Working Paper No. 287

リピート・セールス価格指数における
セレクション・バイアス

唐渡広志

2014 年 3 月 27 日

UNIVERSITY
OF TOYAMA 富山大学

FACULTY OF ECONOMICS
UNIVERSITY OF TOYAMA

リピート・セールス価格指数における

セレクション・バイアス†

平成 26 年 3 月 27 日

唐渡　広志

富山大学経済学部

概要

リピート・セールス法は，複数回売買された商品をサンプルとして選び，ペアになった同一商品の異時点間の対数差分価格を取引時点の時間ダミー変数に回帰させることによって価格指数を推定する手法である。本研究は，ランダム・ウォーク型の誤差構造を想定した Case-Shiller 型のリピート・セールス法がもたらすサンプル・セレクション・バイアスを除去するために，セレクション関数とリピート・セールス回帰モデルとの間の共分散を推定するために逆ミルズ比の修正し，Gatzlaff and Haurin (1997) の改善を行った。さらに，ランダム・ウォークではなく系列相関および分散不均一を持つ誤差構造のもとでのセレクションの除去を行う。東京都世田谷区のデータによる分析結果より，セレクション・バイアスを制御した場合のリピート・セールス回帰モデルとセレクション関数の相関は有意に負であり，２回取引された物件の価格指数は住宅市場の母集団に比べて負のバイアスを持つ可能性がある。

目次

† 本研究は 2010 年度日本不動産学会秋季全国大会（東京大学）における報告論文を基に加筆修正したものである。また，JSPS 科研費 24618008（基盤 C）の助成を受けている。

1. はじめに

これまで住宅価格指数を推定するために, いくつかの手法が検討されてきた。なかでも, Rosen (1974) によって理論的に整理されたヘドニック法や, Bailey, Muth and Nourse (1963), Case and Shiller (1987, 1989) によって精緻化されたリピート・セールス法（以下 RS 法とよぶ）は最も利用されている手法である。ヘドニック法は，英国政府や Halifax 等の英国を代表するモーゲージバンクによって採用されている。米国では，シカゴ・マ−カンタイル市場に上場されている住宅価格指数であるスタンダード・アンド・プアーズ・ケース・シラー価格指数に代表されるように RS 法が利用されている。

ヘドニック法とは，ある不動産価格をその不動産のさまざまな属性（性能や機能，購入者や販売者の特徴）の価値に関する集合体（属性の束）とみなし，主に回帰分析を利用してそれぞれの属性価格を推定する手法である。プーリング・データを利用すれば，品質調整された価格指数を計測することができる。これに対して RS 法は，複数回売買された商品をサンプルとして選び，ペアになった同一商品の異時点間の対数差分価格を取引時点の時間ダミー変数に回帰させることによって価格指数を推定する。

ヘドニック法を利用した不動産市場の実証分析では，住宅の機能や特徴，あるいは住宅購入者の所得やタイプを説明変数として利用したものが多い（例えば，Sirmans *et al*. 2006 によるサーヴェイ）。しかしながら，本来は Rosen (1974) が提案しているように，購入者だけでなく販売者の経済行動や意思決定も市場価格の決定に影響する。そもそもデータとして観察されるためには，販売者がなんらかの意思決定をする結果として市場に供給されるという事実を考慮すべきであろう。

住宅が売買される財として市場に登場する場合，販売者にとってのオファー価格は彼の留保価格を凌駕しているという, 条件を想定することは自然である。Gatzlaff and Ling (1994) および Gatzlaff and Haurin (1998) は，もしも住宅売買の意思決定が，オファー価格と留保価格の決定に影響するならば，実際に販売された住宅はランダム・サンプリングでない可能性があることを検証している。すなわち，観察される取引価格はオファー価格と留保価格を生ぜしめた確率過程に依存していると考え，Heckman (1979) による 2 段階一致推定法 (Heckit) を適用したセレクション・バイアスの除去を行っている。

RS 法もまた, そのデータ発生過程はヘドニック回帰モデルに依存しているため同様の問題が生じる。1 回目と 2 回目の販売時点において，販売者のオファー価格が留保価格を上回っている場合においてのみ，ペアになったデータ・セットとして観察されるため，セレクションされたサンプルを利用した分析にならざるを得ない。

Gatzlaff and Haurin (1997) におけるセレクション・バイアスの修正は Bailey *et al*. (1963) で提案された最もシンプルな RS 回帰モデルを基本として，Heckit を適用している。同論文では，セレクション関数と RS 回帰モデルとの間の共分散が通時的に一定であると仮定している。しかしながら，住宅の売り手が直面する経済状況は年々刻々と変化しているわけであるから，実際には誤差構造が変化している可能性がある。

本研究では，この点を改善してセレクション・バイアスを修正した RS 価格指数の推定方法を提案する。まず第 1 に，Gatzlaff and Haurin (1997) の最も簡単な拡張として，Case and Shiller (1987, 1989) で想定された誤差項におけるランダム・ウォークを仮定し，取引期間を長期化すればするほど，誤差分散が拡大するケースでのセレクションの除去を行う。この場合，Gatzlaff and Haurin (1997) のように RS 回帰モデルとセレクション関数との間の共分散が通時的に一定であったとしても，ランダム・ウォークによって生じる分散不均一を修正した重み付き RS 回帰モデルにおける逆ミルズ比も修正されなければならない。

第 2 に，ランダム・ウォークではなく系列相関および分散不均一を持つ誤差構造のもとでのセレクションの除去を行う。Case and Shiller (1987, 1989) で想定された誤差項におけるランダム・ウォークに関する仮定は，誤差の階差をとることによって分析が容易な定常過程となることを意味している。分析上，このような単純化は他の研究者による再現性が高く，広く受け入れられやすい。ただし，誤差がどのような系列相関を持つのかを検証することなく価格指数を計測することは危険である。本研究では, Hill, Sirmans and Knight (1999) における RS 法の分散不均一特定化手法を利用して，誤差項がランダム・ウォークであるのかどうかを検証するとともにセレクションの除去を行う。

以下では，次節においてセレクション・バイアスを制御した RS 回帰モデルを定義し，3 節において利用するデータの特徴について述べる。4 節において推定結果を示し，セレクション・バイアスの有無を検証する。5 節において結論を述べる。

2.切断された誤差分布を持つリピート･セールス回帰モデル

2.1 リピート･セールス価格指数

物件 i の取引時点 t 期における取引価格 $PRICE_{it}$ の対数値を $p_{it} = \log PRICE_{it}$，その物件の属性ベクトルを $\mathbf{x}_i$ とする。取引価格の対数値を属性ベクトルに回帰させたヘドニック・モデルを次のように書く。

$$p_{it} = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_{it}\,,\ \ \varepsilon_{it} = \alpha + \delta_t + \upsilon_{it} \qquad (1)$$

$\boldsymbol{\beta}$ は属性のパラメタ・ベクトル（属性価格）である。ε_{it} は誤差項を示しており，モデル全体の定数項係数 α，取引時点 t 期の時間効果 δ_t，攪乱項 υ_{it} で構成されているものとする。

Case and Shiller (1987, 1989) の RS 法は攪乱項υ_{it}において次のランダム・ウォークを仮定している。

$$\upsilon_{it} = \upsilon_{it-1} + u_{it},\ \ u_{it} \sim iid.N(0, \sigma^2),\quad E(\upsilon_{it} u_{js}) = 0,\ E(u_{it} u_{js}) = 0$$

$$i, j = 1, \cdots, n;\ t, s = 1, \cdots, T;\ i \neq j;\ t \neq s \tag{2}$$

ここで，u_{it} はホワイト・ノイズを仮定した系列相関のない互いに独立な正規分布にしたがう確率変数である。

Bailey et al. (1963) および Case and Shiller (1987, 1989) は次の仮定を設けて (1) を再定式化している[1]。

［仮定 1］ すべての属性は時間通じて不変である。

［仮定 2］ すべての属性パラメタは時間通じて不変である。

RS 法は同一物件の 1 回目と 2 回目の取引価格の対数差分を回帰モデルで検討する分析手法である。したがって，2 つの仮定がある場合は 2 時点間の価格差は個々の物件の属性やその価格の変化ではなく，市場全体に共通して生じる変化（時間効果）によって説明することができる。2 時点間の価格差に注目する利点として，分析が簡単化できること以外に物件固有のコーホート効果を除去できる点もあげられる。すなわち，住宅建設時点の生産に係る生産要素価格（労務人件費や建設資材価格など）が費用関数に影響している場合，それらの違いを考慮せずに済む[2]。

いま，1 回目の取引が行われた時期を t 期，2 回目の取引が行われた時期を $t + S_i$ 期としよう。ペアになった物件 i の価格の 2 時点間の対数差分は

$$y_i \equiv p_{i,t+S_i} - p_{it} = \delta_{t+S_i} - \delta_t + e_i,\quad (i = 1, 2, \cdots, n;\ S_i > 0)$$

と示すことができる。ここで，誤差項は$e_i = \upsilon_{i,t+S_i} - \upsilon_{it}$ であり，(2) より

[1] 住宅は経年劣化や修繕投資によって価格が変化（住宅の年齢効果）するので，標準的なリピート・セールス法では時間効果に年齢効果が含まれてしまう。しかし，時間ダミーと取引間隔を示す変数との間には完全な線型関係が存在するため，通常の方法では時間効果と年齢効果を識別することができない。今までのところ提案されているのは，時間ダミーと取引間隔変数の間の線型関係を意図的に崩してしまう方法や外部のデータを利用して価格指数に外挿する方法である。日本のデータによる推定例として唐渡・清水・中川・原野 (2012) や Wong, S. K., K.W. Chau, K. Karato and C. Shimizu (2013) がある。

[2] 時間効果，年齢効果および世代効果を表現する 3 つの独立変数の間には完全な線形関係が生じているため，多くの文献では世代効果を除外している。Karato, K., O. Movshuk and C. Shimizu (2010) はリピート・セールス法を用いずにヘドニック・アプローチで価格指数を求めると，世代（コーホート）効果を除外することによるバイアスが生じることを一般化加法モデルにより実証している。

$$\upsilon_{i,t+S_i} = \upsilon_{it} + \sum_{j=1}^{S_i} u_{i,t+j}$$

であり，その期待値は$E(e_i) = 0$である。また，分散は

$$Var(e_i) = Var\left(\sum_{j=1}^{S_i} u_{i,t+j}\right) = \sum_{j=1}^{S_i} E\left(u_{i,t+j}{}^2\right) + \sum_{j \neq k}\sum_{k \neq j} E\left(u_{i,t+j}u_{i,t+k}\right) = S_i\sigma^2$$

となる。

ヘドニック回帰モデル (1) において時間効果を推定するには，取引時点に対応したダミー変数を利用する。したがって RS 回帰モデルにおいても時間効果を推定するために 2 時点間のダミー変数の差分を利用する。すなわち，1 回目の取引時点をt期，2 回目の取引時点を$t + S_i$期とすると，(1) より

$$p_{it} = \mathbf{x}_i'\boldsymbol{\beta} + \alpha + \dot{\mathbf{d}}_i'\boldsymbol{\delta} + \upsilon_{it},$$

$$p_{i,t+S_i} = \mathbf{x}_i'\boldsymbol{\beta} + \alpha + \ddot{\mathbf{d}}_i'\boldsymbol{\delta} + \upsilon_{i,t+S_i},$$

である。ただし， $\dot{\mathbf{d}}_i = \left(\dot{d}_i^1, \cdots, \dot{d}_i^T\right)$ は 1 回目の，$\ddot{\mathbf{d}}_i = \left(\ddot{d}_i^1, \cdots, \ddot{d}_i^T\right)$ は 2 回目の取引時点に対応するダミー変数ベクトルであり，次のように定義する。

$$\dot{d}_i^\tau = \begin{cases} 1 & \tau = t \\ 0 & \tau \neq t \end{cases},\ \ddot{d}_i^\tau = \begin{cases} 1 & \tau = t + S_i \\ 0 & \tau \neq t + S_i \end{cases},\ \left(i = 1, \cdots, n;\ \tau = 1, 2, \cdots, T\right)$$

また，$\boldsymbol{\delta} = \left(\delta_1, \cdots, \delta_t, \cdots, \delta_{t+S_i}, \cdots, \delta_T\right)$ は時間効果ベクトルである。この 2 つの式より対数価格差分は

$$p_{i,t+S_i} - p_{i,t} \equiv y_i = \mathbf{D}_i\boldsymbol{\delta} + e_i,$$

であり，$\mathbf{D}_i = \ddot{d}_i - \dot{d}_i \quad (i = 1, 2, \cdots, n)$ とする。このことを考慮すると，クロス・セクション方向に積み上げた行列表示の RS 回帰モデルは次のように書ける。

$$\mathbf{y} = \mathbf{D}\boldsymbol{\delta} + \mathbf{e} \tag{3}$$

ここで，$\mathbf{y} = (y_1, \cdots, y_i, \cdots, y_n)'$，$\mathbf{D}$ は各取引時点に対応した 2 時点のダミー変数の差分の $n \times T$ 行列であり，1 回目の取引時点のとき-1，2 回目の取引時点のとき 1，それ以外のとき 0 となる。また，$E(\mathbf{ee}') = \sigma^2\,\overline{\mathbf{S}}$，ただし

$$\overline{\mathbf{S}} = \mathrm{diag}[S_i] = \begin{pmatrix} S_1 & & 0 \\ & \ddots & \\ 0 & & S_n \end{pmatrix}$$

であるから，誤差分散は取引時点の間隔に依存しており分散不均一である。

Case and Shiller (1987, 1989) はこれを修正して，推定量の効率性を高めるための一般化最小 2 乗推定を提案している。すなわち $n \times n$ の重み行列

$$\mathbf{Q} = \mathrm{diag}\left[S_i^{-1/2}\right] = \begin{pmatrix} 1\big/\sqrt{S_1} & & 0 \\ & \ddots & \\ 0 & & 1\big/\sqrt{S_n} \end{pmatrix}$$

を利用して，(3) の両辺に $\mathbf{Q}$ を乗じた次の重み付き RS 回帰モデルを推定している。

$$\mathbf{y}^* = \mathbf{D}^*\boldsymbol{\delta} + \mathbf{e}^* \tag{4}$$

ここで，$\mathbf{y}^* = \mathbf{Q}\mathbf{y}$, $\mathbf{D}^* = \mathbf{Q}\mathbf{D}$ および $\mathbf{e}^* = \mathbf{Q}\mathbf{e}$ である。$\overline{\mathbf{S}}^{-1} = \mathbf{Q}\mathbf{Q}'$ であるから，

$$E\left(\mathbf{e}^*\mathbf{e}^{*\prime}\right) = \sigma_u^2 \mathbf{Q}\overline{\mathbf{S}}\mathbf{Q}' = \sigma_u^2 \mathbf{I}$$

となり，最小 2 乗法が適用できる。もしくは，(1) の残差 2 乗値を $S_1, \cdots, S_T$ に回帰してその理論値をウェイトとする実行可能な一般化最小 2 乗法によっても推定できる。

時間効果ベクトル $\boldsymbol{\delta}$ の一般化最小 2 乗推定量は $\hat{\boldsymbol{\delta}} = \boldsymbol{\delta} + \left(\mathbf{D}'\overline{\mathbf{S}}^{-1}\mathbf{D}\right)^{-1}\mathbf{D}'\overline{\mathbf{S}}^{-1}\mathbf{e}$ より，$E(\hat{\boldsymbol{\delta}}) = \boldsymbol{\delta}$，$V(\hat{\boldsymbol{\delta}}) = \sigma^2\left(\mathbf{D}'\overline{\mathbf{S}}^{-1}\mathbf{D}\right)^{-1}$ である。基準時点を 1 期とする t 期の RS 価格指数は，(4) の推定量を利用して $I_{t/1}^{RS} = PRICE_t / PRICE_1 = \exp\left(\hat{\delta}_t - \hat{\delta}_1\right)$ から計測される。しかしながら，取引時点のダミー変数は $\sum_{t=1}^{T} D_{it} = 0$ なる線型関係で結びつけられるので，多重共線性を考慮すると，すべての取引時点をダミー変数として利用できない。多くの文献にならって，取引の初期時点となるダミー変数 D_{i1} をモデルから落とし，$\delta_1 = 0$ を想定する。基準時点を第 1 期とするとき，RS 価格指数系列は

$$I^{RS} = \left\{ \exp(0), \exp\left(\hat{\delta}_2\right), \cdots, \exp\left(\hat{\delta}_T\right) \right\}$$

となる[3]。

(2) より対数価格は正規分布にしたがう確率誤差を伴うので，指数の期待値は $E\left[\exp(\hat{\delta}_t)\right] = \exp\left[\delta_t + \frac{1}{2}\sigma^2\Gamma_{tt}\right]$ である。ただし，Γ_{tt} は $\left(\mathbf{D}'\overline{\mathbf{S}}^{-1}\mathbf{D}\right)^{-1}$ の t 番目の対角要素を示

[3] ヘドニック価格指数，パーシェ指数，ラスパイレス指数などとの相違点については Diewert, Heravi and Silver (2007), Hill and Melser (2008) などにおいて詳しい。

している。価格指数は本質的にバイアスを持つが，RS 回帰モデルから推定される $\sigma^2\Gamma_{tt}$（推定値の標準誤差の 2 乗）はそれほど大きくない

2.2 サンプル・セレクション

観測期間において物件は複数回取引されるものもあれば，1 度だけ取引されるものもある。したがって，RS 法には 2 回取引された物件だけを扱うことによって生じる，サンプリング・バイアスが存在する可能性がある。なお，3 節においても述べるが，本研究では一度も取引されていない物件を扱わない。したがって 1 回目の取引時点におけるセレクションを考慮しない。

Gatzlaff and Haurin (1997) は Bailey *et al*. (1963) による RS 法において，セレクション関数と RS 回帰モデルとの間の共分散が通時的に一定であると想定した上で，セレクション関数を導入することによって，セレクション・バイアスを取り除いた価格指数を計測している。しかしながら，住宅の売り手が直面する経済状況は年々刻々と変化しているわけであるから，実際には誤差構造が変化している可能性がある。本研究では，これに対応するために，前節で示された Case and Shiller (1987, 1989) による誤差項の分散不均一性を考慮し，セレクション・バイアスを修正した新しい RS 回帰モデルを提案する。

物件が市場で売買されるためには，売り手のオファー価格が留保価格を上回る必要があり，そしてそのときだけ取引価格が観察される。したがって，切断されたヘドニック価格の誤差分布の条件付き期待値が 0 でないことにより，ヘドニック価格にはセレクション・バイアスが生じる。

サンプル･セレクションがもたらすバイアスの修正方法は Heckman (1979) で示されている（Heckit とよばれる手法）。この方法は，実際には観察できない 2 回目の取引価格がない場合でも，時間効果の一致推定量をもたらす。いま，$t+S_i$ 期において住宅を売りに出すかどうかの選択を表す閾値を $Z^*_{i,t+S_i}$ とし，次の回帰式でその選択メカニズムが示されるものとしよう。

$$Z^*_{i,t+S_i} = \mathbf{w}'_{i,t+S_i}\boldsymbol{\gamma} + \nu_{i,t+S_i} \tag{5}$$

ここで，$\mathbf{w}_{i,t+S_i}$ は売り手の意思決定に影響すると考えられる住宅の属性やマクロ経済要因を表す変数ベクトル，$\boldsymbol{\gamma}$ は対応するパラメタ・ベクトルであり，$\nu_{i,t+S_i} \sim N(0,1)$ を仮定する。実際には，閾値 $Z^*_{i,t+S_i}$ は観察されないので，$Z^*_{i,t+S_i} > 0$ ならば $Z_{i,t+S_i} = 1$，それ以外の場合 0 となる 2 値変数を考え，これらの確率を

$$\Pr\left(Z_{i,t+S_i} = 1\right) = F\left(\mathbf{w}'_{i,t+S_i}\boldsymbol{\gamma}\right)$$

$$\Pr\left(Z_{i,t+S_i}=0\right)=1-F\left(\mathbf{w}'_{i,t+S_i}\boldsymbol{\gamma}\right)$$

とする。ただし，$F\left(\mathbf{w}'_{i,t+S_i}\boldsymbol{\gamma}\right)$ は標準正規分布の累積確率密度関数である。2 値変数はオファー価格が留保価格を上回るときだけ観察されるので, 次のように定義することができる。

$$Z_{i,t+S_i}=\begin{cases}1 & \text{if } p^O_{i,t+S_i}-p^R_{i,t+S_i}\geq 0\\ 0 & \text{otherwise}\end{cases}$$

本研究では Gatzlaff and Haurin (1997) と異なり，一度も取引されていない物件を扱わないので，1 回目と 2 回目の間のセレクション構造だけを考える。したがって，RS 回帰モデルの誤差構造と (5) におけるセレクションとの間においての相関だけを考慮する。いま，(3) における RS 回帰モデルの誤差 e_i と (5) におけるセレクション関数の誤差 $\nu_{i,t+S_i}$ が 2 変量正規分布にしたがい，その共分散が $\sigma_{e\nu}$ であるものとしよう。観察される重み付き RS 回帰モデルは次の条件付き期待値を持つ。

$$E\left(y_i^*\,\middle|\,Z^*_{i,t+S_i}>0\right)=\mathbf{D}_i^{*\prime}\boldsymbol{\delta}+E\left(e_i^*\,\middle|\,\nu_{i,t+S_i}>-\mathbf{w}'_{it}\boldsymbol{\gamma}\right)=\mathbf{D}_i^{*\prime}\boldsymbol{\delta}+\sigma_{e\nu}\cdot S_i^{-1/2}\lambda_i$$

ここで，$y_i^*=y_i/\sqrt{S_i}$，$\mathbf{D}_i^*=\mathbf{D}_i/\sqrt{S_i}$ であり，λ_i は (5) のプロビット推定から得られる逆ミルズ比である。セレクション・バイアスを制御した重み付き RS 回帰モデルは

$$y_i^*=\mathbf{D}_i^{*\prime}\boldsymbol{\delta}+\sigma_{e\nu}\cdot\frac{\lambda_i}{\sqrt{S_i}}+\eta_i \tag{6}$$

となる。以上のように逆ミルズ比を利用して切断されたデータがもたらす誤差分布の歪みを調整することができる。共分散 $\sigma_{e\nu}$ の推定値と残差分散より (3) と (5) の方程式間の相関係数は$\hat{r}=\hat{\sigma}_{e\nu}/\hat{\sigma}_u^2$ より計算できる。

Heckit を利用した 2 段階の推定プロシージャは次のようになる。

i. 2 回取引された住宅と 1 回だけ取引された住宅を合わせたサンプルについて，(5) のプロビット推定を行い，逆ミルズ比を計算する。このとき，2 回取引された住宅ならば $Z_{i,t+S_i}=1$，それ以外は 0 となる 2 値変数を用いる。

ii. $1/\sqrt{S_i}$ で重み付けを行い，Case and Shiller 型の RS 回帰モデルのセレクション・バイアスを制御した (6) に最小 2 乗法を適用して時間効果 $\boldsymbol{\delta}$ および共分散 $\sigma_{e\nu}$ を推定する。ここで，サンプルは 2 回取引された住宅だけを利用する。

以上のようにして求められる RS 回帰モデルの推定量は次のように書ける。

$$\hat{\boldsymbol{\delta}} = (\mathbf{D}'\mathbf{D})^{-1}(\mathbf{D}'\mathbf{y} - \mathbf{D}'\mathbf{L}\hat{\sigma}_{ev})$$
$$\hat{\sigma}_{ev} = (\mathbf{L}'\mathbf{L})^{-1}(\mathbf{L}'\mathbf{y} - \mathbf{L}'\mathbf{D}\hat{\boldsymbol{\delta}})$$

ここで，$\mathbf{L}$ は (6) 右辺第 2 項の独立変数ベクトル $\left(\lambda_1/\sqrt{S_1},\cdots,\lambda_n/\sqrt{S_n}\right)'$ である。時間効果推定量はこれらを利用して

$$\hat{\boldsymbol{\delta}} = \boldsymbol{\delta} + \left[\mathbf{I}_T - (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}\mathbf{M}_L\mathbf{D}\right]^{-1}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'(\mathbf{I}_n - \mathbf{M}_L)\boldsymbol{\eta}$$

である。ここで，$\mathbf{I}_T : T \times T$ 単位行列，$\mathbf{I}_n : n \times n$ 単位行列，$\mathbf{M}_L = \mathbf{L}(\mathbf{L}'\mathbf{L})\mathbf{L}'$ である。Heckit における 2 段階目の最小 2 乗法による推定値の標準誤差は，(6) の誤差項 η_i の分散が

$$Var(\eta_i \mid Z_i = 1) = \sigma^2 - r^2\sigma^2(d_i/S_i),\ \ d_i = \lambda_i(\lambda_i + \mathbf{w}_i'\boldsymbol{\gamma})$$

となることから，別の分散不均一が生じてしまう。Greene (1981, 2008) より，次の手順で $\boldsymbol{\delta}$ および σ_{ev} の推定値の分散共分散行列の推定値を再計算する。はじめに，セレクション関数の推定結果より $\hat{\lambda}_i$ および $\hat{\boldsymbol{\gamma}}$ を得るので，$\hat{d}_i = \hat{\lambda}_i(\hat{\lambda}_i + \mathbf{w}_i'\hat{\boldsymbol{\gamma}})$ が定義できる。次に (6) の推定結果より，残差 $\hat{\eta}_i$ および修正された逆ミルズ比の係数推定値 $\hat{\sigma}_{ev}^2$ が得られるので，(2) の誤差分散が $\hat{\sigma}^2 = n^{-1}\left(\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} + \hat{\sigma}_{ev}^2\sum\hat{d}_i\right)$ より推定できる。また，誤差分散 $\hat{\sigma}^2$ と係数推定値 $\hat{\sigma}_{ev}^2$ より相関係数が $\hat{r}^2 = \hat{\sigma}_{ev}^2/\hat{\sigma}^2$ と計算できる。最後にこれらの結果を利用して，以下の分散共分散行列の推定値を計算する。

$$Var\left(\hat{\boldsymbol{\delta}},\hat{\sigma}_{ev}\right) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Gamma}\mathbf{X} + \mathbf{R})(\mathbf{X}'\mathbf{X})^{-1}$$

ただし，

$$\mathbf{X} = \left(\mathbf{D}^*, \mathbf{Q}\hat{\boldsymbol{\lambda}}\right),$$
$$\boldsymbol{\Gamma} = \mathrm{diag}\left[1 - \hat{r}^2\,\hat{d}_i\right],$$
$$\mathbf{R} = \hat{r}^2\,\mathbf{X}'\,\mathrm{diag}\left[\hat{d}_i\right]\mathbf{w}Var(\hat{\boldsymbol{\gamma}})\mathbf{w}'\mathrm{diag}\left[\hat{d}_i\right]\mathbf{X}$$

である。

2.3 ランダム・ウォーク仮定の修正とセレクション除去

Case and Shiller (1987, 1989) の攪乱項は(2)で示されたようにランダム・ウォークを仮定している。より一般的には Hill, Sirmans and Knight (1999) で検討されているように次の 1 階自己回帰 (AR1) モデルが想定される。

$$\upsilon_{i,t} = \rho\upsilon_{i,t-1} + u_{it},\ \ u_{it} \sim iid.N\left(0,\sigma_i^2\right),\ \ E(\upsilon_{it}u_{js}) = 0,\ \ E(u_{it}u_{js}) = 0 \tag{7}$$

ここで，ρは自己回帰係数であり，u_{it} はなんらかの分散不均一をもつとしよう。 (7) より

$$\upsilon_{i,t+S_i} \cong \sum_{j=0}^{S_i-1} \rho^j u_{i,t+S_i-j}\,,\ \ |\rho|<1$$

であるから，同一住宅の対数価格差分の誤差 e_i の分散は次のように書ける。

$$Var(e_i) = Var\left(\upsilon_{i,t+S_i}\right) - 2\,\mathrm{Cov}\left(\upsilon_{i,t+S_i}, \upsilon_{i,t}\right) + Var\left(\upsilon_{i,t}\right) = \frac{2\sigma_i^2\left(1-\rho^{S_i}\right)}{1-\rho^2}$$

u_{it} の分散は Harvey (1976) で想定された次の分散 (multiplicative heteroscedasticity) を持つものとしよう。

$$\sigma_i^2 = \sigma_u^2 \exp\left(\mathbf{q}_i'\boldsymbol{\theta}\right) \tag{8}$$

ここで，ベクトル $\mathbf{q}_i$ は誤差分散に影響を与える変数である。(3) で示される対数価格差分はセレクションがあったときに観察されるが，(7) で示される分散不均一を考慮しなければ推定量は非効率になる。切断されたデータに対する最尤法は (3) が観察される確率と観察されない確率からなる尤度を考慮する。(7) の誤差構造に対応した正規分布密度関数, (3) が観察される確率（累積密度関数）, (3) が観察されない確率（累積密度関数）を以下のように書くことができる。

$$\begin{cases} f\left(e_i, \sigma_i^2, \rho\right) = \dfrac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\dfrac{e_i^2}{2} \cdot \dfrac{1-\rho^2}{2\sigma_i^2\left(1-\rho^{S_i}\right)}\right), \text{ただし} Z_i = 1 \\ F\left(\dfrac{\mathbf{w}_i'\boldsymbol{\gamma} + re_i\sqrt{\dfrac{1-\rho^2}{2\sigma_i^2\left(1-\rho^{S_i}\right)}}}{\sqrt{1-r^2}}\right), \text{ただし} Z_i = 1 \\ 1 - F\left(\mathbf{w}_i'\boldsymbol{\gamma}\right), \text{ただし} Z_i = 0 \end{cases}$$

すなわち，対数尤度関数は以下のようになる。

$$\begin{aligned} \ln L = &\sum_{Z=1} \ln f\left(e_i, \sigma_i^2, \rho\right) + \sum_{Z=1} \ln F\left[\left(\mathbf{w}_i'\boldsymbol{\gamma} + re_i\sqrt{\frac{1-\rho^2}{2\sigma_i^2\left(1-\rho^{S_i}\right)}}\right)\left(1-r^2\right)^{-\frac{1}{2}}\right] \\ &+ \sum_{Z=0} \ln F\left(-\mathbf{w}_i'\boldsymbol{\gamma}\right) \end{aligned} \tag{9}$$

ただし，$e_i = y_i - \mathbf{D}_i'\boldsymbol{\delta}$ である。

以上の自己回帰過程を持つ RS 回帰モデルの推定プロシージャは次のようになる。

i. (7) の自己回帰係数および (8) の分散を推定するために，$Z_{i,t+S_i} = 1$ となるサンプルだけを利用して次の対数尤度関数のもとで最尤法を行う。

$$\begin{aligned}\ln L_0 = &-\frac{n}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{n}\ln\left(\frac{2\left(1-\rho^{S_i}\right)\exp\left(\ln\sigma_u^2 + \mathbf{q}_i'\theta\right)}{1-\rho^2}\right)\\ &-\frac{1-\rho^{S_i}}{2}\sum_{i=1}^{n}\ln\left(\frac{e_i^2}{2\left(1-\rho^{S_i}\right)\exp\left(\ln\sigma_u^2 + \mathbf{q}_i'\theta\right)}\right)\end{aligned} \tag{10}$$

ii. (10) の最尤推定値 $\hat{\rho}$ および $\hat{\boldsymbol{\theta}}$ を利用して (3) に対するウェイトを次のように定義する。

$$\hat{\boldsymbol{\omega}} = \mathrm{diag}\left[\hat{\omega}_i\right] = \begin{pmatrix}\hat{\omega}_1 & & 0\\ & \ddots & \\ 0 & & \hat{\omega}_n\end{pmatrix},\quad \text{ただし}\ \hat{\omega}_i = \sqrt{\exp\left(\mathbf{q}_i'\hat{\boldsymbol{\theta}}\right)\frac{2\left(1-\hat{\rho}^{S_i}\right)}{1-\hat{\rho}^2}} \tag{11}$$

iii. (11) を利用して $y_i^{**} = \omega_i y_i$, $\mathbf{D}_i^{**} = \omega_i \mathbf{D}_i$ および $e_i^{**} = y_i^{**} - D_i^{**\prime}\boldsymbol{\delta}$ を定義し，最尤法の初期値を得るために以下の重み付最小 2 乗法を実行する。

$$y_i^{**} = \mathbf{D}_i^{**\prime}\delta + e_i^{**} \tag{12}$$

iv. (9) の e_i を e_i^{**} に置き換えた次の対数尤度関数を利用して最尤推定を行う。

$$\ln L = \sum_{Z=1}\ln f\left(e_i^{**},\sigma_u^2\hat{\omega}_i^2\right) + \sum_{Z=1}\ln F\left[\left(\mathbf{w}_i'\boldsymbol{\gamma} + r\frac{e_i^{**}}{\sigma_u\hat{\omega}_i}\right)\left(1-r^2\right)^{-\frac{1}{2}}\right] + \sum_{Z=0}\ln F\left(-\mathbf{w}_i'\boldsymbol{\gamma}\right) \tag{13}$$

なお，(7) において $\rho = 0$，(8) において $\boldsymbol{\theta} = 0$ が想定される場合，RS 回帰モデルの誤差は均一分散 $Var(e_i) = Var\left(\upsilon_{i,t+S_i}\right) + Var\left(\upsilon_{i,t}\right) = 2\sigma_u^2$ となる。このとき，対数尤度関数は

$$\ln L = \sum_{Z=1}\ln f\left(e_i, 2\sigma_u^2\right) + \sum_{Z=1}\ln F\left[\left(\mathbf{w}_i'\boldsymbol{\gamma} + r\frac{e_i}{\sqrt{2\sigma_u^2}}\right)\left(1-r^2\right)^{-\frac{1}{2}}\right] + \sum_{Z=0}\ln F\left(-\mathbf{w}_i'\boldsymbol{\gamma}\right) \tag{14}$$

となる。

3. データ

本研究で利用するデータは，株式会社リクルートの情報誌である『週刊住宅情報』および『住宅情報タウンズ』に掲載された戸建住宅を対象として，東京都世田谷区において 1986 年 1 月から 2009 年 7 月までの期間に取引された物件である。情報が不完全なものや異常値と思われるものを除いたサンプル・サイズは 30,916 である。このうち 1 度だけ取引された住宅は 28,131 件，2 度取引された住宅は 2,785 件である。

2 度取引された住宅は RS 法のサンプルとして利用するが，本論文では時間効果を推定するための時間の単位を年次にするため，取引間隔が 1 年未満の物件を除いている。このことは，投機目的で短期的に売買が繰り返されるケースを排除して，できるだけ居住目的の価格指数に近い値を計測するための必要な措置である。

第 1 表は住宅属性の記述統計を示している。サンプル・サイズは 30,916 であり，このうち RS 法が適用される 2 度取引された住宅は 2,785 件である。その割合（2 値変数 Z の平均）は 2785/30916＝0.090 である。『平成 18 年度 国土交通白書』によると，既存住宅の流通量のシェアは米国で 77.6%，イギリスで 88.8%，フランスで 66.4%となっているのに対して，日本は 13.1%と極端に低い値になっている。2 度目の取引サンプルは確実に中古物件であるから，本研究で利用するデータは特に偏っているわけではない。

第 1 表 住宅属性の記述統計

変数	単位	平均	標準偏差	最小	最大
Z		0.090	0.286	0	1
都心までの時間距離	[分]	12.8	5.3	1	25
占有面積	[m^2]	111.4	51.4	10.0	1475.4
土地面積	[m^2]	104.5	64.7	13.1	1648.1
最寄駅までの徒歩時間	[分]	10.0	4.8	1	36
建築後年数	[年]	5.6	8.4	0	69
取引年次	[年]	2000.6	5.3	1986	2009
竣工年次	[年]	1995.1	10.4	1929	2009

注. サンプル・サイズは 30,916 であり，このうち RS 法が適用される 2 度取引された住宅は 2,785 件である。その比率（2 値変数 Z の平均）は 2785/30916＝0.090 である。

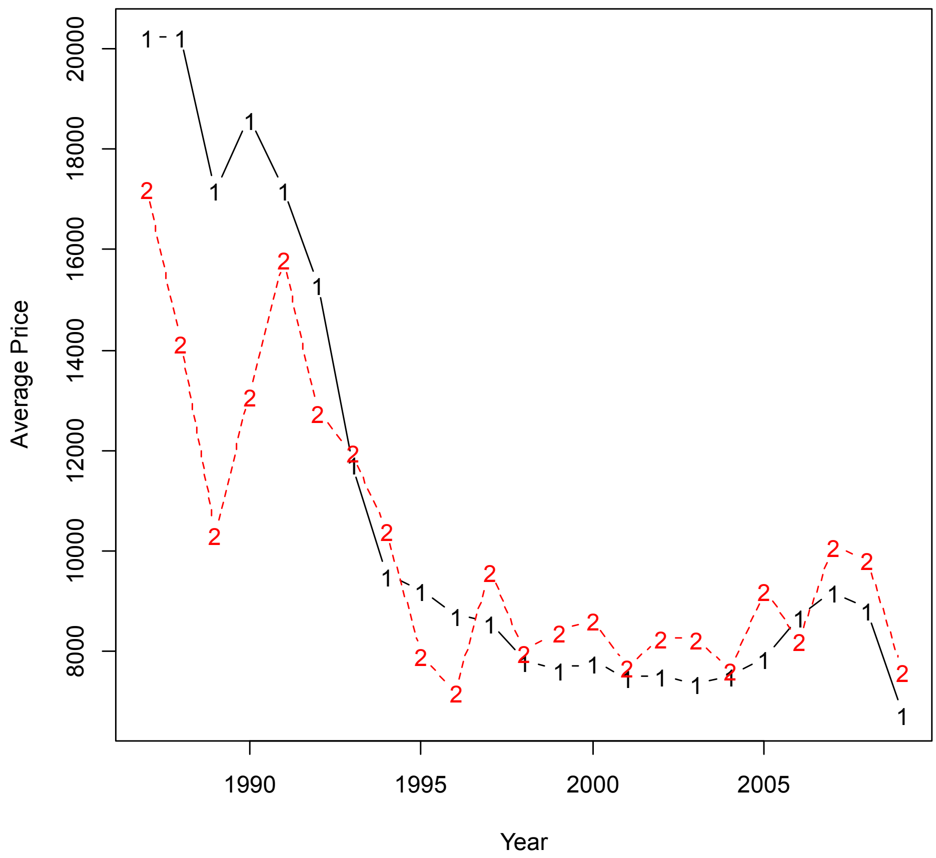
第 2 表 取引価格の記述統計（単位: 万円）

		平均	標準偏差	最小	最大
取引が 1 回の物件		8,848	6,531	1,000	200,410
取引が 2 回の物件	1 回目	10,151	7,047	1,850	87,550
	2 回目	8,756	5,581	1,800	70,000

注. サンプル・サイズは取引が 1 回の物件が 28,131，取引が 2 回の物件は 2,785 件である。

第 2 表は取引価格の記述統計を示している。1 回だけ取引された住宅と 2 回取引された住宅では価格に大きな差異がある。また 2 回取引されたサンプルでは，1 回目に比べて 2 回目の価格は低くなり，ばらつきも小さくなっている。図 1 は 1 回だけ取引された住宅と 2 回取引された住宅のそれぞれの価格の推移を示している。1987 年から 1992 年までは 2 回取引された住宅は 1 回だけ取引されたそれに比べ低い水準で売買されており，1993 年以降は 2 回取引された住宅の方が高い価格で売買されることが幾つかの時点で観察できる[4]。

第 3 表はリピート・セールス価格の取引年次別のサンプル・サイズ，価格，竣工年を示している。我々のデータベースは 80 年代後半の物件が若干少なく，再販売という性質上，2 回目は特に少ない。



第 1 図 平均価格の推移（1: 1 回だけ取引された物件，2: 2 回取引された物件）

[4] 1986 年のデータは除いている。

第 3 表 RS 価格の年次別記述統計

取引年次	1 回目の取引 観測値の数	平均価格	平均竣工年	2 回目の取引 観測値の数	平均価格	平均竣工年
1986	27	6,436	1980	N.A.	N.A.	N.A.
1987	29	15,235	1980	4	17,220	1979
1988	14	15,734	1979	2	14,150	1982
1989	74	13,168	1980	12	10,351	1979
1990	61	16,739	1980	26	13,093	1979
1991	97	15,165	1981	27	15,816	1977
1992	61	16,124	1980	36	12,756	1981
1993	80	13,184	1985	46	11,971	1979
1994	84	9,132	1985	58	10,412	1983
1995	95	10,260	1985	77	7,932	1984
1996	90	9,931	1987	78	7,193	1984
1997	246	9,333	1991	85	9,601	1984
1998	189	9,645	1991	173	7,997	1990
1999	164	10,199	1991	187	8,406	1989
2000	164	9,735	1992	149	8,636	1989
2001	178	9,413	1995	200	7,688	1990
2002	207	8,094	1997	205	8,285	1994
2003	186	7,942	1999	227	8,247	1995
2004	168	9,926	2000	243	7,630	1996
2005	192	8,110	2001	206	9,227	1997
2006	120	10,223	2001	227	8,214	1999
2007	164	11,071	2004	160	10,110	1999
2008	95	8,547	2004	220	9,838	2000
2009	N.A.	N.A.	N.A.	137	7,615	2002

4. 推定結果

4.1 セレクション関数と Case-Shiller 型 RS 回帰モデルの推定結果

第 4 表は (5) 式のプロビット推定の結果を示している。占有面積が広く，建築後年数の古い物件は, 売りに出す確率を有意に高めていることがわかる。また, 失業率が高くなり，経済成長率が低くなる場合にも確率を高める要因になる。売りに出されるかどうかは，マクロ経済の景況に依存しており，景気の悪化は再販売の確率を高めると予想できる。

プロビット推定から逆ミルズ比 λ_i が計測できる。λ_i の平均は 1.75，標準偏差は 0.20 である。また, (6) の RS 回帰モデルにおいて, セレクション・バイアスを制御する $\lambda_i/\sqrt{S_i}$ の平均は 1.46，標準偏差は 0.52 になる。

第 4 表 セレクション関数 (5) 式のプロビット推定

説明変数	推定値	t 値	限界効果
定数項	−2.30	−15.91***	
都心までの時間距離	0.00	1.68*	0.001
対数占有面積	0.13	2.83***	0.020
対数土地面積	−0.04	−1.18	−0.007
最寄駅までの徒歩時間	0.00	−0.61	0.000
建築後年数	0.03	22.43***	0.004
完全失業率	0.09	6.58***	0.014
経済成長率	−0.02	−3.33***	−0.003
観測値の数	30916		
最大対数尤度	−9057		
尤度比	605		
	[.000]		

注. *** は 1%水準，** は 5%，* は 10% で有意であることを示している。尤度比は定数項を除くすべてのパラメタがゼロであるという帰無仮説についての検定統計量であり，[ ] 内はカイ 2 乗分布での p 値を示している。失業率および経済成長率は前年の値を利用している。

第 5 表 RS 回帰モデルの推定結果

	Eq. (6) SEL		Eq. (4) CS	
変数	推定値	t 値	推定値	t 値
D_1987	0.51	11.05***	0.49	10.55***
D_1988	0.52	8.74***	0.51	8.37***
D_1989	0.47	9.53***	0.45	8.83***
D_1990	0.54	10.81***	0.49	9.78***
D_1991	0.52	10.37***	0.47	9.22***
D_1992	0.35	6.90***	0.29	5.52***
D_1993	0.18	3.58***	0.11	2.11**
D_1994	0.08	1.57	0.00	−0.03
D_1995	−0.03	−0.64	−0.12	−2.38**
D_1996	−0.10	−2.02**	−0.20	−3.87***
D_1997	−0.08	−1.57***	−0.18	−3.48***
D_1998	−0.14	−2.61***	−0.26	−4.99***
D_1999	−0.18	−3.53***	−0.32	−6.20***
D_2000	−0.22	−4.24***	−0.36	−7.05***
D_2001	−0.27	−5.00***	−0.42	−8.08***
D_2002	−0.31	−5.85***	−0.48	−9.28***
D_2003	−0.33	−6.04***	−0.51	−9.86***
D_2004	−0.31	−5.57***	−0.51	−9.81***
D_2005	−0.28	−4.91***	−0.50	−9.52***
D_2006	−0.24	−4.13***	−0.49	−9.23***
D_2007	−0.18	−3.05***	−0.45	−8.47***
D_2008	−0.23	−3.82***	−0.54	−9.96***
D_2009	−0.33	−5.14***	−0.66	−12.08***
$\lambda/\sqrt{S}$	−0.02	−10.08***		
$(\hat{\rho})$	−0.1880			
$(\hat{\sigma}_u)$	0.1158		0.1353	
adj. R^2	0.2800		0.2630	

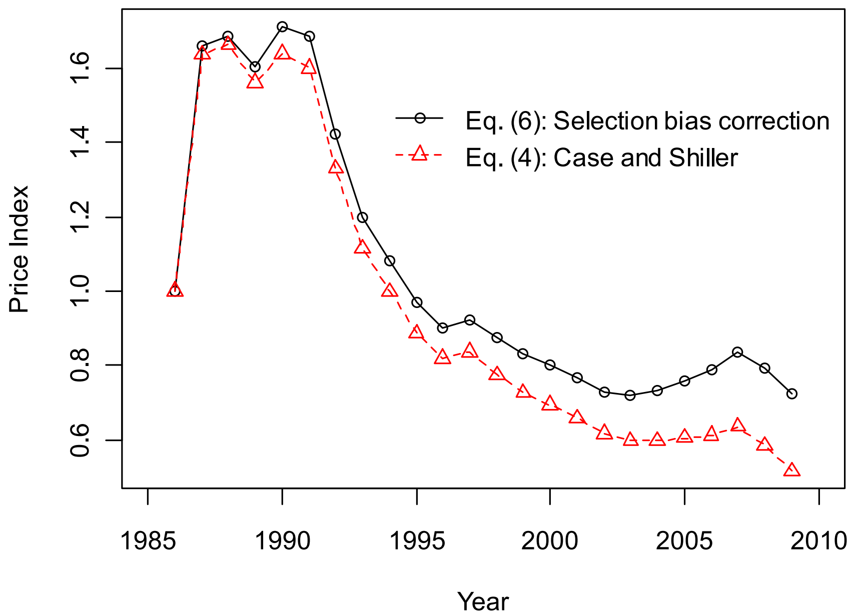
注. *** は 1%水準，** は 5%水準で有意であることを示している。(6) 式の推定値の標準誤差は Greene (1981) にしたがって再計算している。

第 5 表に RS 回帰モデル (6) および (4) の推定結果を示した。RS 回帰モデルはセレクション関数で用いられる変数と比べ大幅に異なっており，最小 2 乗法で (6) を推定したとしても多重共線性の問題はほとんど起こらないと考えられる。(6) はセレクション・バイアスを修正した RS 回帰モデル (SEL) の, (4) は Case-Shiller 型の分散不均一のみを修正した RS 回帰モデル (CS) の結果を示している。

セレクション・バイアスを制御した結果を見ると，重み付きの逆ミルズ比 $\lambda_i/\sqrt{S_i}$ の係数推定値($\hat{\sigma}_{ev}$)は有意に負であり，(3) と (5) の誤差項の相関係数も負になる。このことから，2 回取引された物件固有の性質が価格の変化率に対して負の影響をもたらしていることが予想される。

セレクション・バイアスを制御していない Case-Shiller 型の推定結果を見ると，時間効果はやや低い値で推定されている。(3) と (5) の相関が負であることから，除外変数バイアスによる影響が考えられる。図 2 は時間効果の推定値より計算した 1986 年を基準年とした価格指数である。(4) の価格指数はやや下方にバイアスを持っていることがわかる。価格変化率は平均で−1.45%，最大で−2.94% だけ (6) に比べて過少になる。

2.1 節で示したように価格指数 $\exp(\hat{\delta}_t)$ の期待値は $E[\exp(\hat{\delta}_t)] = \exp[\delta_t + \frac{1}{2}\sigma^2\Gamma_{tt}]$ よりバイアスを持つ。推定値 $\hat{\delta}_t$ の標準誤差より $\sigma^2\Gamma_{tt}$ を推定すると，(6) のケースでは $E[\exp(\hat{\delta}_t)]$ は $\exp[\delta_t + \frac{1}{2}\sigma^2\Gamma_{tt}]$ よりも平均で 0.33% 小さく, (4) のケースでは 0.14% 小さい。この差は (4) と (6) との差に比べて遥かに小さいといえる。



第 2 図 RS 価格指数（1986 年価格を 1 に基準化）

4.2 自己回帰過程をもつ RS 回帰モデルの推定結果

(7) と (8) の誤差構造を持つ RS 回帰モデルの推定結果を第 6 表に示した。分散不均一を考慮して (8) における変数として $\mathbf{q}_i =$ (都心までの時間距離，占有面積，土地面積) を利用した。(10) の推定結果より，自己回帰係数 $\hat{\rho}$ は有意であり，推定値は $|\rho|<1$ を満たす。したがって，RS 回帰モデルの誤差項はランダム・ウォークでない可能性がある。(12) はこれらの結果を利用して実現可能な一般化最小 2 乗法で推定した RS 回帰モデルの推定結果である。

第 6 表 自己回帰過程をもつ RS 回帰モデルの推定結果

	(10) 最尤法		(12) 一般化最小 2 乗法	
変数	推定値	標準誤差	推定値	標準誤差
D_1987	0.571***	0.013	0.564***	0.045
D_1988	0.571***	0.026	0.568***	0.057
D_1989	0.497***	0.016	0.491***	0.045
D_1990	0.560***	0.016	0.549***	0.044
D_1991	0.530***	0.016	0.534***	0.044
D_1992	0.371***	0.016	0.354***	0.045
D_1993	0.200***	0.017	0.189***	0.044
D_1994	0.089***	0.017	0.075*	0.044
D_1995	−0.025	0.016	−0.041	0.043
D_1996	−0.098***	0.017	−0.118***	0.043
D_1997	−0.083***	0.016	−0.100**	0.043
D_1998	−0.161***	0.016	−0.176***	0.043
D_1999	−0.217***	0.016	−0.236***	0.043
D_2000	−0.256***	0.016	−0.273***	0.043
D_2001	−0.301***	0.016	−0.322***	0.043
D_2002	−0.353***	0.017	−0.380***	0.043
D_2003	−0.376***	0.017	−0.405***	0.043
D_2004	−0.366***	0.017	−0.394***	0.043
D_2005	−0.341***	0.017	−0.371***	0.044
D_2006	−0.320***	0.017	−0.350***	0.044
D_2007	−0.269***	0.017	−0.298***	0.044
D_2008	−0.341***	0.017	−0.370***	0.045
D_2009	−0.458***	0.018	−0.490***	0.046
都心までの時間距離 q_1	0.028***	0.002		
対数占有面積 q_2	0.000*	0.000231		
対数土地面積 q_3	0.003***	0.000199		
$\ln \hat{\sigma}_u^2$	−5.669***	0.023		
$\hat{\rho}$	0.709***	0.009		
観測値の数	2785		2785	
最大対数尤度	7349.44		2563.39	
回帰の標準誤差 $\hat{\sigma}_u$	0.059		0.097	

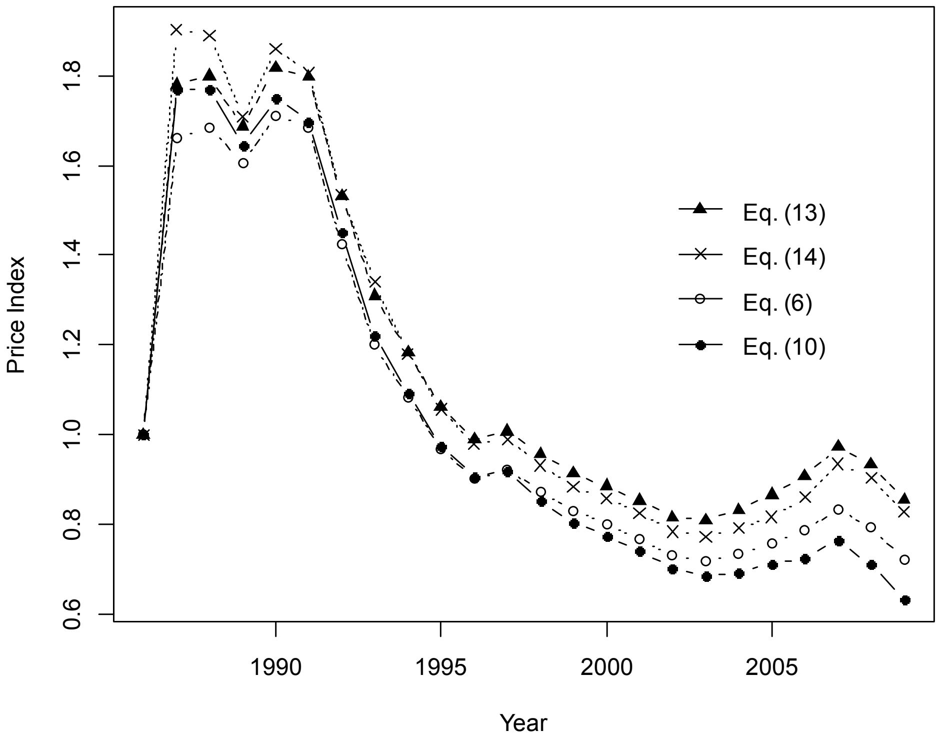
注. *** は 1%水準，** は 5%水準，*は 1%水準で有意であることを示している。

第 7 表はすべてのデータを利用して最尤法でセレクションを除去した推定結果を示している。ただし，(13) は (10) で得られた分散不均一を修正しており，(14) は (7) において $\rho = 0$，(8) において $\boldsymbol{\theta} = 0$ となる制約をおいている。セレクション関数の推定結果に大きな違いはないが, RS 回帰モデルにおける時間効果や方程式間の相関係数 r の推定に若干の相違がある。ただし，どちらの場合でもセレクション関数と RS 回帰モデルの間の相関は有意に負であり，第 5 表の結果とも整合的である。このことから，RS 価格指数には住宅市場の母集団に比べて負のバイアスが生じている可能性があるといえる。

第 7 表 セレクション・バイアスを考慮した自己回帰過程をもつ RS 回帰モデルの推定結果

	(13) 最尤法		(14) 最尤法	
変数	推定値	標準誤差	推定値	標準誤差
プロビット定数項	−2.160***	0.145	−2.180***	0.145
都心までの時間距離	0.003	0.002	0.003	0.002
対数占有面積	0.113**	0.045	0.113**	0.045
対数土地面積	−0.054	0.036	−0.051	0.036
最寄駅までの徒歩時間	−0.002	0.002	−0.002	0.002
建築後年数	0.025***	0.001	0.026***	0.001
完全失業率	0.089***	0.014	0.090***	0.014
経済成長率	−0.022***	0.006	−0.022***	0.006
D_1987	0.577***	0.044	0.644***	0.037
D_1988	0.587***	0.055	0.636***	0.048
D_1989	0.523***	0.043	0.536***	0.033
D_1990	0.598***	0.043	0.620***	0.033
D_1991	0.587***	0.043	0.591***	0.033
D_1992	0.426***	0.044	0.428***	0.034
D_1993	0.270***	0.043	0.292***	0.032
D_1994	0.168***	0.043	0.166***	0.032
D_1995	0.060	0.043	0.055*	0.031
D_1996	−0.010	0.043	−0.022	0.032
D_1997	0.008	0.043	−0.011	0.031
D_1998	−0.044	0.043	−0.071**	0.032
D_1999	−0.090**	0.043	−0.123**	0.032
D_2000	−0.121***	0.044	−0.153***	0.032
D_2001	−0.158***	0.044	−0.191***	0.032
D_2002	−0.203***	0.045	−0.243***	0.032
D_2003	−0.212***	0.045	−0.257***	0.032
D_2004	−0.183***	0.046	−0.234***	0.033
D_2005	−0.143***	0.047	−0.203***	0.033
D_2006	−0.098**	0.048	−0.150**	0.034
D_2007	−0.028	0.049	−0.067*	0.034
D_2008	−0.069	0.050	−0.102***	0.035
D_2009	−0.156***	0.053	−0.190***	0.037
$\hat{\sigma}_u$	0.096***	0.001	0.077***	0.002
$\hat{r}$	−0.197***	0.017	−0.161***	0.014
観測値の数	30916		30916	
最大対数尤度	−6426.34		−7755.30	

注. *** は 1%水準，** は 5%水準，*は 1%水準で有意であることを示している。



第 3 図 RS 価格指数（1986 年価格を 1 に基準化）

(10), (13) および (14) と 4.1 節で示した (6) の時間効果推定値を利用して価格指数を描いたものが第 3 図である。サンプル・セレクションを考慮した (13) または (14) は考慮していない (10) に比べ価格指数の水準は高い。自己回帰係数が $\rho = 1$ である (6) の結果に比べて，$|\rho| < 1$ を想定した (13), (14) の価格指数は高い水準にあり大きな違いがあるといえる。

4.3 セレクション・バイアスの評価

2 回の取引が行われる物件だけをサンプルとして分析するとき，(4) や (10) の結果のように時間効果が負のバイアスを持つという予想にはどのような経済的意味が考えられるであろうか。Clapp and Giaccotto (1992) は，再販売される確率が高い住宅には，粗悪な品質（いわゆる「レモン」）の住宅が混在していることを指摘している。本研究で利用したデータにおいても 2 回取引された物件にそのようなデータが混在している可能性がある。

また，Ooi (2006) や Pope (2008) で分析されているように，一度取引されたことのある中古住宅について，購入者が観察できない不確実性がある場合には，リスク回避的な購入者は期待されるよりも過少に住宅の質を評価する。他の事情が等しければ，住宅の確実性等価が低くなり，リスク中立的な購入者に比べて付け値が低くなる可能性がある。住宅が

再販売されるのは，それなりの理由があるからであり，本研究の結果はRS サンプルが母集団に比べ偏っている可能性があることを示唆している。

複数回取引に出される物件は住宅市場全体でも限られており，通常の RS 法では 1 度しか取引されていないデータは全く利用されない。したがって，セレクション・バイアスの問題だけでなく，サンプル・サイズが減少することによる推定の効率性低下もまねく。McMillen (2012) はそのような問題への対処としてマッチング法を利用した価格指数の推定方法を提案している。ヘドニック回帰モデルにおける時間効果が政策評価における処理効果 (treatment effects) に対応しているものと考えれば，RS 法の価格変化率は「処理後」の代表的な値と「処理前」の代表的な値の差を観察していることがわかる。プロジェクト評価（project evaluation）の手法を用いてサンプル全体の平均としてどの程度の価格変化が生じたのかを測るには，平均処理効果 (Average Treatment Effect; *ATE*) を求める必要がある。平均処理効果を「処理群の平均処理効果」として近似するには，どの観察時期のデータもランダム・サンプリングされたものでなければならない。リピート・セールス法で用いられるデータの場合，処理群のデータは実際に売買された時点でしか観察できない。したがって, 基準年の対照群に対応するデータを各時点においてマッチングする必要がある。マッチングを行うには，まず，時間ダミーをヘドニック回帰モデルで用いられる属性変数にロジット回帰した上で傾向スコア (propensity score) を求める。次に，カーネル・マッチング (Heckman, Ichimura and Todd 1998) により，処理群となる各期のマッチング・データを生成し，最後に処理効果を計算して価格指数を計測する。

McMillen (2012) によればマッチング推定量はヘドニック・アプローチによる推定量がきわめてロバストであることを示している。価格指数を推定するためのリピート・セールス法はマッチング法による推定の極端なヴァージョンであるといえる。必ずしも同一ではないが似ている住宅取引データのマッチングは，価格指数推定においてリピート・セールス法あるいはヘドニック法のどちらと比べてもいくつかの利点がある。標準的なリピート・セールス法と比較すると, 一般的なマッチング法はサンプル･サイズを飛躍的に増大させ, より効率的な推定値を得る傾向が高まる。たとえば，小規模な地域，短い観察期間などのサンプルが少ない場合においてさえ, 価格指数を作成することを可能にするかもしれない。そのような意味でマッチングによる推定は有用な価格指数の計測手法となりえるかもしれない。

5. まとめ

住宅の売り手が直面する経済状況は，年々刻々と変化しているわけであるから，誤差構造が変化していると考えるのは自然な想定である。本研究は，Case-Shiller 型のリピート・セールス法がもたらすサンプル・セレクション・バイアスの修正において，セレクション関数と RS 回帰モデルとの間の共分散を推定するために逆ミルズ比の修正し，Gatzlaff and Haurin (1997) の改善を行った。

東京都世田谷区のデータによる分析結果より，セレクション・バイアスを制御した重み付きの逆ミルズ比の係数推定値は有意に負であり，相関係数も負になる。このことから，2 回取引された物件の価格には、住宅市場の母集団に比べて負のバイアスが生じる可能性がある。

さらに，ランダム・ウォークではなく系列相関および分散不均一を持つ誤差構造のもとでのセレクションの除去を行った。本研究では，Hill, Sirmans and Knight (1999) における RS 法の分散不均一特定化手法を利用して，誤差項がランダム・ウォークであるのかどうかを検証するとともにセレクションの除去を行った。この場合でもセレクション・バイアスを制御した場合のリピート・セールス回帰モデルとセレクション関数の相関は有意に負であり，2 回取引された物件の価格指数は負のバイアスを持つ可能性があることを示した。

ただし，本研究ではデータの制約から 1 度も取引されていない物件を含めることができなかった。実際には、そのような物件は非常に少ないと思われるが、セレクション関数を 1 回目の取引時点についても考慮して分析する必要がある。このような問題点は今後の課題である。

参考文献

[1] Bailey, M. J., R. F. Muth and H. O. Nourse (1963). “A Regression Model for Real Estate Price Index Construction,” *Journal of American Statistical Association* 58, 933−942.

[2] Case, K. E. and R. J. Shiller (1987). “Prices of Single Family Homes since 1970: New Indexes for Four Cities,” *New England Economic Review*, (Sept./Oct.) 45−56.

[3] Case, K. E. and R. J. Shiller (1989). “The Efficiency of the Market for Single−Family Homes,” *The American Economic review*, 79 (1), 125−137.

[4] Clapp, J. M. and C. Giaccotto (1992). “Estimating Price Trends for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods,” *Journal of Real Estate Finance and Economics* 5, 357−374.

[5] Diewert, W. E., Heravi, S. and M. Silver (2007). "Hedonic imputation indexes versus time dummy Hedonic indexes," *IMF Working Paper*, WP/07/234.

[6] Gatzlaff, D., and D. Ling. (1994). "Measuring Changes in Local House Prices: An Empirical Investigation of Alternative Methodologies," *Journal of Urban Economics* 35 (2), 221-244.

[7] Gatzlaff, D. H. and D. R., Haurin (1997). "Sample Selection Bias and Repeat-Sales Index Estimates," *Journal of Real Estate Finance and Economics* 14, 33−50.

[8] Gatzlaff, D. H. and D. R., Haurin (1998). "Sample Selection and Biases in Local House Value Indices," *Journal of Urban Economics* 43, 199−222.

[9] Greene, W. H. (1981). "Sample Selection Bias as a Specification Error: Comment," *Econometrica* 49, 795−798.

[10] Greene, W. H. (2008). *Econometric Analysis* 6th ed. (Chapter 24, 884-887), Prentice Hall.

[11] Harvey, A. C. (1976), "Estimating Regression Models with Multiplicative Heteroscedasticity," *Econometrica* 44, 461–465.

[12] Heckman, J. (1979). "Sample Selectio n Bias as a Specification Error," *Econometrica* 47, 153−161.

[13] Heckman, J., H. Ichimura and P. Todd (1998). "Matching as an econometric evaluation estimator," *Review of Economic Studies* 65, 261–294.

[14] Hill, R.J. and D. Melser (2008). "Hedonic imputation and the price index problem: an application to housing," *Economic Inquiry* 46, 593–609.

[15] Hill, R. C., C. F. Sirmans and J. R. Knight (1999), "A random walk down main street?," *Regional Science and Urban Economics* 29, 89–103.

[16] Karato, K., O. Movshuk and C. Shimizu (2010). "Semiparametric Estimation of Time, Age and Cohort Effects in An Hedonic Model of House Prices," University of Toyama, Working Paper No. 256.

[17] McMillen, D. P. (2012). Repeat sales as a matching estimator. *Real Estate Economics* 40, 745–773.

[18] Pope, J. C. (2008). "Do Seller Disclosures Affect Property Values? Buyer Information and the Hedonic Model," *Land Economics* 84 (4), 551−572.

[19] Ooi, J.T.L., C.F. Sirmans and K.T., Geoffrey (2006). "Price Formation Under Small Numbers Competition: Evidence from Land Auctions in Singapore," *Real Estate Economics* 34, 51−76.

[20] Rosen, S. (1974). "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy* 82, 34−55.

[21] Sirmans, G.S. and MacDonald, L. and Macpherson, D.A. and Zietz, E.N. (2006). "The value of

housing characteristics: a meta analysis," *Journal of Real Estate Finance and Economics* 33, 215–240.

[22] Wong, S. K., K.W. Chau, K. Karato and C. Shimizu (2013). "Separating the Age Effect from a Repeat Sales Index: Land and Structure Decomposition," The University of Tokyo, CSIS Discussion Paper, No.125.

[23] 唐渡広志・清水千弘 (2010). 「サンプル・セレクション・バイアスを除去したリピート・セールス価格指数の計測」, 日本不動産学会秋季全国大会（東京大学）報告論文.

[24] 唐渡広志・清水千弘・中川雅之・原野啓 (2012). 「リピートセールス不動産価格指数における集計バイアス」,『日本経済研究』, 66, 22–50.

[25] 唐渡広志・W. Erwin Diewert・清水千弘 (2013). 「不動産価格指数の理論と推計サーベイ: 品質調整済み不動産価格指数の推計方法」, University of Toyama, Working Paper, No.284.