

時系列データにおける補間方法の分析と考察

辰巳憲一・松葉育雄*

1 はじめに

補間 (interpolation) とは、何らかの理由で存在しないデータを、存在するデータを使って何らかの方法で補って、あたかも存在していたかのようにすることである。

そんなことを敢えてする必要はない、というのが経済学における一部の伝統的な考えである。経済学は、実際の取引行動に注目しており、カレンダー上の経過時間には関心がないとする、見解がとられることが多い。欠損データはやむをえず補間することは認めるが、継続的、体系的に存在するデータ無し状態は補間するまでもない、という考えである。

データが存在しない理由には経済的理由と非経済的理由があるが、極端な考えをとる経済学者の場合、データが存在しない理由を問うことはない、と問答無用である。これは重要な経済分析のいくつかを放棄している、と言わざるをえない。これはまた、本文で後述のように、経過時間の大きさを無視している。

本稿では、このような様々な問題を経済学分野向けにまとめ、今後の研究方向を示唆しておこう。

2 補間の背景

既知の数値データを基にして、そのデータの範囲の内側で予想される数値を求めること、またその手法を補間 (法) という。内挿 (法) とも呼ぶ。なお、補間は、既知のデータを基にしてそのデータの範囲の内側の数値を予測する (内挿) だけでなく、データの範囲の外側の数値を予測する外挿補間 (あるいは単に外挿とも呼ぶ) も行われる。

2-1 データが体系的に存在しない例

補間が必要になるのは、データが体系的に存在しないからである。システム障害や取引事故が発生し、取引所が取引を停止するという誰が考えてもやむをえない事由もある。これら以外の例をあげてみよう。

(1) CMS

今日から丁度1年後、2年後、3年後、…、n年後に満期になる債券でしかも今日価格が付いている債券は、すべて揃って存在しないのが普通である。多種多量の債券が発行されている米国でも、そうであ

*) 前者は学習院大学経済学部教授。後者は千葉大学大学院教授。内容などの連絡先: 〒171-8588 豊島区目白1-5-1 学習院大学経済学部辰巳憲一、TEL(DI):03-5992-4382、Fax:03-5992-1007、E-mail: Kenichi.Tatsumi@gakushuin.ac.jp

る。例えば365日(=1年)に対して数日足りない、数日分日数が多すぎるとというのが普通であろう。本来理論的に望まれるデータは割引債であるので、揃わない可能性はさらに、高まる。

そこで、債券(の市場)価格を内挿して、仮想の1年後、2年後、3年後、…、 n 年後満期債の価格を推定する。これが、CMS(constant maturity securities)と名付けられた仮想債券の価格(金利)体系である。

米国においては前世紀末から、取引所で、さまざまな先期日から先行き3カ月期間の資金貸借のレートが値付けされるようになり、CMSの役割は低下した。しかしながら、十分な取引量がない(特に10年以上期間)場合もあり、一部の価格・レートの信頼度は高くない。

(2) 非同期取引

価格は平均的には何分(minutes)かに一度成立する。それゆえ、分析する時間を細かく採るにつれて、ある特定時点に取引(約定)がない、それゆえ価格が(付か)ないことが頻繁に起こる。これが非同期取引(asynchronicity)問題である。リターンの計算には経過時間の長さを考慮すれば特に問題は起こらないが、他銘柄リターンとの単純な相関係数は計算できなくなる。

「株式リターンの相関係数計測値は、計測期間を細かくとる程その絶対値は小さくなる。」という事実を発見したEpps(1979)が取り扱ったのは10分から3日までの高頻度データであった。いわゆるEpps効果は、補間していないデータに起こる問題を浮き彫りにしている、少なくとも非同期取引問題が引き起こしている、といえる。

単純な分散、標準偏差の計算においてさえも、経過時間がばらばらで疑問を感じる研究者はいる。少なくとも等間隔データによる分散、標準偏差とは違った解釈をしなければならない。推計方法も変えてしかるべきである。

(3) 土日休祭日

取引所、広く市場は土日休祭日には休む。休んでいる間も、様々な出来事が起こる。

休んでいる間、時差などの関係で、他国では市場が開いている場合がある。情報化社会で、しかもグローバルな時代の昨今では、これが様々な問題と課題を我々に突きつけている。

(4) 休みと夜間

典型的には、1日が朝9:00から始まり11:30まで取引が行われ、昼休みを挟み、12:30から15:00まで取引が行われる(米国など海外では昼休みのない取引所がある)。この取引時間中、需給がある限り価格が付けられる。しかし、昼休みに1時間、日の終わり15:00から翌日9:00まで18時間、の間価格付けは基本的に休みになる。

これにさらに連休が入り込むと18の倍数時間の非常に長い間、値付けは休みになる。

これらの休み期間や夜間の間にも、価格を変化させるような情報が絶え間なく流れ込んでくる。場合によってはショッキングな大ニュースも入ってくる。

しかしながら、取引がなされない夜間でも、夜間取引市場が我が国でも導入開設されつつあり、取引は眠らなくなりつつある。夜間取引は海外の取引所をカバーする機能がある。

(5) データ収集間隔のずれ

年一度や5年に一度集計されているデータを月次の分析に用いる場合、補間する(か、それとも分析をやめる)しかない。何らかの方法で月間補間補正数を出して毎月足していくしかない。月間補間補正数としては、例えば、月次で観測される別のデータを参考にして、年間の変化を12カ月や60カ月に按分して計算される。

ちなみに、このような、データ収集間隔のずれに基づく補間はやむをえない、という理解がなされる。

2-2 補間が必要な理由

(1) 補間の経済学的考えとそれに対する反論

株価がなぜ日々変動するかといえば、企業価値に影響を与えるマクロ・ミクロの新しい情報（「ニュース」）が得られることで、投資家が売買するからである。しかし、制度的な理由から土、日曜日にはこのようなニュースの発表、公表はきわめて稀である。それゆえ、経済学的には土、日曜日が存在しない、と考えるほうが適切なのではないか、という考えもある。

一見、物理的な時間の等間隔（equidistance）が大切なのか、経済活動としての等間隔が重要なのか、という問題であるように見える。このような経済学的な考えに基づくと、等間隔の観察頻度にするために、金曜日から月曜日までのリターンを3分割（interpolate）して、それぞれ土曜日リターン、日曜日リターン、（真の）月曜日リターンにするのは、適切ではない、ということになる。

このような考え方は、土日休祭日、さらには昼休みは取引の単なる一時中断であると解釈しているようである。取組の一時中断といえば、大相撲の水入りのようなものを想定されているようだ。

現実の世界は大相撲の水入りとはまったく違う。大相撲の水入りでは、取組を中断させるに際して、行司は、力士双方の足の位置、組み方などをよく観察した上で、両力士に短時間の休憩を促す。再開時にはそれらを入念に水入り前と同じにした後、行司の合図により取組再開となる。現実の経済には、このような役割を果たす行司は存在しない。

水入りの場合でも休止時間での疲労回復力の差、心理状態の変化などで水入りはどちらかに有利に働く。水入りの前と後でまったく同じではないので試合は連続的ではないといえる。水入りを無視して取り組みの結果（試合の結末）を考えると予想を誤る、と思う。

土、日曜日、祭日においては生産や経営企画などの活動は停滞しているが、販売などはむしろ活発で、経済活動の無視できない比率は土、日曜日、祭日に行われている。しかも、企業ニュース（特に、企業にとって悪いニュースは土、日曜日に発表されるという研究報告がかつてあった）は土、日曜日にも発

1) 例えば、5年ごとの国勢調査結果による人口データを月次分析に用いる場合、月次データに変換する補間の算出方法は、当該月の補正後推計人口=前月の補正後推計人口+前月の異動人口+補間補正数、となる。ここで、「前月の補正後推計人口」は前回国勢調査による人口を出発点にする。異動人口（前月1日～末）にはデータが存在し、異動人口=自然動態+社会動態+国籍異動（日本人人口のみ）、となる。補間補正数は前回国勢調査人口を基準として算出した5年後現在推計人口と今回国勢調査人口との差を、各月に均等配分して算出する。例えば、つぎのようになる。

$$\text{各月の補間補正数} = \frac{1}{60} \times (P^{(12)} - P_{12}^{(7)})$$

$P^{(12)}$: 平成12年国勢調査による人口

$P_{12}^{(7)}$: 平成7年国勢調査の人口を基準として算出した平成12年10月1日現在の推計人口

$$= \left[\begin{array}{l} \text{平成7年国勢調査} \\ \text{による人口} \end{array} \right] + \left[\begin{array}{l} \text{平成7年10月～平成12年9月} \\ \text{の異動人口} \end{array} \right]$$

表されるという観察がファイナンス分野では従来から常識化している。

土、日曜日、祭日の新聞やネットには、事実記述型の記事ではなく、要約、展望的な評論記事が増え、読者の考え方・見方に影響し、むしろ行動を誘発する。

(2) 24時間取引の効果

現代の経済、特にグローバルな市場は、周知のように眠らない、のである。24時間世界のどこかで取引が可能なので、日本の午前9時が一日の最初ではない、午後3時が一日の最後ではない。また、月曜日9時が週の最初ではない、金曜日3時が週の最後ではない。

それでは、24時間取引によって先物の在庫調達機能はどう変わるのか。先物市場の役割は強化されるようになった、と考えられる。

その結果、市場リターンや取引量が一日の間にU字型に推移するという日中U字型効果はあいまい化（さらには消滅）していると予想される。しかしながら、この事実は前小節の曜日効果²⁾が消滅する理由にはならない。

(3) 日中高頻度データの課題

ティック毎、2分毎、15分毎の価格、金利、売買高などの典型的な高頻度金融データが使えるようになり、研究者がそれを積極的に使いだし、補間の必要性が増した。

データが高頻度になればなるほど、2つの時系列データの両方あるいは他方にデータが無い（これを非同期観測データ問題という）場合が生じ、相関係数、回帰分析などを用いて変数間の関係を調べる時にはデータの無いところは補間して（これを同期化という）、補うのがひとつの方法になっている。直前の値を使う（previous value interpolation）、線形補間、などがなされる。

補間をせずに、統計量を推定する方法を考える方向にも研究は変化している。この方向への発展を展望することは本稿の目的ではないので、この点はこれ以上触れない。

2-3 補間以外のいくつかの解決策～データや変数の削除

日中高頻度データ分析においては、朝あるいは昼休み明け第一番に、付く価格、なされる取引を、分析から省くという苦肉のデータ処理がなされる。時間の経過は非連続であるがリターンの大きさを等間隔で測り、夜の間あるいは昼休みの間に起こり蓄積した情報の効果を除くためである。

これを正当化するためには、休み中に入ってきた情報の価値は休み明け後の取引によって価値は実現し（続け）、情報自体が内蔵する価値は逡減していくという仮定をおく必要がある。

情報価値逡減のスピードによっては、休み明け後第一番だけでなく、第二、三番などいくつかを除く必要もあるが、いくつ除くべきかは情報カスケード（information cascade）理論などを用いて更に分析する必要があり、現在の研究段階では詳細はわからない。

計量経済学では、変数（データ）削除法（omitted variables method）は、真のモデルと比較して、統計学的性能が劣る、ことを明らかにしている。それゆえ、このような解決策は理論的に支持されない。

2) 曜日効果については、リターン平均という観点以外にも、高次モーメントの標準偏差、歪度、尖度の曜日効果も考えられるが、経済的意味が異なるはずである。しかしながら、歪度と尖度だけでなく、標準偏差の経済的意味についても経済分析はない。4次モーメントまでの曜日別単純統計量と関係付ける研究は数理的にも、また経済的にもまだ提供されていない。

2.4 補間することのメリット

上の2.2小節の理由から補間がなされる。補間することのメリットは次のように要約できる。

(1) 従来から使われてきた統計学、計量経済学の手法が好ましい特性を持つのは、等間隔データに基づいている。これらの特性を維持できるために、補間が適用される。

(2) 欠損などによって存在しないデータの補間をしなければ、経済的に意味の無い周期性を生み、予測精度が悪くなるという、大きな問題が生じることを示せる。

もちろん、意味の無い不適切な補間によっても、意味の無い周期性を生み、予測精度が悪くなる。

(3) 結論を得るために必須ではないが、統計学的検定が有意にでてくるなど、結論を明瞭にするためデータの補間を行うのが、曜日効果を検証した宮野・辰巳 (2004) や Miyano-Tatsumi (2006) の一つの特徴になっている。

3 様々な補間方法

代表的な補間方法、補関数には次がある。

3.1 従来からある補間方法

(1) 近傍補間

近傍補間の例としては最近傍補間がもっとも知られている。最近傍補間は0次補間、最近傍点補間ともいう。直前の値を使う方法が代表的である。

その他の近傍補間の例としては、「ある日の欠損データについて、過去30の日次リターンから最高値1個（あるいは2個）と最低値1個（あるいは2個）を除外した28個（あるいは26個）の単純平均値に代える」、実務で用いられる方法がある。

(2) 線形補間

線形補間は直線補間、1次補間ともいう。

(3) キュービック補間

キュービック補間³⁾は、3次補間ともいう高次補間である。

(4) ランダム補間

ランダム補間はブラウンのブリッジ (Brownian bridge) ともいい、文字どおりランダムに補間する。

(5) その他

ラグランジュ補間とスプライン補間⁴⁾、Sinc関数、Lanczos (ランツォシユ) 補間⁵⁾、などが知られている⁶⁾。

3) 字義的には3次補間という意味であるが、キュービックコンボリューションという言葉が使われ、下記の補関数を用いる3次補間を指すことが多い。 a は補関数の性質を制御するための変数で $-0.5 \sim -2$ 程度の数値が用いられる。

$$h(t) = \begin{cases} (a+2)|t|^3 - (a+3)|t|^2 + 1, & 0 \leq |t| < 1 \\ a|t|^3 - 5a|t|^2 + 8a|t| - 4a, & 1 \leq |t| < 2 \\ 0, & 2 \leq |t| \end{cases}$$

3-2 新しい補間方法

比較的新しい補間法について、詳しく説明しておこう。

(1) フラクタルとは

時系列過程がフラクタルな動きをするかどうかは、ある統計量（自己相関関数、分布 $p(x)$ の原点回帰率 $p(0)$ など）が、時間、あるいはデータのサンプル間隔に対して、べき関数になるかどうかで決まる（松葉（2007）ほか）。

月次、週次、などのデータに観測される（2、3期先までの）変動パターンは、そのパラメータの大きさには違いがありえるが、日次データにも同じように観察されるというのがフラクタル特性である。このようにスケールを変更しても同じような変動パターンが現れる特徴をもつ時系列は、自己相関関数が指数関数的に減衰する時系列と異なり、べき関数で表される統計量をもつ。

(2) フラクタル補間とその妥当性

フラクタル補間は、時系列データが持つ時間的な相関性を維持して補間する、という考えに基づく補間方法の1つである。

フラクタルは金融データでは、1、2秒、長くてもせいぜい数時間以内に実現する株価などの価格の分布であるという見方もある。数日、あるいは数カ月の時間範囲での価格等変化がフラクタルになるといえる、のかどうか。それはどう考えればよいのだろうか。

フラクタル性は、自己相関関数、分布 $p(x)$ の原点回帰率 $p(0)$ 、などが、時間、あるいはデータのサンプル間隔に対して、べき関数になるかどうかで決まる。この時、そのべき依存性が成立する範囲（スケーリング領域）があり、対象によってその範囲が数秒だったり、数年だったりする。ただし、数値的には微妙で、べき関数を当てはめたときの誤差の評価によっては、その範囲が異なってくる。

4) ラグランジュ補間は、N次関数（N次多項式）を使って補間する方法の1つで、次のラグランジュ補間多項式（Lagrange's interpolating polynomial）

$$L(x) = \sum_{k=0}^N L_k(x) y_k$$

$$\text{ただし、} L_k(x) = \prod_{j=0, j \neq k}^N \frac{x - x_j}{x_k - x_j}$$

を用いる。補間にあたっては、値がわかっている既知の点 (x_i, y_i) を上式に代入して、 x の各係数値を求めれば補間関数が得られる。補間したいのが、 x がある値をとる時の y の値とすれば、その x の値を代入すれば、補間値 y が得られる。

次数が高ければ高いほど式が複雑になるため、与えられた全部の点を通る次数の高い式をラグランジュ補間法で求めるのは困難になる。それゆえ、ある数以上の既知の点 (x_i, y_i) が存在する場合ラグランジュ補間法はうまくいかない。

そこで、全ての点を用いるのではなく、ある区間毎に分けてその区間ごとに式を求めることで全ての点を通る式を構成することを考えるのが、区分的多項式を用いる方法であり、その代表例がスプライン法である。

5) Lanczos（ランツォシユ）補間は、 n を補関数の性質を制御するための変数として、

$$h(t) = \begin{cases} \sin c(t) \cdot \sin c\left(\frac{t}{n}\right), & |t| \leq n \\ 0, & n < |t| \end{cases}$$

で表わされる。 $n=2$ とした補関数は Lanczos-2、 $n=3$ とした補関数は Lanczos-3 と呼ばれる。

6) Missing Value Analysis のソフトではこれら以外の補間方法もとられることがあるが、ここでは省略。

フラクタル補間を日次データに適用することが可能かどうか、このような検証の後明らかになる。フラクタル補間をして曜日効果を検証した研究に千葉・松葉（2005）がある。

3-3 最適な補間方法

以上のように複数ある補間方法のうち、どれを選ぶべきだろうか。そもそも補間するのはデータがないからであるので、誰も真の値を、知らない。そこで、シミュレーションで判定するしかない。

まず、時系列データを設定し、等間隔の場合（土日曜日）を含めていくつかデータをそこから抜く。そこに様々な補間方法を適用し、当てはまりの精度を比較すればよい。

当初に設定する時系列データを様々なタイプの確率過程にすれば、この補間方法はどのようなタイプの確率過程に向くか、などの知見も得られるかもしれない。

4 残された課題

補間について、いくつか残された観点を解説しておこう。

4-1 補間データの高予測性とその回避方法について

補間されたデータを、補間が始まる1日前の日から見ると、補間作業では2日先あるいはさらに先の日のデータとその日（補間が始まる1日前）のデータを用いるから、先の将来の日のデータが何分の1かのウエイトで補間データに入り込み、補間データはいわゆる先見バイアス（look-ahead bias）を持つ。補間している日のデータは、先の日のデータを結果として使っているわけで、当然、高い予測精度を示す。

これを回避する方法としては、サンプル出発時点から休祭日前日までだけのサンプルを用いて補間係数値を計測し休祭日のデータを補間する方法が考えられる。サンプル期間における第二の休祭日については、出発点から当該休祭日前日まで（第一の休祭日は先に補間した値になる）だけのサンプルを用いて補間係数値を計測する。以下、同様。

1日先予測精度を調べる場合を例に説明しよう。たとえば1995年1月30日時点が休祭日であるとし、その補間データに対する予測精度を調べる際には、その前日（29日）までのサンプルを用いて30日を補間しておく。そして、29日までのデータで1日先予測をして30日の予測値をだす。その結果と補間された値（休祭日でないふつう日の場合は実データ値になる）と比較して予測誤差とする。ちなみに、1月31日予測精度に関しては、1月30日（補間値）までのサンプルで1日先予測をし、その結果と実データ値の差で予測誤差とすることになる。

以降サンプル終了日1日前まで同様に行い、予測誤差の総和は予測開始日（別途、設定する必要がある）から足し算する。

このような手数をかける方法は、合理的期待仮説（rational expectation hypothesis）を検証する実証研究で確立された。結果の概略が変わらなくても、このような分析を行った後でなくては、結論の妥当性は信頼されなくなるものと思われる。

4-2 あるべき補間のルール

(1) 補間係数の計測ルール

補間ルールのパラメータは、その日（時点）までに観測されたデータのみ用いて計測し、1日ごとに修正していくのが、作業量は極めて多くなるが、本来あるべき方式である。

(2) 変数間関係の活用

他の経済変数を持ち出し、それらが営業時間に当該変数に与える影響を計測し、その効果を非営業時間にも適用するのが一つの正当な解決策である。その候補としては、市場インデックスやマクロ変数がある。しかし、国内の市場インデックスでは同じ時間に値付けされていない。海外変数でも、2分や15分の間隔のマクロ変数データは存在しないのが普通である。存在する唯一の2分、15分データは海外市場インデックスである。

(3) ジャンプ

補間にはもう一つ大きな課題がある。それは時系列データには何かの理由でジャンプが生じていることである。ジャンプは、ふつう考えられる経済的要因以外で、非常に稀に起こる何かの複数の要因が関係している、と考えられる。その稀な要因は時系列データが本来持つ時間的な相関性（特に、経済活動が産み出す、かなり規則的な動き）を壊している可能性があり、しかも分布の fat tail を生じさせている。

そこで、データのジャンプを補正する必要がある。時系列データが持つ時間的な相関性を維持して補正することが望ましい。

①ジャンプの補正を済ませると、等間隔データを得るために、欠損値も同様な方法で補間するべきと考えられる。

②何がジャンプか、一々認定して補正していくというより、自動的に補正していく方法が適切である。しかしながら、すべての人を納得させられる適切なジャンプの定義はない。

ジャンプが起こっているような時系列データについては、標準偏差に代えて1期（ジャンプが起こり1期以内に終了する場合）ずれた自己相関係数を用いる方法が勧められる。場合によっては、その値に1期ラグ修正係数をかけて修正することも必要になる。

相関係数についても、同様な対処方法が考えられる。つまり、他方の変数を1期ずらすのである。この方法が妥当するのは、ジャンプは頻繁に起こらないから、異時点でリターンをかけ算すれば、ジャンプの効果は小さくなる、からである。

4.3 1つの新しい研究の方向

既述のように、データが存在しない理由には、経済的理由と非経済的理由がある。経済的理由のなかでは、データを形成する情報がない、という視点が重要である。

情報は、毎秒、毎分、毎時間流入してくる。すべての情報は投資家によって評価され処理される。そして需給となって市場へ出され、価格が形成される。そしてこれらのプロセスが投資家のメモリーに蓄積される。土日休祭日、昼休み、夜間に蓄積された情報は明けの午前のセッションの価格形成に影響する。それゆえ、もし情報流入プロセスの分布をモデル化でき計測できるならば、そしてその評価をモデル化でき計測できるならば、価格への影響度を計測できるであろう。

これは補間の情報流入仮説と呼べるであろう。理想的には、この仮説は補間係数値を提供してくれる。

5 まとめ

経済分析はデータを適切に補間する必要がある。確かに、補間することによって問題点がなくなるとは思えない。問題を起こさずに補間するのは、単に残された課題というより、新しい大きな課題が残されているというべきである。

参考文献

**）本稿では宮野・辰巳（2004）でリストされている参考文献を除いている。

Epps, T.W., "Comovement of Stock Prices in the Very Short Run," *JASA (Journal of American Statistical Association)*, 74 366, June 1979, pp.291-298.

松葉育雄「経済時系列データのフラクタル分析」T-port投資研究会、2007年7月20日。

千葉隆司・松葉育雄「経済時系列のフラクタル補間法」『信学技報』電子通信情報学会、2005年。

宮野尚哉・辰巳憲一「非線形時系列解析による株価の曜日効果分析」、『電子通信情報学会論文誌A』, Vol.J87, No.9, 2004、pp. 1226-1235。

Miyano, T. and Tatsumi, K., "The Days-of-the-Week Effect and LME Metal Market", Paper presented at APAD Conference in Busan, Korea, June 22, 2006.

辰巳憲一「コモディティ先物の価格形成～LME金属の非線形時系列分析」『証券アナリスト・ジャーナル』, Vol.45, No.11, 2007、pp. 18-29。