



Osaka Gakuin University Repository

Title	クロスセクションデータと時系列データにおける相関関係の非整合性の研究 ～なぜ私たちはクロスセクションで現実の諸現象を説明できると考えているのか～ Research for Difference of Correlation between Cross-sectional Data and Time-series Data
Author(s)	加茂 英司 (Eiji Kamo)
Citation	大阪学院大学 流通・経営学論集 (OSAKA GAKUIN UNIVERSITY REVIEW OF DISTRIBUTION, COMMUNICATION AND BUSINESS ADMINISTRATION), 第 38 巻第 2 号 : 41-72
Issue Date	2013.03.31
Resource Type	ARTICLE/ 論説
Resource Version	
URL	
Right	
Additional Information	

クロスセクションデータと時系列データにおける 相関関係の非整合性の研究

～なぜ私たちはクロスセクションで現実の諸現象を説明できると考えているのか～

加茂 英司

Research for Difference of Correlation between Cross-sectional Data and Time-series Data

Eiji Kamo

ABSTRACT

Many researchers think that the strength of correlation is same between cross-sectional data and time-series data, but it is not because each data does not always reflect same fluctuations. This is why we cannot explain the change of real phenomenon by cross-sectional data. But why do we still keep using cross-sectional data as an important tool for comparison analysis? It is because we cannot identify whether the result of investigation by cross-sectional analysis is correct or not, when time-series data keeps stable positions.

はじめに なぜクロスセクションデータを使っても時系列データを説明できるのか

変数間の関係性を観察する時に、クロスセクションデータを使うことはきわめて一般的な方法である。しかしクロスセクションデータでは強い正の相関関係を観察することができて実際には無相関であったり、反対にクロスセクションデータでは無相関だが時系列データでは強い相関関係であったり、ということを経験した人は少なくないだろう。

変数間には固有の関係性というものがあるがどんなクロスセクションデータでも変数間の関係性を測定できると考える人が多いが、実際はそうではない。クロスセクションデータの集め方しだいで相関関係の強さは変わってしまうだけでなく、同じ時期の同じ変数間であってもクロスセクションデータと時系列データでは相関関係の強さは異なる。

したがって比較分析の問題点とは、クロスセクションデータと時系列データでは相関関係の強さが異なっているにもかかわらず、なぜクロスセクションデータを使っても現実の諸現象における相関関係を説明できるように見えるのか、ということなのである。

そこで筆者が着目したことは、時系列データにおける相関関係の判断方法に問題があるのではないか、ということである。ある条件下の時系列データに限れば、時系列データとクロスセクションデータ間に齟齬があっても表面化しないからである。

本稿の第1節ではクロスセクションデータと時系列データ間に生じる齟齬は、なぜ不可避なのかを説明し、第2節はその齟齬がなぜ表面化しないのかを説明する構成になっている。

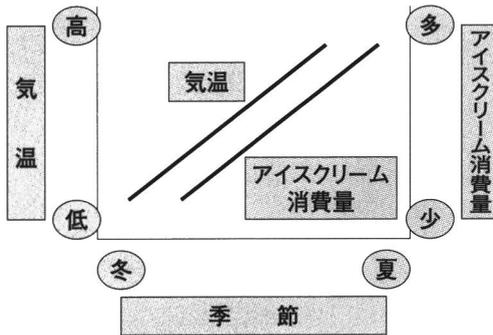
第1節 なぜクロスセクションデータは現実の時系列的変化を説明できないのか

気温とアイスクリームのトリック

暑い日が続く夏にはアイスクリームがよく売れ、対照的に寒い日が続く厳し

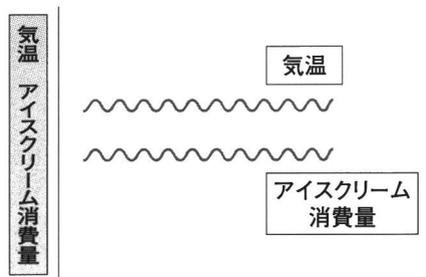
い冬の日にはアイスクリームはあまり売れない。豊かな四季に恵まれている我が国では、アイスクリーム消費量と気温には強い相関関係のあることを誰もが知っている。時系列データで表すと図1のようになる。

図1 気温とアイスクリーム消費量



これを繰り返すと図2のようになり、アイスクリーム消費量と気温には一年を単位とした正の相関関係がある。

図2 1年を単位とした短期変動には正の相関関係



ではシンガポールに住んでいる人は、この事実を知っているのだろうか。表1はある年におけるシンガポールの各月における平均最低気温と平均最高気温

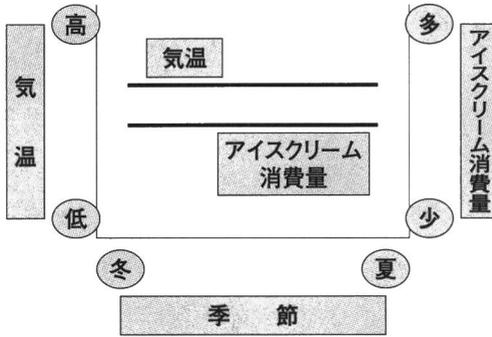
だが、毎月の気温は、驚くほど安定していることがわかる。

表1 シンガポールにおける各月の平均最低気温と平均最高気温

月間平均		
1月	最低24℃	最高30℃
2月	最低24℃	最高31℃
3月	最低25℃	最高32℃
4月	最低25℃	最高32℃
5月	最低26℃	最高32℃
6月	最低26℃	最高32℃
7月	最低25℃	最高31℃
8月	最低25℃	最高31℃
9月	最低25℃	最高31℃
10月	最低25℃	最高31℃
11月	最低24℃	最高31℃
12月	最低24℃	最高30℃

アイスクリーム消費量もまた安定しているとすれば、図3のように2本の時系列データは平行線になる。実際には雨季と乾季があって湿度は変化しているため、年間のアイスクリーム消費量には変化があるだろう。しかしここでは話を簡単にするために年間のアイスクリーム消費量には変化が無いとして話をすすめたい。この場合、シンガポールでも気温とアイスクリーム消費量には正の相関関係があると、人は考えるだろう。

図3 シンガポールにおける気温とアイスクリーム消費量



しかし図1と図3が似て非なることは、散布図を書くときよくわかる。気温をX軸、アイスクリーム消費量をY軸にすると、図1の散布図は図4、図3のそれは図5である。図4の近似線は右肩上がりであり、気温とアイスクリーム消費量に正の相関関係のあることがわかる。ところが図5には近似線は存在せず、ただの点しか存在しない。図6、図7、図8をみればわかるが、ひとつの点だけを見て相関関係の有無は判断できない。気温とアイスクリーム消費量の相関関係を判断するためには、複数の値の異なるサンプルが必要なのである。

図4 我が国における気温とアイスクリーム消費量の散布図（近似線）

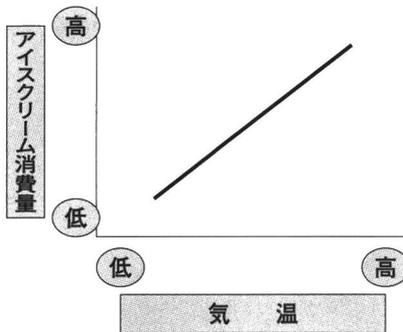


図5 シンガポールにおける気温とアイスクリーム消費量の散布図

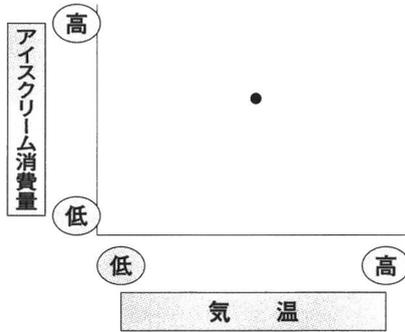


図6 1つのサンプルだけを見て正の相関関係があると考えられる場合

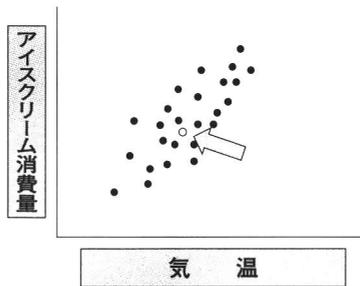


図7 1つのサンプルだけを見て負の相関関係があると考えられる場合

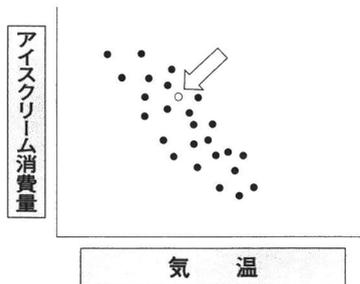
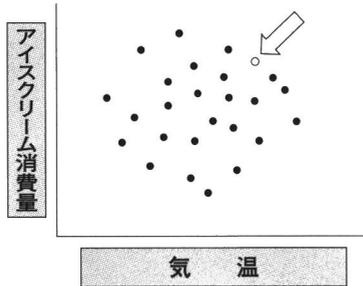


図8 1つのサンプルだけを見て無相関であると考える場合



アイスクリーム消費量と気温の長期変動は無相関

ところがちょっと視点を変えると、シンガポールでもアイスクリーム消費量は次のような事情によって、大きく変化をしていることがわかる。

氷が手に入らないシンガポールでは、太古の時代からアイスクリームを食べてきたわけではない。冷凍技術の普及、アイスクリームの大量生産、冷凍したまま流通させる技術の開発、あるいはいつでも買うことができるコンビニの普及などが必要になる。また外国から来た冷菓を食べられるだけの生活水準の向上も必要である。

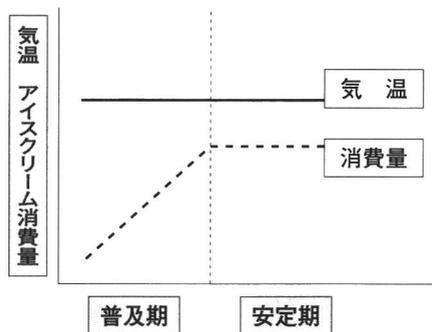
シンガポールにおける伝統的な冷たいお菓子といえばカキ氷なのだが、牛乳を使った冷たいデザートを日常的に食べるようになるためには欧米風の食習慣へ変化すること¹⁾が必要である。

長期的にアイスクリーム消費量が増えている間、気温はとりたてて変化していないため、アイスクリーム消費量と気温の長期変動は無相関ということになる。それをグラフにしたものが図9である。

1) この食習慣という変数は、一般に潜在変数と呼ばれているものである。潜在変数というのは数値化できないが、社会の様々な諸現象を理解するには重要なものである。

安定期のデータだけでは相関関係があるように見えてしまうが、変動期を含むデータ全体を見れば両変数は無相関であることがわかる。

図9 シンガポールにおける気温とアイスクリーム消費量の長期変動

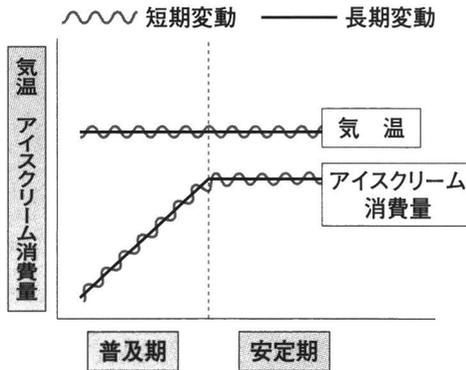


我が国の気温とアイスクリーム消費量の場合

普段、長期変動のあることを私たちが意識することがない理由は、アイスクリーム消費量の変化を語る上でとりたてて社会的に意味がないからとか、意図的に数値化して取り扱わないとか、あるいは食習慣などの潜在変数の場合では数値化が困難な場合があるからである。

長期変動は、シンガポールと同じように我が国にも存在している。したがって我が国のアイスクリーム消費量と気温の時系列データをグラフに表すと図10のようになる。

図10 我が国における短期変動と長期変動



一般に時系列データは複数の変動によって構成されているとされており、統計学のテキストではそのサイクルの長短や規則性の有無によって変動の種類を4つ²⁾、もしくは3つ³⁾に分類している。なお本稿では話を簡単にするために短期変動と長期変動だけを考えて、話をすすめる。

2) (1) 傾向変動（トレンド trend ; T）

上昇もしくは下降などの比較的単調な長期的傾向で、一般には時間軸上の単純な線形関数（直線など）で表現される。

(2) 循環変動（サイクル cyclical ; C）

トレンドのまわりで上下する周期が定まっていない循環の変動で、景気変動などがこれにあたり、次の季節変動とは区別する。

(3) 季節変動（seasonal ; S）

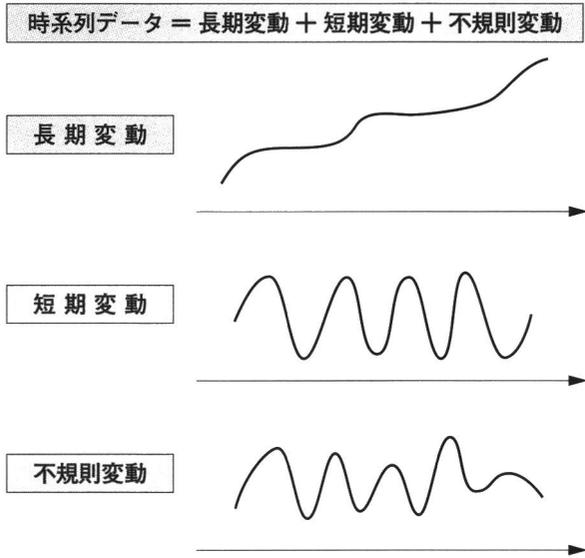
季節によって左右される1年を周期として規則的に繰り返される変動。ただし1年を周期としなくても、同じサイクルで繰り返される日変動のような固定的な変動があれば、季節変動と同様な処理手順が適用できる。

(4) 不規則変動（irregular ; I）

上記以外の説明がつかない突発的な変動や不規則かつ短期間の上下に起こる小変動で、ランダムノイズと言い、期待値（平均値）は0と仮定される。

3) 長期変動、短期変動、不規則変動

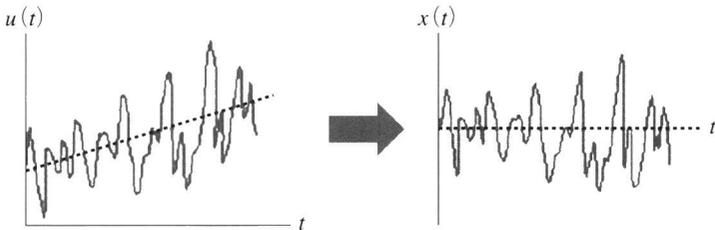
図11 3種類の時系列変動



出典 名古屋大学 サイト

複数の異なる変動が混在している場合、短期変動と長期変動を分離してそれぞれの変動の相関関係を判断するという作業が必要になる。分離する作業は簡単である。例えば図12は、混在した状態から長期変動のみを除去する作業をしているものである。

図12 長期変動を除去する方法



出典 <http://www.astrobio.k.u-tokyo.ac.jp/tajika/class/kiso2/fourier2.pdf>

この作業を経て長期変動だけを抜粋したものが図13であり、短期変動だけを抜粋したものが図14である。短期変動は互いに正の相関関係があるが、長期変動は互いに無相関である。

図13 長期変動だけを抜粋したもの

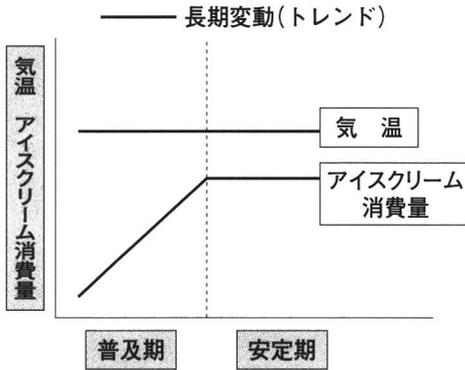
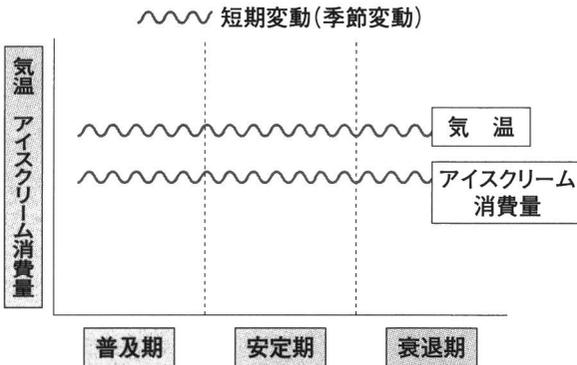


図14 短期変動だけの抜粋

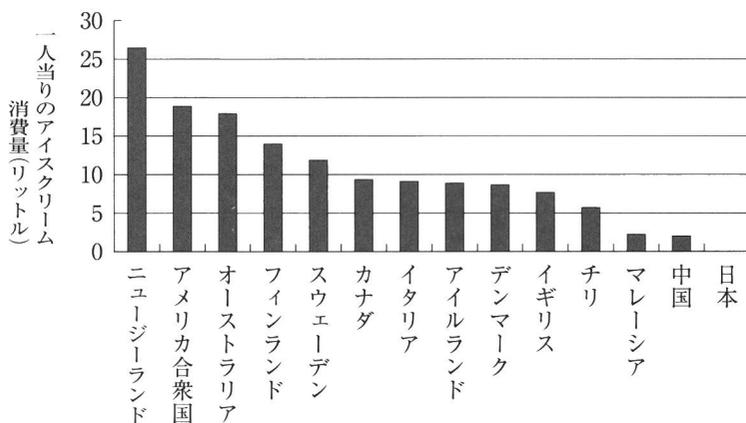


クロスセクションデータは単独の変動だけを反映する

時系列データが複数の変動を常に混在させているのに対し、クロスセクションデータは単独の変動だけを反映するという特徴がある。

グラフ1は国別の1人当たりアイスクリーム消費量である。上位にならぶのは欧米の国であり、肉食を中心とした食文化を持つ国である。それに対して下位に並ぶのは肉食の伝統を持たないアジアの国である。同じ様な気候の国でもアイスクリーム消費量は様々であり、またフィンランドのようなきわめて冷涼な国が上位にあるのに対して、マレーシアのような常夏の国が下位に並んでいる。国別クロスセクションデータではアイスクリーム消費量と気温が無関係になるのを見て、「このデータはおかしいのではないか」と思う人もあるだろうが、普段は隠れている長期変動だけを反映しているからなのである。

グラフ1 我が国の一人当たりアイスクリーム消費量は少ない



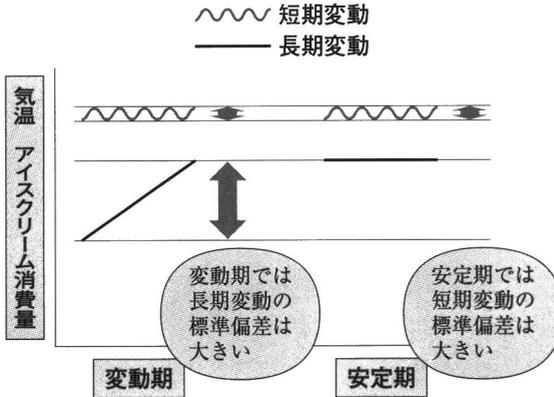
出典 アイスクリーム情報ネット 2002年

隠れた変動を視覚化させるクロスセクションデータ

時系列データでは、その主役の座が長期変動と短期変動とで入れ替わるたび

に、アイスクリーム消費量と気温の相関関係は無相関になったり、強い相関関係となったりする。

図15 時系列データでは強い変動が入れ替わる



それに対してクロスセクションデータでは各変動の標準偏差が安定する⁴⁾ため、国別クロスセクションデータは、いつまでたっても長期変動しか反映しな

4) 古くから「平均回帰」という名前でも知られてきた現象だが、そのメカニズムを人間の身長を例にして説明しよう。戦後、日本人の平均身長は時系列的に高くなったことは良く知られている。1945年20歳の平均身長は男性165cm、女性が153.2cm。ところが1995年になると20歳の平均身長は男性171.1cm女性が158.4cmになる。50年間でみると、男性は6cm以上、女性で5cm以上も伸びている。しかしその間でも各世代の身長の標準偏差は安定していたのは、平均回帰が働くからなのである。かりに平均回帰が働かないと仮定しよう。その場合、夫婦とも身長の高い親から生まれた子どもの身長は、親の身長の周りに分布することになる。そのために一定の割合で親の身長を上回る子どもが誕生することになる。数十世代後、数百世代後には何十メートルもの身長の人間が誕生することになる。最初、平均回帰は種子の大きさや人間の身長などによって確認されたために、生物学的に限られた現象ではないかと考えられてきたが、しだいに一般的な現象であることがわかり、統計学的なものであるものとされるようになった。そのためにクロスセクションデータにおける相関関係の強さは安定することになる。

い。そこでアイスクリーム消費量と気温は無相関であり続けるのである。

クロスセクションデータの持つこの性格は、隠れてはいるものの、変数間の関係性を理解する上で重要な変動を炙り出してくれる。一部の国を除いて、アイスクリーム消費量は既に安定期に入っているため、長期変動の存在はわかりにくくなっている。それにもかかわらず、やはりアイスクリーム消費量には長期変動があるということを教えてくれるのである。

しかし様々な事情によって、いくらクロスセクションデータを使っても、重要ではあるが隠れた変動を明らかにすることのできない場合も少なくない。それを朝食と成績を使って説明したい。

朝食と成績は1日を単位とした超短期変動

朝食を食べることが子供達の心身の発達に良いことは経験的にわかっている。その理由は次のようなものである。

資料1 短期変動のメカニズム

◎なぜ朝食をとることが大切なのか、朝食には3つの大事な役割があります。

1. 体温を上げて眠っていた体を起きた状態にする働き

2. 体とアタマにエネルギーを補給する働き

3. 体調を整える働き

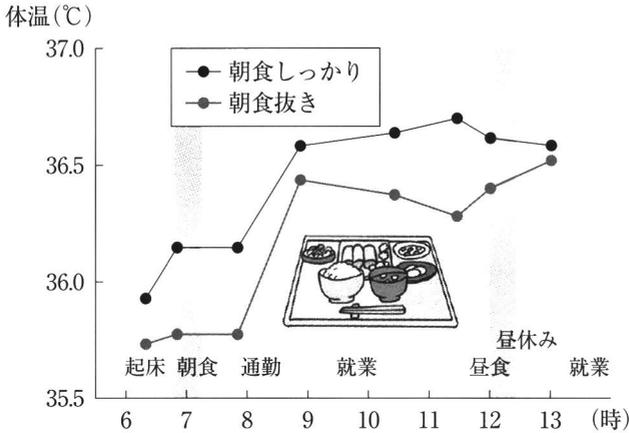
出典 京都市教育委員会サイト

<http://www.city.kyoto.lg.jp/kyoiku/cmsfiles/contents/0000101/101682/11siryo.pdf>

朝食がエネルギーを補給して脳を活性化させることを、科学的に証明しているのがグラフ2である。しかし差異が持続するのは午前中までであり、昼食後の差異は小さい。夜になって寝てしまえば、どちらの場合でも体温は低下するため、つまりエネルギー補給を通して朝食が脳を活性化させる影響は、僅かに

「1日」を単位としたきわめて短い「短期変動」である。

グラフ2 朝食を食べている場合と食べていない場合の午前中の体温の推移



出典 文部科学省サイト

http://www.mext.go.jp/a_menu/shougai/katei/08060902/004.pdf

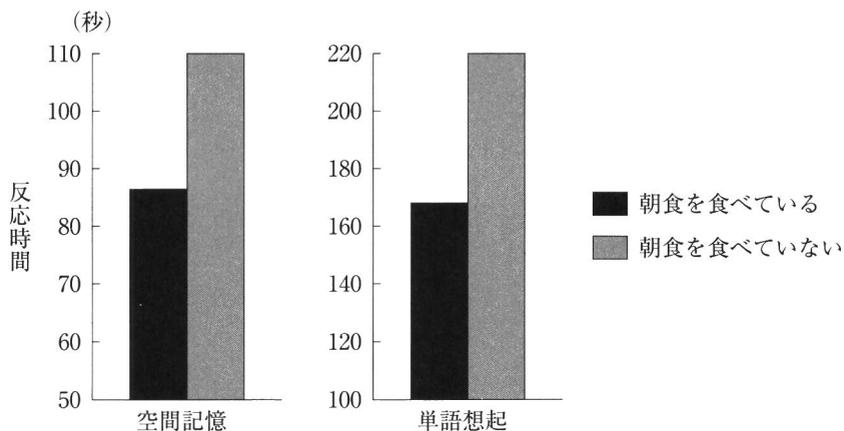
一日を単位とした短期変動を明らかにしたいのであれば、同じ人をサンプルにして朝食を食べた日と朝食を食べない日を比べるという時系列データでもよいが、同じ日に朝食を食べた人と食べていない人を比べるクロスセクションデータでもかまわない。

アメリカの大学で行われた実験だが、被験者はすべて同じ栄養スープを食べる。37.7グラムの糖質、12.2グラムの脂質、18.5グラムの蛋白質を含み、熱量は326キロカロリーであった。テストは「空間記憶」と「単語想起」の2種類。結果は、両方とも朝食を食べたほうが短時間で答えを出した⁵⁾。このクロスセ

5) 空間記憶のテストとは、りんごやイヌなど16種類の図を配置した絵を見せて、あとで位置関係を思い出させるものである。単語想起のテストは、15個の単語を2秒おきに読み上げ、あとで思い出させるものである。

クシオンデータは短期変動を反映しており、朝食摂取率と成績に正の相関関係のあることがわかる。

グラフ3 アメリカの大学で行われた実験結果
朝食摂取の有無と記憶の関係



(対象：平均年齢21.3歳の男女。n = 33)

(David Benton and Pearl Y Parker, Breakfast, blood glucose, and cognition, American Journal of Clinical Nutrition 1998)

出典 テルモ体温研究所サイト
<http://www.terumo-taion.jp/>

ところが朝食摂取率と成績の短期変動の相関関係を観察するために、朝食摂取率別（もしくは朝食欠食率別）クロスセクションデータをとることが一般的に行われる（資料2参照）。

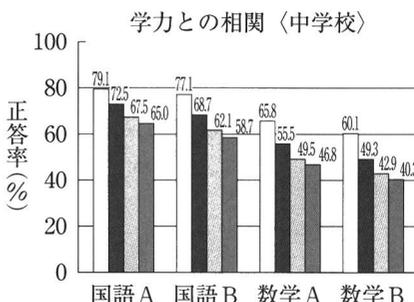
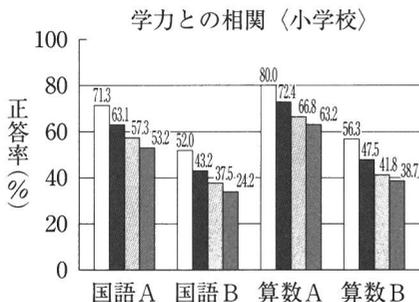
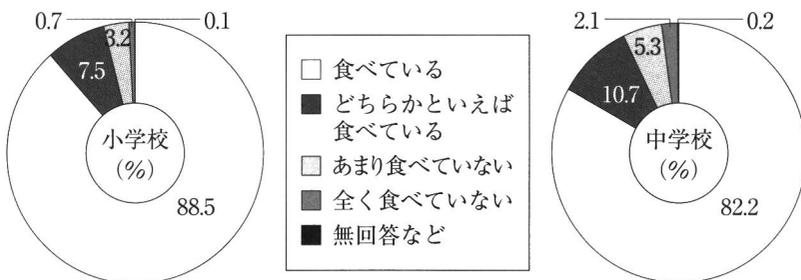
この調査で確認していることは、試験当日に朝食を食べてきたのかどうかではなく、むしろ朝食を食べる「習慣」の有無である。たまたま何らかの事情で試験当日に朝食を食べてきても、朝食を食べる習慣がなければ「食べない」サ

ンプルとして扱われてしまう。反対にたまたま何らかの原因で試験当日に朝食を食べられなかったが、普段は食べる習慣さえあれば「食べる」サンプルとして扱われてしまうからである。

習慣は1日ごとに変化するものではないため、資料2が計測しているものは短期変動とは別の何らかの変動であると考えらるべきである。別の何らかの変動を本稿では便宜的に「長期変動」と名付けるが、短期変動を説明したいにもかかわらず、使っている説明資料は長期変動を反映するクロスセクションデータである。

資料2 朝食摂取率別クロスセクションデータ

■朝食を毎日食べていますか？(全国・小学6年生、全国・中学3年生)



出典 京都市教育委員会サイト

<http://www.city.kyoto.lg.jp/kyoiku/cmsfiles/contents/0000101/101682/11siryo.pdf>

クロスセクションデータは長期変動を炙り出してくれるが、短期変動と長期変動だけでは朝食と成績の関係性を説明することはできない。というのもどちらも正の相関関係なので、朝食と成績が無相関であるという事実を説明できないからである。

多くの方は朝食と成績が無相関であることを経験的に知っている。朝食をよく食べている子供の成績が必ずしも良いというわけでもないし、成績の良い子供が必ずしも朝食をしっかり食べているわけではないからである。そこで朝食と成績に正の相関関係があるという資料を見せられても違和感が残るだけなのである。長期変動はそもそも擬似相関だから、検証結果に強い相関関係が現れても実際には相関関係がないという無理やりな説明をしたがるのだが、間違った理解の仕方である。ではどのように理解すればこの違和感を解消できるかを次に説明したい。

中期変動は無相関

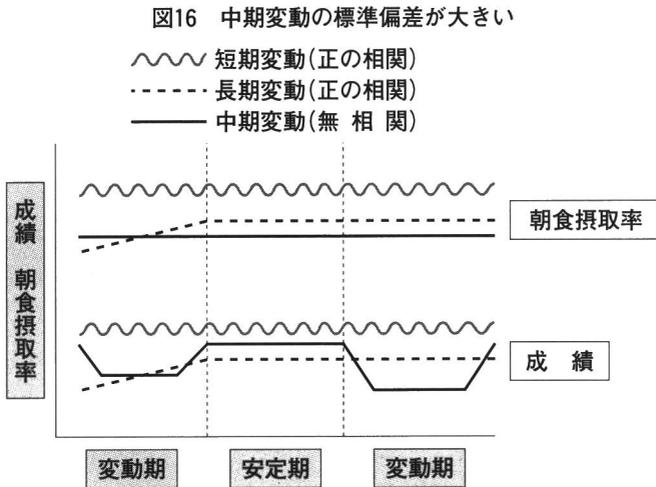
短期変動は実験などでのそのメカニズムもよくわかっているが、長期変動はメカニズムがわかっていないので様々な推測が横行する。家庭環境を原因変数にした擬似相関なので、これが朝食と成績を無相関にしているという説明をする人が少なくない。しかしクロスセクションデータで正の相関関係があるということは、少なくとも時系列データでも正の相関関係があると考えなければならない。

朝食と成績が無相関であるということを説明したいのであれば、短期変動、長期変動以外に、朝食と成績を無相関にする変動が他に存在していることを示すべきなのである。そのひとつの事例としてつぎのような変動の存在を指摘することが出来る。

昨今の子供たちの学力低下が社会問題化している⁶⁾が、とりたてて朝食摂取

6) 小学校では1980年度から2010年度、中学校では、1981年度から2011年度、高校では1982年度から2014年度（数学及び理科は2013年度）まで施行される「ゆとり教育」が原因であると言われている。

率が低下しているわけではない⁷⁾。学力低下の原因としてゆとり教育の導入を指摘されている。朝食摂取率と成績を無相関にしているこの変動を便宜的に「中期変動」と呼べば、短期変動、中期変動、そして長期変動を分離して表したものが図16である。時系列データにおける主役は時系列的に交代するので、中期変動の標準偏差が大きくなれば朝食摂取率と成績の時系列データは無相関になる。比較分析研究では、こうした方法によって朝食と成績が無相関であるという説明ができるのである。



中級変動の時系列データが無相関であれば、それを反映させるクロスセクションデータでも、朝食摂取率と成績が無相関になるのは当然である⁸⁾。

7) データの見方によっては昨今の子供たちの朝食摂取率は低下していると考える人もあるし、若干高くなっていると見る人もいるが、本稿ではそれは誤差の範囲内であり変化してないと考えて検証を進める。

8) そもそもクロスセクションデータというのは、指数帯別に作る必要は無い。指数帯別というのは、朝食摂取率の高い順番という意味である。実際、国際比較などでは国別クロスセクションデータが一般的である。また開業率廃業率では地域別クロスセクションデータ、そして流通研究では業種別クロスセクションデータなど、様々な種類のクロスセクションデータが使われているのである。

例えばゆとり教育が地域別に段階的に導入されていたとすれば、数年前から導入した小学校、導入したばかりの小学校、まだ導入していない小学校を対象にして、それぞれの学校ごとに成績の平均点と、朝食摂取率の平均点を集めることが出来る。

ゆとり教育が成績に大きな影響力を持っているとすれば、導入の早い学校ほど、その平均点は低くなる。一方、朝食摂取率は教育制度改革と関係が無いため、朝食摂取率と成績は無相関になる。

ところが一般的には教育制度改革は全国一斉に導入されるものである。こうした現実の制約のために、教育制度改革の開始年度別クロスセクションデータを作ることは難しい。したがってクロスセクションデータを使っても、重要な変動を必ずしも炙り出すことはできないのである。

実際のところ、様々な現実の制約によって重要な変動を抽出できないことは珍しくない。例えば国際比較をしている研究では国別クロスセクションデータを使わざるをえないが、重要な変動を反映する保証はない。また経済成長率や市場成長率などのマクロデータは直接にアンケート調査をするわけではないから、二次データが分類している分類法に依存せざるをえない。そこで業界別クロスセクションデータなどを使うことになる。

これまでの説明をまとめると次のようになる。朝食と成績には正の相関関係の短期変動と長期変動がある。無相関であることを言いたいのであれば、無相関である中期変動を示すべきである。

重要な変動を炙り出すことができれば、クロスセクション分析は有用な方法である。しかし重要でない変動を炙り出すのであれば、かえってクロスセクション分析は研究をミスリードする方法論となってしまう。クロスセクションデータはいわば「諸刃の剣」なのである。そこで第2節では、クロスセクションデータで何らかの変動を発見しても、その変動がどの程度重要なのかを検証する方法について詳しく話をすすめたい。

第2節 なぜクロスセクションデータで現実の諸現象を説明できると考えてしまうのか

安定期のデータでは相関関係を判断できない

クロスセクションデータが重要でない変動を反映していれば、相関関係の符号は時系列データのそれと異なる。しかしクロスセクションデータと時系列データの符号が異なれば、いくらなんでも見ればわかるだろう。そうであればこれだけ広範囲にクロスセクションデータを使った分析方法が一般的であるわけがない。ところがそれをわからなくさせる相関関係の「トリック」が存在するのである。

第1節でも一部を紹介したが、あらためて時系列データのパターンごとに散布図がどうなるかを使って説明しよう。図17は時系列データのいくつかのパターンを描いたものである。図17-Aの右肩上がり、図17-Bの右肩下がり、ともに正の相関である。散布図では図18-Aになる。図17-Cは負の相関であることを表しており、散布図では図18-Bのようになる。また図17-Dと図17-Eは片方の変数の時系列データが安定しており、もう片方の変数だけが変化をしているために無相関と判断されるパターンである。散布図では図18-Cになる。

ここまでは散布図を描くことが可能だが、図17-Fのように両変数ともに安定して推移していれば、散布図は図18-Dのようにただのひとつの点になるだけである。これでは相関関係の符合どころか、相関関係の有無すら判断できない。

相関関係の有無を判断するためには、少なくとも時系列的に変動するデータが必要である。図19-Aは時系列データが正の相関関係、図19-Bは負の相関関係、そして図19-Cは無相関にあるが、どの場合も安定期の姿だけをみれば全く同じであることに着目してもらいたい。

ところが何らかの事情で安定期のデータしか入手できないことがある。その場合、実際には無相関なのに正の相関関係があると思ってしまうことになる。これこそがクロスセクションデータと時系列データに齟齬のあることがわからない原因なのである。

図17 変数間の関係性を表した6つのパターン

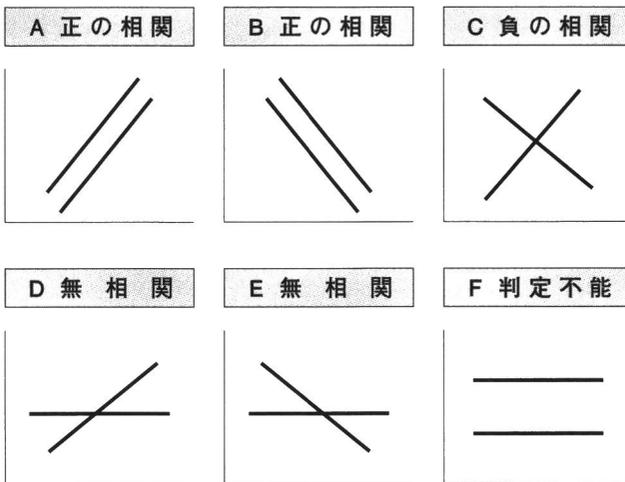


図18 散布図

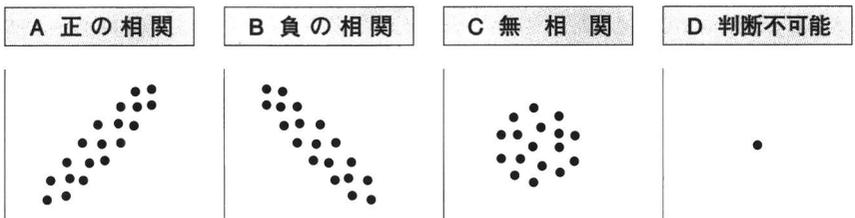
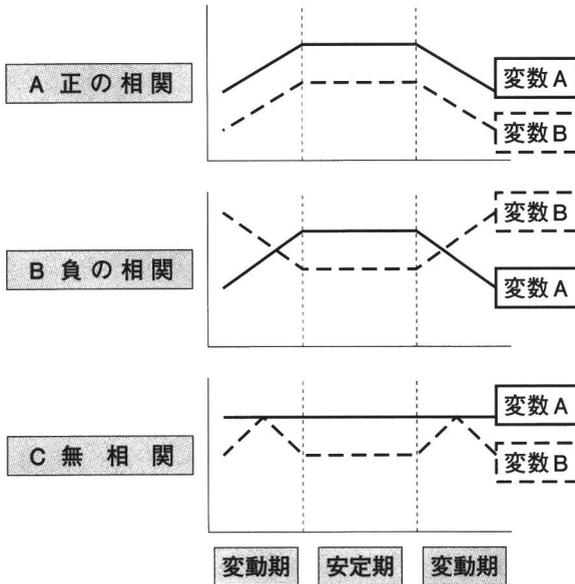


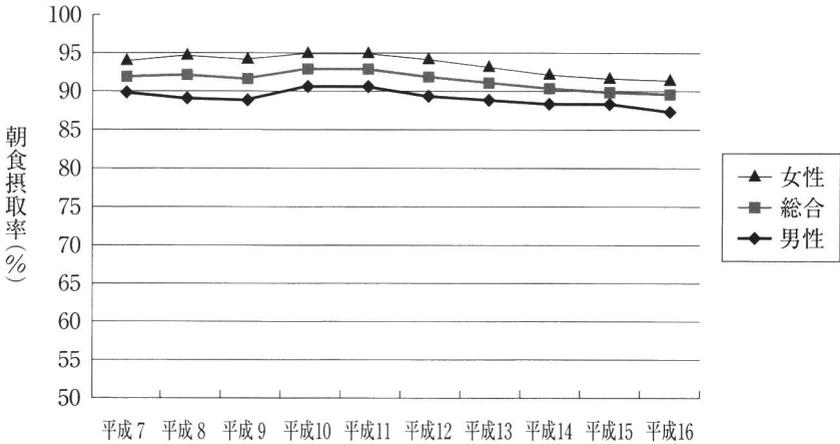
図19 安定期だけのデータだけでは判断不可能である



これを確認するために朝食と成績の話にもどらう。グラフ4は朝食摂取率の時系列データだが、きわめて安定していることがわかる。つまり安定期しか存在しないデータである。成績の時系列データもまた安定していれば、朝食摂取率と成績の本当の関係性がどうなっているかは不明なままである。関係性を正しく判断するためには、変動期のデータが欲しいところである。

その意味では成績の時系列データが変動していることは容易にわかることである。しかし朝食を推進したい人々は、朝食と成績が無相関であるという事実には意図的に目を向けないのである。

グラフ4 朝食摂取率の時系列データ



出典 厚生労働省

開業率と廃業率も変動期を持たないデータである

安定期のデータしか存在しないので、クロスセクションデータを根拠にした仮説化が一般的になっている研究分野はいたるところにある。グラフ5の開業率、廃業率もそのひとつである⁹⁾。

それによれば開業率は「人口や所得の増加率など地域における需要の伸びを示す指標、また人口規模、所得水準等などの需要の水準を表す指標は新規開業を促進することが実証的に裏付けられている」¹⁰⁾(岡室、小林)ということである。要するに、多くの地域別クロスセクションデータでは、需要の伸び率と、

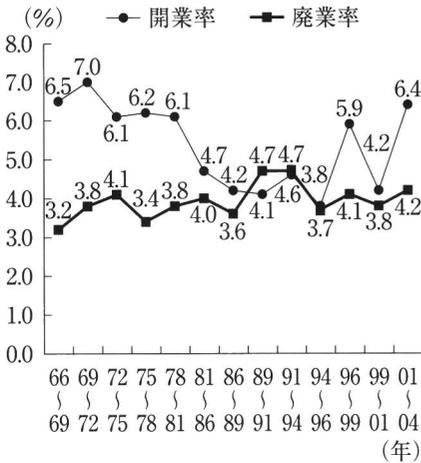
9) 地域別クロスセクションデータを使って需要要因、費用要因、人的資本要因、資金調達要因、産業集積・構造要因、およびその他の要因（企業規模構造、交通アクセス、公共サービス）などの説明変数がどの程度の強さで開業率を決定するかを検証したものである。

10) 「地域データによる開業率の決定要因分析」RIETI Discussion Paper Series 05-J-014

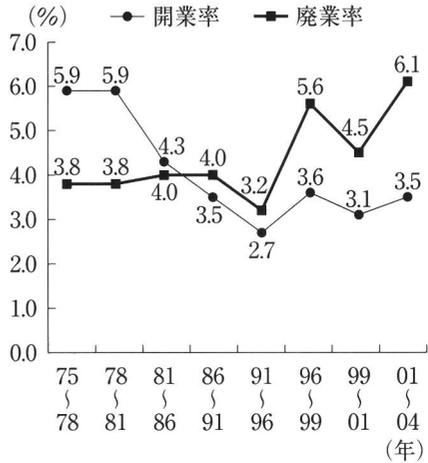
開業率に相関関係のあるという検証結果が出てくるというのである。

グラフ 5 我が国の開業率と廃業率の時系列データ

① 事業所数ベース



② 企業数ベース



資料：総務省「事業所・企業統計調査」

- (注) 1. 図①については、事業所を対象としており、支所や工場の開設・閉鎖、移転による開設・閉鎖を含む。
 2. 1991年までは「事業所統計調査」、1989年及び1994年は「事業所名簿整備調査」として行われた。
 3. 開業率・廃業率の計算方法については、付属統計資料の4表を参照。

しかし筆者が調べたところ、需要の伸び率と、開業率は無相関であることがわかっている。第2節で詳しく説明するが、デフレが続く我が国では90年代以降の市場成長率は一貫して右肩下がりになっている（グラフ7参照）。

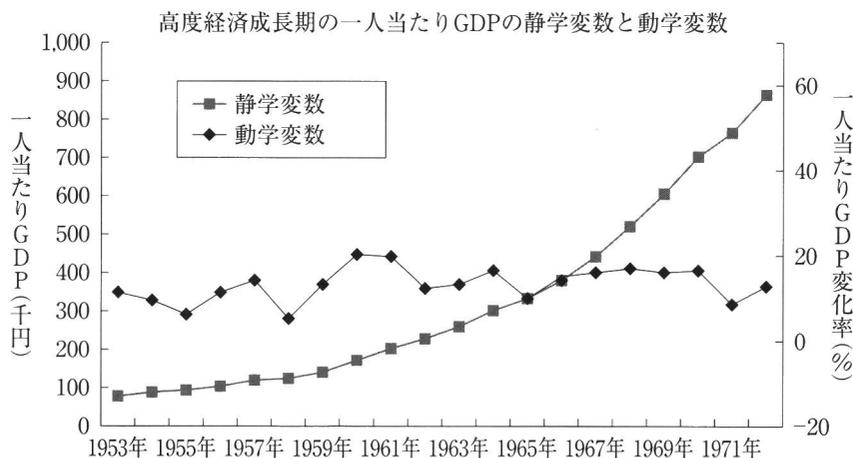
しかし同時期の開業率は安定的に推移しており、開業率と市場成長率に正の相関関係は無相関である。つまり地域別クロスセクションデータは瑣末な変動を反映しているだけであり、現実の現象を説明できていないのである。

動学変数もまた安定期のデータ

しかし安定期のデータしか手に入らないという意味で最も一般的に使われている変数といえば、何ととっても動学変数である。市場成長率が低下しているのは90年代以降に限る話であって、我が国でも長期にわたって動学変数は安定して推移していたからである。

我が国だけでなく一般的に経済指標の多くは右肩上がりに推移する性格を持っているが、それらの変数の変化率をとれば安定的に推移することになる。高度経済成長期の我が国のGDPの推移は右肩上がり（グラフ6）だが、GDP変化率は安定したデータになる。

グラフ6 高度経済成長期における我が国のGDPとGDP成長率の推移



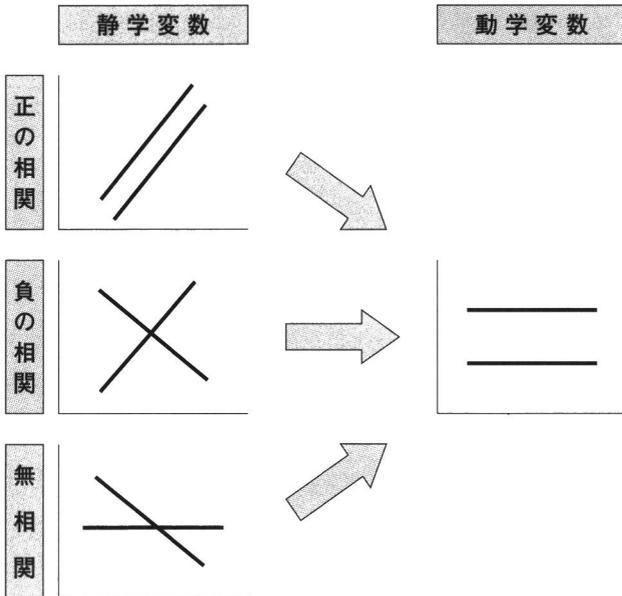
出典 内閣府

動学化すれば相関関係の判断ができなくなる現象は、単に右肩上がりの変数どうしで生じるだけではない。図20を見てもらいたいだが、右肩下がりどうしの変数でも動学化するだけで判定不可能になるのはもちろん、右肩上がりの変数

と右肩下がりの変数でも動学化すれば判定不可能になる。さらには右肩上がり
と水平の変数、右肩下がりの変数と水平の変数でも同じである。このメカニズ
ムこそ、多くの分野で動学比較分析が多用されている原因なのである。

変化率が変化しない限りこのメカニズムは働く。しかし経済指標の変化率が
変化するというのは、歴史的にはきわめてレアなケースである。1930年代のア
メリカ、90年代の我が国などがその事例であるといえ、いかにレアであるか
がわかるだろう。次に実際の事例をつかってこのメカニズムが働いていること
を示したい。

図20 動学化するだけで相関関係が強くなるように見える



動学比較分析の代表的なものといえばなんといっても成長仮説¹¹⁾だが、本稿では筆者の専門領域である流通研究の事例を使って説明したい。

流通研究には二つの動学研究がある。市場スラック仮説¹²⁾は「小売店数変化率」と「市場成長率」、そして自己雇用モデル¹³⁾は「就業者数変化率」と「雇用弾力性」、ともにクロスセクションデータに見られる強い相関関係を利用した仮説である。

グラフ7、グラフ8はこれらの4つの変数の時系列データだが、60年代から70年代のところを見てほしい。すべて安定して推移している。雇用弾力性は分子に就業者数変化率、分母に市場成長率を持つ変数であるため、それぞれの動学変数が安定して推移すれば雇用弾力性もまた安定して推移する。したがって、市場成長率、小売店数変化率、就業者数変化率、雇用弾力性、そしてここには書かれていないが「小売店数弾力性」を含めて、どの変数を組み合わせても生業店仮説を作ることができたのである。

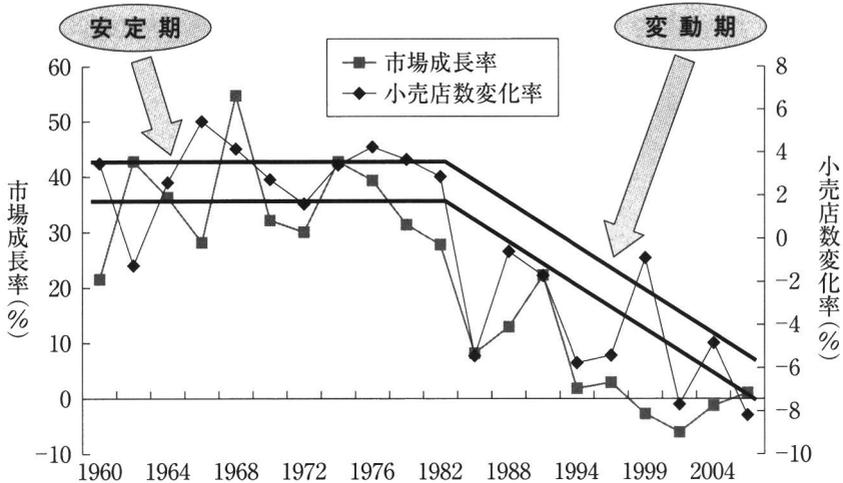
これらの変数間に本当に相関関係があるのかどうかを確認するためには変動期の時系列データを見る必要があるが、我が国の流通構造がどのように出来上がったのかを見るためには、高度経済成長期「以後」の変動期よりも、高度経済成長期「以前」の変動期のデータが必要なのである。なぜなら我が国の流通構造を決定したのは「以後」の変数間の関係からわかるのではなく、「以前」の変数間の関係性からわかるからである。

11) 1980年代後半以降、ハーバード大学のバロー教授とコロンビア大学のサラ・イ・マーティン教授らの成長回帰に基づく研究であるが、その最大の特徴は経済成長率という動学変数を使うものである。国別クロスセクション分析を使えば、ある程度の民主化が進んだ国を除いて、経済成長と民主化に何らかの相関関係のあることを発見したとされる。しかし実際にGDP成長率と民主化に相関関係があるかといえば、答えは明らかである。経済発展ばかり進んでも民主化が一向に進まない国は容易に挙げられることからわかる。それでも時系列データと比較できなければ、容易に仮説を却下できないのである。

12) 田村正紀、(1986)、『日本型市場システム』、千倉書房

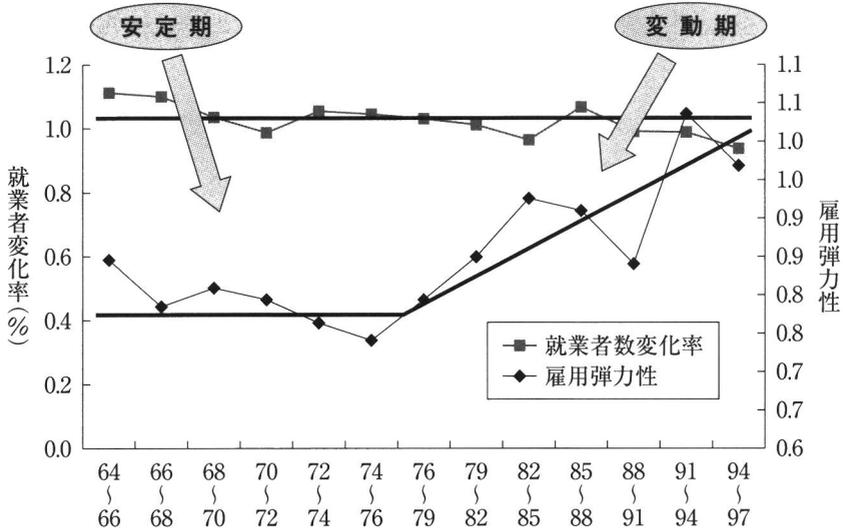
13) 風呂勉、(1960)、「商業における過剰就業と雇用需要の特定」神戸商科大学論集 通巻第37-39号

グラフ7 小売店数変化率と市場成長率の時系列データ



出典 商業統計

グラフ8 就業者数変化率と雇用弾力性の時系列データ

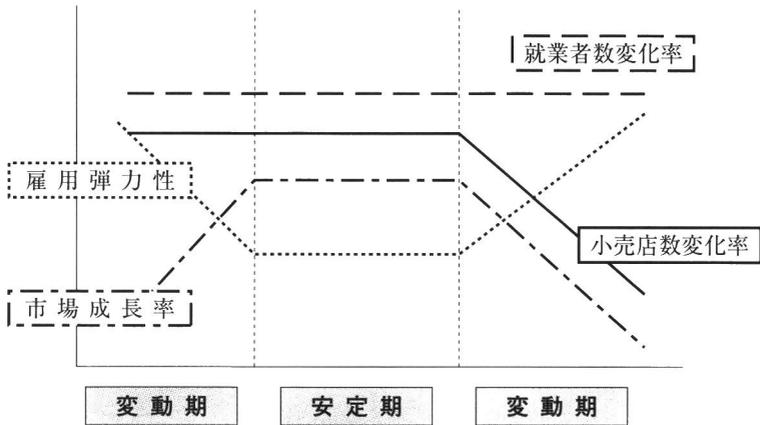


出典 商業統計

高度経済成長期以前のデータを見れば無相関

我が国の小売店数変化率や就業者数変化率の高いことが知られているのは、既に戦前からである。そのために「好況でも小売店が増える、不況でも小売店が増える」と言われてきたのである。そこで戦前から高度経済成長期初期にかけて市場成長率だけが右肩上がりに高くなり、一方で小売店数変化率や就業者数変化率は安定して高い水準で推移していたことを前提にして作ったものが図21である。

図21 小売店数変化率、就業者数変化率、市場成長率、雇用弾力性の時系列データ



小売店数変化率や就業者数変化率が安定しているのに市場成長率だけが右肩上がりであれば、弾力性変数は右肩下がりになる。この場合、就業者数変化率や小売店数変化率のような流通構造を表す変数は、市場成長率とも弾力性変数とも無相関である。したがって流通構造は必ずしも市場成長率や弾力性変数によって単純に説明できる変数ではないことがわかる¹⁴⁾。

14) そもそも変数間の関係性は数字で表現できない潜在変数の影響を受けるから、我が国の流通構造は経済的な要因によって作られただけでなく、社会的な産物である可能性も否定できない。

クロスセクションデータは隠れた変動を浮き彫りにするメカニズムを持っていることは既に指摘したとおりである。つまり流通構造は確かに市場成長率や雇用弾力性と何らかの関係性を持っていることを明らかにしている。しかし少なくとも高度経済成長期までの変化に関する限り、流通構造をそれらの変動によって説明することは難しいのである。

まとめ

時系列データを使って相関関係を検証すれば、私たちの身の回りの変数の多くは無相関である。これは筆者の経験から出てきた率直な感想である。おそらく社会の諸現象は決して単独の変動によって決定されるわけではなく、様々な変動によって複合的に決定されていることが多いからだと思う。これでは単独の変数を使った仮説化は不可能である。

それに比べると、クロスセクションデータでは何らかの相関関係のあることを比較的容易に観察することができる。どの研究者も同じ様な検証結果を得られれば、研究の蓄積が可能になる。蓄積ができれば研究の体系が出来上がる。つまりクロスセクションデータは研究者好みの方法論なのである。

もちろんその最大の弱点は実際の変数間の関係性を表すことができないことである。したがってこの点を指摘されれば、クロスセクションデータを使った検証方法はいつでも崩壊せざるをえない方法論なのである。つまりクロスセクションデータが一般的に使われているということは、なぜクロスセクション分析と時系列データには整合性がないのかを検証されないのか、ということであった。

クロスセクション分析はきわめて広く使われている方法なので、その方法論に対して直球的な問題意識を持つことはかなり勇気が必要であった。しかし安定期の時系列データだけでは相関関係の有無を判断することができないという、

きわめて基本的なことに気付いたとたん、本稿を上梓する決心をすることができた。筆者は流通分野の研究者に過ぎないが、本稿は広く社会科学に対する一般的な問題提議として、多くの分野の研究者に読んでいただきたいと考えている。

以上。

参考資料

- 岡室博之、小林伸生、(2005年3月)「地域データによる開業率の決定要因分析」
独立行政法人経済産業研究所 RIETI Discussion Paper Series 05-J-014
- 田村正紀、(1986)、『日本型市場システム』、千倉書房
- 風呂勉、(1960)、「商業における過剰就業と雇用需要の特定」神戸商科大学論
集 通巻第37-39号

アイスクリーム情報ネット 2002年

京都市教育委員会サイト

商業統計調査

総務省「事業所・企業統計調査」

テルモ体温研究所サイト

内閣府サイト

文部科学省サイト