1−10％では「I (1)」の帰無仮説が棄却できないことを示す[14]。none，μ，trend の意味については脚注を参照されたい。

表 4.1

	Y			C		
	none	μ	μ,trend	none	μ	μ,trend
Y, C	−0.10788	−2.12045	−0.63329	0.685287	−2.99262	−0.71289
$\ln(Y)$, $\ln(C)$	0.26039	−2.33769	−0.64189	0.932341	−3.72824	−1.238

	Y			C		
	none	μ	μ,trend	none	μ	μ,trend
Y, C	＊＊＊	＊＊＊	＊＊＊	＊＊＊	5％で棄却	＊＊＊
$\ln(Y)$, $\ln(C)$	＊＊＊	＊＊＊	＊＊＊	＊＊＊	1％で棄却	＊＊＊

実際には「none」ということはありえないだろうから，「μ」もしくは「μ,trend」が問題になるだろう。C の「μ」モデルについてのみ，一定の有意水準で I (1) 仮説を棄却する。

次に作られたデータについてみてみるが，元のデータと対数をとったデータでは検定結果にはさほど差がないので，元のデータについてのみ結果を記すことにする。まず，順序は導入するが相関は導入しないデータについてみる。上から検定統計量の値，検定統計量の分布，判定

---

13) R のパッケージ urca の ur.df では AIC によるラグ設定も可能だが，計算上ラグ 1 と変わりがない。ADF 検定については山本拓『経済の時系列分析』(1988) 創文社 pp. 239−250，森棟公夫『計量経済学』(1999) 東洋経済新報社 pp. 308−315 など参照。

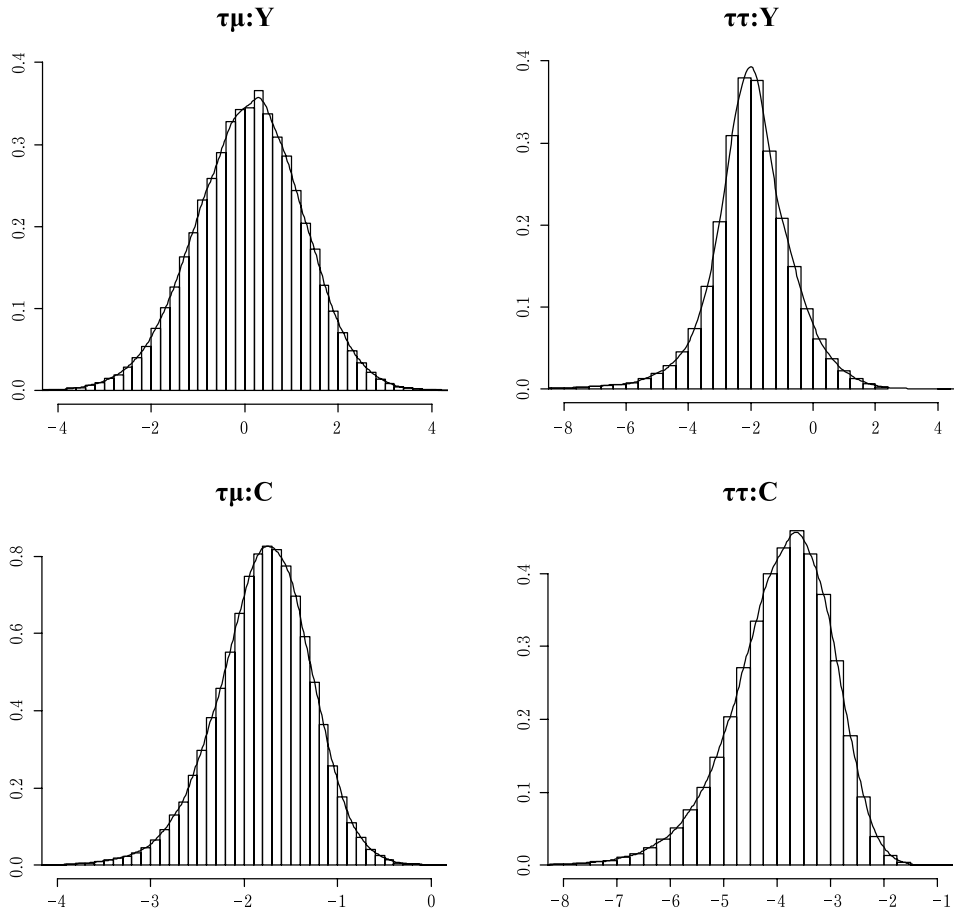
14) 左辺を ΔX_t とし，右辺を $\mu+\alpha_t+\delta X_{t-1}+\Delta X_{t-1}\ldots$ とした ADF 検定において「none」は $\mu=\alpha=0$ を，「μ」は $\alpha=0$ をそれぞれ仮定し，「μ,trend」はこれらの仮定をおかない。

結果である。検定統計量の分布については「none」は省略する[15)]。以下では 30 個のデータの発生回数は常に 100000 回である。

表 4.2

	Y			C		
	none	μ	μ, trend	none	μ	μ,trend
平　均	3.531	0.087	−1.988	0.625	−1.796	−3.922
標準偏差	0.830	1.146	1.334	0.249	0.506	0.932

図 4.1



15)　$\tau\mu$，$\tau\tau$ はそれぞれ「μ」「μ,trend」の検定統計量の分布を示す。境界値は none の場合で 1%，5%，10%がそれぞれ −2.62，−1.95，−1.61，μ の場合で −3.58，−2.93，−2.60，μ，トレンドの場合で −4.15，−3.50，−3.18 である。

表 4.3

	仮　説	none			μ			μ, trend		
Y	有意水準	1%	5%	10%	1%	5%	10%	1%	5%	10%
	非棄却率	100.0%	100.0%	100.0%	99.8%	99.3%	98.7%	95.3%	90.5%	86.2%

	仮　説	none			μ			μ, trend		
C	有意水準	1%	5%	10%	1%	5%	10%	1%	5%	10%
	非棄却率	100.0%	100.0%	100.0%	99.8%	97.9%	93.9%	63.6%	35.2%	21.7%

検定統計量の値そのものは実際のデータからのものとかなり異なる。Y において μ と μ, trend で数値が対照的である。「none」は意味がないと思うが，やや解釈に難があるにせよ「μ」「μ, trend」でもかなりの率で I (1) の帰無仮説を棄却することができない。もちろん棄却できないということは I (1) であるということを直ちに意味するものではない。純粋な順序統計量ではない C についてはやや率が低くなっている。また，μ, trend では Y 以上に棄却できなくなっている。

次に順序を入れた上で，さらに相関を導入しよう。分布の図は省略する。

表 4.4

	Y			C		
	none	μ	μ, trend	none	μ	μ, trend
平　均	3.659	−0.282	−2.156	3.753	−0.344	−2.401
標準偏差	0.791	1.099	0.897	0.754	0.906	0.862

	仮　説	none			μ			μ, trend		
Y	有意水準	1%	5%	10%	1%	5%	10%	1%	5%	10%
	非棄却率	100.0%	100.0%	100.0%	99.7%	98.9%	98.0%	98.1%	94.3%	89.8%

	仮　説	none			μ			μ, trend		
C	有意水準	1%	5%	10%	1%	5%	10%	1%	5%	10%
	非棄却率	100.0%	100.0%	100.0%	99.9%	99.6%	99.1%	97.0%	90.3%	83.6%

Y の μ については多少非棄却率が下がるものがあるが，大体において非棄却率は増加しており，とりわけ C については大幅に増加している。ほぼ Y と C がパラレルになるためだと思われるが，Y と C から計算される値はよく似た傾向を示すことになる。実際のデータと μ と

μ, trend で数値が対照的であることは相関を入れない場合と同じである。

4.2　KPSS 検定

ADF 検定では帰無仮説が I (1) だから，単位根の存在は「棄却できない」という弱い結論である。帰無仮説が定常である KPSS 検定を行ってみよう[16]。

KPSS のトレンド定常を帰無仮説とする検定が適切だと思われるが，レベル定常の検定も行っておく。KPSS 検定は DF，ADF 検定よりは対立仮説と帰無仮説の対比という点で解釈しやすい利点がある。

現実のデータでトレンド定常仮説は Y，C ともに 0.01％より小さい水準で棄却される。これは R で短いトランケーション・ラグ (R の計算上実際には 1) で計算した場合である[17]。長いラグ (実際には 3) の場合には Y，C それぞれ 1.5％以下，1.1％以下で棄却する。あまり意味はないだろうがレベル定常仮説は全て 1％以下で棄却する。

作られたデータで検定してみよう。相関がないものでは検定結果は次のようになる。

表 4.5

	Y				C			
	level		trend		level		trend	
	short	long	short	long	short	long	short	long
1％以下で棄却	100.0％	100.0％	25.1％	0.2％	98.4％	57.8％	0.8％	0.0％
5％以下で棄却	0.0％	0.0％	45.0％	24.3％	1.5％	41.9％	6.9％	4.5％
10％以下で棄却	0.0％	0.0％	14.6％	25.9％	0.0％	0.3％	7.9％	10.3％
棄却しない	0.0％	0.0％	15.3％	49.7％	0.0％	0.0％	84.3％	85.2％

現実的には trend のほうのみを考えればよいだろうが，傾向は ADF 検定と同じではあるものの，KPSS 検定のほうがシビアな結果となっている。Y についてはラグの長さによって異なる傾向が見られる一方，C の trend ではほとんど同じである。

相関を入れ，同様の表を作成してみよう。

---

16)　KPSS 検定については蓑谷千凰彦『計量経済学大全』(2007) 東洋経済新報社 pp. 612−620 参照。

17)　R のパッケージ tseries の kpss.test で算出。level の場合は帰無仮説はレベル定常，trend の場合はトレンド定常になる。short，long の違いは検定統計量において分散を計算する基準の切断ラグの値の違いを意味しており，R の使用したパッケージでは short の場合約 $3\sqrt{n/13}$，long の場合約 $10\sqrt{n/14}$ としている。

表 4.6

	Y				C			
	level		trend		level		trend	
	short	long	short	long	short	long	short	long
1％以下で棄却	100.0％	100.0％	40.8％	0.7％	100.0％	100.0％	32.3％	0.4％
5％以下で棄却	0.0％	0.0％	29.2％	38.8％	0.0％	0.0％	28.3％	35.0％
10％以下で棄却	0.0％	0.0％	11.3％	18.3％	0.0％	0.0％	12.8％	18.1％
棄却しない	0.0％	0.0％	18.7％	42.2％	0.0％	0.0％	26.6％	46.5％

相関を入れた場合，Y と C はかなり似てくる一方で，やや C は「上昇強度」のようなものが弱いという結果が顕著に出てくる。ADF 検定と同じような傾向とはいえるが，全体にシビアに判定されるし，とりわけラグ次数を高くした場合，かなり棄却できないという結果になる。

4.3　共和分検定

4.1 では相関を入れない場合でも仮説，あるいは有意水準によってはかなり I (1) の仮説を棄却できず，相関を入れた場合にはあらゆる場合においてかなりの確率で棄却できなかったことを見た。

ここでは共和分検定を行うことにしよう。共和分検定の方法はいくつかあるが，本稿ではシンプルに C に Y を回帰した残差が I (0) といえるかどうか，ADF 検定の前項の「none」を使って検定する。Y，C それぞれの過程が和分過程かどうかの判定は 4.1 に従う。

現実のデータでは検定統計量の値は −2.309 になるが，これは「微妙」な値で，有意水準 5％なら棄却する，つまり I (0) といえるが 1％では棄却しない。その結果として 1％水準では共和分なし，となるが，よりゆるい水準では，none，μ,trend で共和分あり，μ で共和分なしという結果になる。

作られたデータについてみてみよう。相関がない場合，残差についての検定統計量の値は平均 −4.137，標準偏差 0.969 となる。境界値は先の中で示したが，検定統計量の値はかなり低いから高率で I (1) 仮説を棄却する。有意水準 1％で 94.36％，5％で 99.66％，10％で 99.96％となる。したがって，共和分関係にあるというのはほぼ，Y および C がともに I (1) 仮説を棄却できないときに等しい。以下の表に，Y，C ともに I (1) 仮説を棄却できず，かつ C に Y を回帰した残差が I (0) といえる割合 (判定率) を示す。

表 4.7

C と Y の共和分	仮　説	none			μ			μ, trend		
	有意水準	1％	5％	10％	1％	5％	10％	1％	5％	10％
	判定率	94.4％	99.7％	100.0％	94.0％	97.0％	92.8％	55.2％	31.7％	19.0％

有意水準が低くなるほど存在率が高くなる場合があるのは，C や Y の I (1) 仮説のほうが「棄却できない」という形式であるからである。μ, trend のケースでは共和分と判定される率がある程度低くなっているが，それでも一定程度の存在率はあるといえるだろう。

相関を導入してみよう。残差についての検定統計量の平均は −4.201，標準偏差は 0.953 となり，さらに高率で I (1) 仮説を棄却する。有意水準 1％で 97.63％，5％で 99.94％，10％で 100.00％ (99.995％) となる。同様の表を示す。

表 4.8

C と Y の共和分	仮　説	none			μ			μ, trend		
	有意水準	1％	5％	10％	1％	5％	10％	1％	5％	10％
	判定率	97.6％	99.9％	100.0％	97.3％	98.8％	97.8％	93.7％	88.2％	80.7％

4.1 の内容と，残差についての I (1) 仮説の棄却率の高さから予想されることではあるが，かなりの割合で「共和分が存在する」と判定されることになる。

4.4　ECM

4.1 と 4.3 から，相関を入れるかどうかで差があるにせよ，C と Y が共和分していると判定されることが一定程度はあることが明らかになった。そこで，本項では単一方程式の ECM を推定することにしよう[18)]。

ECM の定式化にはバリエーションがあるが，以下のシンプルなモデルを考えることにする。

$$\Delta C_t = \beta(C_{t-1} - \alpha_0 - \beta_0 Y_{t-1}) + \gamma \Delta Y_t + u \qquad (4.3.1)$$

現実のデータを使用すると以下の結果が得られる。

18)　ECM については蓑谷，前掲著 pp. 654−656 など参照。

$$\Delta C_t = -0.384(C_{t-1} - 9158.0\, Y_{t-1}) + 0.391 \Delta Y_t$$
$$(5.77) \qquad\qquad (7.30)$$
$$S.E.\ 2556.0 \qquad \overline{R}^2\ 0.817 \qquad D.W.\ 1.29$$

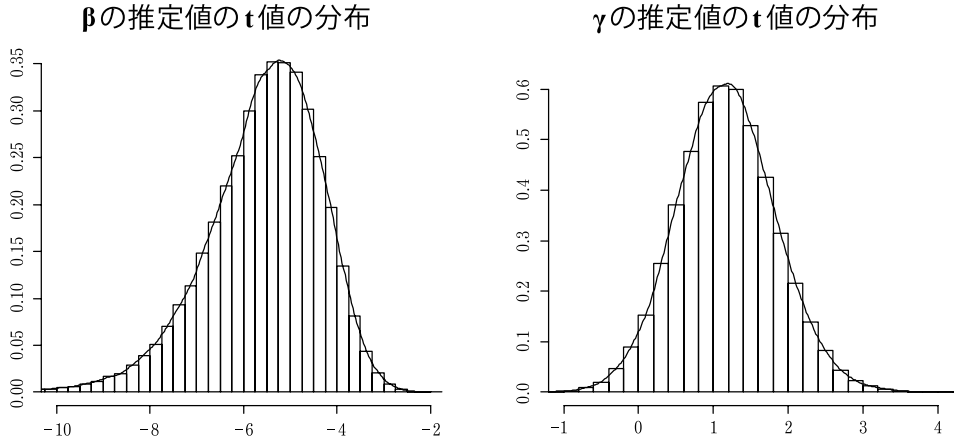
ECM であっても，少なくともこのモデル設定ではまだ DW は低い。ただし，エラーコレクション項や ΔY_t に期待される符号条件は満たされている。

次に，まず相関を入れない作られたデータで分析しよう。すでに共和分関係があることは「発見」されているので，α_0，β_0 は 3 節で使用した単純な OLS で推定されたものを使用する(したがってそれら数値は 3.1 で報告されている)。

表 4.9

	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\beta}$ の t 値	$\hat{\gamma}$ の t 値	$\overline{R}^2$	F	$D.W.$	$S.E.$
平　均	−1.044	0.564	−5.57	1.18	0.505	17.28	1.98	28645.0
標準偏差	0.204	0.309	1.21	0.66	0.105	7.55	0.10	3362.9

図 4.2



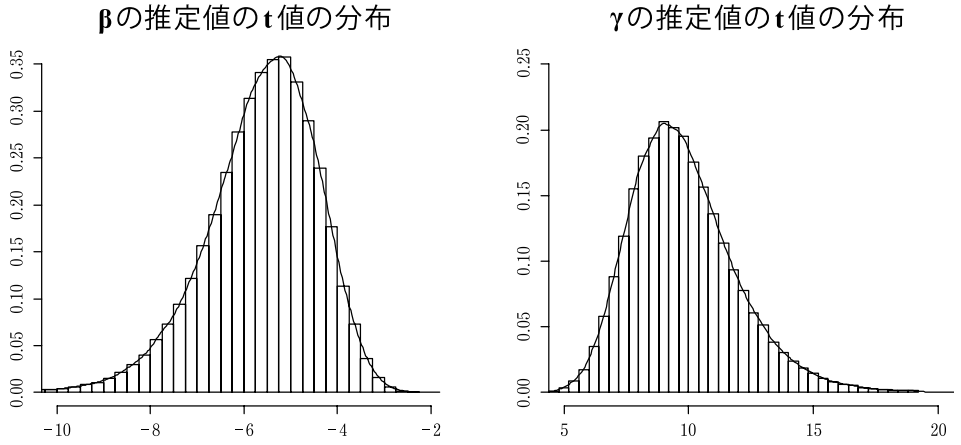
γ の t 値が低いということを除けば，符号条件や β の t 値については問題ない数値になっている。ただし，実際のデータと推定値はかなり異なる。

次に相関を導入し，同様の分析を行おう。

表 4.10

	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\beta}$ のｔ値	$\hat{\gamma}$ のｔ値	$\overline{R}^2$	F	$D.W.$	$S.E.$
平　均	−1.055	0.544	−5.63	9.82	0.806	68.30	1.98	3901.1
標準偏差	0.199	0.041	1.19	2.15	0.056	27.02	0.10	643.8

図 4.3



元のデータに ECM，またはそれを導き出すような何らかの長期均衡に戻るメカニズムがあることを想定してもおかしくない分析結果である。

ΔC と ΔY にはどちらかというと直接的な関係があるといってもよいだろうが，ΔC とエラーコレクション項に，相関があってもなくても安定的な関係が見られるのはやや直接的ではない。

しかし，次のように考えれば「謎」を解くことができる。相関があってもなくても似たような係数の推定値になるのはヒントになる。

β の値はほぼ −1 である。また，γ の値と β_0 の値がよく似ていることに注目しよう。そこで ECM (4.3.1) の右辺において (思い切って) $\beta=-1$，$\gamma=\beta_0$ とおいてしまおう。撹乱項を除いて次のように計算できる。

$$
\begin{aligned}
&\beta(C_{t-1}-\alpha_0-\beta_0 Y_{t-1})+\gamma\Delta Y_t \\
&=C_{t-1+\alpha_0}+\beta_0 Y_{t-1+\beta_0}+\beta_0(Y_t-Y_{t-1}) \\
&=\alpha_0+\beta_0 Y_t-C_{t-1}
\end{aligned}
$$

これはほぼ ΔC_t に等しいだろう。さらにこの事情を検討してみよう。

$\beta=-1+x$，$\gamma=\beta_0+y$ とおき，$\Delta C_t-\beta(C_{t-1}-\alpha_0-\beta_0 Y_{t-1})-\gamma\Delta Y_t$ を計算する。

$$
\begin{aligned}
&\Delta C_t-\beta(C_{t-1}-\alpha_0-\beta_0 Y_{t-1})-\gamma\Delta Y_t\\
&=\Delta C_t-(-1+x)(C_{t-1}-\alpha_0-\beta_0 Y_{t-1})-(\beta_0+y)\Delta Y_t\\
&=C_t-C_{t-1}+C_{t-1}-\alpha_0-\beta_0 Y_{t-1}-\beta_0\Delta Y_t-x(C_{t-1}-\alpha_0-\beta_0 Y_{t-1})-y\Delta Y_t\\
&=(C_t-\alpha_0-\beta_0 Y_t)-x(C_{t-1}-\alpha_0-\beta_0 Y_{t-1})-y\Delta Y_t\\
&=e_t-xe_{t-1}-y\Delta Y_t
\end{aligned}
$$

ここで e_t は t 期の回帰式 $C_t=\alpha_0-\beta_0 Y_t$ の残差である。したがって，ECM における β，γ の最小二乗推定は e_t を e_{t-1} と ΔY_t とに回帰することに実際上等しくなる。仮に e_t と e_{t-1}，ΔY_t の間に関係が弱ければ x や y は 0 に近くなり，したがって $\beta=-1$，$\gamma=\beta_0$ がおよそ成立することになる。そして今まで見たように，作られたデータではこのような時間的関係が弱かった。よって，さきのような結果になったものと考えられるが，一方で現実のデータで x，y を求めると係数がそれぞれ 0.616，−0.153 (これは ECM の式と符合する)，t 値が 4.41，−4.41 となる。

このことを逆に言えば，さほど時間構造がなくても ECM のようなものが推定されたように見えてしまう可能性があるということである。もっとも，このことをもって，ほかの ECM の定式化についても同様の推論が無条件に当てはまることは保証されない。

4.5 Johansen システム検定

以上で単一方程式モデルについてみてきた。本項では Johansen によるランク検定を見てみることにしよう。ここでは検定結果の共和分の数のみに対象を絞ることにする[19]。

R のパッケージ urca の ca.jo を使用するが，これは固有値検定，トレース検定ともに可能であり，コインテグレーションの中に定数およびトレンド項を入れるか入れないか (const，trend) などを設定可能になっている[20]。どちらも入れない場合 (none) は現実的ではないが，結果については参考として掲載しておく。

今までと同じようにまず現実のデータから分析しよう。表中の数字は各検定方法，有意水準に対して得られるランク数，つまり共和分関係の数である。

---

19) 共和分を含めた Johansen システム検定については森棟，前掲著 pp. 364−374，田中勝人『現代時系列分析』(2006) 岩波書店，pp. 251−261 など参照。

20) そのほか，spec (VECM の定式化) は transitory，K (ラグオーダー) は 2 に設定した。

表 4.11

	none			const			trend		
有意水準	10％	5％	1％	10％	5％	1％	10％	5％	1％
固有値検定	1	0	0	2	1	0	1	1	0
トレース検定	0	0	0	2	1	1	1	1	1

一番ハンドリングしやすいのは共和分関係が 1 のときだろうから，const や trend といった現実的な設定のときに，「望ましい結果」が出ているといえる。固有値検定とトレース検定で顕著な違いが出ているともいえないだろう。

作られたデータによって同様の表を掲げるが，表中は各ランクに評価される割合を示す。まず，相関を入れない場合である。

表 4.12

		none			const			trend		
		10％	5％	1％	10％	5％	1％	10％	5％	1％
固有値検定	r＝0	32.81％	48.07％	75.67％	15.83％	29.02％	63.82％	45.33％	60.01％	82.93％
	r＝1	65.37％	51.29％	24.26％	15.83％	24.82％	24.73％	46.65％	35.78％	16.09％
	r＝2	1.81％	0.64％	0.07％	68.34％	46.16％	11.44％	8.03％	4.22％	0.98％
トレース検定	r＝0	44.08％	59.79％	86.08％	4.32％	9.30％	29.08％	47.98％	62.04％	82.21％
	r＝1	53.65％	39.30％	13.80％	19.08％	34.53％	52.72％	42.09％	32.44％	16.42％
	r＝2	2.26％	0.91％	0.13％	76.60％	56.17％	18.20％	9.93％	5.52％	1.38％

比率の評価は分かれるだろうが，共和分関係あり，つまり r＝1 または 2 の比率は const では一定程度高いといえるだろう。ただし，r＝1 に限定すれば，比率はそう高くない。const においては検定の種類によって結果がかなり異なる。

最後に相関を入れて計算しよう。

表 4.13

		none			const			trend		
		10％	5％	1％	10％	5％	1％	10％	5％	1％
固有値検定	r＝0	29.21％	44.48％	72.89％	9.05％	19.70％	54.54％	43.61％	58.42％	82.38％
	r＝1	68.71％	54.65％	26.99％	10.72％	20.88％	28.03％	50.17％	38.97％	17.21％
	r＝2	2.09％	0.87％	0.12％	80.23％	59.42％	17.44％	6.23％	2.61％	0.41％
トレース検定	r＝0	41.11％	56.88％	84.57％	1.46％	3.92％	17.53％	42.71％	58.14％	81.51％
	r＝1	56.39％	41.95％	15.23％	12.44％	27.46％	56.12％	49.18％	38.17％	17.88％
	r＝2	2.50％	1.18％	0.20％	86.10％	68.62％	26.36％	8.11％	3.69％	0.61％

trend の結果は相関のあるなしでさして変わりがない。const においてある程度「共和分あり」の判定が増えている。しかし，r＝2 という判定が多く，必ずしも「望ましい結果」とはいえないだろう。

Johansen のシステム検定についていえば，現実のデータと偽のデータでかなりの差があるといえる。

5. 結　語

一様分布，ないしその 2 つの線形結合によってデータがふってわいてくるなどということは現実にはありえない。去年までのデータがある範囲の一様分布からとられ，しかし，最も大きいデータは今年や来年，さらに次の年などのためにとっておくなどといったことはまったくもって不自然である。しかし，そのようなデータであってもなんらかの統計解析はでき，分析結果をあれこれと評価することはできる。

研究者は何らかの制約されたモデル群の中で考察を行うことが一般的だから，そのデータが全く異なるところから発しているとまでは考えることはないし，その必要もないことになっている。ある場合には，もっともらしい結論が，それに至る多くに正当性を与える。しかし，本稿では必ずしもその正当性が保障されるものではないことを例示した。逆にいうならば，研究者はモデル群の制約の正当性に確信を抱く必要がある。

また，あわせて本稿では GDP データにあらわされるような種類の回帰分析は，実質的にはその構成要素の安定性を示しており，構成要素の比率および (構成要素以外の) 非構成要素との相関係数の 2 つからかなり説明されることを示した。さほど明示はしていないが，実質値を使った分析を支えているものが現代日本経済においては価格という要素であることが考えられることに触れた。

3 節でシンプルな回帰分析よりラグつき変数を含むモデルのほうで，さらに 4 節の時系列的方法が導入されている場合でも作られたデータのいわば「底の浅さ」が露呈した。経済においては「箱の中から玉を取り出す」といったモデルに解消されることが困難な時系列構造が根深くあり，これは経済分析において厄介なものというよりは，経済における一種の構造を示すものと考えることができるだろう。

参 考 文 献

刈屋武昭，矢島美寛，田中勝人，竹内啓『経済時系列の統計』(2003) 岩波書店
田中勝人『現代時系列分析』(2006) 岩波書店
山本拓『経済の時系列分析』(1988) 創文社
間瀬茂『R プログラミングマニュアル』(2007) 数理工学社
蓑谷千凰彦『計量経済学大全』(2007) 東洋経済新報社
森棟公夫『計量経済学』(1999) 東洋経済新報社

Appendix A 順序統計量 Y_t および Y_{t-1} の同時密度関数

本稿 2 節で取り上げた順序統計量 Y_t および Y_{t-1} の同時密度関数の式を明示しておく。

やや煩雑だが，基本的な公式から求められる。

x および y を Y_{t-1} および Y_t の値とする。簡略化のために $f=f(x, y)$ と表記しておく。

x および y を次の表のように分類することができる。

		Y		
		$0 \leqq y < b$	$b \leqq y < a$	$a \leqq y \leqq a+b$
x	$0 \leqq x < b$	A	B	C
	$b \leqq x < a$	D	E	F
	$a \leqq x \leqq a+b$	G	H	I

ここで $x>y$ ならば $f=0$ なので，領域 D，G，H を考慮する必要は事実上ない。また，領域 A，E，I も $x>y$ ならば $f=0$ であるからこれらの領域についてはその部分を除いた領域における密度関数を考慮する。ここで $c=ab$ とする。

領域 A　$\tilde{f}(x, y)=(x^2/(2c))^{t-2}\cdot((xy)/c^2)\cdot(1-(y^2/(2c)))^{n-t}$
領域 B　$\tilde{f}(x, y)=(x^2/(2c))^{t-2}\cdot(x/(ac))\cdot(1-(1-(2y-b)/(2a))^{n-t}$
領域 C　$\tilde{f}(x, y)=(x^2/(2c))^{t-2}\cdot(x(a+b-y)/c^2)\cdot((a+b-y)^2/(2c))^{n-t}$
領域 E　$\tilde{f}(x, y)=((2x-b)/(2a))^{t-2}\cdot(1/a^2)\cdot((a+b-y)^2/(2c))^{n-t}$
領域 F　$\tilde{f}(x, y)=((2x-b)/(2a))^{t-2}\cdot((a+b-y)/(ac))\cdot((a+b-y)^2/(2c))^{n-t}$
領域 I　$\tilde{f}(x, y)=(1-(a+b-x)^2/(2c))^{t-2}\cdot((a+b-x)(a+b-y)/c^2)\cdot((a+b-y)^2/(2c))^{n-t}$

確率密度は上の $\tilde{f}(x, y)$ に $n!/((t-1)!(n-t)!)$ をかけた値になる。

Appendix B 順序統計量などの相関係数の近似計算

基本的な考え方は連続分布を離散化するということである

ここでの例示は n 個の点をベースに計算する方法を示す。本稿では以下の 2 つの方法のうちのひとつで n 個の「分位点」を計算している。

方法 1：分布関数 $F(x)$ について，$F(x)=i/n$ であるような $n+1$ 個の x_i の値を算出する（$i=0\ldots n$）。$(x_i+x_{i+1})/2$ (i=0…n−1) の 20 個の点をもとにする。
方法 2：分布関数 $F(x)$ について，$F(x)=i/n-1/2n=(2i-1)/2n$ であるような x_i の値を算出し

(i=1...n)，これらの点をもとにする。

相関係数の近似値の計算などは，x_i を総当たりで発生させ，その結果の相関係数などを計算する。本稿に即して言えば，k 個の順位統計量は x_i の n 種類のとりうる値をすべて発生させ，Δx_i や x_i の相関係数を計算する。

R での実装は以下のようにすればよい。なお，R の文献ではよく代入として「＜－」が推奨されているが，ここでは一般的な「＝」で表記する。実際上特に問題はない。

一様分布の場合は方法 1 と方法 2 で計算させる値は同じになる。

```
x=c(1:n)
x=x/n-1/(2*n)
```

として

```
for(i1 in x){
 for(i2 in x){
...
この間に相関係数などを計算するための統計量を計算 (for 文はデータの数だけ必要)
...
 }
}
```

とすればよい。

一様分布の和の場合には次のようにすればよい。ここでは和は 0 から 2 の間をとる。また n は偶数とし，n=2m とする。そのように設定しても実際上問題ない。

方法 1 をとる場合は x の計算の部分は次のようになる。

```
y=c(0:m)/m
y=sqrt(y[-1])+sqrt(y[-n])
x=cbind(t(y),t((2-rev(y))))
```

方法 2 をとる場合は以下のようになる。

```
y=c(1:m)
y=sqrt(y/m-1/(2*m))
x=cbind(t(y),t(2-rev(y)))
```

この方法の利点は，かなり込み入った分布関数であっても，複雑な統計量の特性値を容易に求めることができることである。一般的な 2 つの一様分布確率変数の和，本稿でいえば a や b の様々な値についての特性値についても求めることができる。

しかし一方で，この方法には次の問題がある。

データの大きさを k，「詳細さ」をあらわす分割数 (分位点の数) を n としたとき，計算回数は n^k になるから n や k を増やすとかなり計算時間がかかってしまう。n を k 以上にすることも必須である。順序統計量などの計算では，さらに並べ替えもループ内に入ってくるから，計算時間は k の増加とともにますますかかることになる。

中敷領　孝能

Summary

Plausible Data Analysis on Artificial Imitated Statistics, generated by Uniform Random Variables or Order Statistics

When researchers study their subjects using some models, they often have to make their models, whether consciously or unconsciously, in restricted model spaces to some extent. And if constructed models explain those subjects fairly, models are considered to correct and the conclusions of researchers are also thought to be right. However, we have to think that those conclusions could have derived because initial model spaces were restricted.

In this article, I show this apparent problem by analyzing the SNA data. Simple consumption function is popular in statistics or econometrics, I handle this function again. From some statistics derived from real data structures, I generate GDP or consumption data. But those data are generated in the way far from the reality of the world.

From wherever data have come, researchers can analyze those data. And more plausible results convince us of more reliability of conclusions.

I state this article in the following way. After explaining the stochastic characters of generated data, I set forth from simple regression models to times series analysis including integration and cointegration. One subject, relatively slightly uttered, is the role of price in an analysis of modern Japanese economy.