

ネットオークションにおける商品取引の特性分析

石塚 美加[†] 松田 和浩[†] 吾妻 晋^{††} 佐島 敬真^{††} 石井 啓之^{††}

^{††} 東海大学電子情報学部コミュニケーション工学科 〒259-1292 神奈川県平塚市北金目 1117

[†] 日本電信電話株式会社 NTT 情報流通プラットフォーム研究所 〒180-8585 東京都武蔵野市緑町 3-9-11

あらまし インターネットを利用したオークション（以下ネットオークション）は、欲しいものを安く手に入れられるという利点から、インターネット上の代表的なサービスのひとつとなっている。ネットオークションにおける商品取引では、匿名性や非対面性が高いため、ユーザが入手し得る情報には限りがある。従って、取引に関わるユーザ自身が商品取引の妥当性を簡単に判断できる方法が必要とされている。本報告では、このような背景を踏まえ、ネットオークションで行われる商品取引の分析、モデル化を行い、同定モデルと実際の商品取引の差分から商品取引の妥当性を判断することを試みる。特に、今回の特性分析では、情報の入手、分析が容易に行えるとともに、商品取引の特徴を最も的確に表すと考えられる落札価格に着目して分析を行う。

キーワード インターネットオークション、落札価格、時系列分析、正規分布

Statistic Analysis of the Highest Bids in the Internet Auctions

Mika ISHIZUKA[†], Kazuhiro MATSUDA[†],

Susumu AZUMA^{††}, Yoshimasa SAJIMA^{††}, and Hiroshi ISHII^{††}

^{††} Dep. of Communications Engineering, School of Information Technology and Electronics, Tokai University

[†] NTT Information Sharing Platform Laboratories, NTT Corporation

Abstract The Internet-auction is one of the most representative services in the Internet. Though users can get low-priced items through auctions, anonymity and indirect dealings cause uncertainty. Thus, it is important to give auction-users clues to judge the validity of the dealing. For this purpose, we analyze and model the Highest Bits in the Internet auctions. In this report, we focus on time series analysis of the Highest Bits since we consider time series of the Highest Bits stands for characteristics of dealings most appropriately.

Key words Internet auction, highest bits, time series analysis, normal distribution

1. はじめに

インターネット上で提供されるサービスが多様化すると共に、それらのサービスを利用するユーザ数も増加傾向にある。特に、インターネットを利用したオークション（以下ネットオークション）では、欲しいものを安く手に入れられる、不用品を適価で処分できるという利点からインターネット利用者全体の53.9%が利用する代表的なサービスのひとつとなっている[1]。携帯インターネットからの利用も可能になってきたことから、今後もネットオークションを利用するユーザは増えることが予想される。しかし、ネットオークションにおける商品取引では匿名性や非対面性が高いため、ユーザが入手し得る情報には限りがある。従って、取引に関わるユーザ自身が商品取引の妥当性を簡単に判断できる方法が必要とされている。

本報告では、このような背景を踏まえ、ネットオークション

で行われる商品取引の分析、モデル化を行い、同定モデルと実際の商品取引の差分から商品取引の妥当性を判断することを試みる。商品取引の分析方法としては、特定のユーザの過去取引行動を分析する方法と同一商品の取引履歴を分析する方法とが考えられる。ユーザの過去の取引行動の分析は、出品者や入札者の信頼性を測るのに適しているが、一人一人の取引情報を過去に遡って分析するのは、多大な計算時間が必要であるだけでなく、情報の収集自身も困難である。そこで、今回の特性分析では、情報の入手、分析が容易に行える同一の商品の取引履歴を分析する。また、分析項目としては、商品取引の特徴を最も的確に表すと考えられる落札価格の時間変化に着目する。

以下に、本報告の構成を示す。まず、2. 節で、準備として、本報告で用いる分析手法について説明する。3. 節では、データの取得方法、分析手順を説明し、4. 節で分析結果を述べる。最後に、5. で本報告のまとめを行う。

2. 準備

分析にあたっては、落札価格 (Y) そのものではなく、落札価格の対数 ($\ln(Y)$) を分析の対象とする。これは、価格が大きいほど、価格変動の幅も大きいと考えるのが自然なためである。以降の議論では、価格という言葉は特に断らない限り価格の対数を示す。

以下で、本報告で用いる分析手法を簡単に説明する。

2.1 モデルの選定 [2]

次式で表される一次の自己回帰モデルを考える。

$$X_{t+\delta t} = \rho X_t + \mu_t + \epsilon_t \quad (1)$$

ここで、 ϵ_t は独立、同一の分布に従う確率変数である。

このモデルは、 μ_t が時間によらず一定値 μ を取り、特性方程式の根 (ϕ) が絶対値で 1 より小さい場合に定常であるという。モデルが非定常となる場合で、 $|\phi| > 1$ (発散)、 $\phi = -1$ (振動) は時系列データとしては分析の意味をなさないの、通常 $\phi = 1$ の場合のみを問題にする。モデルの定常性を検定するために、式 (1) を次のように変形する。

$$X_{t+\delta t} - X_t = \delta X_t + \mu + \epsilon_t \quad (2)$$

$$\rho = 1 + \delta \quad (3)$$

回帰分析により、式 (2) 中の δ 、 μ を決め、 $\delta = 0$ の仮説を検定することは、定常性の仮説を検定することを意味する。 $\delta = 0$ の検定は、 $\hat{\tau} = \hat{\delta}/s_{\delta}$ ($\hat{\delta}$: δ の推定値、 s_{δ} : $\hat{\delta}$ の標準誤差) を $\hat{\tau}$ 分布の下側 $\alpha\%$ の臨界値と比べることによって行う。 $\hat{\tau}$ 値が臨界値よりも小さい場合、仮説: $\delta = 0$ (非定常であるという仮説) を棄却する。

モデルが非定常の場合は、さらに階差をとって定常性の検定を行う、つまり、 $X_{t+1} - X_t$ の定常性を同様の手順で検定する。この手順を繰り返すことにより、モデルの次数を決める。

2.2 正規分布性の検定 [3]

正規分布性の検定としては、Jarque-Bera 検定、D'Agostino の D 検定の二つを用いる。

以下、それぞれについて説明する。

- Jarque-Bera 検定:

歪度 ($\sqrt{b_1}$)、尖度 (b_2) により、以下の量 (JB) を計算する。

$$JB = \frac{nb_1}{6} + \frac{nb_2^2}{24} \quad (4)$$

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} \quad (5)$$

$$b_2 = \frac{m_4}{m_2^2} \quad (6)$$

ここで m_k は、平均のまわりの k 次モーメントを表す。

歪度、尖度は、正規分布で 0 になる。従って、 JB が 0 に近いほど、正規分布であるという仮説に整合的だといえる。 JB テストの臨界点としては、[3] の数表に記載された値を用いる。

• D'Agostino の D 検定: 式 (7) ~ (9) から計算される Y の 0 からの乖離度によって、正規分布であるという仮説の整合性を判断する。 D テストの臨界点についても [3] の数表に記

載された値を用いる。

$$Y = \frac{\sqrt{n}(D - 0.28209479)}{0.02998598} \quad (7)$$

$$D = \frac{\sum_{i=1}^n c_i X_{(i)}}{n^{3/2} \{\sum_{i=1}^n (X_i - \bar{X})^2\}} \quad (8)$$

$$c_i = i - \frac{1}{2}(n+1) \quad (9)$$

$X_{(i)}$ は、 X_1, \dots, X_n を大きさの順に並べ替えた変数である (つまり、 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$)。

3. 分析方法

本節では、データの取得、分析手順を説明する。

3.1 データの取得方法

オークション統計サイト [4] から、商品名を指定して過去一ヶ月間の落札価格、落札日情報を取得した。なお、統計分析が意味をなす最低限のサンプル数という意味で、分析対象は落札件数 30 件以上の商品に限定した。

3.2 モデルの選定と検証

取得したデータを、以下の手順により分析する。

- ステップ 1: 明らかに落札価格の高低に影響を与えるキーワード^(注1)を含む商品を取り除く。
- ステップ 2: 回帰分析により、式 (2) のパラメータ δ 、 μ を決める。2. 節の手順により、モデルの次数を決める。
- ステップ 3: モデルから計算される価格と実際の価格の差 (残差) の分布について、正規分布性を検定する。選択したモデル、パラメータが妥当であれば、残差は i.i.d. となるので、サンプル数が十分に多い場合、残差の分布は正規分布に従うはずである。

4. 分析結果

前節の手順に従って、14 種類の商品の落札価格 (対数値) の分析を行った。

4.1 モデルの選定と適合性

一ヶ月の落札件数を N とすると、一日の平均落札件数 n は、 $n = \lceil N/30 \rceil$ となる。 $\delta t = 2n$ (2 日相当の意味) とし、 $\hat{\tau}$ を計算した。短期価格変動に着目して評価を行うのは、直近の商品の購入を検討しているユーザにとっては、近未来の落札価格の分布特性を知ることが、より重要であると考えたためである。各々の商品の $\hat{\tau}$ 値を、図 1 に示す。全ての商品で $\hat{\tau}$ は 9.5% 臨界値を下回り、定常性の仮説が支持された。

次に、残差の正規分布性を検定したところ、14 品目中 9 品目が正規分布に適合した。各々の商品について、残差分布の JB 値 (式 (4) より計算)、 Y 値 (式 (7) より計算) を、図 2, 3 に示す。更に、残差が正規分布とならない 5 品目について、価格変動の時間変化を図 4~8 に示す。全ての図において、落札価格が定価をこえる取引があることが分かる。一般的には、商品そのものの価値を超えた価格で落札することはないと考えら

(注1): 未使用品、ジャンク (商品 ID1)、全巻新刊 (商品 ID2)、全巻帯付 (商品 ID7)、ジャンク (商品 ID9)、傷あり、訳あり (商品 ID12)、発売日前の予約販売 (商品 ID13)。

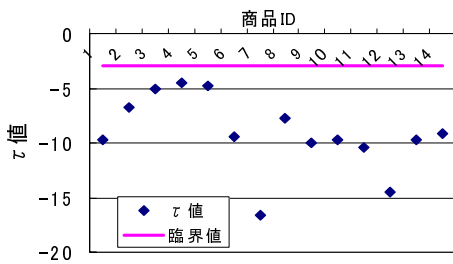


図1 落札価格の対数に関する z 値

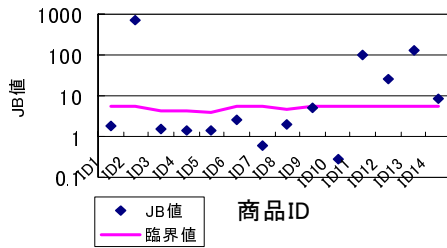


図2 残差に関する JB 値

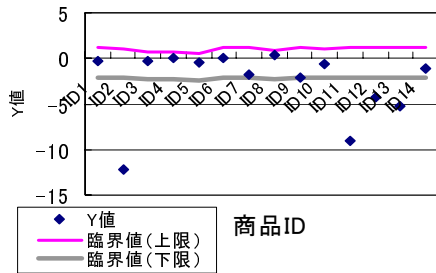


図3 残差に関する Y 値

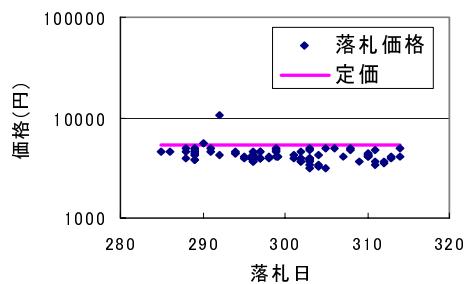


図4 商品 ID2 の落札価格

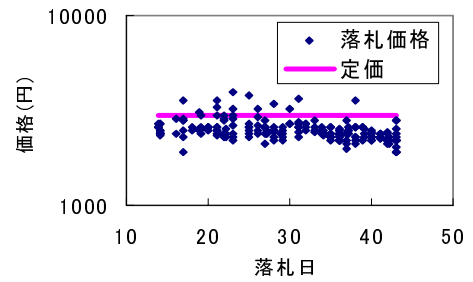


図5 商品 ID11 の落札価格

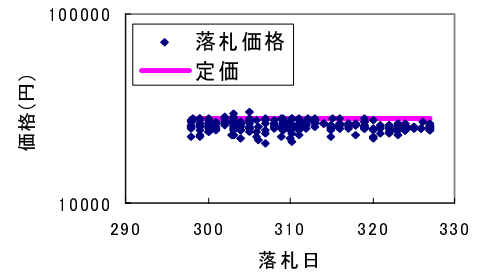


図6 商品 ID12 の落札価格

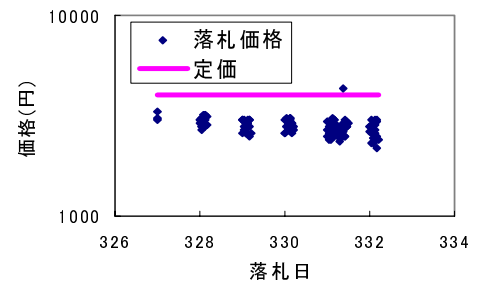


図7 商品 ID13 の落札価格

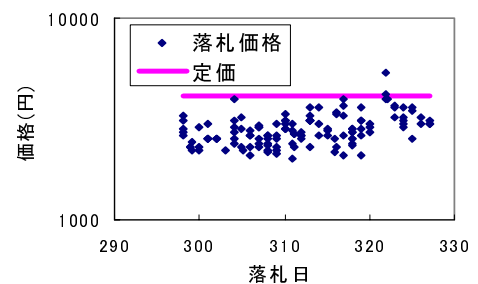


図8 商品 ID14 の落札価格

れるので、落札価格が定価をこえる場合を特異データとして取り除き、3. 節の(ステップ2)(ステップ3)を再度適用した(図9, 10)。図9, 10から分かるように、定価以上の落札価格を除外すると、4品目中、3品目(ID2, 11, 13, 14)が正規分布に適合した。これらの3つの商品については、図5, 7, 8より、定価以上で落札される取引とそれ以外の取引の落札価格が明らかに離れている、つまり、定価以上の落札が特異であることが示唆されている。一方、残差が正規分布に適合しない商品ID12の商品は、落札価格全体が定価近傍に集まっている(図6)。つまり、落札価格が定価以上であることが、必ずしも

特異ではないといえる。そこで、落札価格と定価の距離を、各商品ごとに比較する。指標としては、定価以上で落札された取引を除いた平均落札価格($\hat{\mu}$)と定価(U)の差分を残差の標準偏差(σ)で正規化した量($d = (U - \hat{\mu})/\sigma$)を用いる(表1)。この結果からも、商品ID12の商品は他の商品に比べて定価と落札価格の差が小さいことが確認できる。

定価と落札価格の差が小さい場合に、残差が正規分布となら

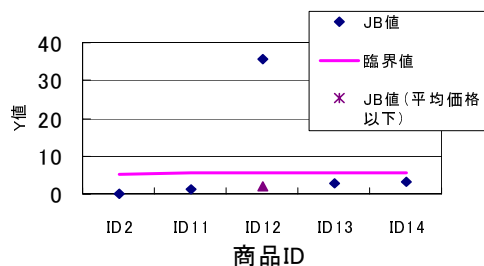


図 9 残差の JB 値 (定価以下の取引)

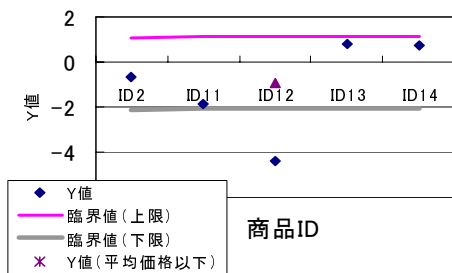


図 10 残差の Y 値 (定価以下の取引)

ID	ID2	ID11	ID12	ID13	ID14
d	2.1	2.4	1.5	4.7	2.5

表 1 平均落札価格と定価の距離

ないのは、落札価格が定価に近い領域では、落札価格の変動が定価で頭打ちになり、価格の変動幅が小さくなるためだと考えられる。そこで、商品 ID12 について、平均落札価格以下の取引のみを取り出して、再度 (ステップ 2) (ステップ 3) により分析したところ残差は正規分布に適合した (図 9, 10)。また、平均落札価格以下の取引のみをモデル化した場合の残差の標準偏差 (0.057) は、落札価格が定価以下の全ての取引についてモデル化した場合の残差の標準偏差 (0.048) よりも大きい。これらの結果から、落札価格の変動が高価格領域で小さくなっていること、落札価格が定価よりも十分離れた領域では、他の商品と同様に、価格変動が定常モデルで記述できるといえる。

5. ま と め

本報告では、ネットオークションにおける落札価格の分析、モデル化を行った。分析の結果、落札価格が定価よりも十分離れた領域では、落札価格は定常モデルとして記述できることがわかった。つまり、同定モデルから計算される値と実際の落札価格の差 (残差) の分布は正規分布となり、残差が落札価格の妥当性を判断する一基準となり得る。

文 献

- [1] 総務省, "情報通信白書平成 17 年度版", <http://www.johotsusintokei.soumu.go.jp/whitepaper/ja/h17/pdf/index.html/>
- [2] 山本 拓著, "経済の時系列分析", 創文社, 1998.
- [3] 蓑谷 千鳳彦著, "ブラックショールズモデル", 東洋経済新報社, 2000.
- [4] オークション統計ページ (仮), <http://www.aucfan.com/>