Volume 8, No.2 November 2007

ISSN 1411 - 4453

Pakar

JURNAL TEKNOLOGI INFORMASI DAN BISNIS

Diterbitkan oleh :
Pusat Pengembangan Sains dan Teknologi
Fakultas Sains dan Teknologi
Universitas Teknologi Yogyakarta

Pakar
Vol. 8, No.2 Nopember 2007
Hal. : 125 – 134

ALGORITMA BOOTSTRAP DALAM INFERENSI MODEL EKONOMETRIK DAN APLIKASINYA UNTUK MEMPREDIKSIKAN BEBERAPA INDIKATOR EKONOMI

Oleh: Suparman

ABSTRACT

When fitting a econometric model to real data, the model parameter often unknown. Our aim is to find estimators of the parameter based on the data.

In this paper the parameter estimation for econometric model is posed within a Classic framework. Within this framework the distribution of the stochastic disturbance are assumed to be unknown. Obtaining the estimators of the parameter are difficult.

Here bootstrap algorithm is used to find estimators of the parameter. The method developed is evaluated in simulation studies on number of synthetic and real data sets.

Keyword: bootstrap algorithm, autoregressive, polynomial regression, multiple regression.

PENDAHULUAN

Algoritma bootstrap adalah metode berbasis komputer yang sangat potensial untuk dipergunakan pada masalah inferensi statistik. Salah satu contohnya adalah penggunaan algoritma bootstrap untuk menginferensi parameter model ekonometri. Dalam metode klasik para peneliti mendasarkan pada asumsi bahwa gangguan stokhastik dalam model ekonometrik dianggap berdistribusi normal. Penyisipan gangguan stokhastik ke dalam model ekonometrik disebabkan oleh karena ketidaksempurnaan spesifikasi bentuk matematis model.

Permasalahannya sekarang adalah bagaimana jika ternyata gangguan stokhastik tersebut tidak diketahui distribusinya. Secara aplikatif, banyak fakta menunjukkan bahwa dalam suatu penelitian kadang-kadang kita mengalami kesulitan menentukan bentuk distribusi dari gangguan stokhastik. Dalam hal inilah metode bootstrap digunakan sebagai alternatif.

Bootstrap sendiri berdasar dari istilah "*pull one self up by one's bootstrap*" yang dapat diartikan berusaha dengan sumber daya yang minimal (Elfron and Tibshirani, 1993). Dalam permasalahan statistik sumber daya yang minimal dapat diartikan sebagai data yang sedikit, data yang menyimpang dari asumsi tertentu bahkan data yang tidak memiliki asumsi apapun tentang distribusinya. Tujuan utama penggunaan bootstrap adalah untuk memperoleh estimasi yang sebaik-baiknya berdasar data yang minimal dengan bantuan komputer. Prinsip dasar bootstrap adalah resampling yaitu pengambilan sampel ulang/buatan dari observasi x_1, x_2, ..., x_n yang telah ada. Hal ini sangat jelas membantu peneliti jika dalam suatu penelitian, sampel yang diperoleh sangat minim dan terdesak oleh kondisi dimana tidak memungkinkan untuk menambah atau memperbanyak sampel penelitian.

Bagian II mengulas secara singkat pendekatan bootstrap. Bagian III mendiskusikan algoritma bootstrap untuk menginferensi parameter dalam model ekonometri. Bagian IV menyajikan implementasi baik pada data sintesis maupun data riil. Bagian V memberikan kesimpulan penelitian.

BOOTSTRAP

Misalkan $\hat{F}$ adalah distribusi empirik yang diambil dengan probabilitas 1/n pada setiap nilai yang diamati x_1, x_2, ... ,x_n. Sampel bootstrap didefinisikan sebagai sampel random berukuran n disusun dari $\hat{F}$, misal sampel bootstrap ke-b (b = 1, 2, ..., B) dinotasikan dengan $x_1^b, x_2^b, \Lambda, x_n^b$. Sampel bootstrap $x_1^b, x_2^b, \Lambda, x_n^b$ adalah sampel random berukuran n yang diambil dengan pengembalian dari populasi x_1, x_2, ... ,x_n. Anggota data bootstrap $x_1^b, x_2^b, \Lambda, x_n^b$ beranggotakan sampel asli x_1, x_2, ... ,x_n, yang muncul sekali, dua kali atau lebih bahkan tidak muncul dalam proses pengembalian ulang sampel asli tersebut.

Perbandingan antara kondisi sebenarnya dan kondisi bootstrap dapat digambarkan sebagai berikut :

Kondisi Sebenarnya	Kondisi Bootstrap
→ Sampel asli $x_1, x_2, \ldots, x_n \sim F$ adalah sampel random berukuran n dari distribusi F yang tidak diketahui → $\theta = \theta(F)$ adalah nilai riil dari suatu parameter yang menjadi perhatian. → Jika T adalah statistik untuk θ maka $\hat{\theta} = T(x_1, x_2, \Lambda, x_n)$.	→ Sampel bootstrap $x_1^b, x_2^b, \Lambda, x_n^b \sim \hat{F}$ adalah sampel buatan berukuran n dari distribusi $\hat{F}$. → $\theta = \theta(\hat{F})$ → $\hat{\theta} = T(x_1^b, x_2^b, \Lambda, x_n^b)$

MODEL EKONOMETRIK

Dalam bagian ini diuraikan inferensi parameter untuk tiga model ekonometrik yaitu : model autoregresif, model regresi ganda dan model regresi polinomial.

Autoregresif

Misalkan $x_1, x_2, \ldots, x_n$ adalah suatu runtun waktu berharga riil. Runtun waktu tersebut dikatakan memiliki model AR dengan orde p, dinotasikan sebagai AR(p), jika memenuhi persamaan stokhastik berikut (Brockwell and Davis, 1991; Box *et al.*, 1994):

$$x_t = \sum_{i=1}^{p} \beta_i x_{t-i} + z_t \qquad t = 1, 2, \Lambda, n \qquad (1)$$

di mana p adalah orde diketahui,

$$\beta = (\beta_1, \beta_2, \Lambda, \beta_p)'$$

adalah vektor koefisien dan z_t adalah suatu barisan gangguan stokhastik dengan mean nol dan variansi σ^2. Data indeks harga saham gabungan (IHSG), data indeks harga konsumen, dan data laju inflasi merupakan beberapa contoh data riil yang dapat dimodelkan oleh model AR.

Selanjutnya model AR disebut stasioner jika dan hanya jika persamaan suku banyak berikut

$$\phi(u) = 1 - \sum_{i=1}^{p} \beta_i u^i$$

bernilai nol untuk nilai u di luar lingkaran dengan jari-jari sama dengan satu (Brockwell and Davis, 1991).

Berdasarkan data x_t (t = 1, 2, …, n), selanjutnya kita akan berusaha untuk menaksir harga β dan σ^2. Penaksir kuandrat terkecil untuk β dan σ^2 diperoleh dengan meminimalkan jumlah kuadrat kesalahan. Penaksir kuadrat terkecil untuk β adalah

$$\hat{\beta} = (X'X)^{-1} X'Y$$

di mana $Y = (x_1, x_2, \Lambda, x_n)^t$ dan

$$X = \begin{pmatrix} x_0 & x_{-1} & x_{-2} & \Lambda & x_{-p} \\ x_1 & x_0 & x_{-1} & & x_{1-p} \\ x_2 & x_1 & x_0 & & x_{2-p} \\ M & & & O & \\ x_n & x_{n-1} & x_{n-2} & & x_{n-p} \end{pmatrix}$$

Sedangkan penaksir kuadrat terkecil untuk σ^2 adalah

$$\hat{\sigma}^2 = \frac{Y'Y - \hat{\beta}'X'Y}{n-p}$$

Apabila taksiran tersebut disubtitusikan ke dalam model, maka modelnya dapat digunakan untuk memprediksi nilai $\hat{x}_{t+1}$. Langkah-langkah komputasi untuk menentukan interval kepercayaan $100(1-\alpha)\%$ untuk meramalkan nilai $\hat{x}_{t+1}$ adalah sebagai berikut :

- Hitung nilai $\hat{\beta}$ dan $\hat{\sigma}^2$ dari data asli.
- Hitung $\hat{z}_t$ dengan menggunakan persamaan $z_t = x_t - \sum_{i=1}^{p} \beta_i x_{t-i}$
- Untuk b = 1, 2, …, B .
  - Resampling $\hat{z}_t^{(b)}$.
  - Hitung $x_t^{(b)}$ dengan menggunakan persamaan $x_t = \sum_{i=1}^{p} \hat{\beta}_i x_{t-i} + \hat{z}_t^b$.
  - Hitung $\hat{\beta}^{(b)}$, $\hat{\sigma}^{2(b)}$ dan $\hat{x}_{t+1}^{(b)}$.
- Hitung $\hat{\beta}_{boot}$, $\hat{\sigma}^2_{boot}$ dan $\hat{x}_{(t+1)(boot)}$.
- Hitung interval kepercayaan $100(1-\alpha)\%$ untuk $\hat{x}_{t+1}$.

Regresi Polinomial

Misalkan y_t adalah variabel tak bebas, x_t variabel yang menjelaskan, z_t adalah gangguan yang stokhastik, dan t menyatakan pengamatan yang ke-t, maka model regresi polinomial orde m bisa ditulis sebagai (Gujarati, 1978)

$$y_t = a_0 + a_1 x_t + a_2 x_t^2 + \Lambda + a_m x_t^m + z_t \qquad (2)$$

untuk t = 1, 2, 3, …, n. Di mana m adalah orde diketahui, $a = (a_0, a_1, \Lambda, a_m)^t$ adalah vektor koefisien dan z_t adalah barisan gangguan stokhastik dengan mean 0 dan variansi σ^2. Kita notasikan.

Data laju inflasi vs data indeks harga konsumen dan data laju inflasi vs data kurs valuta asing merupakan beberapa contoh data riil yang dapat dimodelkan oleh model regresi polinomial.

Berdasarkan data x_t (t = 1, 2, …, n), selanjutnya kita akan berusaha untuk menaksir harga

a, dan σ^2.

Penaksir kuadrat terkecil untuk $a = (a_0, a_1, \Lambda, a_m)$ adalah

$$\hat{a} = (X'X)^{-1} X'Y$$

di mana

$Y = (y_1, y_2, \Lambda, y_n)^t$ dan

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \Lambda & x_1^k \\ 1 & x_2 & x_2^2 & & x_2^k \\ 1 & x_3 & x_3^2 & & x_3^k \\ M & & & O & \\ 1 & x_n & x_n^2 & & x_n^k \end{pmatrix}$$

Penaksir LSE untuk σ^2 adalah

$$\hat{\sigma}^2 = \frac{Y'Y - \hat{a}'X'Y}{n - m - 1}$$

Langkah-langkah komputasi untu menentukan interval kepercayaan $100(1-\alpha)\%$ untuk $\hat{y}_{t+1}$ adalah sebagai berikut :

- Hitung $\hat{a}$ dan $\hat{\sigma}^2$ dari data asli.
- Hitung $\hat{z}_t$ dengan menggunakan persamaan $z_t = y_t - a_0 - \sum_{i=1}^{m} \hat{a}_i x_t^i$
- Untuk b = 1, 2, …., B :
  - Resampling $\hat{z}_t^{(b)}$.
  - Hitunglah $y_t^{(b)}$ dengan menggunakan persamaan $y_t^{(b)} = a_0 + \sum_{i=1}^{m} \hat{a}_i x_t^i + \hat{z}_t^{(b)}$.
  - Hitunglah $\hat{\beta}^{(b)}$, $\hat{\sigma}^{2(b)}$ dan $\hat{y}_{t+1}^{(b)}$.
- Hitunglah $\hat{\beta}_{boot}$, $\hat{\sigma}^2_{boot}$ dan $\hat{y}_{(t+1)(boot)}$
- Hitunglah interval kepercayaan $100(1-\alpha)\%$ untuk $\hat{y}_{t+1}$

Regresi Ganda

Misalkan y_t adalah variabel tak bebas, x_{t1}, x_{t2}, ..., x_{tk} adalah variabel yang menjelaskan, z_t adalah gangguan yang stokhastik, dan t menyatakan pengamatan yang ke-t, maka model regresi ganda bisa ditulis sebagai (Johnston, 1972; Gujarati, 1978)

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \Lambda + \beta_k x_{kt} + z_t \qquad 3)$$

untuk t = 1, 2, 3, …, n. Di mana $b = (b_1, b_2, \Lambda, b_k)'$ adalah vektor koefisien dan z_t adalah barisan gangguan stokhastik dengan mean 0 dan variansi σ^2. Data laju inflasi vs data kurs valuta asing ($ dolar, Euro, Yen) merupakan beberapa contoh data riil yang dapat dimodelkan oleh model regresi ganda.

Berdasarkan data x_{2t}, x_{3t}, x_{4t}, …, x_{kt} (t = 1, 2, …, n), akan ditaksir harga b dan σ^2. Penaksir kuandrat terkecil untuk b adalah

$$\hat{b} = (X^t X)^{-1} X^t Y$$

di mana $Y = (y_1, y_2, \Lambda, y_n)^t$ dan

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \Lambda & x_{k1} \\ 1 & x_{12} & x_{22} & & x_{k2} \\ 1 & x_{13} & x_{23} & & x_{k3} \\ M & & & O & \\ 1 & x_{1n} & x_{2n} & & x_{kn} \end{pmatrix}$$

Sedangkan penaksir kuadrat terkecil untuk σ^2 adalah

$$\hat{\sigma}^2 = \frac{Y'Y - \hat{b}'X'Y}{n-k-1}$$

Langkah-langkah komputasi untuk menentukan interval kepercayaan $100(1-\alpha)\%$ untuk $\hat{y}_{t+1}$ adalah sebagai berikut :

- Hitunglah $\hat{b}$ dan $\hat{\sigma}^2$ dari data asli.
- Hitunglah $\hat{z}_t$ dengan menggunakan persamaan $\hat{z}_t = \hat{y}_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1t} - \hat{\beta}_2 x_{2t} - \Lambda + \hat{\beta}_k x_{kt}$.
- Untuk b = 1, 2, …, B :
  - Resampling $\hat{z}_t^{(b)}$.
  - Hitunglah $y_t^{(b)}$ dengan persamaan $y_t^{(b)} = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \Lambda + \beta_k x_{kt} + \hat{z}_t^{(b)}$.
  - Hitunglah $\beta^{(b)}$, $\hat{\sigma}^{2(b)}$ dan $\hat{y}_{t+1}^{(b)}$.
- Hitunglah $\hat{\beta}_{boot}$, $\hat{\sigma}^2_{boot}$ dan $\hat{y}_{(t+1)(boot)}$.
- Hitunglah interval kepercayaan $100(1-\alpha)\%$ untuk $\hat{y}_{t+1}$

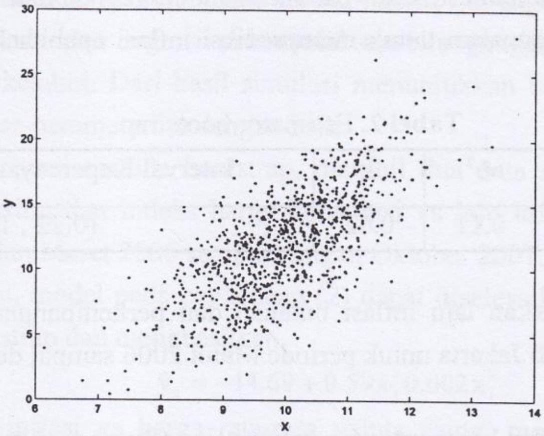
IMPLEMENTASI

Sebagai ilustrasi, kita akan menerapkan algoritma bootstrap untuk menginferensi parameter data AR sintesis dan data riil. Studi simulasi (Law and Kelton, 2000) ditempuh untuk mengkonfirmasi kinerja dari algoritma bootstrap apakah dapat bekerja dengan baik. Sedangkan studi kasus diberikan untuk memberikan contoh penerapan penelitian dalam memecahkan permasalahan dalam kehidupan sehari-hari.

Di sini resampling dilakukan sebanyak B = 2000 dan probalilitas kesalahan jenis 1 $\alpha = 0{,}05$.

Data Sintesis

Gambar 1 menunjukkan 1000 data sintesis regresi polinoial orde 2.. Nilai x ditentukan sedangkan nilai y dibuat dengan menggunakan persamaan (3) di atas. Nilai koefisien regresi dan variansi gangguan stokhastik adalah $a_0 = 1.57$, $a_1 = -0.85$, $a_2 = 0.19$ dan $\sigma^2 = 9$.



Gambar 1. Data sintesis

Berdasarkan data sintesis tersebut, selanjutnya mengestimasi koefisien regresi polinomial dan variansi σ^2 dengan menggunakan algoritma bootstrap. Hasilnya adalah $\hat{a}_0 = 1.26$, $\hat{a}_1 = -0.80$, $\hat{a}_2 = 0.19$ dan $\hat{\sigma}^2 = 8.76$. Apabila nilai parameter dan nilai estimatornya baik untuk koefisen regresi dan variansi terlihat bahwa algoritma bootstrap dapat bekerja dengan "baik" dalam mengestimasi parameter berdasarkan data sintesis.

Data Riil

Tabel 1 menunjukkan laju inflasi bulanan (y) dan indeks harga konsumen (x) di Indonesia untuk periode Maret 2006 sampai dengan Oktober 2007. Inflasi adalah indikator yang memberikan informasi tentang dinamika perkembangan harga barang dan jasa yang dikonsumsi masyarakat. Sedangkan indeks harga konsumen (IHK) adalah angka/indeks yang menunjukkan perbandingan relative antara tingkat harga (konsumsi/eceran) pada saat bulan survey dan harga tersebut pada bulan sebelumnya.

Tabel 1. Laju inflasi dan indeks harga konsumen. (Sumber : http://www.bps.go.id)

Periode	Mar	Apr	Mei	Juni	Juli	Agus	Sept	Okt	Nop	Des
Inflasi	0.03	0.05	0.37	0.45	0.45	0.33	0.38	0.86	0.34	1.21
IHK	139.57	139.64	140.16	140.79	141.42	141.88	142.42	143.65	144.14	145.84
Periode	Jan	Peb	Mar	Apr	Mei	Juni	Juli	Agus	Sept	Okt
Inflasi	1.04	0.62	0.24	-0.16	0.10	0.23	0.72	0.75	0.80	0.79
IHK	147.41	148.32	148.67	148.43	148.58	148.92	149.99	151.11	152.32	153.53

Data pada Tabel 1 dicocokkan terhadap regresi polinomial orde 2. Algoritma bootstrap digunakan untuk mendapatkan estimator parameter model regresi dan variansi σ^2. Selanjutnya model yang diperoleh digunakan untuk memprediksi inflasi apabila IHK = 153.53. Hasilnya disajikan pada Tabel 2.

Tabel 2. Estimator bootstrap

$\hat{a}$	$\hat{\sigma}^2$	$\hat{y}$	Interval Kepercayaan 95% untuk $\hat{y}$
(-44.69, 0.59, -0.002)'	0.11	0.62	(0.22 , 1.00)

Tabel 3 menunjukkan laju inflasi bulanan dan perkembangan harga rata-rata valuta asing (US $, Euro, Yen) di Jakarta untuk periode Maret 2006 sampai dengan Oktober 2007.

Tabel 3. Laju inflasi dan harga rata-rata valuta asing (Sumber : http://www.bps.go.id)

Periode	Mar	Apr	Mei	Juni	Juli	Agus	Sept	Okt	Nop	Des
Inflasi	0.03	0.05	0.37	0.45	0.45	0.33	0.38	0.86	0.34	1.21
US $	9.12	8.83	9.21	9.35	9.12	9.27	9.22	9.31	9.17	9.20
Euro	11.02	10.95	11.83	11.75	11.47	11.62	11.66	12.64	12.07	12.27
Yen	77.50	76.50	81.50	80.50	77.50	86.00	78.50	89.50	79.50	83.00
Periode	Jan	Peb	Mar	Apr	Mei	Juni	Juli	Agus	Sept	Okt
Inflasi	1.04	0.62	0.24	-0.16	0.10	0.23	0.72	0.75	0.80	0.79
US $	9.09	9.15	9.13	9.12	9.21	9.10	9.09	9.35	9.14	9.10
Euro	11.76	12.07	12.19	12.25	11.83	12.23	12.35	11.98	12.89	13.13
Yen	74.50	76.50	77.50	76.50	82.00	73.50	75.50	85.50	79.00	79.00

Data pada Tabel 3 dicocokkan terhadap regresi ganda. Algoritma bootstrap digunakan untuk mendapatkan estimator parameter model regresi dan variansi σ^2. Selanjutnya model yang diperoleh digunakan untuk memprediksi nilai inflasi ($\hat{y}$) apabila US \$ =9.10, Euro = 13.13 dan Yen = 74.50. Hasilnya disajikan pada Tabel 4.

Tabel 4. Estimator bootstrap

$\hat{a}$	$\hat{\sigma}^2$	$\hat{y}$	Interval Kepercayaan 95% untuk $\hat{y}$
(-5.71, 0.15, 0.31, 0.01)'	0.09	0.81	(0.52 , 1.10)

SIMPULAN

Uraian di atas menunjukkan bahwa betapa sederhananya algoritma bootstrap dapat digunakan untuk menghasilkan taksiran parameter/koefisien hubungan estimator dalam model autoregresif, regresi ganda dan regresi polinomial apabila gangguan stokhastiknya adalah distribusi yang tidak diketahui. Dari hasil simulasi menunjukkan bahwa algoritma bootstrap dapat menaksir parameter-parameter itu dengan baik.

Sebagai implementasi metode bootstrap, diambil dua data riil yaitu : laju inflasi vs harga rata-rata valuta asing dan indeks harga konsumsi vs laju inflasi. Kedua data tersebut merupakan data dari bulan Maret 2006 sampai dengan Oktober 2007. Untuk data laju inflasi vs harga rata-rata konsumsi, model pada persamaan (2) dapat diselesaikan dengan menggunakan prosedur algoritma bootstrap dan menghasilkan

$$\hat{y}_t = -44.69 + 0.59x_t\, 0.002x_t^2$$

Sedangkan untuk data inflasi vs harga rata-rata valuta asing, model persamaan (3) dapat diselesaikan dengan menggunakan prosedur algoritma bootstrap dan menghasilkan

$$\hat{y}_t = -5.71 + 0.15x_{1t} + 0.31x_{2t} + 0.01x_{3t}$$

Taksiran ini sangat bermanfaat bagi pengambilan keputusan, misalnya untuk memprediksi indikator ekonomi (laju inflasi, indeks harga konsumsi) pada bulan November 2007 dan seterusnya.

Meskipun dalam artikel ini hanya dibahas tiga model ekonometri (autoregresif, regresi polinomial dan regresi ganda), tetapi algoritma bootstrap dapat diterapkan juga pada model-model ekonometrik yang lainnya. Dalam artikel ini, resampling dilakukan terhadap gangguan stokhastik. Pendalaman dan perluasan algoritma bootstrap dapat ditempuh dengan melakukan resampling terhadap pasangan variabel y dan variabel x (Mackinnon, 2006).

DAFTAR PUSTAKA

[1] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994) Time Series Analysis : Forecasting and Control, Prentice Hall, New Jersey.

[2] Brockwell, P.J. and Davis, R.A. (1991) Times Series : Theory and Methods, Springer, New York.

[3] Efron, B. and Tibshirani, R. (1993) An Introduction to the Bootstrap, Chapman & Hall, New York.

[4] Gujarati, D (1978) Ekonometrika Dasar, Erlangga, Jakarta.

[5] http://www.bps.go.id/

[6] Johnston, J. (1972) Econometric Methods, McGraw-Hill, New York

[7] Law, A.M. and Kelton, W.D. (2000) Simulation Modeling and Analysis. McGraw-Hill, Singapore.

[8] Mackinnon, J.G. (2006) Bootstrap Methods in Econometrics, The economic record, Volume 82.