



## Kriging による 公示地価の推定

西郷 浩(早大・政経)  
増成敬三(早大・経研)  
2004年10月29日・応用統計WS(東大)

1



## 報告の流れ

- (官庁)統計データにおける空間統計学的手法の有効性
- 本報告の目的
- 公示地価の予測
  - 公示地価の分布
  - 公示地価の推定モデル
  - 推定結果
- 今後の課題

2



## (官庁)統計データにおける 空間統計的手法の有効性(1)

- 官庁統計における空間情報の利用
  - 標本設計の段階:
    - 層別の基準: 都道府県、商業圏・住宅圏
    - 抽出単位: 市町村、調査区
    - ロータション・サンプリング  
調査効率・調査客体の負担・推定効率への考慮。
  - 推定の段階:
    - 空間的情報を十分に活用してはいない。

3



## (官庁)統計データにおける 空間統計的手法の有効性(2)

- 労働力調査:
  - 都道府県別の失業率の試算を開始(H14)
    - 労働力調査のサンプル・サイズ:
      - 全国表章が基準
      - 都道府県別の表章にはたえないこともある。
    - サンプルサイズの不足をどう補うか?:
      - 小地域統計的手法の利用: 「空間情報を生かす」という発想は強調されていない。
      - 空間統計的手法の利用: 「空間情報である」ことを生かしながら近隣地域のデータを補助情報として利用する。

4

## (官庁)統計データにおける 空間統計的手法の有効性(3)

- 土地価格の推定
  - 全国消費実態調査:
    - 最近隣3点平均法
  - 国土交通省の研究(土地基本調査):
    - 回帰式の利用
  - 国民生活基礎調査の研究会:
    - 回帰式の利用空間統計的手法を活用してはいない。

5

## (官庁)統計データにおける 空間統計的手法の有効性(4)

- 欧米での状況:
  - M. E. Thompson (1997), *Theory of Sample Surveys*, Chapman&Hall, Chap. 7: Sampling Strategies in Time and Space
  - S. K. Thompson (2002), *Sampling*, Wiley, Chap. 20: Spatial Prediction or Kriging; Chap. 21: Spatial Designs
  - 実際の標本設計に生かされているかどうかは不明。

6

## 今回の報告の目標

- 空間統計的手法による(公示)地価の予測
  - 応用を中心とした研究
    - 使いやすく、かつ自然な手法の模索
    - 理論的な水準は高くありません。
- 注:
  - 以下の報告は、増成敬三の修士論文「空間統計学による地価の推計」に相当の部分を負っている。

7

## 地価に関するデータ

- 公示地価(地価公示制度で公表された地価)
  - 実施機関:国土交通省
  - 調査時点:毎年1月1日
  - 公表時期:毎年3月下旬
  - 評価対象:1平方メートルの(更地の)「正常な」価格
  - 評価方法:不動産鑑定士などの判断を総合
    - 「実勢価格とズレがある」という(よくある)批判あり。
- その他:基準地価、路線価

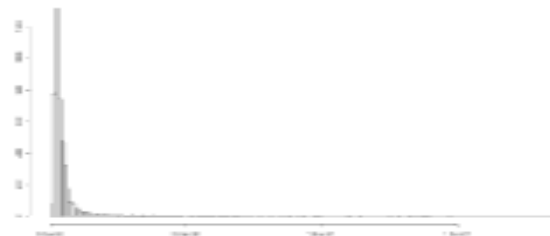
8

## 地価公示(2002年)地点



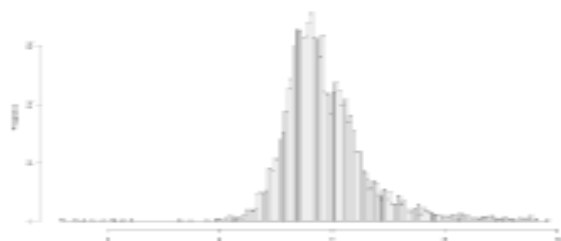
9

## 公示地価<sub>2002</sub> のヒストグラム



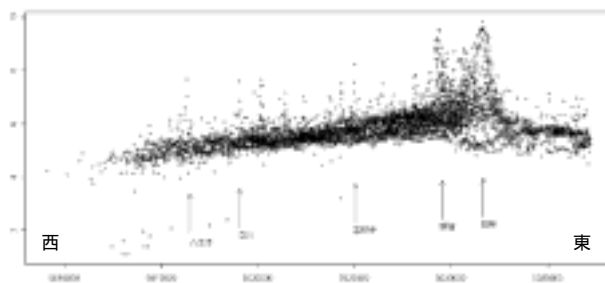
10

## $\log(\text{公示地価}_{2002})$ のヒストグラム



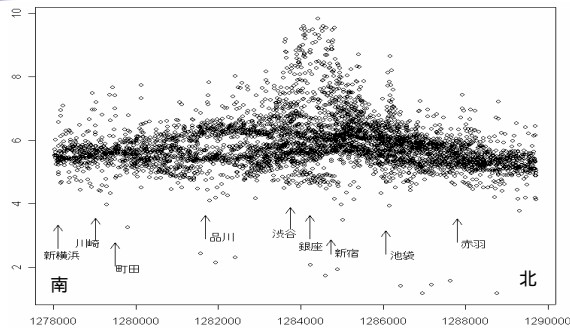
11

## $\log(\text{公示地価}_{2002})$ の空間的様相(1)



12

## log(公示地価<sub>2002</sub>)の空間的様相(2)



13

## log(公示地価<sub>2002</sub>)の空間的様相(3)

- 地価(の対数値)の平均的な水準:
  - 都心(銀座・新宿)から遠くなるほど低下する。
- 空間的な相関:
  - 地価の高い箇所が集中的に発生する傾向がある。
- バラツキ:
  - 意外と安定的(ちょっと楽観的に過ぎるか?)

14

## 公示地価の推定モデル(1)

- Kriging (誤差に空間的相関をもつ回帰モデル)
  - 南アフリカの鉱山技師 Krige に由来する。
  - Matheron によって1950年代後半に理論化。
  - Goldberger (1962) の BLUP を空間的な予測に応用したもの。
  - 推定の流儀が、計量経済学とは若干異なる。

15

## 公示地価の推定モデル(2)

- 推定モデル:

$$Z(s) = \mu(s) + \delta(s), \quad s \in D$$

$$\text{mean: } \mu(s) = \sum_{j=0}^p \beta_j f_j(s)$$

intrinsically stationary

$$\text{error: } E(\delta(s) - \delta(t)) = 0$$

$$V(\delta(s) - \delta(t)) = 2\underbrace{\gamma(s-t)}_{\text{semivariogram}}$$

16

## 公示地価の推定モデル(3)

### ■ Note:

- もし、 $\delta(s)$  が2次定常、

$$E(\delta(s)) = \mu_\delta (= 0) \quad \text{covariogram}$$

$$\text{Cov}(\delta(s), \delta(t)) = C(s - t)$$

であれば、

$$E(\delta(s) - \delta(t)) = 0$$

$$\gamma(s, t) = C(0) - C(s - t)$$

17

## 公示地価の推定モデル(4)

- Kriging では、(semi)variogram によって空間的な相関を表現することが多い。

### ■ そのひとつの理由:

- $s-t$  が等しい観測地点が  $n'$  組のあるとする。また、 $\mu(s) = \mu$  とする (ordinary kriging の想定)。

$$\hat{\gamma}(s-t) = n'^{-1} \sum (Z(s_i) - Z(t_i))^2 \quad \text{unbiased}$$

$$\hat{C}(s-t) = n'^{-1} \sum (Z(s_i) - \bar{Z})(Z(t_i) - \bar{Z}) \quad \text{biased}$$

18

## 公示地価の推定モデル(5)

### ■ optimal linear unbiased predictor:

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i) \quad \begin{array}{l} \text{予測地点} \quad \text{観測地点の数} \end{array}$$

such that

$$\min E(Z(s_0) - \hat{Z}(s_0))^2$$

$$\text{subject to } E(Z(s_0)) = E(\hat{Z}(s_0)) \quad \forall \beta$$

19

## 公示地価の推定モデル(6)

### ■ すなわち:

$$\begin{aligned} \min E & \left( Z(s_0) - \sum_{i=1}^n \lambda_i Z(s_i) \right)^2 \\ & - 2 \sum_{j=0}^p m_j \left( \sum_{i=1}^n \lambda_i f_j(s_i) - f_j(s_0) \right) \end{aligned}$$

Lagrange 乗数

20

## 公示地価の推定モデル(7)

- 仮定:
  - intrinsically stationary  $\delta(s)$
  - $\gamma_0(s)=1$  (定数項) ( $\sum \lambda_j = 1$  となる。)
- 以下の最小化問題を解けばよい。

$$\min - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(s_i - s_j) + 2 \sum_{i=1}^n \lambda_i \gamma(s_0 - s_i) - 2 \sum_{j=0}^p m_j \left( \sum_{i=1}^n \lambda_i f_j(s_i) - f_j(s_0) \right)$$

21

## 公示地価の推定モデル(8)

- 結局、以下の方程式を解けばよい:

$$\begin{bmatrix} \Gamma & X \\ X' & O \end{bmatrix} \begin{bmatrix} \lambda \\ m \end{bmatrix} = \begin{bmatrix} \gamma^0 \\ f^0 \end{bmatrix}$$

where

$$\Gamma = [\gamma(s_i - s_j)] \quad X = [f_j(s_i)] \quad \lambda = (\lambda_1 \dots \lambda_n)';$$

$$m = (m_0 \dots m_p)'; \quad \gamma^0 = (\gamma(s_0 - s_1) \dots \gamma(s_0 - s_n))';$$

$$f^0 = (1 \ f_1(s_0) \dots f_p(s_0))'$$

22

## 推定の手順(1)

- 推定にあたっての問題点:
  - $\delta(s)$  は観察不可能である。
  - しかも、 $\Gamma$  と  $\gamma^0$  の要素を  $n$  個の観察値から推定しなければならない。
- ひとつの解決策:
  - $\mu(s)$  を推定し、残差を  $\delta(s)$  の推定値とする。
  - semivariogram モデルを想定し、残差からこれを推定して、 $\Gamma$  と  $\gamma^0$  の推定にもちいる。

23

## 推定の手順(2)

- $\mu(s)$  の推定
  - 方針:
    - 説明変数を選択し、回帰式を推定する
      - 説明変数: 予測領域全体についても利用可能なもの。
      - 回帰式: OLS ← 問題あり。
        - $\delta(s)$  の相関を無視している。
        - $\delta(s)$  が intrinsically stationary であることに、完全には対応できていない。
        - 外れ値や影響力の大きい観察点が多い。
        - どれだけ工夫しても「本当の誤差と残差とは異なる」という本質的限界は残る。

24

## 推定の手順(3)

- semivariogram モデルの推定
  - 等方性 isotropic:  $\gamma(s-t) = \gamma(|s-t|)$
  - 経験セミバリオグラムの計算:
    - 観察地点  $s_i (i=1,2,\dots,n)$  のすべての組み合わせについて、 $r_{ij} = |s_i - s_j|$  を求める。
    - $r_{ij}$  について階級  $I_k (k=1,2,\dots,K)$  を設定する。
      - $N_k = \# I_k > 30$
      - $r_{ij} > \max r_{ij} / 2$  はバリオグラムモデル推定でもちいない。

25

## 推定の手順(4)

- 古典的推定式:

$$2\hat{\gamma}(h_k) = \frac{1}{N_k} \sum_{|s_i - s_j| \in I_k} (\hat{\delta}(s_i) - \hat{\delta}(s_j))^2$$

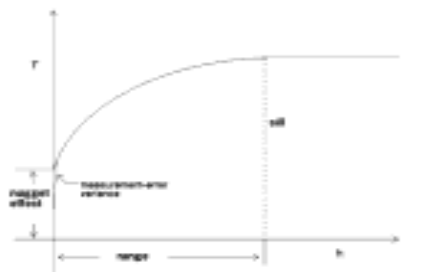
- 頑健な推定式(Hawkins and Cressie (1980)):

$$2\bar{\gamma}(h_k) = \left\{ \frac{1}{N_k} \sum_{|s_i - s_j| \in I_k} |\hat{\delta}(s_i) - \hat{\delta}(s_j)|^{1/2} \right\}^4 \times (0.475 + 0.494 / N_k)^{-1}$$

26

## 推定の手順(5)

- semivariogram モデルの当てはめ:



27

## 推定の手順(6)

- さまざまな理論モデル: 非負の  $h$  について

- 白色雑音モデル:  $\gamma(h; \theta) = 1(h > 0) \theta_0$
- 線形モデル:  $\gamma(h; \theta) = 1(h > 0) (\theta_0 + \theta_1 h)$
- べき乗モデル:  $\gamma(h; \theta) = 1(h > 0) (\theta_0 + \theta_1 h^{\theta_2})$
- 球形モデル: 
$$\gamma(h; \theta) = 1(h > 0) \left\{ \theta_0 + \theta_1 \left( 1 + 1(h < \theta_2) \left( -1 + \frac{3}{2} \left( \frac{h}{\theta_2} \right) + \frac{1}{2} \left( \frac{h}{\theta_2} \right)^3 \right) \right) \right\}$$
- 指数型モデル: 
$$\gamma(h; \theta) = 1(h > 0) \{ \theta_0 + \theta_1 (1 - \exp(-h/\theta_2)) \}$$
- ガウス型モデル: 
$$\gamma(h; \theta) = 1(h > 0) \{ \theta_0 + \theta_1 (1 - \exp(-(h/\theta_2)^2)) \}$$

28

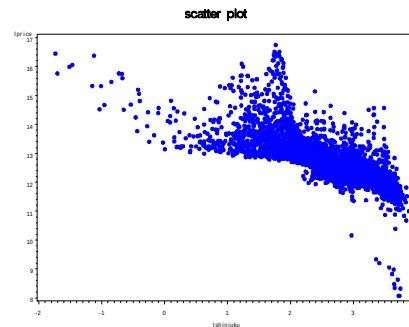
## 推定の手順(7)

- (西郷の)私見:
  - 理論的な semivariogram モデルの物理的な意味づけは、(少なくとも経済データへの利用に関しては)それほど重要でない。経験 semivariogram をながめて、適当と思われるものを採択すればよい。ある程度の近似の精度が保たれれば、妥当なウェイト  $\lambda_j$  が求まるはずである。
  - むしろ、等方性の仮定などについて、残差分析を入念におこなうことが大切である。

29

## $\mu(s)$ の推定(1):

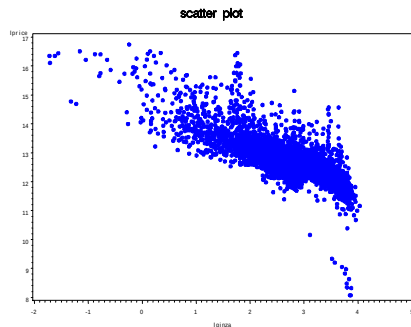
説明変数の選択:  $\log(\text{新宿からの距離})$



30

## $\mu(s)$ の推定(2):

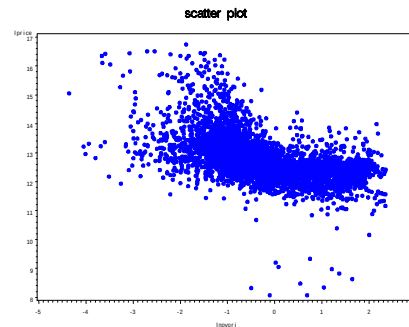
説明変数の選択:  $\log(\text{銀座からの距離})$



31

## $\mu(s)$ の推定(3):

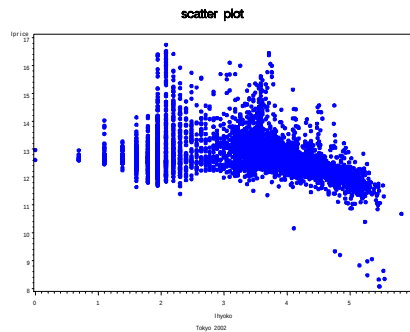
説明変数の選択:  $\log(\text{最寄り駅からの距離})$



32

## $\mu(s)$ の推定(4):

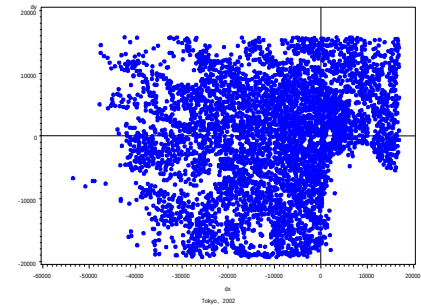
説明変数の選択:  $\log(\text{標高}+5)$



33

## $\mu(s)$ の推定(5):

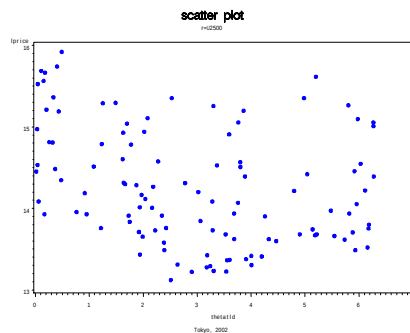
説明変数の選択: 都心からの角度



34

## $\mu(s)$ の推定(6):

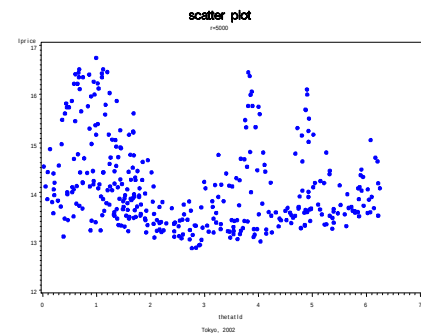
説明変数の選択: 都心からの角度 ( $r < 2.5\text{km}$ )



35

## $\mu(s)$ の推定(7):

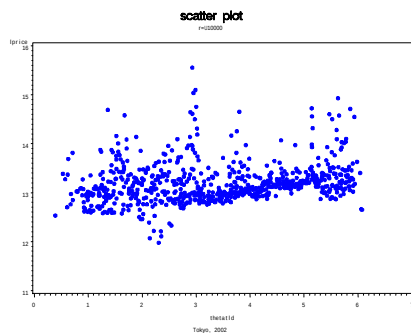
説明変数の選択: 都心からの角度 ( $2.5\text{km} < r < 5.0\text{km}$ )



36

## $\mu(s)$ の推定(8):

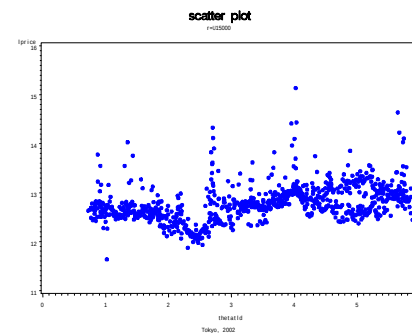
説明変数の選択: 都心からの角度 ( $5.0\text{km} < r < 10.0\text{km}$ )



37

## $\mu(s)$ の推定(9):

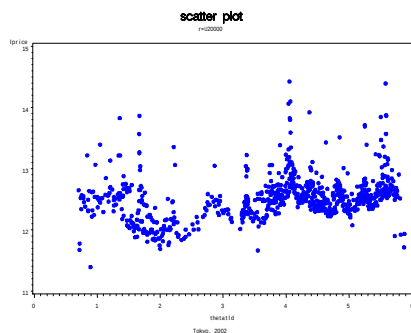
説明変数の選択: 都心からの角度 ( $10.0\text{km} < r < 15.0\text{km}$ )



38

## $\mu(s)$ の推定(10):

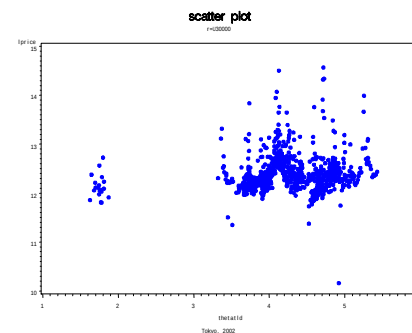
説明変数の選択: 都心からの角度 ( $15.0\text{km} < r < 20.0\text{km}$ )



39

## $\mu(s)$ の推定(11):

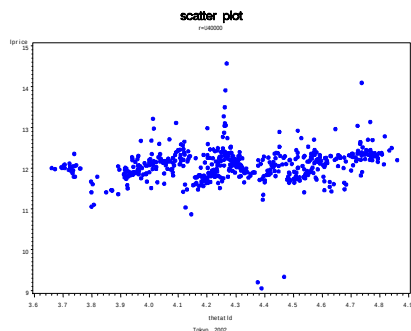
説明変数の選択: 都心からの角度 ( $20.0\text{km} < r < 30.0\text{km}$ )



40

## $\mu(s)$ の推定(12):

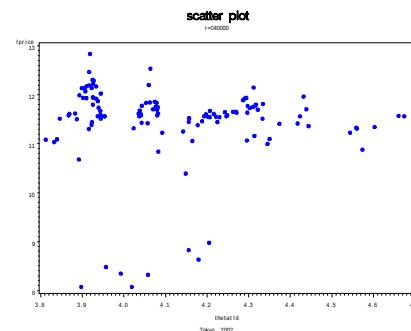
説明変数の選択: 都心からの角度 ( $30.0\text{km} < r < 40.0\text{km}$ )



41

## $\mu(s)$ の推定(13):

説明変数の選択: 都心からの角度 ( $40.0\text{km} < r$ )



42

## $\mu(s)$ の推定(14):

推定モデル

$$\mu(s) = \beta_0 + \beta_1 \log(\min(d_s, d_g)) + \beta_2 \log(d_{st}) \\ + \beta_3 \tilde{\theta} + \beta_4 x + \beta_5 y$$

where

$d_s$  = (新宿からの距離);  $d_g$  = (銀座からの距離);

$d_{st}$  = (最寄り駅からの距離);

$\tilde{\theta}$  = (都心からの角度);  $(x, y)$  = (平面座標)

43

## $\mu(s)$ の推定(15):

平均値モデルの推定(1)

| ANOVA |      |        |       |        |
|-------|------|--------|-------|--------|
| 要因    | df   | 平方和    | 平均平方  | F      |
| 回帰    | 5    | 1926.2 | 385.2 | 2095.3 |
| 残差    | 4451 | 818.3  | 0.18  | $R^2$  |
| total | 4456 | 2744.5 |       | 0.70   |

44

## $\mu(s)$ の推定(16):

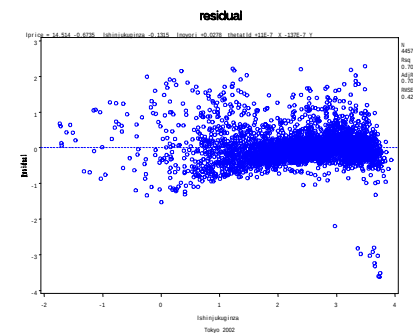
平均値モデルの推定(2)

|      | 推定値    | SD    | t     | VIF  |
|------|--------|-------|-------|------|
| 定数項  | 14.514 | 0.069 | 208.9 |      |
| 距離   | -0.674 | 0.012 | -56.2 | 2.54 |
| 最寄り駅 | -0.132 | 0.007 | -17.9 | 1.58 |
| 角度   | 0.028  | 0.007 | 4.0   | 2.16 |
| x    | 0.000  | 0.000 | 1.7   | 2.29 |
| y    | -0.000 | 0.000 | -14.7 | 1.63 |

45

## $\mu(s)$ の推定(17):

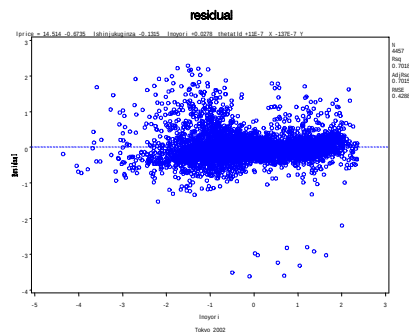
残差プロット(都心からの距離)



46

## $\mu(s)$ の推定(18):

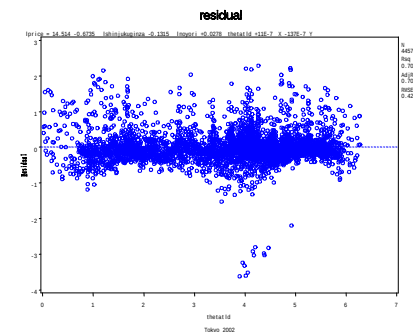
残差プロット(最寄り駅からの距離)



47

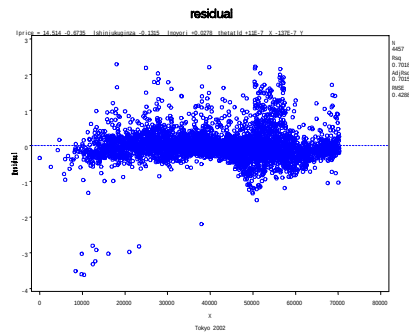
## $\mu(s)$ の推定(19):

残差プロット(角度)



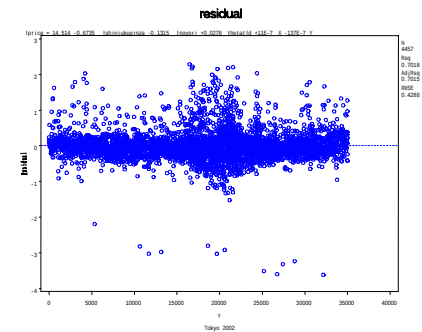
48

## $\mu(s)$ の推定(20): 残差プロット(x)



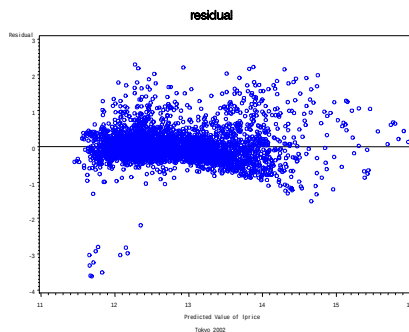
49

## $\mu(s)$ の推定(20): 残差プロット(y)



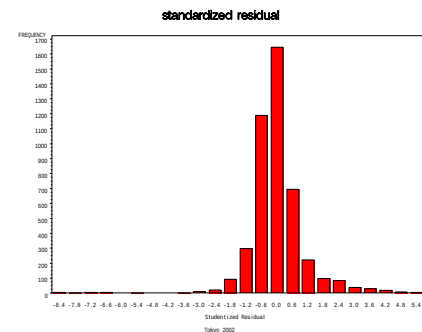
50

## $\mu(s)$ の推定(21): 残差プロット(当てはめ値)



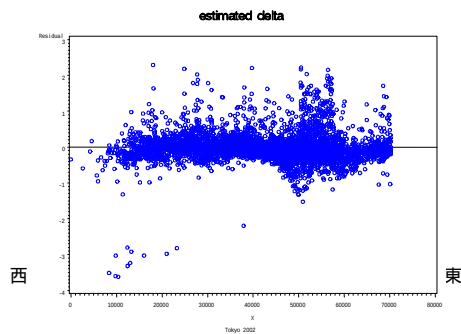
51

## 推定された $\delta(1)$ : スチューデント化残差のヒストグラム



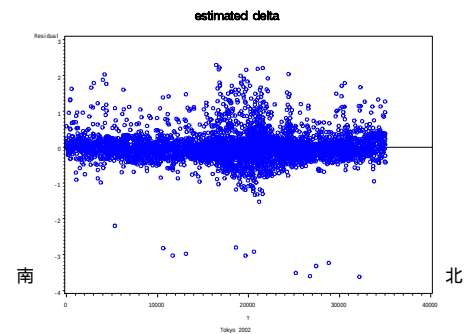
52

## 推定された $\delta(s)$ (2)



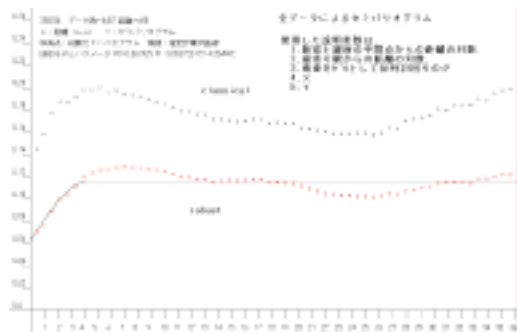
53

## 推定された $\delta(s)$ (3)



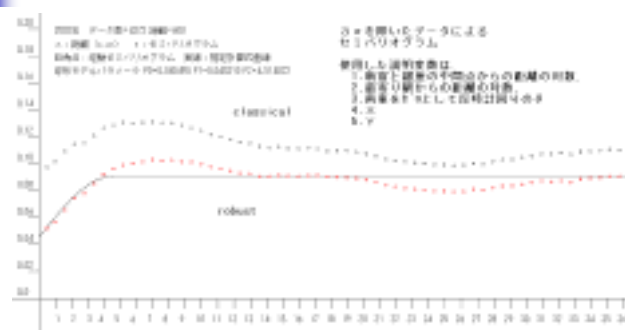
54

## 経験セミバリオグラム(1)



55

## 経験セミバリオグラム(2)



56

## セミバリオグラム・モデルの推定

- 古典的推定式 と 頑健な推定式 との相違

- 外れ値の影響が大きい  
頑健な経験セミバリオグラムを採用。

- 推定モデル:

- 球形モデル:

$$\gamma(h; \theta) = 1(h > 0) \left\{ \theta_0 + \theta_1 \left( 1 + 1(h < \theta_2) \left( -1 + \frac{3}{2} \left( \frac{h}{\theta_2} \right) + \frac{1}{2} \left( \frac{h}{\theta_2} \right)^3 \right) \right) \right\}$$

$$\theta_0 = 0.063; \theta_1 = 0.054; \theta_2 = 4.355$$

57

## 推定された公示地価



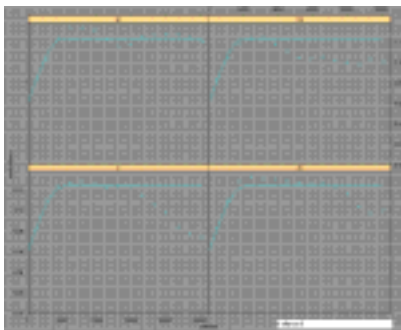
58

## 異方性の検証

異方性セミバリオグラム (robust)

90:東; 135:南東

0:北; 45:北東



59

## 今後の課題(1)

- $\mu(s)$  の推定

- 説明変数の追加 (GISデータの活用など)

- 推定手法

- 外れ値への対処
- 局所的な当てはまりの重視

- 複数年の同時推定

- $t = (s, \text{time})$  として、 $Z(t)$  をモデル化する。
- $Z(s) = (Z_{2000}(s), Z_{2001}(s), Z_{2002}(s))$  として、cokriging を適用する。← こちらの方が有望か？
  - バラツキの違いが簡単に処理できる but 欠測処理が必要。

60

## 今後の課題(2)

1990年

西

東

2002年

西

東

61

## データの取得とソフトウェア

### ■ 今回使ったデータ

- オンラインで無償入手できるものだけ。
  - 詳しくは増成が説明します。

### ■ ソフトウェア

- 計算は増成作成のプログラムで実行した。
- R
  - 岡田昌史編(2004)『The R Book』九天社
    - 第10章 Rで空間統計学
    - 第11章 Rによる空間疫学分析

62

## 参考文献

- Cressie, N. A. C. (1993), *Statistics for Spatial Data*, revised ed., Wiley.
- 間瀬茂・武田純(2001)『空間データ モデリング: 空間統計学の応用』共立出版
- Wackernagel, H. (1998), *Multivariate Geostatistics*, 2<sup>nd</sup> ed., Springer (邦訳 地球統計学研究会 訳編/青木謙治 監訳 (2003)『地球統計学』森北出版)

63