日本大学経済学部経済科学研究所研究会

【第213回】
2021年7月31日

2019～2020年度共同研究B成果報告

「数理・統計解析によるリスク資産の分析」

〈発表者〉

日本大学経済学部専任講師　　戸　塚　英　臣

日本大学経済学部准教授　　生　亀　清　貴

日本大学経済学部教授　　三　井　秀　俊

「ハミルトニアン・モンテカルロ法を用いた安定分布のベイズ推定」

日本大学経済学部　　戸塚　英臣

（資料2　左）そもそもハミルトニアン・モンテカルロ法とはどんな話か，安定分布がどんな分布なのか，話さなければいけないのかもしれません．皆さん，正規分布とかt分布とかγ分布とかいろいろな分布をご存知だと思うんですが，その中でも安定分布というのは，確率変数X_1，X_2，定数a_1，a_2，これの線型結合を考えた場合に，それがもとの確率分布に従う確率変数Xと定数a，bのこのような一次関数で表されるようなかたちの性質を持っている分布のことを言います．t分布だとか指数分布，幾何分布，このあたりは再生性を示さないので多分安定分布にはならないかと思いますが，正規分布とかγ分布とか，われわれが通常使っている確率分布はほとんど安定分布に属すると思います．

（資料2　右の図）この安定分布は特性関数をフーリエ変換というかフーリエ積分したもので表されておりまして，特性関数がちょっと複雑なこんなかたちをしております．大事なのは4つのパラメーターで特徴づけられる分布でありまして，その4つのパラメーターで特徴づけられる特性関数をフーリエ積分することによって得られるのが安定分布の確率密度関数になっております．この安定分布の確率密度関数はフーリエ積分で表されておりますので，一般的には閉じたかたちで表すことができない関数となっております．

4つのパラメーターはα，β，γ，δとありまして，αは形を特徴づけている特性指数と言われていて，これが一番大事なパラメーターです．下のグラフでαが0よりも大きくて2以下の値をとるパラメーターですが，赤がαが0.5の場合で，αの値を0.5から2に近づけていくと徐々に正規分布に近づいていきます．ですから，αの値が2に近ければこの分布は正規分布に近い．2から離れるに従って徐々にすそが厚い，尖った分布になっていくというパラメーターです．βは歪度を表す指数ですので，歪みを表すパラメーターになっています．γは尺度母数と言われていて，許容偏差（？）のようなものに対応します．δは位置母数ですので，平均値に対応する値になっていて，大事なのはαとβの値がどうなるか．αによって分布の形状が決まってきます．βによって，プラス側に歪んでいるのかマイナス側に歪んでいるのか，ということを表します．

この安定分布は特別な場合がありまして，α2，β0の場合がガウス分布，α1，β0の場合がローレンツ分布，α0.5の場合がLevy分布と言われる分布になっております．ただし，全てβが0の場合なのでかなり特殊な場合なんですけれども，この安定分布はいろいろな分布を内包している，広く一般的に使える分布になっています．

（資料3）この分布を使って何がされているか，ここからはぜひ皆さんに教えていただきたい部分ですが，これはNatureに載った論文で，アメリカのS＆P500の1分間隔のデータを用いて収益率の分布を調べたもので，ちょっと古いんですが，1995年に出ております．

ちょっと補正されているんですが，この分布の形がちょうど先ほどの安定分布がフィットして，α＝1.4という値が導かれるという話です．ところが，かなり疑問な部分がありまして，形に合わせている．つまり，本来ならば収益分布なので確率変数の分布なのにもかかわらず，最尤推定法だとか何かの推定法をしたのではなくて，ある変換則に対して不変になるということから，先ほどの安定分布でβ＝0の場合には中のスケール変換に対して不変だという性質が導かれていて，そのスケール変換に対して不変だということから，もともと1分間隔のデータなんですが，1分の場合，5分の場合，10分の場合としてサンプリングレートを変えてヒストグラムを描いていったときに，同じようなグラフになる．それと安定分布のスケール変換に対する不変性というこの2つのことからαの特性指数の値を推定しているという推定方法なんですが，「いやいや，ちょっと待って．それは単にパラメーターフィッティングするわけじゃないの？」というのが，そもそもぼくの疑問点です．

もともとは収益率分布なので関数変数の話なはずなのに，単にヒストグラムの形に関数の形をフィッティングさせてパラメーターを推定しているというやり方にしかぼくには思えない．これでちゃんと推定しましたという論文がnatureに出ているので，「えっ？　ほんとにこんなことやって

いいのかな」というのが，そもそもぼくの疑問点でした．

（資料4　左の図）　これは実は物理の世界ではかなり行なわれている手法で，先ほどのS＆P500を1分間隔，3分間隔，10分間隔，32分間隔と時間間隔を変えていってもーーこれはスケール変換しているので，同じグラフではないんですが，ある一定の同じ時間で変換をしているとちゃんと同じような分布に従うという結果ですので，同じようなかたちになっている．ここから α が大体1.4ぐらいになるだろうという推定をしている．

（資料4　右の図）　同じことを東京証券取引所の100銘柄についても行われています．ただし，こちらはアローヘッドが導入された後のデータを使っているので，一番短いのだと0コンマ何秒のデータがあるんですが，さすがにそこまでは使っていなくて，5秒間隔，15秒間隔，30秒間隔で，それぞれ時間を合わせてヒストグラムを描くと，同じようにちゃんと1つのグラフ上に載ってくる．同じように先ほどの安定分布の特性指数の値を推定すると大体1.4ぐらいになるだろうということで，S＆P500でも特定指数の値は1.4，東京証券取引所に上場されている100銘柄について全部調べてみても1.4ぐらいの値が出る，というのがこれまでの先行研究でした．

ただし，先ほども言いましたように，これでパラメーター推定と言ってよいのか．本来確率変数の分布であるはずなのに，分布関数の形で関数をフィッティングして，それってほんとにパラメーター推定したことになるのかなというのが，そもそもこの研究を始めるぼくのモチベーションというか疑問点でした．

統計を専門にやられている先生からすると，「物理ってこんないい加減なことしてるの？」と思われるかもしれないんですが，意外と物理は何かの変換に対して普遍というのが大好きで，今回も「時間スケールの変換に対してデータが普遍だ．そしたらそれを表す分布関数も普遍な関数だ」というユニバーサリティーという考え方が物理学者は大好きなので，物理に片足突っ込んでいる人間とすると，まあまあそれはそれで面白いなと思うんですけど，統計的な考え方からすると，これってまずいんじゃない？というのがかなりあります．

（資料5）　これはかなり高頻度のデータで1分間隔とか5秒間隔とかいう間隔ですが，これをもう少し時間間隔を延ばして日次のデータにしてみると，こんな感じです．この特徴とすると，首が細いNarrow-neck，Fat-tailで経済物理屋さんが大好きな形をしているんですが，これが日経平均株価の日次収益率，こちらがS＆P500の日次収益率で，2014年から2019年のデータです．日次のデータなので時間スケールとしてはかなり長いスケールですが，それでもやはり同じような特性を持っています．ですので，日次データでも高頻度データでも同様な特徴を示すのだなあということは分かるんですが，推定方法としてほんとにスケール変換から求めていいのというところがあったので，ではもうちょっとちゃんと推定してみましょう，というのが今回のモチベーションです．

（資料6）　推定方法は，先ほどのスケール変換でやっているんですが，それ以外に，エルゴード性といって時間平均と位相平均を等しく置いたような仮定を入れて導く方法だとか，ベイズ推定も行なわれています．先ほどの特性関数をポアソン級数展開して推定する補法や非線形ポピュレーションモンテかる法などいろいろなやり方がされているんですが，高頻度データをやろうとするとベイズ推定は難しいのであまりやられていない．日次だったらできるのかなあと思ったんですが，日次収益率のような高頻度とは言えない低頻度データを使って安定分布の推定というのはあまり行なわれていない．最終的には高頻度のデータをやりたいんですが，いきなり高頻度のデータを取り扱うといってもデータ量が莫大なのでフーリエ積分するときに時間がかなりかかると思いますので，まずはとっかかりとして，日次収益率ぐらいのデータ数で安定分布の分布がちゃんと求められるのか，そこをターゲットに研究をしてみようと思っています．

（資料7）　今回の研究の目的ですが，先行研究との比較として，株価指数の本当は高頻度データを使いたいんですが，データ数が莫大なのですぐはできない．日次収益率だと5年分で1200個ぐらいのデータですから，これだったら積分できそうなのでこれを用いて，ハミルトニアン・モンテカルロ法を使ってMCMCでベイズ推定を行なうということをやってみました．

その場合，特性指数が大体どれぐらいの値が出てくるのか調べたいというのがまず最初の目的です．もう1つ，せっかく安定分布のベイズ推定するのだから，それ以外に歪みがある正規分布だとか歪みがあるt分布なども同じハミルトニアン・モンテカルロ法を使って推定してみて，周辺尤度を使ってどの分布関数が一番当てはまりがよいのか調べてみようというのが，この研究の2つ目の目的です．

（資料8） 推定方法ですが，ハミルトニアン・モンテカルロ法はもともと物理学のほうで使われたMCMCの方法で，説明するとかなり時間がかかるので，やり方を簡単に言うと，運動方程式を立てて，運動方程式の中の運動量と位置という2つの量があって，位置の部分を推定したよいモデルのパラメーターにします．それに仮想的な運動量を仮定して，その2つから運動方程式をつくり，その運動方程式でパラメーターを変更していく．その後，メトロポリス法を使って，変更したパラメーターを採択するのか採択しないのか判定しながらパラメーター更新をしていく．基本はベイズ推定ですが，その推定するときの方法が運動方程式を使ってパラメーターを更新していく．更新されたパラメーターを採択するかどうかはメトロポリス法で判定する．物理屋さんはハミルトニアン・モンテカルロと言いますが，ほかにハイブリッド・モンテカルロと言われている方法も同じようなやり方です．

推定条件ですが，ベイズ推定ですので必ず事前分布があるんですが，あまり偏った分布を仮定したくなかったので，無情報事前分布にしてあります．バーインの回数は1000回ぐらい．確率標本のほうが1万回ぐらい発生させて，それをチェーン数3つぐらい走らせました．間幾つ抜くかですが，依存性が現れてしまうので2つに1回ぐらいデータを取っていますので，1万回サンプリングして，その中の半分ずつ取っていますので，1万回サンプリングしてその中から半分の大体5000回ぐらいサンプリングしたことになります．3つのチェーンがあって，5000回を3回ずつですので，5000個，5000個，5000個ぐらいのデータで判定しました．

後でお見せしますが，確率標本の時系列プロットによる目視による収束判定．それからGelman-Rubinの統計量と言われる量がありますので，これが1に近ければ収束しているだろうという判定基準になりますので，その2つを使ってMCMCのサンプリングが収束しているかどうか判定しました．

もう1つ，フーリエ積分しなければいけないので，ここが一番肝なんですが，二重指数関数型数値積分公式という京都大学の大村先生がつくられたフォーマリズムを使って，なるべく高速かつ高精度に積分できるようにしようという方法です．通常数値計算するとかなり重いので，このあたり少し計算時間を短縮してあげようというので，この方法を導入してあります．

（資料9） データですが，2015年1月から2019年12月で，コロナの時期は除いてあります．データ数は今回，日本の株価指数である日経平均株価とアメリカのS＆P500．先行研究で使われていた同じ銘柄になりますので，そのデータをターゲットにやってみた．これがそれぞれのヒストグラムになっています．

（資料10） 推定の結果ですが，4つのパラメーターがありましたので，それぞれ日経平均株価のαとβとγとδ，こちらがS＆P500です．見ていただくと，上が標本自己相関になっていまして，過去のデータに相関がないので，まあまあよいサンプリングができている気がいたします．真ん中が標本経路で，ちゃんとぐじゃぐじゃっとなっていますので，依存性が表われていない．これが上に行ったり下に行ったりじぐざぐになったりすると乱数の発生としてはよくないと思いますので，どちらも，どのパラメーターも，それなりによい推定がされているかなと思います．一番下が事後確率密度になっておりまして，それなりにちゃんと分布らしい形の分布になっていますので，安定分布のパラメーターとしてはちゃんと推定がうまくいっているのかなあという結果になっております．

（資料11） 各パラメーターの推定結果ですが，上が日経株価で，事後平均と標準偏差，95％信用区間，Gelman-Rubin統計量と，非効率性因子というのは何回に1回サンプリングすればよいかという話で，この値が小さければ小さいほど効率よくMCMCが動いているということになりますが，1桁ですので，この方法だと効率よくサンプリングができているという話になります．

αの値が日経平均株価では1.593，S＆P500でも1.531なので，1.4までは行かない．95％信用区間

を見ていただいても，日経平均株価は1.492から1.692，S＆P500も1.409から1.65で，ぎりぎり1.4に入るか入らないかぐらいの値で，まあこんなものなのかなという結果になっており，そこそこよい推定結果かなという気がいたします．

歪み度ですが，βを見ていただくと，日経平均株価は95％信用区間が0.025から0.444なのでプラス側ですので，日経平均株価は歪みがある分布だということが分かります．一方，S＆P500のほうは－0.079から0.308で，マイナスとプラスを含んでいる．どちらも0を中に含んでいる信用区間ですので歪みはないのかなということが，この推定結果から分かります．

（資料12） せっかくベイズ推定を用いた安定分布の推定ができたので，もう少しこの方法を安定分布だけではなくて歪みがある正規分布と歪みがあるt分布に拡張して，どの分布が一番，日経平均株価とかS＆P500のあてはまりがよいのか調べてみました．

（資料13） 歪んだ正規分布と歪んだt分布があるんですが，こちらが歪んだ正規分布の確率密度分布，後ろが累積密度分布になっています．中が，ωが標準偏差，δが平均に対応する量・期待値，λが歪みを表すパラメーターになっています．ちょっと意味合いが違ったので安定分布と同じ名前をつけなかったんですが，ζとωとλという3つのパラメーターになっています．歪んだt分布のほうは，ωとζ，こっちが確率密度分布，こっちが累積密度分布になっていまして，t分布ですのでこの中に自由度のνが入っているところが違います．このあたりはちょっと複雑な係数がかかっていて，歪んだ分布はいろいろあるんですが，今回はAzzaliniがつくった関数についてパラメーター推定をしてみました．

（資料14） これが日経平均株価の歪んだ正規分布と歪んだt分布の推定結果です．見ていただくと，歪んだ正規分布はあてはまりがあまりよくない．どちらかと言うと歪んだt分布のほうがあてはまりがよい結果なのでこちらだけ見ていただくと，歪んではおります．正規分布のほうは－0.620から－0.036が95％信用区間になっていますので，ちゃんと歪んでいます．自由度は3前後になっています．（資料15）S＆P500の場合には歪んでいないので，このあたりは安定分布と整合性がとれた結果になっています．

（資料16） 最後に，ではどの分布が一番あてはまりがよいかということで，周辺尤度という値をそれぞれ求めて，それを比較することで，この値が一番大きい値を持っている分布が当てはまりがよいという結論になります．日経平均株価とS＆P500で比較していただくと，こちらの表になっていまして，安定分布，歪んだ正規分布，歪んだt分布で，赤で示したんですが，日経平均株価もS＆P500もどちらも歪んだt分布が一番あてはまりがよいという結果になっています．

事後平均を使ってグラフを描いたものがこれで，ちょっと見づらくて申しわけないんですが，緑が歪んだt分布，赤が安定分布で，このあたりは形が似ているんですが，トップの部分が歪んだt分布のほうがなんとなくヒストグラムを表しているんだろうなという感じがします．こちらがS＆P500のほうですが，こちらもこのあたりの形は安定分布よりも歪んだt分布がいいのかな．ですので，歪んだt分布を推定するプログラムを頑張ってつくって，その後，歪んだt分布と歪んだ正規分布もつくってみたんですが，べつに歪んだt分布で十分なのかなというのが今回の研究の結果です．

（資料17） まとめますと，ハミルトニアン・モンテカルロ法を使ってMCMCの1種類の株価指数の日次収益率を用いた安定分布のベイズ推定を行なってみました．本研究の安定分布の推定結果が1.53と1.59という値なんですが，先行研究では1.4だったので，若干大きい値になるかな．ただ，高頻度データで1分間隔ぐらいのデータが本当はやりたいんですけれども，そこまでのデータではなかったので，そのあたりで差が出たのかなという気はいたします．

モデル選択ですが，歪んだ正規分布と歪んだt分布と安定分布で，それぞれ日経平均株価とS＆P500の推定をしてみたんですが，結果とすると歪んだt分布が一番あてはまりがよいという，まあ妥当かなという結果になりました．

今後の課題ですが，やってみたいのは高頻度データです．プログラムがりがりにすれば動かないこともないかなというのはあるんですけど，まあやってみたいなと思います．

簡単ではありますが，以上です．

ハミルトニアン・モンテカルロ法を用いた安定分布のベイズ推定

Bayesian Inference of Stable Distribution using Hamiltonian Monte Carlo method

日本大学経済学部
戸塚英臣　三井秀俊

College of Economics, Nihon University
Hideomi Totsuka and Hidetoshi Mitsui

日本大学経済学部経済科学研究所・共同研究B
「研究テーマ: 数理・統計解析によるリスク資産の分析」
(平成31年度) より助成を受けている.

資料1

安定分布

- 安定分布とは
  - 次の性質を持つ確率変数が従う分布

$$a_1X_1 + a_2X_2 \sim aX + b$$

- 安定分布の確率密度関数

$$f(x|\alpha,\beta,\gamma,\delta) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\psi(t|\alpha,\beta,\gamma,\delta)e^{-ixt}dt$$

$$\psi(t|\alpha,\beta,\gamma,\delta) = \exp\left[it\delta - |\gamma t|^{\alpha}\left(1 - i\beta\,\mathrm{sgn}(t)\Phi(t,\alpha,\gamma)\right)\right]$$

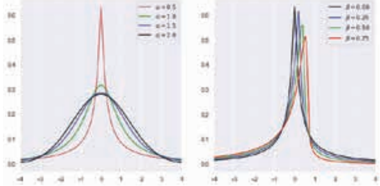
$$\Phi(t,\alpha,\gamma) = \begin{cases} \left(|\gamma t|^{1-\alpha} - 1\right)\tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ -\frac{2}{\pi}\log|\gamma t| & \alpha = 1 \end{cases}$$

  - 安定分布の確率密度関数は閉じた式で表すことができない.

- 安定分布のパラメータ

α:特性指数　γ:尺度母数
β:歪度指数　δ:位置母数

βを変えた場合



特別な場合
$\alpha = 2$:　ガウス分布
$\alpha = 1$:　ローレンツ分布
$\alpha = 0.5$: Levy分布

資料2

安定分布を用いた収益率の分析

- S&P500 の1分間隔の高頻度データを用いて，安定分布のスケール変換の性質を用いてパラメータの推定が行われた．

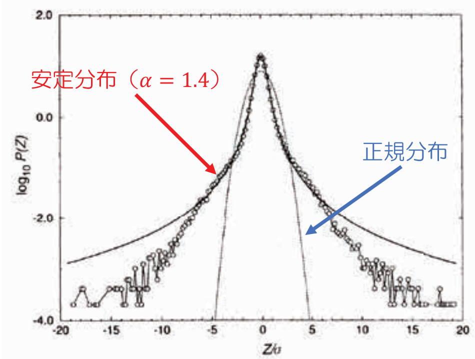
$$L_\alpha(Z,\Delta t)=\frac{1}{\pi}\int_0^\infty e^{-\gamma\Delta t q^\alpha}\cos(qZ)\,dk$$

$$L_\alpha(Z_s,1)=\frac{L_\alpha(Z,\Delta t)}{(\Delta t)^{-1/\alpha}}$$

$$Z_s=\frac{Z}{(\Delta t)^{1/\alpha}}$$

- 特性指数 $\alpha = 1.4$ の安定分布に従うことが報告されている
- 収益率$R(t)$は株価指数$S(t)$の変化率

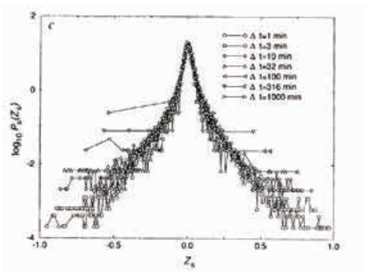
$$R(t)=\ln S(t)-\ln S(t-1)$$



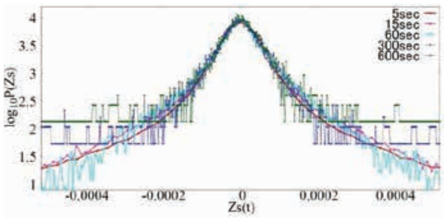
資料3

高頻度データによる収益率分布

- S&P500（分足）



- 東証100銘柄（5秒足）



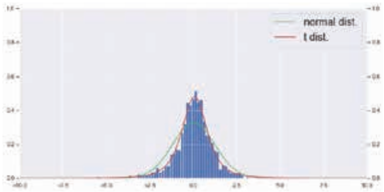
IEICE Technical Report, CCS2016-7(2016)

安定分布のスケール普遍性を用いて指数の推定が行われ，
特性指数 $\alpha = 1.4$ の 安定分布に従うことが報告されている． ➡ パラメータ推定と言えるの？

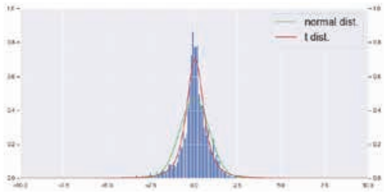
資料4

株価指数の日次収益率分布

- 日経平均株価



- S&P500の日次収益率



  - Narrow-neck
  - Fat-tail

対数最尤推定値

		t
日経平均株価	-1981.0	-1842.8
S&P500	-1575.0	-1455.6

資料5

その他の推定手法

[1] JPSJ, 89, (2020) pp.024802
[2] Digital Signal Processing, 47, (2015) pp. 96
[3] Computational Statics and Data Analysis, 95 (2016) pp.57

- エルゴード性を用いた推定[1]
  - 長時間平均が空間平均に等しいというエルゴード性を仮定し，十分な長さの確率変数列 $X_n (n = 1,2,\ldots,N)$に対して，特性関数$\varphi(t)$について次式が成り立つ．

$$\varphi(t) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} e^{itX_n}$$

  - 特性指数 $\hat{\alpha}$ は異なる時刻における特性関数を用いて次の様に推定される．

$$\hat{\alpha} = \frac{\log(-\log|\hat{\varphi}(t_1)|) - \log(-\log|\hat{\varphi}(t_2)|)}{\log t_1 - \log t_2}$$

- ベイズ推定
  - Poisson級数展開を用いたベイズ推定[2]
  - Nolinear population Monte Carlo 法[3]

資料6

研究の目的

- 先行研究との比較
  - 株価指数の日次収益率を用いて，ハミルトニアン・モンテカルロ法により，安定分布のベイズ推定を行う．特に特性指数(α)の値に注目する．

- モデルの選択
  - ハミルトニアン・モンテカルロ法を用いて，安定分布，歪正規分布，歪 t 分布の株価指数の日次収益率分布の当てはまりを調べる．

資料7

推定方法

- ハミルトニアン・モンテカルロ法[1]
  - 推定方法
    - パラメータの更新はハミルトン方程式で行う．
    - 更新されたパラメターの採択はメトロポリステストで採択・棄却を決める．
  - 推定条件
    - 無情報事前分布
    - 稼働検査期間：1000
    - 確率標本数：10000
    - チェーン数：3
    - 間伐期間の回数：2
    - 収束判定：
      - 確率標本の時系列プロットによる目視
      - Gelman-Rubin統計量

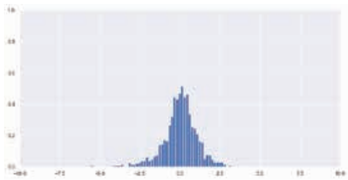
- 安定分布のフーリエ積分[2]
  - 二重指数関数型数値積分公式

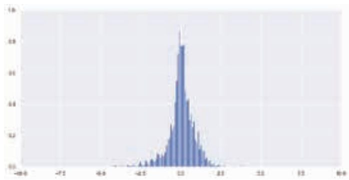
資料8

パラメータ推定に用いた株式のデータ

- 要約統計量
  - 期間：2015年１月～2019年12月

	観測個数	平均	標準偏差	歪度	超過尖度	最大値	最小値
日経平均株価	1221	0.02	1.23	-0.34	6.15	7.43	-8.25
S&P500	1257	0.04	0.85	-0.52	3.84	4.84	-4.18

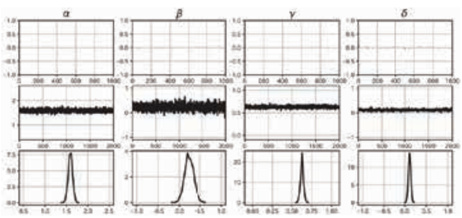




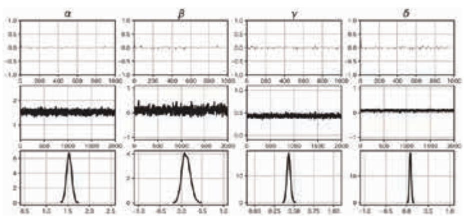
資料9

推定結果：
標本自己相関関数，標本経路，事後確率密度

- 日経平均株価



- S&P 500



資料10

安定分布のパラメータの推定結果

• 日経平均株価

	α（特性）	β（歪度）	γ（尺度）	δ（位置）
事後平均	1.593	0.228	0.627	0.113
（標準偏差）	(0.051)	(0.107)	(0.021)	(0.029)
95 % 信用区間	[1.492, 1.692]	[0.025, 0.444]	[0.585, 0.669]	[0.056, 0.171]
GR 統計量	1.00	1.00	1.00	1.00
非効率性因子	2.92	3.09	3.07	6.53

• S&P500

	α（特性）	β（歪度）	γ（尺度）	δ（位置）
事後平均	1.531	0.104	0.431	0.089
（標準偏差）	(0.062)	(0.099)	(0.023)	(0.022)
95 % 信用区間	[1.409, 1.653]	[-0.079, 0.308]	[0.387, 0.476]	[0.047, 0.135]
GR 統計量	1.00	1.00	1.00	1.00
非効率性因子	4.92	8.67	6.95	6.32

資料11

ベイズ推定を用いた分布関数の選択

• 分布関数の選択
  • 安定分布以下の分布関数についてベイズ推定を行い，どの分布関数の当てはまりが良いか調べる．

• 対象となる分布関数
  • 安定分布
  • 歪正規分布
  • 歪 t 分布

資料12

歪正規分布と歪 t 分布

- 歪正規分布

$$p(x|\zeta,\omega,\lambda) = \frac{2}{\omega}\phi\left(\frac{x-\zeta}{\omega}\right)\Phi\left(\frac{\lambda(x-\zeta)}{\omega}\right)$$

$\phi(\cdot)$は確率密度分布、$\Phi(\cdot)$は累積密度分布

- 歪 t 分布

$$p(x|\zeta,\omega,\lambda,\nu) = \frac{2}{\omega}t_\nu\left(\frac{x-\zeta}{\omega}\right)T_{\nu+1}\left(\frac{\lambda(x-\zeta)}{\omega}\sqrt{\frac{\nu+1}{\nu+\frac{(x-\zeta)^2}{\omega^2}}}\right)$$

$t_\nu(\cdot)$は確率密度分布、 $T_{\nu+1}$ $(\cdot)$は累積密度分布

資料13

日経平均株価による
歪正規分布と歪 t 分布のパラメータの推定結果

- 歪正規分布

	ζ（位置）	ω（尺度）	λ（歪み）
事後平均 （標準偏差）	0.909 (0.009)	1.517 (0.004)	-1.112 (0.025)
95 % 信用区間	[0.705, 1.075]	[1.390, 1.636]	[-1.382, -0.795]
GR 統計量	1.00	1.00	1.00
非効率性因子	2.63	2.22	4.09

- 歪 t 分布

	ζ（位置）	ω（尺度）	λ（歪み）	ν（自由度）
事後平均 （標準偏差）	0.288 (0.009)	0.798 (0.002)	-0.324 (0.022)	3.082 (0.122)
95 % 信用区間	[0.094, 0.480]	[0.725, 0.886]	[-0.620, -0.036]	[2.485, 3.856]
GR 統計量	1.00	1.00	1.00	1.00
非効率性因子	3.56	5.76	3.62	3.82

資料14

S&P500による
歪正規分布と歪 t 分布のパラメータの推定結果

- 歪正規分布

	ζ（位置）	ω（尺度）	λ（歪み）
事後平均	0.721	1.093	-1.331
（標準偏差）	(0.003)	(0.002)	(0.026)
95 % 信用区間	[0.610, 0.828]	[1.012, 1.179]	[-1.630, -1.044]
GR 統計量	1.00	1.00	1.00
非効率性因子	1.92	4.66	3.48

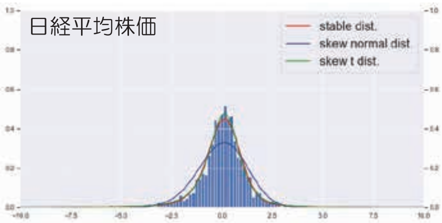
- 歪 t 分布

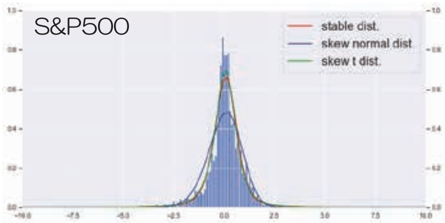
	ζ（位置）	ω（尺度）	λ（歪み）	ν（自由度）
事後平均	0.139	0.542	-0.154	2.807
（標準偏差）	(0.004)	(0.001)	(0.019)	(0.115)
95 % 信用区間	[0.012, 0.272]	[0.478, 0.593]	[-0.440, 0.116]	[2.236, 3.579]
GR 統計量	1.00	1.00	1.00	1.00
非効率性因子	5.14	3.92	5.06	4.69

資料15

推定結果のまとめ

- 日次収益率分布と各分布関数の比較





- 周辺尤度

$\theta^{(j)}$

$f\left(y|\theta^{(j)}\right)$：尤度関数
$\theta^{(j)}$：j番目のパラメータの推定値

対数周辺尤度の比較

	安定分布	歪正規分布	歪 t 分布
日経平均株価	-1851.3	-1972.6	-1842.3
S&P500	-1473.3	-1559.3	-1457.3

資料16

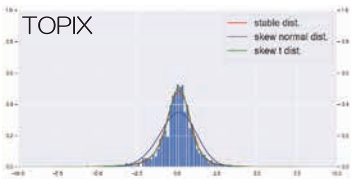
まとめと今後の課題

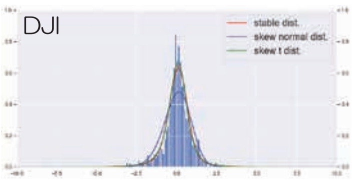
- まとめ
  - ハミルトニアン・モンテカルロ法による株価指数の日次収益率を用いた安定分布のベイズ推定を行った.
  - 先行研究と比較
    - 本研究の推定値 $\alpha = 1.53$ と 1.59 は，先行研究の値 $\alpha = 1.4$ よりも大きい.
    - 95%信用区間は，先行研究の値を含まない.
  - モデルの選択
    - 歪正規分布、歪 t 分布と対数周辺尤度を比較した.
    - すべての株価指数の収益率分布で歪 t 分布の当てはまりが良い.
- 今後の課題
  - 高頻度データを用いた推定
  - 時系列モデルの誤差関数への応用

資料17

ちなみに、TOPIXとDJIによる推定結果

- 日次収益率分布と各分布関数の比較





- 対数周辺尤度

			t
TOPIX	-1775.8	-1911.9	-1767.2
DJI	-1490.9	-1568.6	-1475.0

t 分布の当てはまり良い

資料18

ハミルトン力学

ハミルトン力学では，系の状態は一般化された座標と運動量$(q_i, p_i), (i = 1,2,\dots,n)$が張る位相空間の１点に相当し，系の時間発展は位相空間の軌道で与えられる．系の時間発展を表す軌道の方程式が次のハミルトンの運動方程式である．

$$\frac{dq_i}{dt} = \frac{\partial H(q,p)}{\partial p}$$
$$\frac{dp_i}{dt} = -\frac{\partial H(q,p)}{\partial q}$$

ここでHはハミルトニアンと呼ばれ，系のエネルギーを表し，

$$H = K + U$$

と定義される．ここでKは運動エネルギー， Uはポテンシャルエネルギーである．

資料19

ベイズの定理

事後分布$f(\theta|y)$は，尤度関数$f(y|\theta)$と事前分布$f(\theta)$を用いて

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{\int f(\theta|y)f(\theta)}$$

と表される．

パラメータθの共役運動量pを導入し，共役運動量は正規分布に従う．事後分布と共役運動量の確率密度関数の同時分布$f(\theta,p|y)$は以下のようになる．

$$f(\theta,p|y) = f(\theta|y)\frac{1}{\sqrt{2\pi}}e^{-\frac{p^2}{2}}$$

上記のベイズの定理より

$$f(\theta,p|y) \propto f(y|\theta)e^{-\frac{p^2}{2}} = e^{-\frac{p^2}{2}+\log f(y|\theta)} = e^{-H(\theta,p)}$$

資料20

ハミルトニアン・モンテカルロ法

パラメータθと共役運動量pは，ハミルトンの運動方程式に従い時間発展する.

$$\frac{d\theta}{d\tau}=\frac{\partial H(\theta,p)}{\partial p}=p$$
$$\frac{dp}{d\tau}=-\frac{\partial H(\theta,p)}{\partial\theta}=\frac{\partial \mathrm{log} f(y|\theta)}{\partial\theta}$$

ここでτは仮想的な時間である．この式でパラメータを時間発展させて，次の乱数を生成する.

新しい乱数の候補は，次の確率で選択される.

$$r=\min\left[1,\exp\left(-H_{fin}+H_{init}\right)\right]$$

これをメトロポリステストという.

資料21

歪正規分布と歪 t 分布

- 歪正規分布

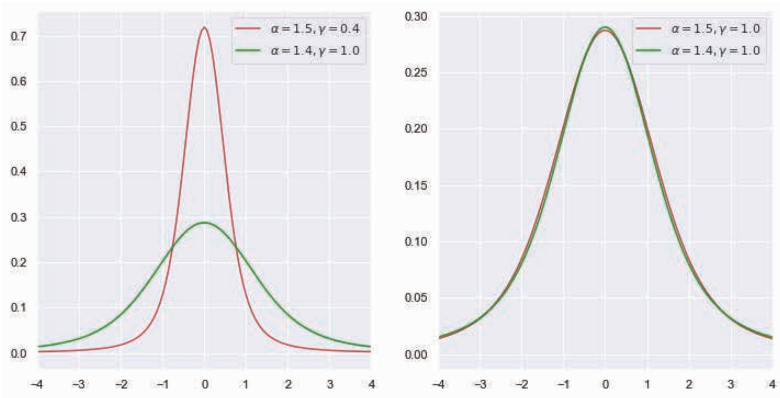
$$p(x|\zeta,\omega,\lambda)=\frac{2}{\omega}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\zeta)^2}{2\omega^2}}\int_{-\infty}^{\frac{\lambda(x-\zeta)}{\omega}}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du$$

- 歪 t 分布

$$p(x|\zeta,\omega,\lambda,\nu)$$
$$=\frac{2}{\omega}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{(x-\zeta)^2}{\nu\omega^2}\right)^{-\frac{\nu+1}{2}}\int_{-\infty}^{\frac{\lambda(x-\zeta)}{\omega}\sqrt{\frac{\nu+1}{\nu+\frac{(x-\zeta)^2}{\omega^2}}}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{(x-\zeta)^2}{\nu\omega^2}\right)^{-\frac{\nu+1}{2}}du$$

資料22

尺度母数の影響



資料23

「経時測定データを用いた多元分割表に関する対称性のモデル」

日本大学経済学部　　生亀　清貴

今回，経時測定データということで株価データを用いているんですが，私は経済出身ではなくて，こちらに来るまでは経済というものにほとんど触れてこなかった人間ですので，経済の専門家の方からするとあまりなじみのない話になってしまうかもしれません．そのへんに関してはいろいろコメントやご意見をいただければと思っております．

本日の内容ですが，研究背景を話しまして，株価データ解析について少しお話しをします．それに対してモデル提案をして，実データ解析，そして最後にまとめと今後の課題という流れでいきたいと思っています．

私はもともと正方分割表というものを長く研究しておりまして，表1が代表的なデータで，イギリス人女性の左右の裸眼視力データを表したものです．右目の視力が4段階，左目の視力が同様に4段階ありまして，7477人からとったデータで，たとえば「右目の視力がよくて，左目の視力がよい」人は1520人というデータになっております．

こういったデータに対して多くの人は独立性というものに関心があります．初等の統計学の教育でも，適合度検定において独立性の検定というものがよく言われておりまして，こういった分割表に対しては独立性というものに通常興味があるわけです．

ただし，この表1のデータのように行分類と列分類が同じ分類である場合，たとえば主対角線の「右目もよい，左目もよい」とか「右目がややよくて，左目もややよい」といった人たちが通常多くなるわけです．視力を考えたときに，基本的には主対角線付近に多くの観測値が集中して，逆に「右目はよいのに，左目が悪い」とか「右目が悪いのに，左目がよい」とかいうがちゃ目の人はあまり多くはない．こういった同じ分類から成る分割表においては当然，行と列に相関があるわけで，独立性というものは成り立たない．

こういったときには対称性に関心があって，右目と左目の視力にどういった関連性があるか，どういった対称構造があるかということに興味が移るわけです．二元分割表，R×R分割表に対して，行変数をX，列変数をYと置きまして，ijセル確率をpijと置いたとします．これは分割表の対称性に関する研究の一番最初のお話ですが，1948年にBowkerという人が対称性の検定をまとめた論文で，pij＝pjiというかたちで「対称モデル」を提示しております．右下の表で言いますと，p_{12}というセル確率とp_{21}という確率構造が等しく，p_{13}とp_{31}が等しい．p_{14}とp_{41}が等しい．ほかも対称的な部分は確率が等しいという構造になっております．ここを一番最初として，さまざまな対称性に関するモデルがこれまで提案されてきました．

先ほどのは二次元の分割表ですが，多次元の分割表を考えることもできて，表2は三元の分割表になります．これは2003年の広島，東京，札幌における毎日の気温をまとめたデータです．各変数が3つのカテゴリーに分かれておりまして，1が平年よりも低い，2が平年並，3が高いとなっております．本来ですと3×3×3のキューブ型の表になるんですが，これをうまく二次元で表した形になっております．たとえば，東京も広島も札幌も，3地点全てが平年より低かったのは365日中37日間あったことを示しております．この27マスの数値を合計すると1年の日数，つまり365になるというデータになっております．こういったかたちで3つの変数が全て同じ分類で，さらに「低い」「平年並」「高い」のように順序のある三元分割表というものが通常得られます．

こういった多元分割表に関しても対称性に関するモデルが提案されております．今回は話をシンプルにするためにとりあえず三元でお話を進めていくんですが，確率変数をX，Y，Zと置いて，ijkセル確率をpijkととります．この多元の対称モデルは1990年ですから30年ぐらい前のお話ですが，ijkを全て並べ替えた6パターン，その確率が等しいというのが「対称モデル」となっています．下の表で言いますと，たとえば赤で囲った6つの部分の確率が全て等しいという構造を示しております．省略していますが，ほかの部分に関しても同様の同等性が成り立つ．それが三元分割表における対称モデルの定義になっております．

多元分割表に関して対称性のモデルは最近でもいろいろ提案されていて，たとえば表2において

対称モデルが成り立ったときには次のような解釈が得られます．東京，札幌，広島の3地点あって，「気温が低い」「並」「高い」という確率を全て並べ替えた確率が全て等しいという解釈が得られます．

これは解釈の1つですが，こういった解釈が得られたときに私が常々感じているのは，多元分割表に関する対称性のモデルというのは数学的には二元正方分割表の自然な拡張であるんですけれども，モデルの解釈を考えたときに，ちょっと理解がしにくい．ピンと来ないと言いますか，数学に特化した人が考えたねというモデルになっておりまして，そこがちょっと問題点かなと常々考えていました．

今回，経済学部に来てから研究を進めていたんですけれども，多元分割表に関して解釈が容易な対称性に関するモデルを提案したいというのが最初のモチベーションです．経済学部に来たということで，株価のような経時測定データに適用して，それに対して何か有用で分かりやすい解釈を得たい．この2つのモチベーションがあったので，今回こういう研究を進めることになりました．

では株価という経時測定データから分割表をどうやって作成したか，少しお話ししたいと思います．ある企業のあるひと月に注目して，そこから3カ月間さかのぼって週ごとの終わり値を13週間にわたってプロットすることを考えます．それがこの赤い直線で，13個の折れ線グラフになっていると考えてください．

その3か月の中で一番高かったところを最高値，一番低かったところを最低値と考えます．その幅を3等分して，上側を「高い」，真ん中を「普通」，下側を「低い」と解釈することにします．この注目したひと月において4週間あるわけですが，その週ごとの終わり値をX_1，X_2，X_3，X_4と置きます．この企業に関しては1から4にかけて「低い」「低い」「普通」「普通」と推移していると考えられますので，X_1は3（低い），X_2も3（低い），X_3は2（普通），X_4も2（普通）ということで，1つの企業からこういう4つの値をとることを考えます．

こういった処理を東証1部上場企業全てに対して行ないました．今回は2016年8月のデータを用いましたので，そのときに該当した企業は全部で2009社ありました．それを，4元の表ですが，3×3×3×3の表を無理やり二次元で表した表になっていて，X_1，X_2，X_3，X_4ということで81マスのデータになっております．

先ほどの3，3，2，2という企業であれば，ここの7という部分に1つカウントされることになります．これを2009社に対して行なったので，セルの総和を足すと2009になることになります．今回はこういった分割表を対象としてモデルを提案していこうと思います．

モデルの提案に移る前に少し記法を示したいと思います．今回は3×3×3×3というかなり限定された状況ですけれども，まずは少し小さめの表で考えておきます．X_1からX_4まで考えて，ijklのセル確率をp_{ijkl}と置きますので，p_{1111}からp_{3333}まであるという状況です．

これからモデルを7つ考えるんですけれども，順番に紹介しておきたいと思います．まずはモデルM1で，これはijklという確率とそれを時間の推移に対して入れ換えたlkjiという確率と等しいよということを表すモデルです．

ijklに関しては必ず単調に増加するというものがありまして，最終的にiよりもlのほうが大きいということなので，どこかで不等号が成り立っているという状況です．たとえばこの左側の1, 1, 1, 2（高い，高い，高い，普通）という確率と，2, 1, 1，1（普通，高い，高い，高い）という確率が等しいという構造を表しております．ほかにも，1, 1，1，3と3，1，1，1のように対応する部分が赤と青で1つずつあって，その部分の確率が全て等しいということになります．それ以外の何も色がついていない部分に関しては何の制約も課していないので飽和モデルということになります．

ijklというふうに単調に増加するあるいは単調に減少するものは変わらないんですけれども，iとlの絶対値を考えて，絶対値が同じものは1つにまとめてしまおうという，少し制約を緩くしたモデルになります．たとえばこの赤い部分は最終的には1から2で1つ上がっています．こちらも1つ上がっていますので，6個の赤い部分は全て足し合わせてしまって，それが2から1に1つ下がっている青い部分の合計と等しいという構造です．

1が2の部分もあって，1，1，1，3のようにiとlで2つ変わっている．こちらもです．6個のオレンジのセル確率の和が黄緑の6個の和と等しいとい

う構造になっておりますので，少し制約を緩めてグループ化して，その同等性を述べたモデルになります.

モデルM3はさらにグループ化を緩くして，赤い部分は全て同じグループにしてしまおう，青い部分も同じグループで考えてしまおうという，1本の制約式になります.

もう少し制約の範囲を増やしたものが，同じようにijklとlkjiが等しいんですけれども，条件はiよりも1のほうが大きいということなので，たとえば1，3，1，2のように，一回下がってまた上がったりするものも，最終的に1から2に増加していればよい．これに対して2，1，3，1が等しいので，赤で囲った部分にはそれぞれ青で囲った部分どれかに対応して，その1つ1つのセル確率が等しいという構造になっています．こちらに関してもグループで分けて，絶対値が1の赤いグループと青いグループの合計が等しいというものと，絶対値が2，差が2の部分，オレンジの9マス分の合計と黄緑の9マス分の合計が等しいというモデル.

さらに最終的にtが1，2というものも区別せずに，赤い部分は全て和にしてしまったもの，青い部分も全て合計してしまったものが等しいという1本の制約式で表されるモデルをM3Dとします.

最後にM1DDですが，たとえば1，2，1，1のように最終を見たときには1から1で変わっていないんですが，これに関しても1，2，1，1であれば1，1，2，1が時系列に関しては対称的な確率になっていますので，そこに関しても同等性を加えたものがM1DDになります．M1DDに関しては，何も色がついていない部分はかなり少なくなって全部で9セルしかないんですけれども，そこ以外に関しては必ずどこか対応するセルが存在するといったかたちになります.

いまモデルを足早に紹介してしまったんですけれども，包含関係がありまして，この矢印で表されるような関係になっております．たとえばA→Bであれば，モデルAが成り立ったならばモデルBも成り立つことを示しております．M1DDが一番制約が厳しくて，右であったり下に行けば行くほど制約は緩くなるという性質を持っております.

このモデルを提案したので，そのモデルがデータに適合しているか，適合度検定によって調べます．適合度検定は帰無仮説としてモデルMが成り立つ．対立仮説としては，モデルMが成り立たない，すなわち飽和モデルですよということで仮説を立てます．尤度比カイ二乗統計量はおなじみのかたちとなっております．n_{ijkl}はセル観測度数，つまりデータの個数です．それに対してm_{ijkl}ハットはセル期待度数の最尤推定量となっております．今回の最尤推定量に関しては全てクローズドフォームで求めることができますので，それに関しては近似解を求めたりする必要はありませんでした.

このカイ二乗統計量が自由度kのカイ二乗分布に従うんですけれども，自由度kが幾つかというのをまとめたのがこの表になります．各モデルの検定の自由度kになります．今回は3×3×3×3という非常に単純な構造の分割表でやっておりますので，まとめて本数を数えるだけで全て出てきて，それぞれこんなかたちになっております.

今回，適合度検定統計量のほかにAICでもう少しモデルの選択をしてみようということでAICの定義を載せておきました．こちらは先ほどのデータで，2016年8月の株価の推移です．これに対して7つのモデルを適用してみますと，G^2値が求まりまして，それに対してp値が求まります．M2DとM3Dがアクセプトされているという状況ですけれども，AICを使って比較をすると，M3Dが一番小さいということで，この中ではM3Dが一番よいのではないかという判断になりました.

M3Dによってちょっと解釈をしてみますと，こちらがM3Dのもとでの期待度数の最尤推定値になります．M3Dというのは，この赤い部分の確率――この場合は最尤推定値ですが――の合計と青い部分の27マスの合計が等しいということになります．実際に計算してみますと，赤い部分が433.5で，青い部分が433.5で，等しくなっている.

そこから得られる解釈としては，こちらは1か月で悪くなった．最初の初週X_1に対してX_4は4週目ということですが，そこを比較したときに，悪くなっている企業の合計とよくなっている企業の合計が同等である，対応がとれている，同じ数だということで，景気の変動はこの月はあまりなかったのではないかという解釈ができます．逆に言うと，このモデルが当てはまらない場合は，よくなっているか悪くなっているかどちらかだとい

う評価が推測されるということです．

駆け足になってしまいましたけれども，まとめに移りたいと思います．今回は株価経時測定データを分割表データに変換しました．それに対して，適用可能で，かつ解釈が比較的容易な対称性に関するモデルを提案しました．

今回は初めてなので3×3×3×3という限定的なサイズに限定されたんですけれども，今後の課題としては，n次正方行列でカテゴリー数がRみたいな，より一般的なサイズの分割表に対してもモデルを拡張する必要があるかなと感じております．

そういった場合に，セルの数が増えます．今回は見て分かる通り，0だったり1のセルが多いので，こういう状況ですと適合度検定統計量のカイ二乗近似があまりうまくいかない場合がありますので今回はAICを使ったんですが，それとはまた何か別の解析手法を提案する必要があるのではないかと感じております．

今回は株価データに応用しましたけれども，分割表データはいろいろな分野に使えるデータですので，医療とかほかの分野のデータに適用して何か有用な解釈を得られるのではないか，得てみよう，というのが今後の課題となっております．

以上が発表になります．ご清聴ありがとうございました．

「Stochastic Volatilityモデルのベイズ推定によるリスク資産分析への応用」

日本大学経済学部　　三井　秀俊

戸塚先生との共同研究で「Stochastic Volatilityモデルのベイズ推定によるリスク資産分析への応用」というテーマで中間報告をさせていただきます．わたしはファイナンスが専門で，数学・統計は皆さんの方が得意だと思いますので，きょうはファイナンスの話を中心にしていきたいと思います．

発表の構成ですが，まず初めに研究の背景と目的に関して述べていきます．次に連続時間SVモデルと離散近似について説明します．標題にはSVモデルと書いてありますが，厳密には連続時間SVモデルと離散時間SVモデルに分かれます．最近の論文を読むと，あまりそのへんのことが書いてないので，きょうは補足説明として連続時間SVモデルを説明したいと思います．次にSVモデルとベイズ推定です．ここのSVモデルはデータを使って実証研究に使うモデルとなっていますので，厳密には離散時間のSVモデルになりますが，表題としてはSVモデルとベイズ推定となっています．次にリスク資産分析モデルの拡張です．SVモデルでリスク資産の分析をする際にどうやって拡張していったらいいのか，できるだけ現実の金融市場の特徴を捉えるようなモデルにするにはどうしたらいいのかを説明します．最後に，まとめと今後の展望に関して述べたいと思います．

まず研究の背景ですが，ここは詳しく説明します．今回のこのプロジェクトの題にありますように，リスク資産の価格変動の分析に関して今回のプロジェクトは動いております．リスク資産と書いてありますけれども，実際に金融市場では，個別株式とか株価指数，外国為替相場，ドル・円とかユーロ・円，デリバティブ取引など，それをまとめてリスク資産と呼んでいます．学部の授業でもよく言うんですが，リスクというと，日本人の方は危ないとか危険とか思っている方がいるんですが，実際の金融市場でのリスク資産のリスクは危ないとか危険という意味ではなくて，価格の変動が上下運動しながら動いていくということで，危険とか危ないということではありません．統計学を詳しく勉強した人は，標準偏差あるいは分散でリスクを測るということを学部時代に勉強したと思います．ここでは「収益率の変動を標準偏差で測ったもの」を「リスク」を定義します．金融業界ではそれを，略して「ボラ」と言う場合もあるんですが，一般的に「ボラティリティ」と呼んでいます．ボラティリティとは何かといったら，それは「リスク資産の収益率の変動を標準偏差あるいは分散で測ったもの」ととらえれば間違いはないと思います．

問題は，時間が進むとボラティリティがどんどん変動していく．そのボラティリティをどうやって時間変動の中でとらえていくかというのがファイナンスの世界では問題になっています．それが前提としてあります．さっきからファイナンスと言っていますが，実際にはファイナンスの中でもいろいろな分野があるんですが，ここでは時系列分析を用いたファイナンスの実証研究に焦点を当てています．

ボラティリティの時間変動を研究する際には2つのカテゴリーに分けることができます．1つはARCH型モデルを使うカテゴリーと，もう1つはStochastic Volatility（SV）モデルを使って分析を行なうカテゴリーと，2つに分かれています．ARCHというのはAutoregressive conditional heteroscedasticityの略で，ARCHモデルというのはEngleが景気変動の時系列特性を捉えるために最初に使ったモデルです．景気変動をとらえるARCHモデルが，リスク資産，特に株価の動きをうまくとらえることが分かったので，ファイナンスの時系列の世界で使用されるようになりました．ここでARCH型モデルと書いてあるのは，もともと最初はARCHモデルというのがあったんですが，それをBollerslevという人が一般化してGeneralized ARCHモデルに拡張しました．その後，NelsonがExponential ARCHモデルに拡張したわけです．ARCH型モデルというのはいままでかなり研究され尽くしています．なぜ研究し尽くされているかというと，推定するのが容易というのと，モデルの拡張が簡単にできるので，ここ10年，20年でやり尽くされた感じがします．きょう説明するのはもう1つ方のStochastic Volatilityモデルを中心に話を進めていきます．

ARCH型モデルとStochastic Volatilityモデルに関してもう少し予備知識を説明したいと思います．まずARCH型モデルですが，先ほど説明しましたように，推定が容易にできます．基本的には最尤法を使う場合が多いですが，もちろん別の推定量でもできます．もう1つの特徴として，ARCH型が最初にあって，その後，GARCHとかEGARCHとか，何種類ものARCH型モデルが発展形で開発されてきました．モデルの拡張が非常に容易だということです．なぜモデルの拡張が簡単かというと，推定が容易にできるというところにあると思います．モデルを拡張してもパラメータの推定ができないと意味がないので，普通はそうそう拡張はできないわけですが，ARCH型モデルの場合は簡単にパラメータの推定ができるので，モデルの拡張も，簡単とは言いませんが，可能であるということです．

もう1つのARCH型モデルの特徴として，もともとARCH型モデルというのは離散時間モデルからつくられています．出発点が離散時間モデルなので，時系列分析をするうえで非常に便利なわけです．そのためARCH型モデルは時系列分析で非常に用いられるようになりました．

ARCH型モデルの欠点ですが，先ほどやり尽くした感があると言ったのはここにも関係しているんですけど，時系列分析する場合には問題ないのですが，いろいろな金融の分析をしたいときに，金融分析モデルへの応用が難しいわけです．もともとあったARCHモデルから始まって，モデル自体を拡張するのは簡単なんだけど，金融の特徴とか金融分析に見られるような現象をとらえる場合に，ちょっと難しい．それは出発点が離散時間モデルにあると思います．

離散時間や連続時間を考えてみると，離散系のモデルというのは数学的な扱いが難しいというか，いろいろ複雑なので，それで拡張が難しいわけです．実際には金融分析のモデルとしては，ただ単に時系列分析だけではなくて，オプション理論とか新しい金融商品の開発とか，あるいはヘッジ手法などをいろいろ知りたいわけです．後でSVモデルの説明をしますが，こういう場合は連続系のモデルでないと式の展開とか数学的な扱いが難しくなるので，それでARCH型モデルは最近かなり下火になってしまったということがあります．

SVモデルですが，連続時間SVモデルと離散時間SVモデルと2つに分けられます．もともとSVモデルというのは連続時間から始まっています．出発点が連続時間なので，モデルを金融分析に当てはめるときに数学的に非常に扱いやすいので，金融分析に応用する場合にはいろいろな応用方法ができるので，研究者としてはSVモデルを使いたいわけです．ただ，SVモデルにも欠点があって，SVモデルを使ってパラメータの推定をするときに，一度，離散時間SVモデルに直さなければいけないわけです．そうすると，連続系のモデルを離散系のモデルに変換するときに，そこに誤差が生まれたり，想定しなかった問題が起こったり，いろいろあるわけです．SVモデルの欠点として一番言われているのが，推定するのが難しいことです．SVモデルのいままでの推定法というのは一般化積率法とか最尤推定が使われたんですが，最尤推定の場合にも厳密にはできませんので，疑似最尤推定を使っていままではしていました．例えば，カルマンフィルターなど使って推定する方法です．

今回はそういう問題を解決するためにベイズ推定法を使うということで，それが表題に書いてある理由です．連続時間SVモデルだと数学的な扱いがし易いので，特にオプション理論で非常に使われています．もともとオプション理論はSVモデルから始まっているので，SVモデルを勉強するとオプションの話もよく理解できると思います．

SVモデルの長所としては，金融分析モデルへの応用が多用である．初めに連続時間のSVモデルで考えて，オプション理論や新しい金融商品，ヘッジ手法を連続時間で最初に考える．ある程度数学的なモデルを考えて，実際に実証研究する場合には離散時間モデルに近似します．その後，パラメータ推定やプライシングをすればよいということで，金融分析をする際にはARCH型モデルよりもSVモデルのほうが応用範囲が広いので，こちらを使いたいわけです．ただ，今回の研究テーマにもなっているんですが，推定が難しいというのが非常にネックになっているわけです．それで今回，戸塚先生との共同研究で，ベイズを使ってSVモデルを推定してやればよいのではないかと

いうことで，それが今回の研究の背景にあります．目的としましては，SVモデルというのは推定が難しいんですが，最近その問題が解決されました．マルコフ連鎖モンテカルロ法によるベイズ推定を使うことによって，SVモデルの推定が可能となりました．その際に，マルコフ連鎖モンテカルロ法の手法として，Gibbs Sampling, Metropolis-Hastings, Hamiltonian Monte Carloをよく使います．

いままでのファイナンスでのSVモデルの実証研究では，初期はGibbs Sampling, Metropolis-Hastings，あるいはこれを発展させたMCMCの方法で実証研究するというのが多かったんですが，ここ最近，Hamiltonian Monte Carloを使ってSVモデルの推定を行なう研究が出てきました．調べてみると，ここにありますように，高石先生がやっていました．ここに書いてあるのはHamiltonian Monte Carloに関する研究ではなくて，SVモデルをHamiltonian Monte Carloでベイズ推定した論文になっています．いままでは，Gibbs SamplingとかMetropolis-Hastings，あるいはMetropolis-Hastingsを改良したかたちのMCMC法でベイズ推定をしていますが，先ほど戸塚先生の発表にもありましたように，Hamiltonian Monte Carloを使ってSVモデルの推定をしてやろうということです．

今回の目的としては，ベイズ推定によってSVモデルを用いたリスク資産の価格変動分析，それに関してサーベイをまず行ないます．プラス，株価収益データを使って例を提示したいと思っています．どの辺りをサーベイするかというと，非対称SVモデルと，先ほど戸塚先生の研究発表にもありましたが，裾の厚い分布を使った場合．あと長期記憶モデルというのがありますので，それをSVモデルの中に取り入れて長期記憶SVモデルとして推定したいということです．長期記憶というのは論文でもあります．もう1つはRealized SVモデルで，先ほどの戸塚先生の報告にもあったんですが，高頻度データを使ってボラティリティを推定して，それを使ってSVモデルに取り込んでモデルのパラメータを推定しようという方法です．もう1個はオプション価格付けです．

きょうは時間がないのですべてはできないと思いますが，非対称SVモデルと，裾の厚い分布．オプション価格付けはちらっとしかできないと思います．長期記憶SVモデルとRealized SVモデルは，割愛させていただきます．

まず簡単に連続時間SVモデルと離散近似を見ていきたいと思います．ご存知の方も多いと思うんですが，一応，簡単にご説明します．まず連続時間でリスク資産価格の変動をとらえようとする場合には，幾何ブラウン運動を使う場合が多いです．リスク資産価格というのがちょっとイメージつかない方は株価だと思ってください．株価の動きを記述します．μはドリフト項，dtは時間の微小変化分で，σが標準偏差．ここでσが標準偏差と書いてありますが，これが最初のところに出てきたボラティリティになります．dz_1がウィナー過程を示しています．μはドリフト項ですが，ドリフトというのがいまいちピンと来ない方は，長期のトレンドを微小変化にして，かなり短い期間のトレンドと考えてもよいと思います．ドリフトのほうが自然な人はドリフトでかまいません．

ここでよく聞かれるんですが，「ボラティリティ」の厳密な定義はあるけれども，結構広い意味で使われていて，大体2つに分かれます．ファイナンス理論ではσそのものをボラティリティと呼ぶ場合が多いです．計量経済学とかファイナンスの計量分析をやっている人はσ^2をボラティリティと呼ぶことが多いので，論文を読むときとか研究発表を聞いているときに，「ボラティリティ」と研究者が言っているのはどっちかなというのをよく理解しておくと，その研究内容がよく分かると思います．二乗がついているかついていないかの違いですけど，結構間違いやすいというか，話がこんがらがる場合があるので，気をつけたほうがよいと思います．

株価の過程は，もう1つ，σ^2の連続時間の動きを考えます．dz_2というのはウィナー過程になります．ここでファイナンスの世界でよくやるのは，株価の動きのところのdz_1とボラティリティの変動のdz_2，ここに相関があると言う人もいるし，ないと言う人もいるので，ないと置いてもいいです．ただ研究論文としては通常は相関を持つというふうにして話を進めていきます．

ボラティリティの過程を一番シンプルに置いたかたちで，ファイナンスの研究では実際にはOrustein Uhlenbeck過程のような平均回帰性を持つような連続時間モデルにして定式化するのが通常です．ここでθは長期的な平均回帰になって，κ

が調整速度を表すものになります．実際には，dz_1とdz_2が相関を持つというふうにしてモデル化するのが一般的です．連続時間だとモデルのパラメータの推定ができないので，離散近似しなければいけません．離散近似の方法はただ単に式変形すればいいだけですが，離散近似すると両辺をSで割って式変形すれば簡単になります．昔はこのモデルがよく使われていたんですが，最近では，このモデルだとここのスケール・パラメータの問題とかが出てくるので，このスケール・パラメータを入れ込んで離散時間SVモデルをつくる方法がとられます。言っていることはそんなに大したことではないんですけど，リスク資産分析ではytが収益率を表すことになります．収益率というのはボラティリティ×誤差項という単純な式になっています．ボラティリティの過程が一次のAR（1）モデルみたいなかたちで，平均回帰性を捉えられるモデルになっています．

最尤法でできれば何の問題もなく万々歳なんですが，SVモデルの場合の尤度関数を考えると，この積分が解けないわけです．それはしょうがないので，MCMC法によるベイズ推定法によりパラメータの推定を行なうことになります．ここらへんは本当はもっと詳しく説明しなきゃいけないんですが，時間がないので問題点だけにします．

先ほども言ったようにSVモデルは推定が難しい．尤度関数がこういうかたちになっていて積分が解けないので推定が難しいということなので，何か工夫をして推定しなければいけない．今回はMCMC法によるベイズ推定で，特に，MCMC法はHamiltonian Monte Carloを使いますよということです．

ここはベイズ推定の話なので飛ばします．

いま紹介したSVモデルをベイズ推定する場合に，また1つ問題が出てきます．ボラティリティというのをどうやって扱うかということです．実際にボラティリティというのは観測可能ではないので，それを潜在変数として考えてやります．ベイズ推定する場合に，ボラティリティも未知パラメータとして考えて，MCMCでサンプリングします．そうすると，SVモデルに関してはベイズ推定するときにボラティリティのサンプリングをどうやって効率的に行なうか，というのが論文を書くときに勝負になります．いままでボラティリティのサンプリングで使われた方法は，Gibbs SamplingやMetropolis-Hastingsを使う場合にはsingle-move sampler, multi-move sampler, mixture samplerを使用しています．ただ，これだと遅いので，Hamiltonian Monte Carloを使ったほうがよいのではないか．ただ，こっちのほうが効率的かどうかまだ分からないので，ここに？をつけているんですが，Gibbs SamplingやMetropolis-Hastingsを使うよりかはHamiltonian Monte Carloで一括してやったほうがいいのではないかというのが戸塚先生の意見で，それで共同研究をしているということです．

いままでの説明はSVモデルでベイズ推定をする場合にどこが問題になっているかということの説明でした．今度はリスク資産分析をする場合にちょっとモデルを拡張してみましょうという話です．ここも簡単にいきます．1つは非対称性を考えている．何の非対称性を考えるかというと，さっきの株価の変動のところの誤差項とボラティリティの変動の誤差項，この2つの誤差項同士が相関を持つというモデルが1つ考えられるということです．それをすることによって何が分かるかというと，株価の動きで，株価収益率が下落すると，その次の日にはボラティリティが上昇する傾向がよくあります．逆の傾向もあります．これはレバレッジ・エフェクトと言うんですが，これを捉えることができます．それで非対称SVモデルはボラティリティのレバレッジ・エフェクトをとらえるときによく使われるモデルで，これがSVモデルの次の応用としては一番オーソドックスなかたちのモデルだと思います．

もう1つは，先ほどの戸塚先生のところでもあったんですけれども，分布の裾の厚いモデルです．ファイナンスの世界ではt分布をよく使います．もちろん別の分布でも良いんですが，推定するときにt分布は利用可能というのと，先ほどの戸塚先生の図でも分かるように，リスク資産の分布というのは裾が厚くて，平均のところには頻度が高い，尖度が高というそういう特徴があります．

では分布の両裾が厚いという性質をSVモデルに入れてみましょう．そうすると，最尤法とかする場合にはt分布でいいんですけれども，ベイズでする場合にはそうではなくて逆ガンマ分布を使

います．そうするとベイズを使ってうまく推定することができます．これもレポートでは書いていますので，興味のある先生はそちらを読んでみてください．

もう1個はオプション価格付けへの応用です．いままでの話だとリスク資産の時系列の特徴をとらえるだけだったんですけど，ファイナンスの世界ではやっぱり価格付けをやりたいわけです．ファイナンスの世界で一番価格づけが難しいのはオプションで，オプション価格への応用ができると非常に有効なわけです．

ここにあるのはオプション価格付けに応用する場合にどうやってやればよいか書いてあるんですが，説明する時間がないのでポイントだけですけど，SVモデルを使って，かつベイズの推定法を使うとオプション価格にも応用できます．従来の方法よりも，ベイズでしたほうが効率性が良い．直感的に説明すると，昔はパラメータを推定して，パラメータを推定した後，モンテカルロ・シミュレーションでオプション価格を出したんですが，ベイズでやると1回で終わります．いままではパラメータを推定してオプション価格を出すシミュレーションという2段階だったのが，ベイズを使うと一気にオプション価格まで導出することができるというのが利点です．先ほどのGibbs SamplingやMetropolis-Hastingsでもできるんですけど，Hamiltonian Monte Carloでもできるんじゃないかというのがぼくと戸塚先生の提案です．

最後にまとめと今後の展望で，Hamiltonian Monte Carloを利用したリスク分析が有効であるということです．もう1つは，Hamiltonian Monte Carloも有効なんだけど，ほかのGibbs SamplingやMetropolis-Hastingsとの効率性の比較などを行なったほうがよいのではないかということです．いままで見てきましたが，ファイナンスへの応用として，裾の厚い分布を使って実証研究する，長期記憶SV, Realized SVモデル，高頻度データを使ってもっと時間間隔を短くして分析していく．もう1個はオプション価格，デリバティブに応用していくということです．今回の研究としてはSVのサーベイを中心に書いて，余裕があればちょっとした実証研究を付け加えて論文にしたいと思います．

以上です．

きょうは皆さん，研究会に参加いただいて，どうもありがとうございます．またいろいろコメントもいただいてありがとうございます．今後の研究に活かしていきたいと思います．今回，3人でリスク資産分析に関して実証分析あるいはサーベイをしているわけですが，戸塚先生も生亀先生もファイナンスの先生ではないのですが，頑張って株価データを使って実証研究していただいて，非常にありがたいと思います．ぼくの個人的な意見としては，やっぱりどうしても経済学部出身の先生だと視点とか見る観点が同じになっちゃって，新しい発想が浮かばないんです．今回，戸塚先生，生亀先生と共同研究をしてみて，やはり発想が違うなと思って，ぼくも非常に勉強になりました．生亀先生の分割表を使って株価の分析をするのも，初め聞いたとき「おーっ！」と思ったので，今後も日大経済の中で新しい観点からリスク資産，株とかドル・円の通貨に関して研究していけたらいいなという，ぼくの思いがあります．

きょうは皆さん，どうもありがとうございました．