

実数値遺伝的アルゴリズムを用いた パラメータ推定における変数選択手法の提案

A Study of Variable Selection
through Parameter Optimization with Real-Coded Genetic Algorithms

小畑 崇弘*¹ 小野 功*² 倉橋 節也*¹
OBATA Takahiro ONO Isao KURAHASHI Setsuya

*1 筑波大学大学院ビジネス科学研究科
Graduate School of Business Sciences, Universe of Tsukuba

*2 東京工業大学大学院総合理工学研究科
Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

The purpose of this paper is to report the result of applying the way of variable selection within real-coded genetic algorithms(RCGA) to concrete issues. In recent researches, It was shown that variable selection is possible in linear regression models by using the variances of genes in RCGA. Moving on the next step, the way is applied to a concrete issue, the dividend policies matter which has been discussed for a long time in financial field.

1. はじめに

変数選択を行う場合、どの説明変数を削除もしくは採用するかを判断する変数選択基準とどのような順番で変数の組み合わせを比較するかを決める変数選択アルゴリズムが重要である。変数選択基準としては赤池情報量基準やベイズ情報量基準、 t 値などがある。変数選択アルゴリズムとしては変数減少法や変数増減法などが頻繁に用いられている。変数選択の分野に遺伝的アルゴリズム(GA)を応用した研究では変数選択アルゴリズムの代わりにGAを用いたものが多かった。しかし、実数値GAによってパラメータ推定をする際の遺伝子の分散に着目し、分散の値を加工すれば変数選択基準として活用できることが確認されている。

本論文では実数値GAの遺伝子の分散を活用した変数選択手法を金融分野における企業配当に関する研究に応用した事例について報告する。なお、本論文では実数値GAの遺伝子の分散を変数選択基準としてではなく、説明変数間の有意性の比較のために利用した。

2. 実数値GAと変数選択

2.1 実数値GA

実数値GAとは遺伝的アルゴリズムの中で遺伝子型に実数値ベクトルを採用したものである。0と1の2値で遺伝子型を表す従来のGAに比べて、実数値を扱う問題でより効率的な解の探索が可能である。実数値GAの解の探索能力は、交叉モデルと子個体生成モデルにどのような手法を組み合わせるかによって決まる。

2.2 実数値GAによる変数選択

実数値GAの枠組みだけで変数選択を行うことに成功した研究に[小畑09]がある。以下、当研究について少し詳しく説明する。

2.2.1 遺伝子の分散と標準誤差

[小畑09]は関数モデル内におけるあるパラメータの変化が目的関数値に与える影響の大きさは、実数値GAでパラメータ

連絡先: 小畑崇弘, 筑波大学大学院ビジネス科学研究科,
〒101-0051 東京都千代田区神田神保町3-25,
03-3942-6918, obata@gssm.otsuka.tsukuba.ac.jp

タ推定をする際の遺伝子の分散の大小に反映されている可能性があると考えた。そこで(1)式のような線形回帰モデルを用いて実験を行った。

$$y = a + a_1 \cdot x_1 + \dots + a_p \cdot x_p \quad (1)$$

(1)式の y は目的変数、 $x_1 \sim x_p$ は説明変数、 a_0 は定数項、 $a_1 \sim a_p$ は回帰係数、 p は説明変数の数である。この式の定数項および回帰係数を実数値GAによって推定し、その際の遺伝子の分散の推移を分析した。利用した実数値GAの交叉モデルは REX^{star} 、子個体生成モデルはJust Generation Gap(JGG)である。各個体は定数項および各回帰係数に対応する遺伝子を持つ。遺伝子の分散についてはパラメータ毎に各個体の対応する遺伝子のみを取り出し、その遺伝子群の値を基に分散を算出した。以後、遺伝子の分散を Vg と表す。例えば回帰係数 a_1 に対応する遺伝子の分散は Vg_{a1} と表す。

実験結果の分析から、GAによるパラメータの推定が進み遺伝子の値が収束しつつある状態よりも以降をみると、あるパラメータに対応する Vg の値と重回帰分析を行った際の当パラメータの標準誤差がほぼ比例する関係を見出した。この関係は(2)式のように表せる。

$$\frac{Vg_a}{STD_a} \simeq \frac{Vg_{a1}}{STD_{a1}} \simeq \dots \simeq \frac{Vg_{ap}}{STD_{ap}} \simeq K \quad (2)$$

ここで STD_i はパラメータ i の標準誤差を表す。

2.2.2 I値の導入と変数選択

重回帰分析において変数選択の基準として頻繁に利用されているF値は t 値の2乗であり、 t 値は回帰係数を標準誤差で除算したものである。このことを勘案して、(3)式で定義される値を導入し、これをI値と呼んだ。I値が大きいほどモデル内における当該変数の有意性が高いことを示す。

$$I_j = \frac{a_j}{Vg_j} \quad (3)$$

I_j はパラメータ j のI値、 a_j はパラメータ j の推定値(つまり定数項または回帰係数の値)、 Vg_j は対応する遺伝子の分散である。まずI値は統計上のF値とほぼ比例した値となることが確認された。その後、I値を変数選択基準に用いた変数選

択を実行した。具体的には I 値が小さい変数から順に削除していく変数減少法を用いた。変数選択の結果は赤池情報量基準を変数選択基準とした場合と類似したものとなり、線形モデルという限定はあるが、I 値が変数選択基準となり得ること、および実数値 GA の枠組みだけで変数選択を実行できることが確認された。

3. 企業の特異性資産と配当政策

3.1 企業配当に関する従来の理論

従来の配当政策の分析においては株主と経営者の意思が注目されてきた。この背景にあるエージェンシー理論では、経営者を株主とは異なる選好と目的を持った独立した経済主体と認識する。その結果、経営者の行動が必ずしも株主の利益に合致する保証はなくなる。こうした世界では、社内に蓄積されたキャッシュを配当として株主に還元することが経営者にとっては株主に対する忠誠を表明するシグナルになるとともに、株主にとっては経営者の機会主義的行動を抑止する力となる。このような配当を支払うことが株主と経営者の間のエージェンシーコストを低減させ、企業価値の向上を促すという考え方をフリーキャッシュフロー仮説 [Jensen86] と呼び、広く受け入れられてきた。

3.2 企業の特異性資産に注目した実証研究

近年、フリーキャッシュフロー仮説では説明できない事象が報告されるようになってきた。その一つに企業の特異性資産に着目した実証研究がある [宮川 09]。特異性資産とは経営者が持つ経験やノウハウ、従業員が持つ技術力や知識など人的資産に依存して蓄積される無形の資産のことである。フリーキャッシュフロー仮説に従えば企業利益の増加に応じて配当額も増加することになるが、現実には特異性資産の効率的な投資回収を実現している企業では利益と配当額の感応度が低下すると宮川は説く。そしてこの点を実証するため [宮川 09] は (4) 式の重回帰モデルを用いた。

$$\Delta DIV_i = \alpha + \beta_1 \cdot Sales_i + \beta_2 \cdot PG_i + \beta_3 \cdot DER_i + \beta_4 \cdot ROA_i + \beta_5 \cdot ROA_i \cdot RDED_i + \tilde{u}_i \quad (4)$$

各変数の具体的な定義は以下の通り。目的変数は企業 i の 5 年間の平均年率配当変化率である。Sales _{i} は売上高成長率、PG _{i} は当期利益成長率の 5 年間の平均成長率であり、配当変化率を含めた以上の 3 変数は幾何平均で算出している。DER _{i} は負債比率、ROA _{i} は総資産利益率であり、いずれも 5 年間の単純平均値を用いている。RDED _{i} は研究開発効率 (RDE) のダミー変数であり、全サンプル企業の RDE の中央値を採り、中央値を超える企業を 1、そうでない企業は 0 の値をとる。なお RDE は分子が直近および一前期の営業利益の合計、分母が二期前および三期前の研究開発費の合計である。企業が行った研究開発投資が利益を生むまでにある程度の時間がかかると考えられるため上記のような設定としている。 \tilde{u}_i は誤差項である。なお分析期間は 2001 年から 2005 年の 5 年間で、対象企業は当該期間中に東京証券取引所第一部に継続して上場し、分析に必要な財務データが収集可能な企業 835 社である。

分析の結果、売上高成長率 (Sales)、当期利益成長率 (PG)、総資産利益率 (ROA) の回帰係数は有意にプラス、負債比率 (DER) の係数は有意にマイナスの値となった。ここまではフリーキャッシュフロー仮説と整合的である。一方、総資産利益率と研究開発効率のダミー変数のクロス項である ROA _{i} · RDED _{i} の係数は有意にマイナスの値となった。こうして研究開発効率

の高い企業は利益に対する配当額の感応度が低下することが確認され、フリーキャッシュフロー仮説に批判的な宮川の見解が実証された。

4. 問題設定

4.1 [宮川 09] に対する改良

前述のように [宮川 09] ではダミー変数を使っているため、研究開発効率が中央値より高い企業における ROA の重回帰モデル内での有意性と中央値より低い企業のそれとを比較することが出来ない。中央値前後で企業サンプルを二つに分けて、サンプル毎に重回帰分析を行って ROA の位置づけを調べることは出来るが、他の変数の回帰係数の値が変化してしまう可能性があるためサンプル間の値を比較するには調整が必要になると考えられるし、何より手間がかかる。そこで本論文では構造変化のある事例に対する分析手法を応用した (5) 式を用いて分析を行う。

$$\Delta DIV_i = \begin{cases} \alpha + f_i + \beta_4 \cdot ROA_i & (i \leq \gamma) \\ \alpha + f_i + \beta_5 \cdot ROA_i & (i > \gamma) \end{cases} \quad (5)$$

$$f_i = \beta_1 \cdot Sales_i + \beta_2 \cdot PG_i + \beta_3 \cdot DER_i$$

γ は企業サンプルを研究開発効率の高低に分ける分割点である。他の記号は (4) 式と同様。なお、本論文では [宮川 09] と同様に γ を全サンプルの研究開発効率の中央値となる企業とした場合と、 γ を可変にしてパラメータの一つとみなした場合のそれぞれについて分析を行った。

4.2 データセット

分析対象データは [宮川 09] と出来るだけ同じになるようにした。すなわち 2001 年から 2005 年まで継続して東京証券取引所一部に上場し、分析に用いる財務データが取得可能な企業である。但し、用いたデータベースが異なることから分析対象企業の中身までが正確に一致している訳ではない。これは企業の合併等の結果、データベースのバージョン毎にデータベースに含まれる企業が相違してしまうことが影響している。なお、基データは証券コード順に並んでいたが分析の際に研究開発効率の数値の小さい順に並び替えを行った。

4.3 分析に用いた実数値 GA の設定

4.3.1 交叉モデルおよび子個体生成モデル

[小林 09] を参考にして交叉モデルに REX^{star}、子個体生成モデルに Just Generation Gap を用いた。本論文で扱う問題は、(1) 解が遺伝子の初期化領域外にある、(2) 悪スケール性がある、(3) 変数間依存性が強いといった特徴をもつ可能性がある。そのため、こうした点に対応できる上記モデルの組み合わせを採用した。

4.3.2 個体の設定

各個体の遺伝子数は 7 つである。それぞれが (5) 式における $\alpha, \beta_1 \sim \beta_5, \gamma$ に対応する。個体の適応度には各個体の表すモデルによる推測値と観測値の二乗誤差の累計を用いた。適応度の値は小さいほど高評価とした。

4.3.3 サンプル分割点に対応する遺伝子の扱い

サンプル分割点は整数値である。そのため、 γ に対応する遺伝子の小数点以下を切り捨てた値を分割点とした。例えば当該遺伝子の値が 14.5 であれば研究開発効率が 14 番目までの企業を研究開発効率の低い企業のサンプルに含めて処理する。なお、研究開発効率の中央値で企業サンプルを分割する場合は γ に対応する遺伝子の値を 418 に固定して処理を行った。

5. 結果

表 1 に γ に対応する遺伝子の値を 418 に固定した試行の結果をまとめた。なお、表 1 の GA に関する値は 5 回の試行の平均値を掲載した。定数項や売上高成長率 (Sales), 当期利益成長率 (PG), 負債比率 (DER), 総資産利益率 (ROA($i \leq \gamma$)) に関する推定値は同じデータに対する (4) 式の重回帰分析を行って得られた値と完全に一致した。総資産利益率 (ROA($i > \gamma$)) の値も重回帰分析の総資産利益率 (ROA) の回帰係数とクロス項 (ROA·RDED) の回帰係数を合計した値に一致した。一方、重回帰分析における総資産利益率 (ROA) の t 値とクロス項 (ROA·RDED) の t 値には大きな差があり両変数の有意性の水準に差があるように見えるが、I 値をみると総資産利益率 (ROA($i \leq \gamma$)) と総資産利益率 (ROA($i > \gamma$)) は互いに近い値となっている。また他の変数の当該値と比べて大きな値になっていることが分かる。このことから I 値を基に判断すれば、研究開発効率の高い企業のサンプルでも低い企業のサンプルでも総資産利益率 (ROA) はモデルの中で有意性の高い説明変数であることが分かる。

表 1: RDE の中央値で分割したサンプルの試行結果

	回帰係数	t 値	GA の	
			推定値	I 値
定数項	-5.6435	-3.2083	-5.6435	5.92×10^{14}
Sales	0.7559	5.3301	0.7559	1.50×10^{15}
PG	0.0084	0.9162	0.0084	3.69×10^{13}
DER	-0.0035	-4.1172	-0.0035	6.70×10^{14}
ROA	2.4566	7.8509		
ROA($i \leq \gamma$)			2.4566	3.48×10^{15}
ROA·RDED	-0.4868	-1.9275		
ROA($i > \gamma$)			1.9698	3.10×10^{15}

次にサンプル分割点を可変にして GA を実行した。20 回の試行の結果、サンプル分割点が 406 となった試行が 16 回、121 となった試行が 4 回であった。サンプル分割点が 406 のときの結果は中央値でサンプルを分けた場合と比べて最少二乗誤差は若干小さくなるものの、それ以外の値はほぼ同じだった。そのため表 2 にはサンプル分割点が 121 の場合の結果だけを掲載した。表中の値は 4 回の試行結果の平均値である。今回の試行結果でも重回帰係数と GA による推定値は同じだったことから表 2 では GA による推定値と I 値のみを掲載した。

表 2: RDE が 121 番目の企業で分割したサンプルでの結果

	GA の推定値	I 値
定数項	-4.3818	2.53×10^{14}
Sales	0.7501	1.65×10^{15}
PG	0.0078	2.51×10^{13}
DER	-0.0036	7.18×10^{14}
ROA($i \leq \gamma$)	1.0953	1.22×10^{14}
ROA($i > \gamma$)	2.0566	2.35×10^{15}

表 2 をみると総資産利益率 (ROA($i \leq \gamma$)) に関する推定値が 1.0953 と中央値で分割したサンプルでの総資産利益率 (ROA($i > \gamma$)) の値よりも小さくなった。一方で表 2 の総資産利益率 (ROA($i > \gamma$)) は 2.0566 であり、表 1 の総資産利益率 (ROA($i > \gamma$)) に関する推定値よりも若干大きくなっている。I 値をみても総資産利益率 (ROA($i \leq \gamma$)) の値は総資産利益率 (ROA($i > \gamma$)) の値と比べて一桁以上小さい。こ

れらの点を踏まえると研究開発効率の非常に低い企業も利益に対する配当額の感応度が全体平均に比べて低い傾向にあることが分かる。但し、感応度については企業ごとの変動も大きいと言える。

最後に経営開発効率が 122 番目に低い企業から 406 番目までの企業を集めた企業群とそれ以外をまとめた企業群の二つにサンプルを分割し、(1) 式を用いた重回帰分析を行った。その結果を表 3 に示す。総資産利益率 (ROA) やクロス項 (ROA·RDED) の回帰係数をみるといずれも表 1 の値に比べて絶対値が大きくなっている。また t 値も同様に絶対値が大きくなっていることから、利益と配当額の間関係を調べるためには経営開発効率が中程度の企業とそれ以外に分割した方がより適切だと言える。

表 3: RDE が 122 ~ 406 番目までの企業とそれ以外に分割したサンプルでの重回帰分析の結果

	回帰係数	t 値
定数項	-5.4014	-3.1143
Sales	0.7598	5.3718
PG	0.0086	0.9369
DER	-0.0036	-4.1283
ROA	2.6311	8.4205
ROA·RDED	-0.7404	-2.8539

6. 結論

本稿では実数値 GA の遺伝子の分散を活用した変数選択手法を金融分野の実問題に応用した結果について述べた。金融分野の分析では本稿で採り上げたようなダミー変数を使った分析が多数見られるが、ダミー項の有意性の検証に I 値が活用できる可能性を示した。また本旨からは外れるものの、サンプルの分割点を予め中央値などに固定してしまうのではなく、可変なパラメータとして検証することで新たな知見が得られる可能性を示した。

本論文ではサンプル分割点に対応する遺伝子の値も実数値を用いたが、分割点は本来、整数値で十分なはずである。値が整数値の場合の遺伝子の分散および I 値の処理法について検討していくとともに、今回扱ったような実数値と整数値 (離散値) が混じった問題に対するより適切な手法を開発することが今後の課題である。

参考文献

- [小畑 09] 小畑 崇弘, 倉橋 節也: 実数値 GA を用いた変数選択手法と株価共変動の分析への応用, 進化計算シンポジウム 2009 講演論文集 (2009)
- [小林 09] 小林重信: 実数値 GA のフロンティア, 人工知能学会誌 24(1) pp.128-143 (2009)
- [宮川 09] 宮川 壽夫: 企業の特異性資産と配当政策の粘着性, 証券アナリストジャーナル 47(8), pp. 42-55 (2009)
- [Jensen86] Jensen Michael: agency Costs of Free Cash Flow, Corporate Finance, and Takeovers, American Economic Review, pp.323-329 (1986).