

12-DA01

計算集約的統計手法による大規模経済データの実証分析

大西 立顕（東京大学）

概要 経済現象の本質である非ガウス性・非一様性・非定常性に注目し，超並列計算による計算集約的統計手法を用いて，大規模な経済データを実証分析した．まず，住宅価格のデータを用いて不動産バブルを分析した．バブル期は人々が近視眼的になるため物件間の価格格差が大きくなり，価格が同質とみなせる裁定が成立する地域の広さ（同一需給圏）が小さくなる．このことを用いて，住宅価格が対数正規分布から乖離する度合いから同一需給圏の大きさを定義し，バブルの度合いを定量化する手法を開発した．実データにこの手法を適用し，バブルの時空間ダイナミクスを明らかにした．また，企業間取引ネットワークやニュースアーカイブデータについても分析を行い，計算集約的統計手法の有効性を提示した．

1. 研究の目的と意義

情報の電子化が進んだことにより，日々，様々な分野で膨大な量のデータが蓄積されるようになってきている．こうしたビッグデータを利活用することで，経済・社会現象を実証科学的に研究し，社会やビジネスに役立ったりする試みが活発になってきている．また，コンピューターの計算能力の急速な向上に伴い，今まではほとんど不可能に近かった，大量のデータを用いて大量の計算を行うような分析が可能になり，計算集約的な統計手法による厳密な分析ができるようになってきている．

経済システムは，売り手と買い手のお金やモノの交換（相互作用）として捉えることができる．相互作用があっても個々がランダムにゆらげば，全体のゆらぎは正規分布に従う．しかし，現実の経済システムでは，常に競争が働き相互作用が強いため，全体のゆらぎは極めて大きく，関連する変数はベキ分布（図1）に従う（非ガウス性）．また，経済現象では，要素（経済主体）の異質性・多様性が強く，作用・反作用の法則のような対称性が存在しないため，関係性を多体の方向つき相互作用として捉える必要がある（非一様性）．さらに，変動の性質や規則性は時々刻々変化するため，時期・環境の状態変化を考慮する必要もある（非定常性）．つまり，非ガウス性・非一様性・非定常性が現象の本質であり，既存のほとんどの統計的・数理的手法はこれらを前提としないため，必ずしも有効なものになっていない．特に，経済・社会現象においては大変動（非ガウス性）の影響が非常に大きいいため，本質的に重要になるのは金融危機や大震災のリスク評価のような大変動が支配する現象であ

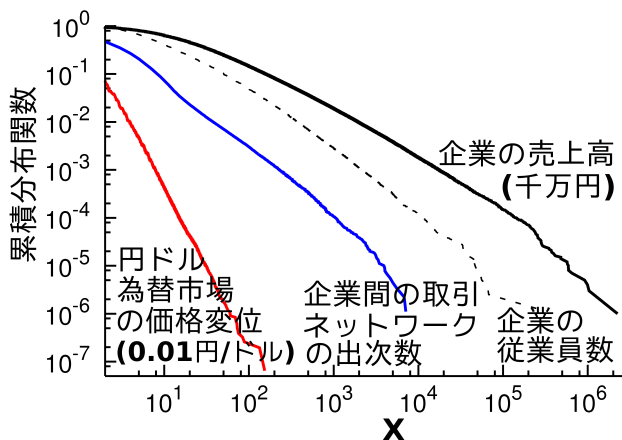


図 1: 経済現象に関連する様々な変数の累積分布．

る．このような非ガウス性・非一様性・非定常性を厳密に考慮した解析を行うには，データドリブン（データ駆動）の思考で実データを基盤とした分析を行う必要がある．

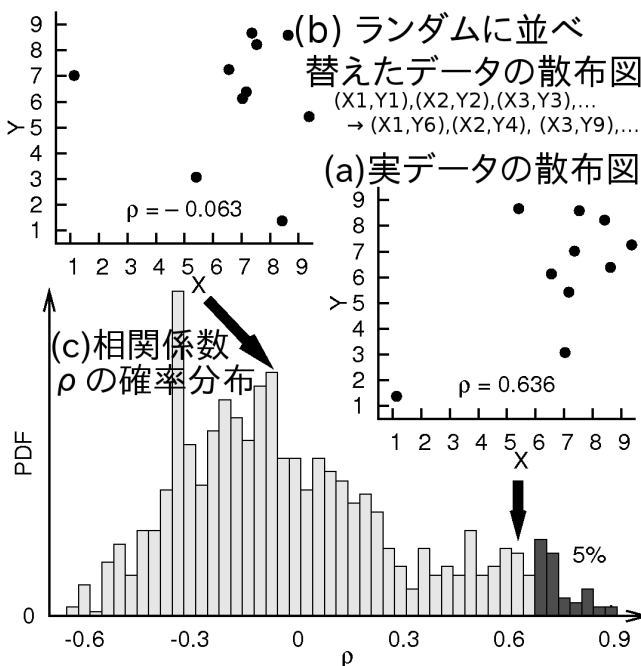


図 2: 計算集約的統計手法の例 (並べ替え検定) .

実データを基盤とした分析の例として，2変数 X, Y の間の相関を求める問題を考える．実データから図 2(a) のような散布図が得られたとする．通常相関係数の検定では， X の分布も Y の分布も近似的に正規分布に従うと仮定するため，有意な相関であると判断されてしまう．しかし，実際には，このデータの X の分布と Y の分布を保存したままで，並べ替えて対の作り方をランダムにしたデータを何個も作成し（一例を図 2(b) に示す），それらの相関係数の分布（図 2(c)）を用いて（分布を仮定せず）正確に検定を行うと，実際には無相関であることが判明する．このような並べ替え検定により，あらゆる統計検定は正確に行うことができる．つまり，莫大な量のランダムデータの計算に集約することで，仮定・近似を入れずに簡潔・明解に正確な分析を行う

ことが可能になる．値を順位に粗視化する順位相関係数などの通常のノンパラメトリック手法では，情報の削ぎ落としにより検出力が低下するが，並べ替え検定では値をそのまま扱うため検出力の低下は生じない．このような計算集約的統計手法を用いることで，非ガウス性・非一様性・非定常性を厳密に扱った解析が可能になる．

計算集約的統計手法は莫大な量の計算を必要とするため，これまで現実にはあまり行われていない．本研究は，スーパーコンピュータによる超並列計算を用いてこれを行うものである．この手法では，全く同じ計算をパラメータや乱数の種だけを変えて何回も行う．そのため，基本的に各コアに独立に計算を行わせればよいから，並列計算に適しており実装も容易である．本研究では，主に東京大学情報基盤センター FX10 スーパーコンピュータシステムを用いて，MPI による並列化により 12~252 ノードで計算を行った．

2. 当拠点公募型共同研究として実施した意義

- (1) 共同研究を実施した大学名と研究体制
 - 東京大学（代表者：大西 立顕）
 - 東京大学（副代表者：渡辺 努）
 - 麗澤大学（清水 千弘）
 - 筑波大学（水野 貴之）
 - 金沢学院大学（藤本 祥二）
 - スイス連邦工科大学（久野 遼平）
- (2) 共同研究分野
 - 超大規模データ処理系応用分野
- (3) 当公募型共同研究ならではの事項など
 - 東京大学情報基盤センター FX10 スーパーコンピュータシステムを使用して，効率的に大規模計算を行うことができた．当拠点シンポジウムにおいて，他分野の研究者と情報交

換することができ、大規模計算機利用に関して有意義な知識が得られた。

3. 研究成果の詳細と当初計画の達成状況

(1) 研究成果の詳細について

【不動産バブルの検出手法】

経済の大きな変動は不動産価格の大きな変動が原因となって生じている。たとえば、1980年代後半の日本のバブル景気は不動産価格の高騰によるものであり、2008年頃の世界金融危機は米国の住宅バブル崩壊によるものである。不動産価格は、個人・企業における資産選択や政府・中央銀行の政策運営に大きな影響を及ぼすため、不動産バブルの性質を知ることが重要な研究課題である。不動産バブルが起きているかどうかを事前に、あるいは、リアルタイムに知ることは困難である。バブルがはじけてはじめて、それがバブルだったと認識できるというのが通説である。不動産バブルとは、不動産価格がファンダメンタルズ(本来あるべき価格水準)から乖離することであると定義できる。ファンダメンタルズを正しく推計できれば、実際の価格との乖離を計算することで、バブルかどうか判定することが原理的には可能である。しかし、ファンダメンタルズを精度高く推計することはできないため、現実にはこれは非常に難しい。

本研究では、ファンダメンタルズを推定するアプローチではなく、物件の価格分布の地域間格差からバブルを定義する方法を考える。バブル時には、すべて物件の価格が一律に上がるのではなく、急速に価格が上がる物件とそうでない物件があるという点に注目する。つまり、バブルが発生すると価格の地域間格差が高まると考えられる。住宅の価格はさまざまな要素を合計したものとみなせる。たとえ

ば、住宅の広さ、どこの沿線か、最寄り駅から何分か、南向きか否かなどさまざまな要素のそれぞれに価格があり、それらを積み上げた結果として住宅の価格が決まっている。これは住宅価格のヘドニック分析の根底にある考え方である。住宅価格がこのようなさまざまな要因の和として決まっているとすれば、中心極限定理により住宅価格のクロスセクション分布は対数正規分布に従うはずである。つまり、バブルでない平常時には住宅価格は対数正規分布に従う。この意味で対数正規分布は重要なベンチマークである。これに対してバブル期には物件間の価格格差が大きくなることに伴い、このベンチマークから乖離する。つまり、物件間の価格格差が存在しない地理的な範囲の広さ(同一需給圏)がバブル期には大きくなると考えられる。

【首都圏の住宅地価格】

実データから上記のことを確認するために、首都圏の住宅地に関する地価公示価格のデータ(1974~2008年の年次、地点数は年に約2000~6000地点)を年毎に分析した。平均0、分散1に規格化した対数価格：

$$\frac{\log P - \langle \log P \rangle}{\sqrt{\langle (\log P - \langle \log P \rangle)^2 \rangle}}$$

は、バブル期を除き正規分布に従うことが判明した(図3)。しかし、バブル期は対数正規分布から統計的に有意に乖離しており、対数正規分布より裾の長いベキ分布(パレート分布)：

$$Pr(\geq P) \propto P^{-\alpha}, \quad \alpha = 2$$

に従うことが判明した(図4)。つまり、価格は通常は対数正規分布に従っているが、バブル期には物件間の価格裁定の及ぶ範囲(同一需給圏)が狭くなるため、ベキ分布に従っている。

この性質を用いて、価格分布から同一需給圏の大きさ(異質性の度合い)を推定すること

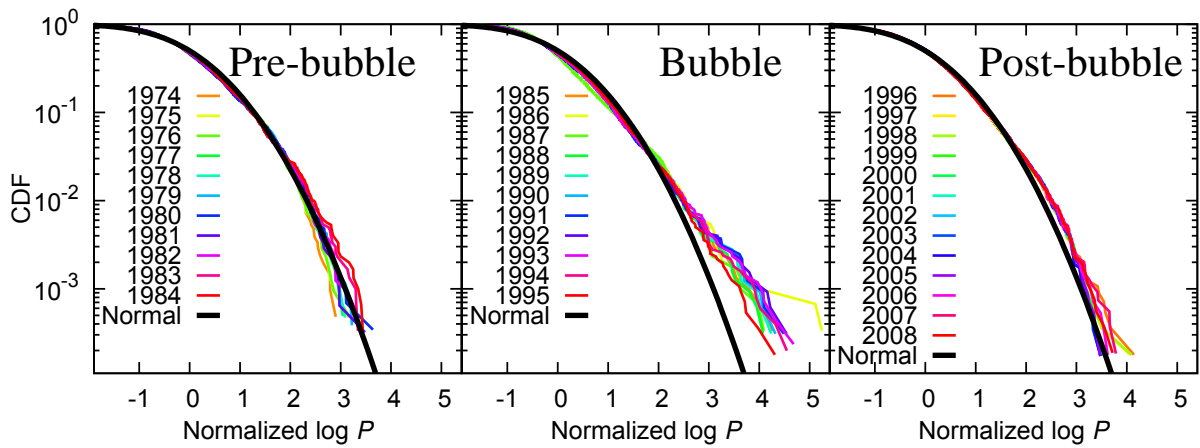


図 3: 首都圏の住宅地について、年別に観測した規格化した対数価格の累積分布。

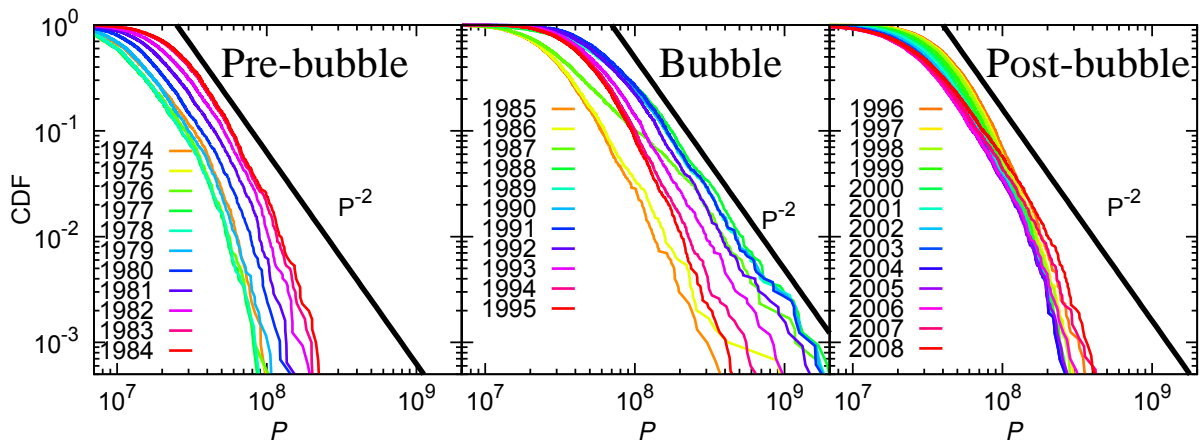


図 4: 首都圏の住宅地について、年別に観測した価格の累積分布。

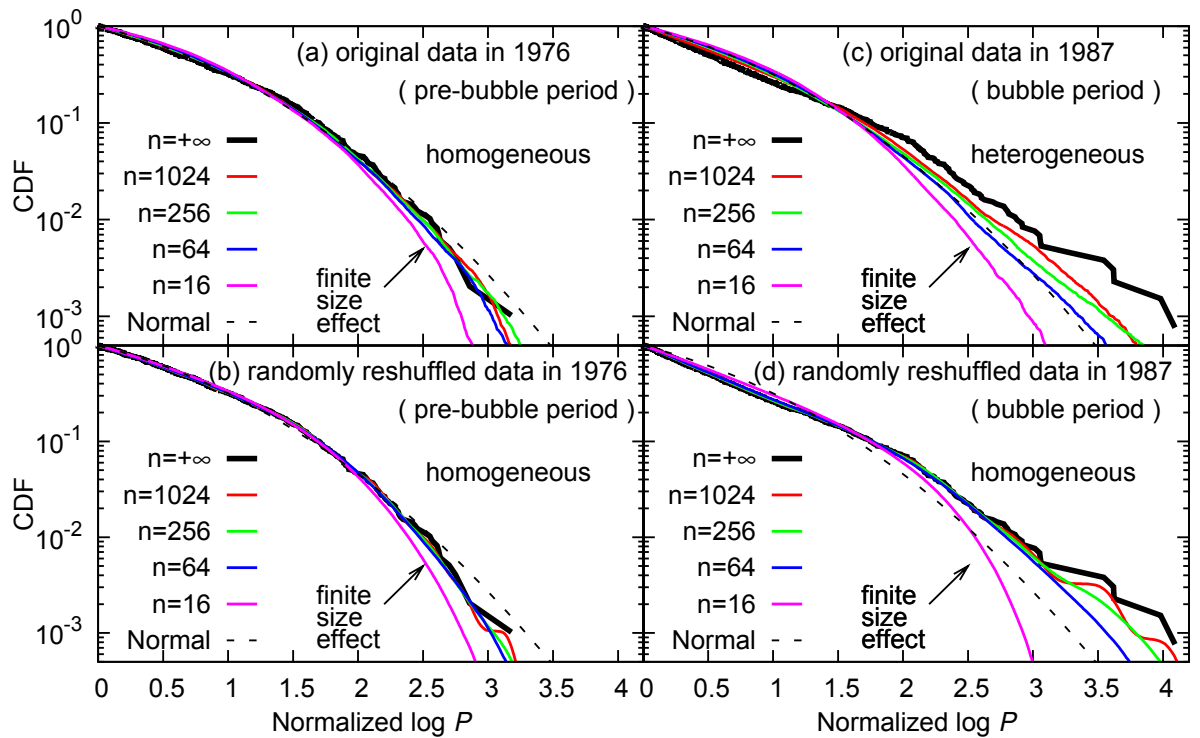


図 5: 首都圏の住宅地について、(a) バブル前と (c) バブル期での規格化した対数価格の累積分布。(b) と (d) は位置 (緯度・経度) をランダムに入れ替えたデータでの結果。

を考える．ランダムに住宅地を一つ選び，その最近接住宅地を $n - 1$ 個取り出し，これら n 個の住宅地を一つの地域と定義するという操作を繰り返すことにより，全領域を複数の地域（一つの地域は n 個の住宅地で構成される）に分割し，各地域内における価格分布を調べた．この操作による領域の分割は一意には定まらないため，乱数の種を変えて 128 通りの分割の仕方について分析した．その結果，近接住宅地数 n を小さくして狭い領域に限定してみれば，バブル期においても価格は対数正規分布に従うことが分かった（図 5(c)）．つまり，バブル期には同一需給圏の大きさ（裁定が成立し一物一価になっている，価格が同質とみなせる地域の広さ） n が小さく，地域の異質性が強く，各地域内では価格は同質だが異なる地域間では異質になっている．このように算出した同一需給圏の大きさ n はバブルの度合いを表す指標になると考えられる．

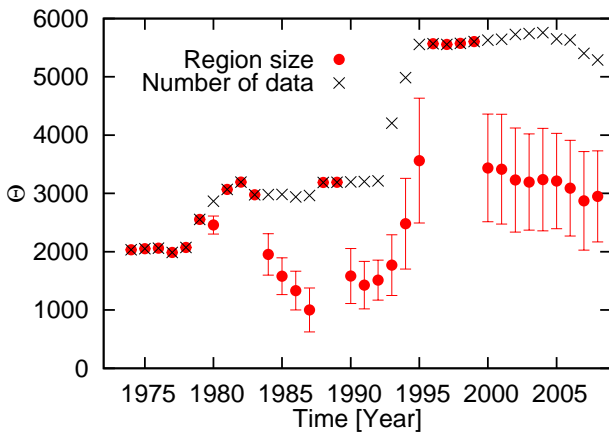


図 6: バブルの度合い θ の年次推移 (●)． \times は全地点数を示す．

そこで， $n = (\text{全物件数})$ の場合からこの分析をはじめ，徐々に n の値を小さくしていくとき，はじめて価格分布が対数正規分布になる n の値（つまり，対数正規分布とみなせる最大の n ）をバブルの度合い θ と定義する．ここで， θ の値は，平常時は大きくなり（最大値は全物件数），バブル期は小さくなるように定

義されていることに注意する．分布が対数正規分布とみなせるかどうかは，コルモゴロフ・スミルノフ検定（有意水準 5%）を用いて判定した．算出した θ の値の年次推移は図 6 のようになり，バブルといわれる時期ほど小さな値をとることが確認できた．

【米国の住宅価格】

このように定義したバブルの度合いが現実のバブルの程度を反映していることをさらに確認するために，Zillow.com から取得した米国の住宅価格のデータ（2011 年 9 月の 1 時点，全物件数は約 11 万件）を用いて，州の違いに注目してバブルの度合い θ を州別に計算した．得られた θ の値と 2011 年 10 月時点でのその州の物件の差し押さえ率（その州の全物件数をその州の差し押さえ物件数で割った量）の関係を調べた．ここで，差し押さえ率は，差し押さえが多いほど小さい値をとることに注意する．各州の θ をその州の全物件数で割った量は，差し押さえ率と相関しており（図 7），バブルになっている州ほど物件の差し押さえが多い傾向がある．したがって， θ は現実のバブルの程度を反映した量になっている．

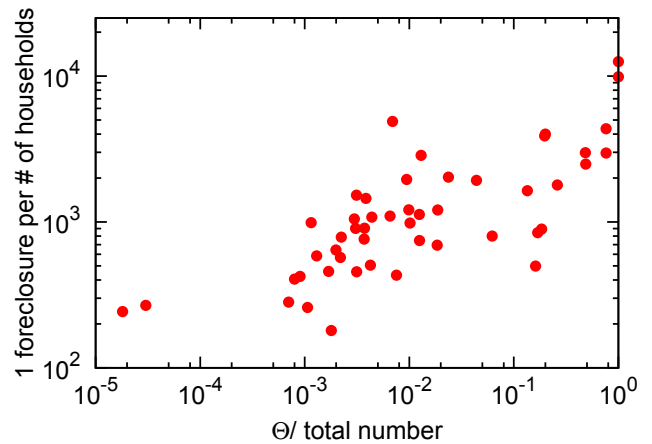


図 7: 各州について，差し押さえ率とバブルの度合い θ の散布図．

【首都圏の住宅取引価格】

バブルが空間的にどのように発生し，伝播・波及し，崩壊していくかを観測するため

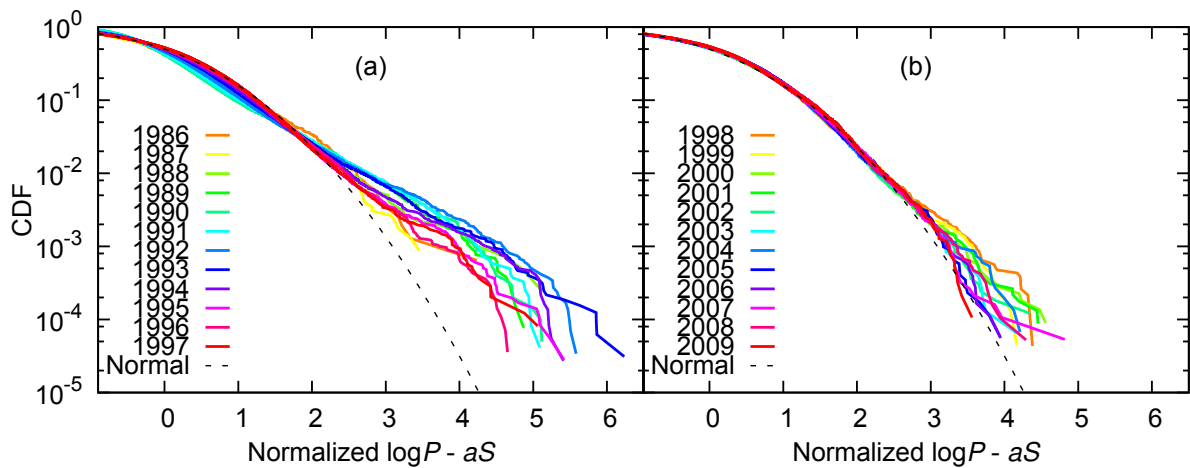


図 8: 首都圏の住宅取引について，年別に観測した規格化した面積調整価格の累積分布．

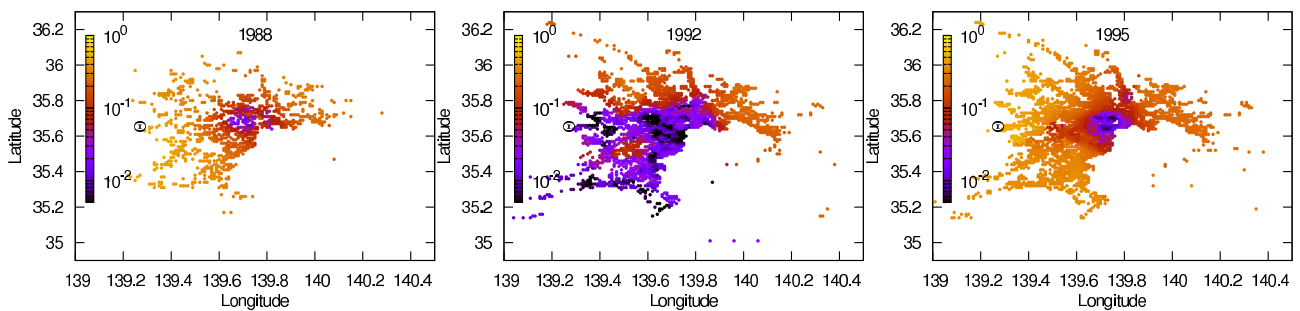


図 9: バブルの度合いの空間分布 (1988 年，1992 年，1995 年)．

に，リクルート社提供の物件別住宅取引データ (1986~2008 年の週次，全物件数は約 72 万件) を年毎に分析した．まず，価格 P と物件属性 (専有面積，築後日数，階，構造など) との関係調べた結果，価格は専有面積 S と強く相関し，専有面積が広がるほど住宅価格は指数的に高くなることを発見した．この関係を用いて導入した面積調整価格 $Q = \log P - aS$ は，バブル期はベキ分布に，バブル後は正規分布に従うことを明らかにした (図 8)．さらに，一週間毎に分布を観測しても同じ結果が得られることを確認した．

いま， Θ の空間分布を計算するために，ある特定の地域についての局所的なバブルの度合いを定義する．ある一つの物件 i に注目し，物件 i と物件 i に近接する n_i 個の物件をまとめて一つのグループとすると，このグループ内の物件の面積調整価格の分布が正規分布にな

るかどうかを検定する．計算コストを節約するため，正規性の検定にはギアリー (Geary) の検定 (有意水準 5%) を用いる．まず， $n_i = N$ (その年の全物件数) の場合からはじめ，徐々に n_i の値を小さくしていくとき，はじめて分布が正規分布になる n_i の値を n_i^* とするとき，物件 i のバブルの度合いを $\Theta_i = n_i^*/N$ と定義する．あらゆる物件についてこの計算を行うことで，バブルの度合いの空間分布を求めた (図 9)．1988 年頃に都心で発生したバブルが，都心の南西部 (神奈川県付近) へ波及し，1992 年にバブルの度合いが一番強くなり，その後，南西部からバブルが終息し，1998 年以降は完全にバブルが消失し，首都圏全域が同一需給圏になる様子が定量的に観測できた．

【企業間取引ネットワーク】

日本企業約 100 万社の取引関係・企業財務データを各ノード (企業) が多変量情報を持つ

100万ノード・400万リンクの有向ネットワークとして分析した。企業間ネットワークは、同じ業種同士でつながりやすい傾向がある。そこで、業種の異質性を考慮した上で統計的有意なつながりを特定するために、各業種で次数(○←, ○→, ○↔の数)を保存してランダムにつなぎ替えたランダムイズドネットワークを作成した(図10, 11)。隣接する3企業間の取引関係を表現する構造(部分グラフ)に注目し、実ネットワークとランダムイズドネットワークを比較した。その結果、3企業は図12のような構造を取りやすく、V字構造は出現しにくいことを明らかにした。さらに、各業種の経済学的機能を反映する特徴的なモチーフ構造を抽出した。

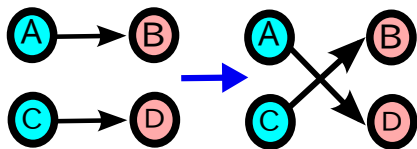


図 10: 各業種で次数を保存したつなぎ替え。

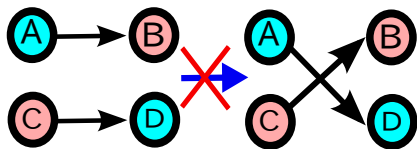


図 11: 各業種で次数が保存されないつなぎ替えは行わない。

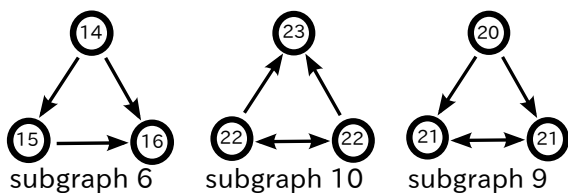


図 12: 統計的有意に出現する部分グラフ(モチーフ)。

【ニュースアーカイブ】

市場価格変動の高頻度時系列にみられる非定常性は、ニュースのような外乱により引き起こされると考えられる。市場がニュースに反

応して変動していることを調べるために、ロイター社のニュースアーカイブのデータ(2003~2011年の約2億件の記事)を言語処理の手法を用いて分析した。ニュースをその内容を表わすトピックに分類した上で、ニュースが株式市場の株価に及ぼす影響を解析した結果、株価のボラティリティの多くはニュースで説明つくことを明らかにした。

(2) 当初計画の達成状況について

超並列計算を用いた計算集約的統計手法により、大規模な経済データを実証分析し、従来の手法では得ることのできなかった新しい経済学的知見を明らかにすることができた。計算集約的統計手法による分析方法を提示し、その有効性を検証するという当初計画を達成することができた。

4. 今後の展望

このような不動産価格の分析から、同一需給圏の広さを用いてバブルの度合いを定量化することができた。今後、住宅取引価格のデータを用いて、より短時間の時間解像度(週次)で観測し、詳細な分析を行う。また、全国や海外のデータについても分析を進め、理解を深める。さらに、対数正規分布のどのような重ね合わせでベキ分布が生じているかを調べ、不動産バブルのメカニズムを考察する。

これまでに行ってきた解析から、計算集約的統計手法の考え方をういた経済データ分析の有効性が確認できた。外国為替・株式・先物などの金融市場データ、ニュース・新聞記事の文字データ、商業統計メッシュデータ、人口メッシュデータなどの様々な実データについても分析を適用し、新たな知見の発見を試みる。

5. 研究成果リスト

(1) 学術論文

- 大西立顕, 水野貴之, 清水千弘, 渡辺努, ”米国における住宅の価格分布と差押え率”, 統計数理研究所共同研究レポート (印刷中)
- Ryohei Hisano, Didier Sornette, Takayuki Mizuno, Takaaki Ohnishi, Tsutomu Watanabe, ”High quality topic extraction from business news explains abnormal financial market volatility”, arXiv preprint arXiv:1210.6321
- 大西立顕, ”非定常性に挑む外国為替市場の科学”, システム/制御/情報:システム制御情報学会誌, vol.56, no.10, pp.550–555
- Takayuki Mizuno, Kazumasa Takei, Takaaki Ohnishi, Tsutomu Watanabe, ”Temporal and Cross Correlations in Business News”, Progress of Theoretical Physics Supplement, no.194, pp.181–192

(2) 国際会議プロシーディングス

- Takaaki Ohnishi, Takayuki Mizuno, Chihiro Shimizu, Tsutomu Watanabe, ”Power Laws in Real Estate Prices during Bubble Periods”, International Journal of Modern Physics: Conference Series, vol.16, no.1, pp.61–81

(3) 国際会議発表

- Takaaki Ohnishi, Hideki Takayasu, Misako Takayasu, ”Colored Motifs in Japanese Inter-firm Network”, 4th Workshop on Complex Networks, Berlin, Germany
- Takaaki Ohnishi, Takayuki Mizuno, Chihiro Shimizu, Tsutomu Watanabe, ”Using Property Price Distribution to Detect Real Estate Bubbles”, 6th

CSDA International Conference on Computational and Financial Econometrics (CFE 2012), Oviedo, Spain

- Takaaki Ohnishi, ”Spatial Heterogeneity and Real Estate Bubble”, Econophysics-Kolkata VII, Kolkata, India
- Takaaki Ohnishi, Hideki Takayasu, Misako Takayasu, ”Structure and function of inter-firm relationships revealed by network motif analysis”, XXXII Dynamics Days Europe, Gothenburg, Sweden
- Takaaki Ohnishi, Hideki Takayasu, Misako Takayasu, ”Using Network Motif Analysis to Characterize Inter-firm Transactions”, NetSci2012, Evanston, Illinois

(4) 国内会議発表

- 大西立顕, 水野貴之, 清水千弘, 渡辺努, ”日本における住宅バブルの形成と崩壊の時空間構造”, 日本物理学会第68回年次大会, 広島県
- 大西立顕, 水野貴之, 清水千弘, 渡辺努, ”Characteristics of real estate bubble”, 日本経済学会2012年度秋季大会, 福岡県
- 大西立顕, 水野貴之, 清水千弘, 渡辺努, ”住宅市場における投機行動とバブル生成”, 日本物理学会2012年秋季大会, 神奈川県
- 大西立顕, ”住宅バブルを特徴づける価格の空間相関”, 日本行動計量学会第40回大会, 新潟県
- 大西立顕, 高安秀樹, 高安美佐子, ”ネットワークモチーフが明らかにする企業間の取引関係”, 平成24年度統数研共同研究集会「経済物理学とその周辺」第1回研究会, 東京都

(5) その他 (特許, プレス発表, 著書等) 該当なし