9.3　回帰分析への応用

y_i を被説明変数 (スカラー)，x_i を説明変数 ($1 \times k$ ベクトル) とする。

回帰モデルは，関数形を特定化しないで，次のように表される。

$$y_i = m(x_i) + u_i, \qquad i = 1, 2, \cdots, n \tag{45}$$

$m(\cdot)$ が未知 → $m(\cdot)$ を推定 → ノンパラメトリック回帰

ただし，誤差項 u_i は平均 0，分散 $\sigma^2(x)$ のある分布に従うものとする。

y と x の密度関数を $f_{yx}(y, x)$，x の密度関数を $f(x)$ とするとき，y の条件付期待値は,

$$\mathrm{E}(y|x) = m(x) = \int \frac{y f_{yx}(y, x)}{f(x)} \mathrm{d}y$$

となり，$m(x)$ によって表される。

$f(x)$ は,

$$f(x) = \int f_{yx}(y, x) \mathrm{d}y$$

となる。$\longrightarrow$ $f(x)$ は周辺分布

2 つの密度関数 $f_{yx}(y,x)$，$f(x)$ の推定量を $\hat{f}_{yx}(y,x)$，$\hat{f}(x)$ として，

$$\hat{f}_{yx}(y,x) = \frac{1}{nh^{k+1}} \sum_{i=1}^{n} K_{yx}\Big(\frac{y-y_i}{h}, \frac{x-x_i}{h}\Big)$$

$$\begin{aligned}
\hat{f}(x) \equiv \int \hat{f}_{yx}(y,x)\mathrm{d}y &= \int \frac{1}{nh^{k+1}} \sum_{i=1}^{n} K_{yx}\Big(\frac{y-y_i}{h}, \frac{x-x_i}{h}\Big)\mathrm{d}y \\
&= \frac{1}{nh^{k+1}} \sum_{i=1}^{n} \int K_{yx}\Big(\frac{y-y_i}{h}, \frac{x-x_i}{h}\Big)\mathrm{d}y = \frac{1}{nh^{k+1}} \sum_{i=1}^{n} h \int K_{yx}\Big(u, \frac{x-x_i}{h}\Big)\mathrm{d}u \\
&= \frac{1}{nh^{k}} \sum_{i=1}^{n} K\Big(\frac{x-x_i}{h}\Big)
\end{aligned} \tag{46}$$

とする。

$\hat{f}(x)$ の 4 番目の等式では，$u = \dfrac{y-y_i}{h}$ として変数変換。

$\hat{f}(x)$ の 5 番目の等式では，$\displaystyle\int K_{yx}(z)\mathrm{d}y = K(x)$ を利用。

$z = (y, x)$ として，$K_{yx}(z)$ と $K(x)$ の性質は,

$$\int K_{yx}(z)\mathrm{d}z = 1, \qquad \int zK_{yx}(z)\mathrm{d}z = 0, \qquad \int zz'K_{yx}(z)\mathrm{d}z = \Omega_*,$$
$$\int K(x)\mathrm{d}x = 1, \qquad \int xK(x)\mathrm{d}x = 0, \qquad \int xx'K(x)\mathrm{d}x = \Omega,$$
$$\int K_{yx}(z)\mathrm{d}y = K(x)$$

として表される。

Ω_* は $(k+1)\times(k+1)$ 行列，Ω は $k\times k$ 行列である。Ω_* と Ω との関係は,

$$\Omega_* = \begin{pmatrix} \omega_y^2 & \cdots \\ \vdots & \Omega \end{pmatrix}$$

となるものとする。

実証分析では，ω_y^2 は $y_i, i = 1, 2, \cdots, n$ から得られた標本分散，Ω は $x_1, x_2, \cdots, x_n$ から計算された $(1/n)\sum_{i=1}^n x_i' x_i$ とする。

このとき，$m(x)$ の推定量 $\hat{m}(x)$ は，$f_{yx}(y,x)$ と $f(x)$ を $\hat{f}_{yx}(y,x)$ と $\hat{f}(x)$ で置き換えると，

$$\hat{m}(x) = \int \frac{y\hat{f}_{yx}(y,x)}{\hat{f}(x)}\mathrm{d}y = \frac{\dfrac{1}{nh^{k+1}}\displaystyle\sum_{i=1}^{n}\int yK_{yx}\Big(\frac{y-y_i}{h},\frac{x-x_i}{h}\Big)\mathrm{d}y}{\dfrac{1}{nh^k}\displaystyle\sum_{i=1}^{n}K\Big(\frac{x-x_i}{h}\Big)}$$

$$= \frac{\dfrac{1}{nh^{k}}\displaystyle\sum_{i=1}^{n}\int (y_i+hu)K_{yx}\Big(u,\frac{x-x_i}{h}\Big)\mathrm{d}u}{\dfrac{1}{nh^k}\displaystyle\sum_{i=1}^{n}K\Big(\frac{x-x_i}{h}\Big)}$$

$$= \frac{\dfrac{1}{nh^{k}}\displaystyle\sum_{i=1}^{n}y_iK\Big(\frac{x-x_i}{h}\Big)}{\dfrac{1}{nh^k}\displaystyle\sum_{i=1}^{n}K\Big(\frac{x-x_i}{h}\Big)} = \frac{\hat{g}(x)}{\hat{f}(x)} = \frac{\displaystyle\sum_{i=1}^{n}y_iK\Big(\frac{x-x_i}{h}\Big)}{\displaystyle\sum_{j=1}^{n}K\Big(\frac{x-x_j}{h}\Big)} = \sum_{i=1}^{n}y_i\omega_i(x) \tag{47}$$

となる。

2 行目では $(y - y_i)/h = u$ として変数変換を行っている。

3 行目では $\int uK_{yx}(u, v)\mathrm{d}u = 0$ が用いられている。

ただし，分子の $\hat{g}(x)$ を

$$\hat{g}(x) = \frac{1}{nh^k}\sum_{i=1}^{n} y_i K\left(\frac{x - x_i}{h}\right)$$

と定義する。

また，

$$\omega_i(x) = \frac{K\left(\frac{x - x_i}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{x - x_j}{h}\right)}$$

とする。

$\hat{m}(x)$ の漸近的性質は，$nh^k \longrightarrow \infty$ のとき，次のようになることが知られている。

$$(nh^k)^{1/2}(\hat{m}(x) - m(x)) \sim N\left(0, \frac{\sigma^2(x)}{f(x)} \int K^2(u)\mathrm{d}u\right) \tag{48}$$

ただし，

$$\sigma^2(x) = \int \frac{(y - m(x))^2 f(y, x)}{f(x)} \mathrm{d}y$$

とする。

このように，$\hat{m}(x)$ は $m(x)$ の一致推定量となり，漸近的に正規分布に従うことになる。

言い換えると，n が大きいとき，近似的に，

$$\hat{m}(x) \sim N\left(m(x),\ \frac{\sigma^2(x)}{f(x)}\frac{1}{nh^k}\int K^2(u)\mathrm{d}u\right)$$

となる。

実証分析において，$\hat{m}(x)$ の漸近分散は，(48) 式の $\sigma^2(x)$ と $f(x)$ をその推定量 $\hat{\sigma}^2(x)$，$\hat{f}(x)$ で置き換えて，求められる。

すなわち，

$$\hat{m}(x) \sim N\left(m(x),\ \frac{\hat{\sigma}^2(x)}{\hat{f}(x)}\frac{1}{n\hat{h}^k}\int K^2(u)\mathrm{d}u\right)$$

が利用される。

実践では，$m(x)$ の $100(1-\alpha)\%$ 信頼区間は，

$$\left(\hat{m}(x) - z_{\alpha/2}\Big(\frac{1}{n\hat{h}^k}\frac{\hat{\sigma}^2(x)}{\hat{f}(x)}\int K^2(u)\mathrm{d}u\Big)^{1/2},\quad \hat{m}(x) + z_{\alpha/2}\Big(\frac{1}{n\hat{h}^k}\frac{\hat{\sigma}^2(x)}{\hat{f}(x)}\int K^2(u)\mathrm{d}u\Big)^{1/2}\right)$$

となる。

$z_{\alpha/2}$ は $N(0,1)$ の上側 $100\alpha/2\%$ 点とする ($\alpha=0.05$ のとき，$z_{\alpha/2}=1.96$)。

ただし，$\hat{f}(x) = \dfrac{1}{nh^k}\displaystyle\sum_{i=1}^{n} K\Big(\frac{x-x_i}{h}\Big)$, $\qquad \hat{m}(x) = \dfrac{\displaystyle\sum_{i=1}^{n} y_i K(\frac{x-x_i}{\hat{h}})}{\displaystyle\sum_{j=1}^{n} K(\frac{x-x_j}{\hat{h}})}$

$$\hat{\sigma}^2(x) = \frac{\displaystyle\sum_{i=1}^{n}(y_i-\hat{m}(x_i))^2 K(\frac{x-x_i}{\hat{h}})}{\displaystyle\sum_{j=1}^{n} K(\frac{x-x_j}{\hat{h}})}, \qquad \hat{h} = 1.06 s_{x_i} n^{-1/5}$$

とする (上の $\hat{h}$ は 1 つの例，s_{x_i} は i 番目の x 変数の標準偏差)。

さらに，$K(\cdot)$ を $N(0,1)$ と仮定すれば，

$$
\begin{aligned}
\int K^2(u)\mathrm{d}u &= \int\Big(\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)\Big)^2\mathrm{d}u = \int\frac{1}{2\pi}\exp(-u^2)\mathrm{d}u \\
&= \frac{1}{2\sqrt{\pi}}\int\frac{1}{\sqrt{2\pi(1/2)}}\exp(-\frac{1}{2(1/2)}u^2)\Big)^2\mathrm{d}u = \frac{1}{2\sqrt{\pi}}
\end{aligned}
$$

となる。

最後の等式は $N(0,1/2)$ を利用。

(*) $N(\mu,\sigma^2)$ の密度関数は，

$$
f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2\sigma^2}(x-\mu)^2)
$$

である。

よって，この場合，$m(x)$ の $100(1-\alpha)\%$ 信頼区間は，

$$\left(\hat{m}(x) - z_{\alpha/2}\Big(\frac{1}{n\hat{h}^k}\frac{\hat{\sigma}^2(x)}{\hat{f}(x)}\frac{1}{2\sqrt{\pi}}\Big)^{1/2},\quad \hat{m}(x) + z_{\alpha/2}\Big(\frac{1}{n\hat{h}^k}\frac{\hat{\sigma}^2(x)}{\hat{f}(x)}\frac{1}{2\sqrt{\pi}}\Big)^{1/2}\right) \tag{49}$$

再度，まとめると，$\hat{f}(x)$，$\hat{m}(x)$，$\hat{\sigma}^2(x)$，$\hat{h}$ は，

$$\hat{f}(x) = \frac{1}{nh^k}\sum_{i=1}^{n} K\Big(\frac{x-x_i}{h}\Big), \qquad \hat{m}(x) = \frac{\sum_{i=1}^{n} y_i K(\frac{x-x_i}{\hat{h}})}{\sum_{j=1}^{n} K(\frac{x-x_j}{\hat{h}})}$$

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^{n} (y_i - \hat{m}(x_i))^2 K(\frac{x-x_i}{\hat{h}})}{\sum_{j=1}^{n} K(\frac{x-x_j}{\hat{h}})}, \qquad \hat{h} = 1.06 s_{x_i} n^{-1/5}$$

上の $\hat{h}$ は 1 つの例，s_{x_i} は i 番目の x 変数の標準偏差

さらに，$K(u)$ を次のように $N(0, \Omega)$ の多変数正規分布を仮定する ($\Omega = I_k$ でも構わない)。

$$K(u) = (2\pi)^{-k/2}|\Omega|^{-1/2} \exp\Bigl(-\frac{1}{2}u'\Omega^{-1}u\Bigr) \tag{50}$$

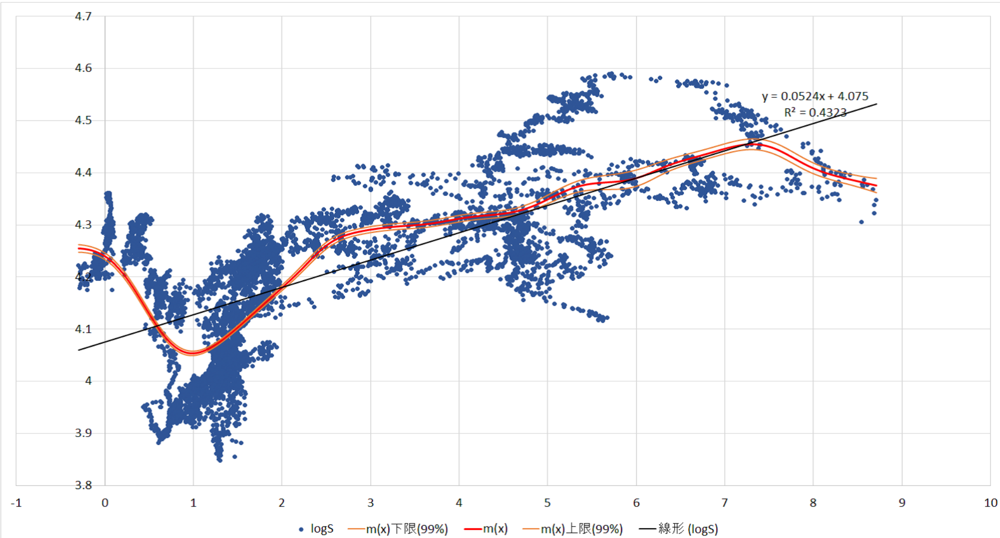
このとき，(48) 式の積分値は，

$$\begin{aligned}\int K^2(u)\mathrm{d}u &= \int (2\pi)^{-k}|\Omega|^{-1} \exp\bigl(-u'\Omega^{-1}u\bigr)\mathrm{d}u \\ &= (2\pi)^{-k/2}|2\Omega|^{-1/2} \int (2\pi)^{-k/2}\Bigl|\frac{1}{2}\Omega\Bigr|^{-1/2} \exp\Bigl(-\frac{1}{2}u'\Bigl(\frac{1}{2}\Omega\Bigr)^{-1}u\Bigr)\mathrm{d}u \\ &= 2^{-k}\pi^{-k/2}|\Omega|^{-1/2}\end{aligned}$$

として計算される。

日経平均株価指数の対数 vs 国債利回り：　推定期間：1986.2.1～2017.12.29,

データ数：$n = 7932$，$h = 1.06\hat{\sigma}n^{-1/5} \approx 0.343565$，$\hat{\sigma}^2 = 1.95245$



横軸：国債利回り，縦軸：日経平均株価指数の対数

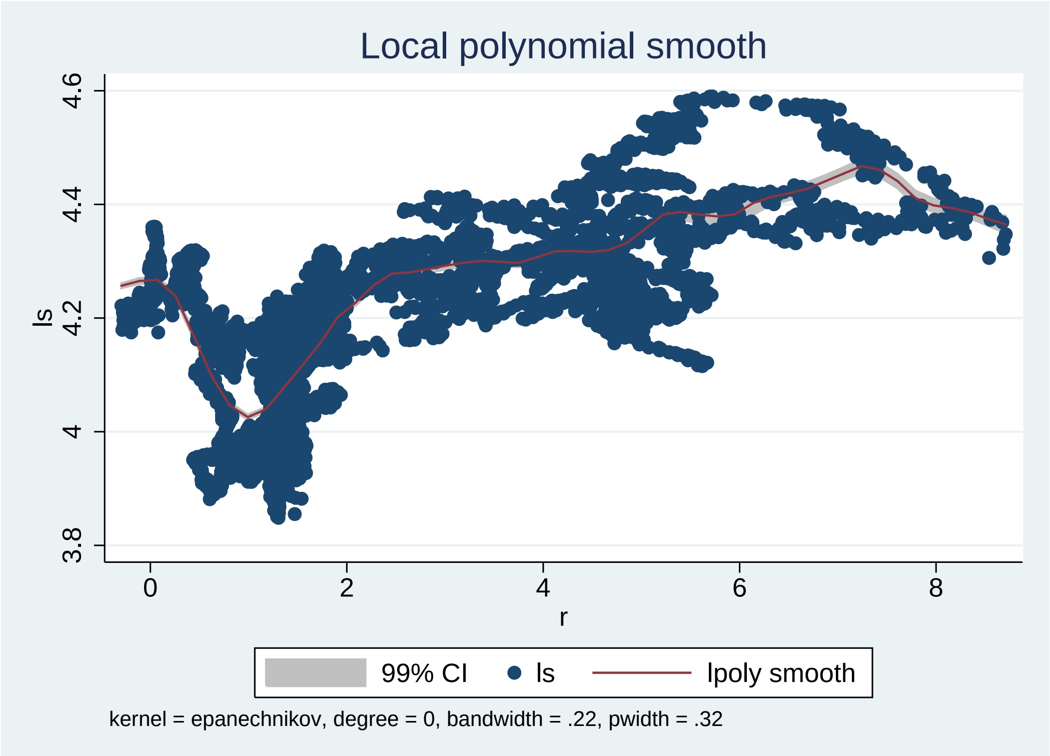
Stata による結果：

```
. lpoly ls r, ci l(99)
```

`ls` は日経平均株価椎の対数，`r` は国債利回り

`ci` はオプションで，95 %信頼区間 (Confidence Interval)

`l(99)` は `ci` のオプションで，99 %信頼区間 (無ければ，デフォルトで 95 %信頼区間)



横軸：国債利回り，縦軸：日経平均株価指数の対数