統計検定 (社会科学)

田淵進

November 9, 2025

第 I 部
2022 年

1

略

2

2.1

2.2

一般に，価格および数量ラスパイレス指数はそれぞれ

$$P_L = \frac{\sum_i^n p_{1i}n_{0i}}{\sum_i^n p_{0i}n_{0i}} \times 100$$
$$Q_L = \frac{\sum_i^n p_{0i}n_{1i}}{\sum_i^n p_{0i}n_{0i}} \times 100$$

で与えられ，パーシェ指数はそれぞれ

$$P_P = \frac{\sum_i^n p_{1i}n_{1i}}{\sum_i^n p_{0i}n_{1i}} \times 100$$
$$Q_P = \frac{\sum_i^n p_{0i}n_{0i}}{\sum_i^n p_{0i}n_{1i}} \times 100$$

で与えられる．よって表 1 からの計算にっよって，ラスパイレス価格指数は

$$\begin{aligned} P_L &= \frac{10363 \times 43.12 + 2272 \times 118.17}{10363 \times 44.74 + 2272 \times 119.86} \times 100 \\ &= 0.9718 \times 100 = 97.18 \end{aligned}$$

となる．

2.3

前問の定義から

$$
\begin{aligned}
P_L &= \bar{x} \\
Q_L &= \bar{y}
\end{aligned}
$$

である.

2.4

いま,

$$
\begin{aligned}
s_{xy} &= \sum_{n=1}^{n} w_i(x_i - \bar{x})(y_i - \bar{y}) \\
&= \sum_{i=1}^{n} w_i x_i y_i - \bar{y}\sum_{i=1}^{n} w_i x_i - \bar{x}\sum_{i=1}^{n} w_i y_i + \bar{x}\bar{y}\sum_{i=1}^{n} w_i \\
&= \sum_{i=1}^{n} w_i x_i y_i - P_L Q_L
\end{aligned}
$$

とかける．ここで,

$$
\begin{aligned}
\sum_{i=1}^{n} w_i x_i y_i &= \frac{\sum_i^n p_{1i} n_{1i}}{\sum_i^n p_{0i} n_{0i}} \\
&= \frac{\sum_i^n p_{1i} n_{1i}}{\sum_i^n p_{0i} n_{1i}} \times \frac{\sum_i^n p_{0i} n_{1i}}{\sum_i^n p_{0i} n_{0i}} \\
&= P_P Q_L
\end{aligned}
$$

となるので,

$$
\begin{aligned}
s_{xy} &= P_P Q_L - P_L Q_L \\
&= Q_L(P_P - P_L)
\end{aligned}
$$

と示される.

2.5

P_L と Q_L の大小関係は s_{xy} の符号と等しい．一般的に，価格と数量の間には負の相関がある場合が多いため，P_P は P_L より大きくなる傾向がある.

3

3.1

$t=1,T$ のみのデータを使った場合，モデルの LSE は

$$Y_t = \tilde{\alpha} + \tilde{\beta} X_t + \epsilon_t \quad (t=1,T)$$

を満たすので，

$$\begin{aligned}\tilde{\beta} &= \frac{Y_T - Y_1}{X_T - X_1} \\ \tilde{\alpha} &= \frac{1}{2}(Y_1 + Y_T) - \tilde{\beta}\frac{1}{2}(X_1 + X_T) & &= \frac{X_1}{X_T - X_1}Y_1 - \frac{X_T}{X_T - X_1}Y_T\end{aligned}$$

となるので，

$$\begin{aligned}V[\tilde{\alpha}] &= \frac{X_1^2 + X_T^2}{(X_T - X_1)^2}\sigma^2 \\ V[\tilde{\beta}] &= \frac{2}{(X_T - X_1)^2}\sigma^2\end{aligned}$$

である．

3.2

いま，$m = (X_1 + X_T)/2$ とおいて

$$\sum_{t=1}^{T}(X - \bar{X})^2 = \frac{(X_T - X_1)^2}{2} + 2\big(\bar{X} - m\big)^2 + \sum_{t=2}^{T-1}(X_t - \bar{X})^2$$

と変形できるので，条件は

$$\begin{aligned}\bar{X} &= m \\ X_2 = \cdots = X_{T-1} &= m\end{aligned}$$

である．

3.3

前問の値を用いて，

$$\begin{aligned}V\left[\tilde{Y}_{T+1}\right] &= V[\tilde{\alpha} + \tilde{\beta}X_{T+1}] \\ &= V\left[\frac{X_{T+1} - X_T}{X_T - X_1}Y_1 + \frac{X_{T+1} - X_1}{X_T - X_1}Y_T\right] \\ &= \frac{(X_{T+1} - X_T)^2 + (X_1 - X_{T+1})^2}{(X_T - X_1)^2}\sigma^2\end{aligned}$$

3.4

モデルが正しいと仮定したとき，3 点で測る際に最も $V[\tilde{\beta}]$ を小さくするには，X_c を X_1 または X_T と同じ値にするのが合理的である．

一方，単回帰であるかどうかを検証するためには，直線からのズレが大きくなる可能性が大きい $X_c = (X_1 + X_T)/2$ が適当である．

4

4.1

4.2

t 期と $t-4$ 期の関係を見て，A, B, C がそれぞれ $a = 0.8, 0, -0.8$ である事がわかる．

4.3

AR モデルを MA モデルとして表現すると，

$$\begin{aligned} y_t &= ay_{t-4} + v_t \\ &= a(ay_{t-8} + v_{t-4}) + v_t \\ &= v_t + av_{t-4} + a^2 vt - 8 + a^3 v_{t-12} + \cdots \end{aligned}$$

となる．あるいは，恒等作用素を I，ラグ作用素を L とおくと，$|a| < 1$ のもとで $I - L^4$ は可逆であり，

$$\begin{aligned} y_t &= \left(I - aL^4\right)^{-1} v_t \\ &= (I + aL^4 + a^2L^8 +)v_t \end{aligned}$$

として，同じ結果を得る．

また，条件より yt は定常であるので，$\mu \coloneqq E[y_t]$ とおくと，$\mu = a\mu$ すなわち $\mu = 0$ であり，$\psi \coloneqq V[y_t]$ とすると，

$$\begin{aligned} \psi =& a^2\psi + \sigma^2 \\ \psi =& \frac{\sigma^2}{1-a^2} \end{aligned}$$

となる．また，共分散は

$$\begin{aligned} \mathrm{Cov}(y_t, y_{t-k}) &= E\left[\left(\sum_{i=0}^{\infty} a^i v_{t-4i}\right)\left(\sum_{i=0}^{\infty} a^i v_{t-k-4i}\right)\right] \\ &= \begin{cases} \sum_{i=1}^{\infty} a^{2i+k/4}\sigma^2 & (k \in 4\mathbb{Z}) \\ 0 & (\text{otherwise}) \end{cases} \\ &= \begin{cases} \frac{a^{k/4}\sigma^2}{1-a^2} & (k \in 4\mathbb{Z}) \\ 0 & (\text{otherwise}) \end{cases} \end{aligned}$$

であるので，自己相関係数は，

$$\begin{aligned} \rho_k &= \mathrm{Cov}(y_i, y_{i-k})/\psi \\ &= \begin{cases} a^{k/4} & (k \in 4\mathbb{Z}) \\ 0 & (\text{otherwise}) \end{cases} \end{aligned}$$

4.4

自己回帰モデルにおける SLE は，$t=5,\dots,T$ までの $T-4$ 個の関係式

$$y_t = a y_{t-4} + v_t$$

から，平方和 $\sum_{t=5}^{T}(y_t - a y_{t-4})^2$ を最小化する値として求められる．これは通常の単回帰と同じで，SLE として

$$\hat{\alpha} = \frac{\sum_{t=5}^{T} y_t y_{t-4}}{\sum_{t=5}^{T} y_{t-4}^2}$$

が求まる．このとき，残差の自由度は $T-5$ になるので，最小化された平方和から計算される分散 σ^2 の推定量としては,

$$\hat{\sigma^2} = \frac{1}{T-5}\sum_{t=4}^{T}(y_t - \hat{a} y_{t-4})^2$$

と取れる．このとき，分母にも確率項があり，一般に推定量が不偏とは限らない．

一方で，大数の法則より[*1],

$$\frac{1}{T-4}\sum_{t=5}^{T} v_t y_{t-4} \to 0 \frac{1}{T-4}\sum_{t=5}^{T} y_{t-4} y_{t-4} \qquad \to a$$

が成り立ち，連続写像定理から $\hat{\alpha}$ は a に確率収束する．

同様に，$e_t \coloneqq y_t - \hat{a} y_{t-4} = v_t + (a-\hat{a}) y_{t-4}$ として大数の法則を用いることで，一致推定量であることが示される．

4.5

$k = 4(n-1)+i, i \in \{1,2,3,4\}$ とおくと，

$$\hat{y}_{T+k} = \hat{a}^n y_{T-4+i}$$

と推定できる．このときの予測誤差は

$$y_{T+k} - \hat{y}_{T+k} = (a^n - \hat{a}^n) y_{T-4+i} + \sum_{j=0}^{n} a^j v_{T+k-4j}$$

となる．

[*1] 大数の法則がちゃんと使えることを示すのはむずい気がする

5

5.1

分布の再生成より，$D \sim N(0, 2\sigma^2)$ となるので，

$$\begin{aligned} E[D] &= 0 \\ V[D] &= 2\sigma_X^2 = 72 \\ P(D \leq -4) &= P(Z \leq -4/\sqrt{72}) && \approx 0.32 \end{aligned}$$

となる．ここで，Z は標準正規分布に従う確率変数とする．

5.2

条件より，$Y \mid X \sim N(\alpha + \beta X, \sigma^2)$ となるので，$\beta = 3/4, \alpha = 120 - 120\beta$ を用いて

$$\begin{aligned} E\left[Y \mid X = 132\right] &= \alpha + 132\beta \\ &= 129 \end{aligned}$$

となり，$V[Y \mid X = 132] = \sigma^2 = \sigma_Y^2 - \sigma_{XY}^2/\sigma_X^2 = 63$ である．

ここで，平均への回帰効果とは，母平均が同じであるとすれば，一回目で 132 を記録しても，2 回目の測定値の期待値は 132 よりも小さい値になることを指しており，平均への回帰分は 3mmHg とみなされる．

5.3

Tower Property などから，

$$\begin{aligned} E\big[(X-\mu)^2\big] &= E\big[E\left[(X-\mu)^2 \;\middle|\theta\right]\big] \\ &= E\big[E\left[(X-\theta)^2 + (\theta-\mu)^2 + 2(X-\mu)(X-\theta) \;\middle|\theta\right]\big] \\ &= \psi^2 + \tau^2 \end{aligned}$$

と計算できる．同様に $V[Y] = \psi^2 + \tau^2$ であり，$\mathrm{Cov}(X, Y) = \tau^2$ と計算できる．これに $\psi^2 + \tau^2 = 144, \tau^2 = 3(\psi^2 + \tau^2)/4$ を代入して，

$$\begin{aligned} \tau^2 &= 108 \\ \psi^2 &= 36 \end{aligned}$$

を得る．

5.4

条件から，θ と ε はそれぞれ独立に正規分布に従うので，

$$\begin{pmatrix}\theta\\X\end{pmatrix}=\begin{pmatrix}1&0\\1&1\end{pmatrix}\begin{pmatrix}\theta\\\varepsilon_1\end{pmatrix}\sim N\left(\begin{pmatrix}\mu\\\mu\end{pmatrix},\begin{pmatrix}1&1\\0&1\end{pmatrix}\begin{pmatrix}\tau^2&0\\0&\psi^2\end{pmatrix}\begin{pmatrix}1&0\\1&1\end{pmatrix}\right)$$

となる．特に共分散行列は

$$\begin{pmatrix}1&1\\0&1\end{pmatrix}\begin{pmatrix}\tau^2&0\\0&\psi^2\end{pmatrix}\begin{pmatrix}1&0\\1&1\end{pmatrix}=\begin{pmatrix}\tau^2+\psi^2&\psi^2\\\psi^2&\psi^2\end{pmatrix}$$

である．このとき，$\theta\mid X$ の分布は，$X=132$ とすると

$$\begin{aligned}E\left[\theta\mid X\right]&=\mu+\frac{\mathrm{Cov}(\theta,X)}{V[X]}(X-\mu)\\&=129\\V\left[\theta\mid X\right]&=V[\theta]-\frac{\mathrm{Cov}(\theta,X)^2}{V[X]}\\&=27\end{aligned}$$

に従う正規分布となる．

5.5

$\varepsilon_1,\varepsilon_2$ は互いに独立であるので，$\theta+\varepsilon_1\mid X$ の分布は $N(129,63)$ となる．よって，

$$P(Y\leq 128\mid X=132)=P(Z\leq -1/\sqrt{63})\approx 0.45$$

となる．ここで Z は標準正規分布に従う確率変数とする．

第 II 部

2023 年

6

6.1

いま、与えられた条件から

$$\begin{aligned}\bar{x} &\approx 3.7327\\ \bar{y} &\approx 27.49\\ \bar{x^2} &\approx 16.15\\ \bar{y^2} &\approx 855.27\\ \bar{xy} &\approx 108.95\end{aligned}$$

よって、$n = 33$ であることから、不偏分散は

$$\begin{aligned}s_x^2 &= \frac{33}{32}\left(\bar{x^2} - \bar{x}^2\right) \approx 2.33\\ s_y^2 &= \frac{33}{32}\left(\bar{y^2} - \bar{y}^2\right) \approx \frac{33}{32}\left(855.27 - 27.49^2\right) \approx 102.68\\ s_{xy} &= \frac{33}{32}(\bar{xy} - \bar{x}\bar{y}) \approx \frac{33}{32}(108.95 - 3.7327 \times 27.49) \approx 6.54\end{aligned}$$

また、相関係数は

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}} \approx \frac{6.54}{\sqrt{2.33 \times 102.68}} \approx 0.42$$

6.2

デルタ法とは、観測値の関数 $g(t_1, \ldots, t_n)$ で表される統計量の分散を、g の各変数の平均値周りでの 1 次近似によって求める手法である。いま、

$$\begin{aligned}\hat{R} = g(\bar{x}, \bar{y}) &\coloneqq \frac{\bar{y}}{\bar{x}}\\ &\approx g(\mu_X, \mu_X) + \frac{\partial g(\mu_X, \mu_X)}{\partial \mu_X}(\bar{x} - \mu_X) + \frac{\partial g(\mu_X, \mu_Y)}{\partial \mu_Y}(\bar{y} - \mu_Y)\\ &= \frac{\mu_Y}{\mu_X} - \frac{\mu_Y}{\mu_X^2}(\bar{x} - \mu_X) + \frac{1}{\mu_X}(\bar{y} - \mu_Y)\end{aligned}$$

と近似されるので、求める分散は

$$\begin{aligned}V\left[\hat{R}\right] &= \frac{1}{\mu_X^2} V[\bar{y} - R\bar{x}]\\ &= \frac{1}{n\mu_X^2}\left(E\left[(Y - RX)^2\right] - 0\right)\end{aligned}$$

となる。

6.3

前問の結果を用いて

$$
\begin{aligned}
V[\hat{\mu_Y}] &= \mu_X^2 V[\hat{R}] \\
&= \frac{1}{n}\bigl(V[Y] + R^2 V[X] - 2R\,\mathrm{Cov}(X,Y)\bigr) \\
&= \frac{\mu_Y^2}{n}\left(\frac{\sigma_Y^2}{\mu_Y^2} + \frac{\sigma_X^2}{\mu_Y^2}\frac{\mu_Y^2}{\mu_X^2} - 2\frac{\mathrm{Cov}\,(X,Y)}{\mu_X\mu_Y}\right) \\
&= \frac{\mu_Y^2}{n}\bigl(\mathrm{cv}(Y)^2 + \mathrm{cv}(X)^2 - 2\rho\mathrm{cv}(X)\mathrm{cv}(Y)\bigr)
\end{aligned}
$$

と求まる。

6.4

$\hat{R} = \bar{y}/\mu_X$ としたときの分散は $V_1 \coloneqq \sigma_Y^2/\mu_X^2$ である。一方、$\bar{y}/\bar{x}$ とした場合の分散は、前問での計算から

$$
V\left[\hat{R}\right] \coloneqq V_2 = \frac{1}{n\mu_X^2}\bigl(V[Y] + R^2 V[X] - 2R\,\mathrm{Cov}(X,Y)\bigr)
$$

となるので、$V_1 > V_2$ となるのは

$$
R\sigma_X^2 < 2\frac{\sigma_{XY}}{\sigma_X\sigma_Y} \tag{1}
$$

$$
\frac{\mathrm{cv}(X)}{2\mathrm{cv}(Y)} < \rho \tag{2}
$$

となるときである。よって、観測値ゕら求まる $\mathrm{cv}(X), \mathrm{cv}(Y), \rho$ を参考に、どちらを使うか切り替えるのが望ましい。