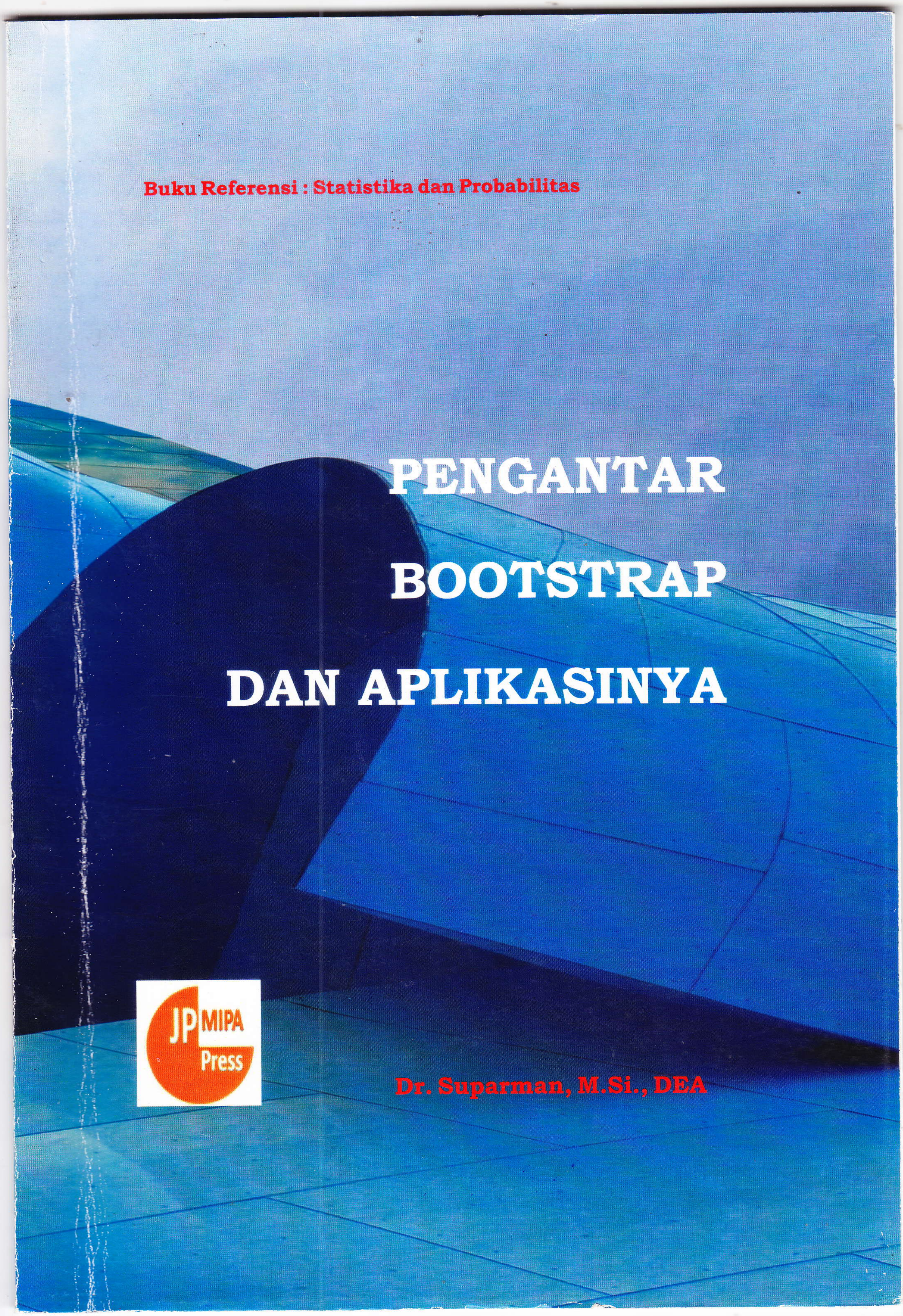
Buku Referensi : Statistika dan Probabilitas



PENGANTAR BOOTSTRAP DAN APLIKASINYA

JP MIPA Press

Dr. Suparman, M.Si., DEA

PENGANTAR
BOOTSTRAP DAN APLIKASINYA

Dr. Suparman, M.Si., DEA

JPMIPA FKIP UAD Press
Yogyakarta

PENGANTAR BOOTSTRAP DAN APLIKASINYA

Oleh : Dr. Suparman, M.Si., DEA

Penerbit :
JPMIPA FKIP UAD Press
Jl. Prof. Dr. Soepomo, SH, Warungboto
Yogyakarta

Edisi pertama
Cetakan pertama, 2012

Editor : Sugiyarto, M.Si., Ph.D
Desain Cover : Magistera Laningratum
Setting : Ayudea Az Zahra Zulfa

Dr. Suparman, M.Si., DEA
Pengantar Bootstrap dan Aplikasinya, ________ Yogyakarta : JPMIPA FKIP UAD Press, 2012
vi+60 hlm; 18.5 x 26.5 cm
ISBN : 978-602-18282-3-6
Statistika : Buku Referensi

KATA PENGANTAR

Buku ini disusun berdasarkan penelitian dan pengajaran yang penulis lakukan selama lima tahun. Di samping itu, penulis juga telah mengkaji berbagai literatur dan hasil penelitian. Buku ini ditulis sebagai buku referensi untuk para peneliti, para pengajar di Perguruan Tinggi dan para mahasiswa dalam memahami Metode Bootstrap. Dalam buku ini disertakan beberapa aplikasinya sehingga membuat buku ini cocok juga untuk para praktisi yang ingin memahami Metode Bootstrap dan permasalahan yang bisa diselesaikan dengan metode ini.

Dalam buku ini dibahas berbagai hal, yaitu : 1) Metode Bootstrap dalam Inferensi Model Ekonometrik, 2) Metode Bootstrap dalam Inferensi Model Subset Lag yang Didistribusikan 3) Metode Bootstrap dalam Model Regresi Ganda, 4) Model Bootstrap dalam Model Bootstrap dalam Inferensi Model Regresi Polinomial dan 5) Metode Bootstrap dalam Uji Hipotesis Mengenai Dua Mean Populasi.

Karena saya tidak mungkin menyelesaikan buku ini sendirian, daya ingin mengucapkan banyak terima kasih pada berbagai pihak yang telah mendukung kelancaran penulisan buku ini. Akhirnya penulis tetap mengharapkan berbagai masukan, kritik dan saran demi perbaikan karya di masa yang akan datang.

Yogyakarta, September 2012

Penulis

DAFTAR ISI

KATA PENGANTAR		**iii**
DAFTAR ISI		**v**
BAB 1	**METODE BOOTSTRAP DALAM INFERENSI MODEL EKONOMETRIK**	**1**
	1.1 Rumusan masalah	1
	1.2 Bootstrap	2
	1.3 Model Ekonometrik	2
	1.4 Hasil dan Pembahasan	6
	1.5 Kesimpulan	9
BAB 2	**METODE BOOTSTRAP DALAM INFERENSI MODEL SUBSET LAG YANG DIDISTRIBUSIKAN**	**11**
	2.1 Rumusan Masalah	11
	2.2 Metode	12
	2.3 Hasil Dan Pembahasan	15
	2.4 Kesimpulan	20
BAB 3	**METODE BOOTSTRAP DALAM INFERENSI MODEL REGRESI GANDA**	**23**
	3.1 Rumusan Masalah	23
	3.2 Metode	24
	3.3 Hasil Dan Pembahasan	28
	3.4 Kesimpulan	31
BAB 4	**METODE BOOTSTRAP DALAM INFERENSI MODEL REGRESI POLINOMIAL**	**33**
	4.1 Rumusan Masalah	33
	4.2 Metode Penelitian	34
	4.3 Hasil Dan Pembahasan	38
	4.4 Kesimpulan	43
BAB 5	**METODE BOOTSTRAP DALAM UJI HIPOTESIS MENGENAI DUA MEAN POPULASI**	**47**

5.1 Rumusan Masalah 47
5.2 Metode 48
5.3 Hasil Dan Pembahasan 52
5.4 Kesimpulan 56

DAFTAR PUSTAKA **57**

BAB 1
METODE BOOTSTRAP DALAM INFERENSI MODEL EKONOMETRIK

1.1 RUMUSAN MASALAH

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.

1.2 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 $x_1^b, x_2^b, \cdots, x_n^b$. Sampel bootstrap $x_1^b, x_2^b, \cdots, x_n^b$ adalah sampel random berukuran n yang diambil dengan pengembalian dari populasi x_1, x_2, ... ,x_n. Anggota data bootstrap $x_1^b, x_2^b, \cdots, x_n^b$ 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, ... ,x_n ~ F adalah sampel random berukuran n dari distribusi F yang tidak diketahui	✈ Sampel bootstrap $x_1^b, x_2^b, \cdots, x_n^b$ ~ $\hat{F}$ adalah sampel buatan berukuran n dari distribusi $\hat{F}$.
✈ $\theta = \theta(F)$ adalah nilai riil dari suatu parameter yang menjadi perhatian.	✈ $\theta = \theta(\hat{F})$
✈ Jika T adalah statistik untuk θ maka $\hat{\theta} = T(x_1, x_2, \cdots, x_n)$.	✈ $\hat{\theta} = T(x_1^b, x_2^b, \cdots, x_n^b)$.

1.3 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, ...,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):

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

di mana p adalah orde diketahui,

$$\beta = (\beta_1, \beta_2, \cdots, \beta_p)^t$$

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 $Y = (x_1, x_2, \cdots, x_n)^t$ dan

$$X = \begin{pmatrix} x_0 & x_{-1} & x_{-2} & \cdots & x_{-p} \\ x_1 & x_0 & x_{-1} & & x_{1-p} \\ x_2 & x_1 & x_0 & & x_{2-p} \\ \vdots & & & \ddots & \\ x_n & x_{n-1} & x_{n-2} & & x_{n-p} \end{pmatrix}$$

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 $z_t = x_t - \sum_{i=1}^{p} \hat{\beta}_i x_{t-i}$.
- ✈ 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 $\hat{\beta}_{boot}$, $\hat{\sigma}_{boot}^2$ 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)

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

untuk t = 1, 2, 3, ..., n. Di mana m adalah orde diketahui, $a = (a_0, a_1, \cdots, a_m)^t$ 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 $a = (a_0, a_1, \cdots, a_m)^t$ adalah

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

di mana

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

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & & x_2^k \\ 1 & x_3 & x_3^2 & & x_3^k \\ \vdots & & & \ddots & \\ 1 & x_n & x_n^2 & & x_n^k \end{pmatrix}$$

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 $\hat{\beta}_{boot}$, $\hat{\sigma}_{boot}^2$ 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)

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

untuk t = 1, 2, 3, ..., n. Di mana $b = (b_1, b_2, \cdots, b_k)'$ 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

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

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

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

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 $\hat{z}_t = \hat{y}_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1t} - \hat{\beta}_2 x_{2t} - \cdots + \hat{\beta}_k x_{kt}$.
- Untuk b = 1, 2, ..., B :
  - Resampling $\hat{z}_t^{(b)}$.
  - Hitunglah $y_t^{(b)}$ dengan persamaan $y_t^{(b)} = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \hat{z}_t^{(b)}$.
  - 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 $\hat{y}_{t+1}$

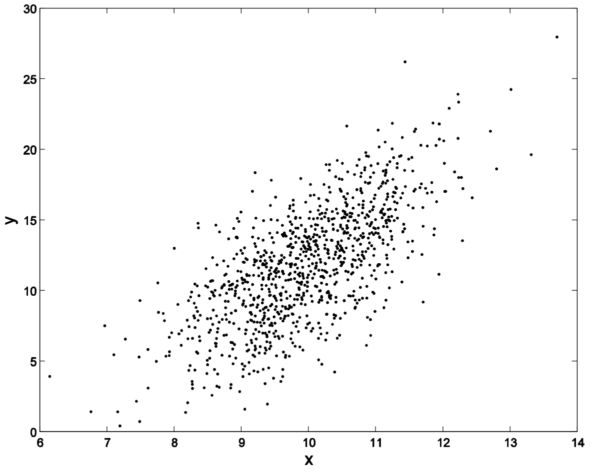
1.4 HASIL DAN PEMBAHASAN

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 1.1 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 1.1 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 1.1 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.

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

Tabel 1.1 Laju inflasi dan indeks harga konsumen. (Sumber : http://www.bps.go.id)

Data pada Tabel 1.1 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 1.2.

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

Tabel 1.2. Estimator bootstrap

Tabel 1.3 menunjukkan laju inflasi bulanan dan perkembangan harga rata-rata valuta asing (US $, Euro, Yen) di Jakarta untuk periode Maret 2006 sampai dengan Oktober 2007.

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

Tabel 1.3. Laju inflasi dan harga rata-rata valuta asing
(Sumber : http://www.bps.go.id)

Data pada Tabel 1.3 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 1.4.

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

Tabel 1.4. Estimator bootstrap

1.5 KESIMPULAN

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).

BAB 2
METODE BOOTSTRAP DALAM INFERENSI MODEL SUBSET LAG YANG DIDISTRIBUSIKAN

2.1 RUMUSAN MASALAH

Dalam analisis regresi yang melibatkan data deretan waktu, jika model regresi memasukkan tidak hanya nilai variabel yang menjelaskan saat ini, tapi juga nilai masa lalu (lagged), model tadi disebut model lag yang didistribusikan (Greene, 2003; Baltagi, 2008; Wooldridge, 2009). Model tadi disebut juga model lag yang didistribusikan penuh. Dalam kasus data deretan waktu menunjukkan beberapa perilaku periodik, pemodelan lag yang didistribusikan penuh sering menghasilkan koefisien yang mendekati nol pada beberapa lag. Koefisien ini perlu dihilangkan melalui konsep subset sehingga menghasilkan model subset lag yang didistribusikan.

Model subset lag yang didistribusikan adalah teknik yang digunakan untuk memprediksi nilai variabel yang tak bebas dari variabel yang menjelaskan pada berbagai lag. Misalnya, hasil penjualan per tahun tergantung pada biaya pemasaran 1 tahun sebelumnya, biaya pemasaran 2 tahun sebelumnya dan biaya pemasaran 5 tahun sebelumnya. Jika kita tertarik untuk mempelajari pengaruh gabungan dari variabel yang menjelaskan pada berbagai lag, kita dapat menggunakan teknik model subset lag yang didistribusikan ini.

Misalkan y_t adalah variabel terikat atau tak bebas, x_t, x_{t-1} , x_{t-2} , …., x_{t-k} adalah variabel yang menjelaskan, z_t adalah gangguan yang stokhastik atau galat, dan t menyatakan pengamatan yang ke-t, maka model subset lag yang didistribusikan bisa ditulis sebagai :

$$y_t = \beta_0 + \beta_{n_1} x_{t-n_1} + \cdots + \beta_{n_k} x_{t-n_k} + z_t \qquad (1)$$

untuk t = 1, 2, 3, …, n. Di mana $\{n_1, n_2, …, n_k\}$ adalah himpunan bagian dari {0, 1, 2, …. ,k}, $\beta = (\beta_0, \beta_{n_1}, \cdots, \beta_{n_k})'$ adalah vektor koefisien dan z_t adalah barisan gangguan stokhastik dengan mean 0 dan variansi σ^2. Data indeks harga saham gabungan vs data kurs USD, data konsumsi vs pendapatan, dan data jumlah uang vs laju inflasi merupakan beberapa contoh data riil yang dapat dimodelkan oleh model subset lag yang didistribusikan. Contoh yang lain, model subset lag yang didistribusikan digunakan untuk memodelkan data pemasaran (Leeflang, P.S.H *et al,* 2000; Soetharaman, 2004) dan data kinerja perusahaan (Lee and Kim, 2006).

Berdasarkan data x_t, x_{t-1}, x_{t-2}, x_{t-3}, ..., x_{t-k} (t = 1, 2, ..., n), pertama akan ditaksir harga β, σ^2 dan k. Selanjutnya akan ditentukan nilai prediksi untuk variabel terikat y_t untuk t = n+1.

2.2 METODE PENELITIAN

Penaksir Kuadrat Terkecil

Persamaan (1) merupakan bentuk ringkas untuk sekumpulan n persamaan simultan berikut :

$$\begin{aligned} y_1 &= \beta_0 + \beta_{n_1} x_{1-n_1} + \cdots + \beta_{n_k} x_{1-n_k} + z_1 \\ y_2 &= \beta_0 + \beta_{n_1} x_{2-n_1} + \cdots + \beta_{n_k} x_{2-n_k} + z_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_{n_1} x_{n-n_1} + \cdots + \beta_{n_k} x_{n-n_k} + z_n \end{aligned} \tag{2}$$

Dalam bentuk matriks, persamaan (2) menjadi

$$Y = X\beta + z \tag{3}$$

di mana

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_{n_1} \\ \vdots \\ \beta_{n_k} \end{bmatrix}, z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix},$$

dan

$$X = \begin{bmatrix} 1 & X_{1-n_1} & X_{1-n_2} & \cdots & X_{1-n_k} \\ 1 & X_{2-n_1} & X_{2-n_2} & \cdots & X_{2-n_k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n-n_1} & X_{n-n_2} & \cdots & X_{n-n_k} \end{bmatrix},$$

Untuk mendapatkan taksiran kuadrat terkecil dari β, mari kita mula-mula menuliskan model subset lag yang didistribusikan sampel :

$$Y_t = \hat{\beta}_0 + \hat{\beta}_{n_1} X_{t-n_1} + \cdots + \hat{\beta}_{n_k} X_{t-n_k} + e_t \tag{4}$$

untuk t = 1, 2, 3, ..., n, yang dapat ditulis secara ringkas dalam notasi matriks sebagai :

$$Y = X\hat{\beta} + e \tag{5}$$

di mana

$$Y=\begin{bmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{bmatrix},\hat{\beta}=\begin{bmatrix}\hat{\beta}_0\\\hat{\beta}_{n_1}\\\vdots\\\hat{\beta}_{n_k}\end{bmatrix},e=\begin{bmatrix}e_1\\e_2\\\vdots\\e_n\end{bmatrix},$$

dan

$$X=\begin{bmatrix}1 & X_{1-n_1} & X_{1-n_2} & \cdots & X_{1-n_k}\\1 & X_{2-n_1} & X_{2-n_2} & \cdots & X_{2-n_k}\\\vdots & \vdots & \vdots & & \vdots\\1 & X_{n-n_1} & X_{n-n_2} & \cdots & X_{n-n_k}\end{bmatrix}.$$

Di sini, $\hat{\beta}$ adalah suatu vektor kolom dari penaksir kuadrat terkecil koefisien model subset lag yang didistribusikan dan e adalah suatu vektor kolom dari n residual.

Menurut metode kuadrat terkecil, penaksir kuadrat terkecil diperoleh dengan meminimumkan

$$\sum_{t=1}^{n} e_t^2=\sum_{t=1}^{n}(y_t-\hat{\beta}_0-\hat{\beta}_{n_1}X_{t-n_1}-\cdots-\hat{\beta}_{n_k}X_{t-n_k})^2 \tag{6}$$

Ini dicapai dengan menurunkan (6) secara parsial terhadap $\hat{\beta}_0$, $\hat{\beta}_{n_1}$, …, $\hat{\beta}_{n_k}$ dan menyamakan hasil yang diperoleh dengan nol. Proses ini menghasilkan k+1 persamaan simultan dalam k+1 variabel yang tidak diketahui.

$$\begin{aligned}
&n\hat{\beta}_0+\cdots+\hat{\beta}_{n_k}\sum_{t=1}^{n}X_{t-n_k}=\sum_{t=1}^{n}y_t\\
&\hat{\beta}_0\sum_{t=1}^{n}X_{t-n_1}+\cdots+\hat{\beta}_{n_k}\sum_{t=1}^{n}X_{t-n_1}X_{t-n_k}=\sum_{t=1}^{n}x_{t-n_1}y_t\\
&\hat{\beta}_0\sum_{t=1}^{n}X_{t-n_2}+\cdots+\hat{\beta}_{n_k}\sum_{t=1}^{n}X_{t-n_2}X_{t-n_k}=\sum_{t=1}^{n}x_{t-n_2}y_t\\
&\cdots\cdots\cdots\\
&\hat{\beta}_0\sum_{t=1}^{n}X_{t-n_k}+\cdots+\hat{\beta}_{n_k}\sum_{t=1}^{n}X_{t-n_k}^2=\sum_{t=1}^{n}x_{t-n_k}y_t
\end{aligned} \tag{7}$$

Dalam bentuk matriks, persamaan (7) dapat disajikan sebagai :

$$(X'X)\hat{\beta}=X'Y \tag{8}$$

di mana

$$Y=\begin{bmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{bmatrix},\hat{\beta}=\begin{bmatrix}\hat{\beta}_0\\\hat{\beta}_{n_1}\\\vdots\\\hat{\beta}_{n_k}\end{bmatrix},$$

$$X = \begin{bmatrix} 1 & X_{1-n_1} & X_{1-n_2} & \cdots & X_{1-n_k} \\ 1 & X_{2-n_1} & X_{2-n_2} & \cdots & X_{2-n_k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n-n_1} & X_{n-n_2} & \cdots & X_{n-n_k} \end{bmatrix},$$

dan

$$X'X = \begin{bmatrix} n & \cdots & \sum_{t=1}^{n} X_{t-n_k} \\ \sum_{t=1}^{n} X_{t-n_1} & \cdots & \sum_{t=1}^{n} X_{t-n_1} X_{t-n_k} \\ \vdots & \vdots & \vdots \\ \sum_{t=1}^{n} X_{t-n_k} & \cdots & \sum_{t=1}^{n} X_{t-n_k}^2 \end{bmatrix}$$

Kalau invers dari (X'X) ada, katakan $(X'X)^{-1}$, maka dengan mengalikan di muka kedua sisi dari (8) dengan invers ini, kita memperoleh

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

atau

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

Penaksir kuadrat terkecil untuk $\beta = \left(\beta_0, \beta_{n_1}, \cdots, \beta_{n_k}\right)^t$ adalah

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

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

$$X = \begin{bmatrix} 1 & X_{1-n_1} & X_{1-n_2} & \cdots & X_{1-n_k} \\ 1 & X_{2-n_1} & X_{2-n_2} & \cdots & X_{2-n_k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n-n_1} & X_{n-n_2} & \cdots & X_{n-n_k} \end{bmatrix},$$

Apabila invers dari (X'X) tidak ada, maka invers dari (X'X) dengan invers semu dari (X'X).

Dengan menggunakan penaksir kuadrat terkecil untuk β, selanjutnya ditentukan penaksir kuadrat terkecil untuk σ^2, yaitu :

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

Statistik C_k

Untuk memilih model subset lag yang didistribusikan terbaik digunakan kriteria statistik C_k. Model subset lag yang didistribusikan terbaik dipilih adalah model subset lag yang didistribusikan yang memiliki nilai C_k terkecil. Nilai C_k untuk masing-masing model dihitung dengan menggunakan persamaan berikut (Elfron and Tibshirani, 1993) :

$$C_k = \frac{\sum_{t=1}^{n} (y_t - \hat{\beta}_0 - \hat{\beta}_{n_1} x_{t-n_1} - \cdots \hat{\beta}_{n_k} x_{t-n_k})^2}{n} - \frac{2k\hat{\sigma}^2}{n}$$

2.3 HASIL DAN PEMBAHASAN

Sebagai ilustrasi, kita akan menerapkan metode kuadrat terkecil dan statistik C_k untuk menentukan nilai prediksi pada data sintesis (studi simulasi) dan data riil (studi kasus). Studi simulasi (Law and Kelton, 2000) ditempuh untuk mengkonfirmasi kinerja dari pendekatan yang diusulkan apakah dapat bekerja dengan baik. Sedangkan studi kasus diberikan untuk memberikan contoh penerapan penelitian dalam memecahkan permasalahan dalam kehidupan sehari-hari. Komputasi ditulis dalam bahasa pemrograman MATLAB (Hanselman and Littlefield, 1997).

Data Sintesis

Tabel 1 menunjukkan 18 data sintesis model subset lag yang didistribusikan. Nilai x_t ditentukan sedangkan nilai y_t dibuat dengan menggunakan persamaan (1) di atas. Nilai koefisien model subset lag yang didistribusikan dan variansi gangguan stokhastik adalah β_0 = -7.1951, β_1 = 0.6360, dan β_3 = 0.6695 dan σ^2 = 1.

Tabel 1 : Data sintesis

t	Y_t	X_t
1	28.3768	28.736
2	27.0613	27.280
3	31.3883	30.219
4	30.9428	30.796
5	31.5399	30.896
6	35.6736	33.113
7	36.9593	35.032
8	38.6815	37.335
9	42.6640	41.003
10	46.5120	44.869
11	49.6113	46.449
12	55.5498	50.282
13	57.3752	53.555
14	62.2702	52.859

15	64.0868	55.917
16	67.7507	62.017
17	76.7172	71.398
18	86.5862	82.078

Berdasarkan data sintesis tersebut, selanjutnya mengestimasi koefisien model subset lag yang didistribusikan dan variansi σ^2 dengan menggunakan metode kuadrat terkecil. Pemilihan model terbaik, dilakukan dengan melihat nilai statistik C_k untuk le-7 model.

Tabel 2 : Nilai statistik C_k

Variabel Terikat	Variabel Bebas	Statistik C_k
Y_t	X_t	4.4345
Y_t	X_{t-1}	2.7848
Y_t	X_{t-2}	6,1276
Y_t	X_t, X_{t-1}	2,2424
Y_t	X_t, X_{t-2}	0,7555
Y_t	X_{t-1}, X_{t-2}	2,3930
Y_t	X_t, X_{t-1}, X_{t-2}	0,8452

Dari Tabel 2 terlihat bahwa nilai statistik C_k terkecil dicapai oleh persamaan model subset lag yang didistribusikan ke-5. Dengan demikian, model subset lag yang didistribusikan ke-5 inilah yang merupakan model subset lag yang didistribusikan terbaik.

Berdasarkan model subset lag yang didistribusikan terbaik ini, selanjutnya diestimasi parameter model subset lag yang didistribusikan yang bersesuaian. Hasilnya adalah $\hat{\beta}_0 = -7.3446$, $\hat{\beta}_1 = 0.6175$, $\hat{\beta}_3 = 0.7025$ dan $\sigma^2 = 0.7111$. Apabila nilai parameter dan nilai estimatornya baik untuk koefisen regresi maupun variansi dibandingkan, terlihat bahwa metode kuadrat terkecil dan statistik C_k dapat bekerja dengan "baik" dalam memilih model dan mengestimasi parameter dari data sintesis. Prediksi untuk nilai y_{18} jika x_{18} = 82.0789 dan x_{16} = 62.0170 adalah 86.9025.

Data Riil

Tabel 3 menunjukkan indeks harga saham gabungan (y_t) dan kurs USD (x_t) dari tanggal 4 Januari 2010 sampai dengan tanggal 9 Juni 2010 (http://www.finance.yahoo.com dan http:// www.bi.go.id).

Tabel 3 : Indeks Harga Saham Gabungan dan kurs USD

t	Y_t	X_t
1	2533.9	9377
2	2575.6	9355
3	2605.5	9355
4	2603.5	9274
5	2586.8	9286
6	2615.6	9176
7	2632.3	9231
8	2657.9	9226
9	2633.6	9196
10	2646.7	9251
11	2645.4	9276
12	2642.4	9271
13	2666.6	9321
14	2664.7	9366
15	2637	9435
16	2609.7	9387
17	2597.4	9362
18	2577.9	9427
19	2565.4	9455
20	2619.3	9412
21	2610.6	9442
22	2588.3	9417
23	2580.7	9392
24	2604.8	9372
25	2592.4	9440
26	2518.6	9460
27	2474.7	9435
28	2490.3	9397
29	2483.6	9407
30	2508.2	9418
31	2533.7	9387
32	2517.7	9384
33	2558.6	9326
34	2581.4	9372
35	2559.9	9405
36	2554.8	9338
37	2564.2	9365
38	2582.4	9368

39	2579.3	9382
40	2548.8	9360
41	2554.5	9321
42	2577	9323
43	2567	9311
44	2566.1	9311
45	2579	9246
46	2626.3	9244
47	2656.9	9234
48	2670.1	9231
49	2676.1	9229
50	2666.4	9221
51	2669.4	9195
52	2756.5	9166
53	2738.8	9171
54	2742.7	9162
55	2702.6	9165
56	2721.2	9166
57	2773.3	9184
58	2799.2	9182
59	2813.5	9135
60	2795.2	9115
61	2798.7	9161
62	2777.7	9120
63	2831	9100
64	2888.8	9090
65	2881.4	9082
66	2898.2	9109
67	2850.6	9094
68	2845.6	9048
69	2881	9065
70	2884.9	9054
71	2885.1	9049
72	2900.8	9063
73	2878	9091
74	2840.6	9073
75	2891.5	9052
76	2912.7	9072
77	2926.1	9061

78	2924.9	9046
79	2943.7	9058
80	2939	9068
81	2903.5	9067
82	2927.3	9057
83	2971.8	9075
84	2961.3	9062
85	2958.5	9098
86	2846	9251
87	2808.7	9339
88	2739.9	9166
89	2850.8	9118
90	2813	9161
91	2847.1	9139
92	2858	9191
93	2820.5	9179
94	2833.7	9214
95	2730.3	9251
96	2692.7	9382
97	2623.7	9315
98	2609	9382
99	2514.9	9420
100	2696.3	9385
101	2714.3	9226
102	2796.7	9256
103	2725	9281
104	2734.2	9236
105	2810.9	9250
106	2820.9	9341
107	2750.4	9311
108	2779.8	9284

Data pada Tabel 3 dicocokkan terhadap model subset lag yang didistribusikan. Metode kuadrat terkecil digunakan untuk mendapatkan estimator parameter model regresi dan variansi σ^2. Pemilihan model, dilakukan dengan melihat nilai statistik C_k untuk ke-15 model.

Tabel 4 : Nilai statistik C_k

Variabel Terikat	Variabel Bebas	Statistik C_k

Y_t	X_t	4716.6
Y_t	X_{t-1}	3658.7
Y_t	X_{t-2}	4342.7
Y_t	X_{t-3}	4531.5
Y_t	X_t, X_{t-1}	3599.0
Y_t	X_t, X_{t-2}	3601.6
Y_t	X_t, X_{t-3}	3471.5
Y_t	X_{t-1}, X_{t-2}	3485.7
Y_t	X_{t-1}, X_{t-3}	3101.9
Y_t	X_{t-2}, X_{t-3}	4002.2
Y_t	X_t, X_{t-1}, X_{t-2}	3420.7
Y_t	X_t, X_{t-1}, X_{t-3}	3107.6
Y_t	X_t, X_{t-2}, X_{t-3}	3413.5
Y_t	X_{t-1}, X_{t-2}, X_{t-3}	3159.5
Y_t	X_t, X_{t-1}, X_{t-2}, X_{t-3}	3167.4

Dari Tabel 4 terlihat bahwa nilai statistik C_k terkecil dicapai oleh persamaan model subset lag yang didistribusikan ke-9. Dengan demikian, model subset lag yang didistribusikan ke-9 inilah yang merupakan model subset lag yang didistribusikan terbaik.

Berdasarkan model subset lag yang didistribusikan terbaik ini, selanjutnya diestimasi parameter model subset lag yang didistribusikan yang bersesuaian. Hasilnya adalah $\hat{\beta}_0 = 12094.66$, $\hat{\beta}_3 = -0.62$, $\hat{\beta}_5 = -0,4$ dan $\sigma^2 = 3072,63$. Prediksi untuk nilai y_{108} jika $x_{106} = 9311$ dan $x_{104} = 9250$ adalah 2670.57.

2.4 KESIMPULAN

Uraian di atas menunjukkan bahwa betapa sederhananya metode kuadrat terkecil dan statistik C_k dapat digunakan untuk menghasilkan taksiran parameter dalam model subset lag yang didistribusikan dan menentukan nilai prediksi untuk variabel terikat dalam model subset lag yang didistribusikan. Dari hasil simulasi menunjukkan bahwa pendekatan yang diusulkan dapat menaksir parameter dan menentukan nilai prediksi dengan baik.

Sebagai implementasi, diambil data indeks harga saham gabungan (y_t) dan kurs USD (x_t) dari tanggal 4 Januari 2010 sampai dengan 9 Juni 2010. Dengan menggunakan metode kuadrat terkecil dan statistik C_k, diperoleh model matematik yaitu :

$$Y_t = 12094.66 - 0.62X_{t-2} - 0.4X_{t-4}$$

Model matematik ini sangat bermanfaat bagi pengambilan keputusan, misalnya untuk memprediksikan nilai variabel Y_t di masa mendatang.

Dalam artikel ini, dibahas prediksi titik untuk variabel terikat. Pendalaman dan perluasan dapat ditempuh dengan mengadopsi algoritma Bootstrap (Efron and Tibshirani, 1993) untuk mendapatkan prediksi selang dari variabel terikat.

BAB 3
METODE BOOTSTRAP DALAM INFERENSI MODEL REGRESI GANDA

3.1 RUMUSAN MASALAH

Regresi ganda adalah teknik yang digunakan untuk memprediksi nilai variabel yang tak bebas dari dua atau lebih variabel yang menjelaskan. Misalnya, hasil padi per hektar tergantung pada kualitas benih, kesuburan tanah, pupuk yang digunakan, temperatur, dan curah hujan. Jika kita tertarik untuk mempelajari pengaruh gabungan dari semua variabel pada hasil padi, kita dapat menggunakan teknik regresi ganda ini. Keuntungan tambahan teknik regresi ganda ini juga memungkinkan kita untuk mempelajari pengaruh individual variabel-variabel terhadap hasil padi.

Misalkan y_t adalah variabel terikat atau tak bebas, x_{t1} , x_{t2} , ..., x_{tk} adalah variabel yang menjelaskan, z_t adalah gangguan yang stokhastik atau galat, dan t menyatakan pengamatan yang ke-t, maka model regresi ganda bisa ditulis sebagai (Allison, 1999; Freund and Wilson, 1998) :

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + z_t \qquad (1)$$

untuk t = 1, 2, 3, ..., n. Di mana $\beta = (\beta_0, \beta_1, \cdots, \beta_k)'$ 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 ($ dollar, Euro, Yen) merupakan suatu contoh data riil yang dapat dimodelkan oleh model regresi ganda (Suparman, 1997). Contoh lain dapat ditemukan diberbagai literatur (Dupont, 2002; Heij *et al.*, 2004; Sen and Srivastawa, 1990).

Berdasarkan data x_{t1}, x_{t2}, x_{t3}, ..., x_{tk} (t = 1, 2, ..., n), pertama akan ditaksir harga β, σ^2 dan k. Selanjutnya akan ditentukan selang prediksi untuk variabel terikat y.

3.2 METODE PENELITIAN

Penaksir Kuadrat Terkecil

Persamaan (1) merupakan bentuk ringkas untuk sekumpulan n persamaan simultan berikut :

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + z_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + z_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + z_n \end{aligned} \qquad (2)$$

Dalam bentuk matriks, persamaan (2) menjadi

$$Y = X\beta + z \qquad (3)$$

di mana

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}, \ \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \text{ dan } z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}.$$

Untuk mendapatkan taksiran kuadrat terkecil dari β, mari kita mula-mula menuliskan regresi ganda sampel :

$$Y_t = \hat\beta_0 + \hat\beta_1 X_{t1} + \hat\beta_2 X_{t2} + \cdots + \hat\beta X_{tk} + e_t \qquad (4)$$

untuk t = 1, 2, 3, ..., n, yang dapat ditulis secara ringkas dalam notasi matriks sebagai :

$$Y = X\hat\beta + e \qquad (5)$$

di mana

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}, \hat\beta = \begin{bmatrix} \hat\beta_0 \\ \hat\beta_1 \\ \vdots \\ \hat\beta_k \end{bmatrix}, \text{ dan } e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}.$$

Di sini, $\hat\beta$ adalah suatu vektor kolom dari penaksir kuadrat terkecil koefisien regresi ganda dan e adalah suatu vektor kolom dari n residual.

Menurut metode kuadrat terkecil, penaksir kuadrat terkecil diperoleh dengan meminimumkan

$$\sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n} (y_t - \hat\beta_0 - \hat\beta_1 X_{t1} - \hat\beta_2 X_{t2} - \cdots - \hat\beta_k X_{tk})^2 \qquad (6)$$

Ini dicapai dengan menurunkan (6) secara parsial terhadap $\hat{\beta}_0$, $\hat{\beta}_1$, …, $\hat{\beta}_k$ dan menyamakan hasil yang diperoleh dengan nol. Proses ini menghasilkan k+1 persamaan simultan dalam k+1 peubah yang tidak diketahui.

$$
\begin{aligned}
& n\hat{\beta}_0 + \hat{\beta}_1 \sum_{t=1}^{n} X_{t1} + \hat{\beta}_2 \sum_{t=1}^{n} X_{t2} + \cdots + \hat{\beta}_k \sum_{t=1}^{n} X_{tk} && = \sum_{t=1}^{n} y_t \\
& \hat{\beta}_0 \sum_{t=1}^{n} X_{t1} + \hat{\beta}_1 \sum_{t=1}^{n} X_{t1}^2 + \hat{\beta}_2 \sum_{t=1}^{n} X_{t1} X_{t2} + \cdots + \hat{\beta}_k \sum_{t=1}^{n} X_{t1} X_{tk} && = \sum_{t=1}^{n} x_{t1} y_t \\
& \hat{\beta}_0 \sum_{t=1}^{n} X_{t2} + \beta_1 \sum_{t=1}^{n} X_{t2} X_{t1} + \hat{\beta}_2 \sum_{t=1}^{n} X_{t2}^2 + \cdots + \hat{\beta}_k \sum_{t=1}^{n} X_{t2} X_{tk} && = \sum_{t=1}^{n} x_{t2} y_t \\
& \cdots\cdots\cdots \\
& \hat{\beta}_0 \sum_{t=1}^{n} X_{tk} + \beta_1 \sum_{t=1}^{n} X_{tk} X_{t1} + \hat{\beta}_2 \sum_{t=1}^{n} X_{tk} X_{t2} + \cdots + \hat{\beta}_k \sum_{t=1}^{n} X_{tk}^2 && = \sum_{t=1}^{n} x_{tk} y_t
\end{aligned} \tag{7}
$$

Dalam bentuk matriks, persamaan (7) dapat disajikan sebagai :

$$(X'X)\hat{\beta} = X'Y \tag{8}$$

di mana

$$
Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad
X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}, \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix},
$$

dan

$$
X'X = \begin{bmatrix}
n & \sum_{t=1}^{n} X_{t1} & \sum_{t=1}^{n} X_{t2} & \cdots & \sum_{t=1}^{n} X_{tk} \\
\sum_{t=1}^{n} X_{t1} & \sum_{t=1}^{n} X_{t1}^2 & \sum_{t=1}^{n} X_{t1} X_{t2} & \cdots & \sum_{t=1}^{n} X_{t1} X_{tk} \\
\vdots & \vdots & \vdots & & \vdots \\
\sum_{t=1}^{n} X_{tk} & \sum_{t=1}^{n} X_{tk} X_{t1} & \sum_{t=1}^{n} X_{tk} X_{t2} & \cdots & \sum_{t=1}^{n} x_{tk}^2
\end{bmatrix}
$$

Kalau invers dari (X'X) ada, katakan $(X'X)^{-1}$, maka dengan mengalikan di muka kedua sisi dari (8) dengan invers ini, kita memperoleh

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

atau

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

Penaksir kuadrat terkecil untuk $\beta = (\beta_0, \beta_1, \cdots, \beta_k)^t$ adalah

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

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

$$
X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix},
$$

Sedangkan penaksir kuadrat terkecil untuk σ^2 adalah :

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

Statistik C_k

Untuk menaksir orde p digunakan kriteria statistik C_k. Orde terbaik dipilih adalah orde yang memiliki nilai C_k terkecil. Nilai C_k dihitung dengan menggunakan persamaan berikut (Elfron and Tibshirani, 1993) :

$$C_k = \frac{\sum_{t=1}^{n} (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{t1} - \cdots \hat{\beta}_k x_{tk})^2}{n} - \frac{2k\hat{\sigma}^2}{n}$$

Metode Bootstrap

Algoritma bootstrap adalah metode berbasis komputer yang sangat potensial untuk dipergunakan pada masalah inferensi statistik (Gentle, 2002). Salah satu contohnya adalah penggunaan algoritma bootstrap untuk menginferensi parameter model regresi ganda. 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 $z_1, z_2, \ldots, z_n$ 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.

Misalkan $\hat{F}$ adalah distribusi empirik yang diambil dengan probabilitas 1/n pada setiap nilai yang diamati $z_1, z_2, \ldots, z_n$. Sampel bootstrap didefinisikan sebagai sampel random berukuran n disusun dari $\hat{F}$, misal sampel bootstrap ke-b (b = 1, 2, ..., B) dinotasikan dengan $z_1^b, z_2^b, \cdots, z_n^b$. Sampel bootstrap $z_1^b, z_2^b, \cdots, z_n^b$ adalah sampel random berukuran n yang diambil dengan pengembalian dari populasi $z_1, z_2, \ldots, z_n$. Anggota data bootstrap $z_1^b, z_2^b, \cdots, z_n^b$ beranggotakan sampel asli $z_1, z_2, \ldots, z_n$, 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 $z_1, z_2, \ldots, z_n \sim F$ adalah sampel random berukuran n dari distribusi F yang tidak diketahui	• Sampel bootstrap $z_1^b, z_2^b, \cdots, z_n^b \sim \hat{F}$ adalah sampel buatan berukuran n dari distribusi $\hat{F}$.
• $\theta = \theta(F)$ adalah nilai riil dari suatu parameter yang menjadi perhatian.	• $\theta = \theta(\hat{F})$
• Jika T adalah statistik untuk θ maka $\hat{\theta} = T(z_1, z_2, \cdots, z_n)$.	• $\hat{\theta} = T(z_1^b, z_2^b, \cdots, z_n^b)$.

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

- Hitunglah $\hat{\beta}$ dan $\hat{\sigma}^2$ dari data asli.
- Hitunglah $\hat{z}_t$ dengan menggunakan persamaan

  $$\hat{z}_t = \hat{y}_t - \hat{\beta}_0 - \hat{\beta}_1 x_{t1} - \hat{\beta}_2 x_{t2} - \cdots + \hat{\beta}_k x_{tk}.$$
- Untuk b = 1, 2, ..., B :
  - Resampling $\hat{z}_t^{(b)}$.
  - Hitunglah $y_t^{(b)}$ dengan persamaan

$$y_t^{(b)} = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \hat{z}_t^{(b)}.$$

- 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 $\hat{y}_{t+1}$

3.3. HASIL DAN PEMBAHASAN

Sebagai ilustrasi, kita akan menerapkan algoritma bootstrap untuk menentukan selang prediksi pada data sintesis (studi simulasi) dan data riil (studi kasus). 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 $\alpha = 0.05$. Algoritma ditulis dalam bahasa pemrograman MATLAB (Hanselman and Littlefield, 1997).

Data Sintesis

Tabel 1 menunjukkan 20 data sintesis regresi ganda orde 2. Nilai x_1 dan x_2 ditentukan sedangkan nilai y dibuat dengan menggunakan persamaan (3) di atas. Nilai koefisien regresi dan variansi gangguan stokhastik adalah β_0 = 6, β_1 = 3, dan β_2 = 5 dan $\sigma^2 = 1$.

Tabel 1 : Data sintesis

Y	X_1	X_2
45.2944	3	6
51.6638	4	7
52.7143	7	5
46.6236	3	6
56.3082	7	6
42.858	7	3
40.254	6	3
27.4063	1	4

31.559	4	3
38.5711	9	1
29.6001	3	3
63.69	4	9
39.8156	6	3
32.7119	2	4
41.2902	3	5
48.6686	4	6
31.1908	3	3
38.7975	8	2
48.9802	6	5
36.8433	7	2

Berdasarkan data sintesis tersebut, selanjutnya mengestimasi orde, koefisien regresi polinomial dan variansi σ^2 dengan menggunakan algoritma bootstrap. Estimasi orde, dilakukan dengan melihat nilai statistik C_k untuk tiga model.

Tabel 2 : Nilai statistik C_k

Variabel Terikat	Variabel Bebas	Statistik C_k
Y	X_1	239.0859
Y	X_2	109.1284
Y	X_1, X_2	*2.8143*

Dari Tabel 2 terlihat bahwa nilai statistik C_k terkecil dicapai oleh persamaan regresi ganda ketiga. Dengan demikian, regresi ganda ketiga inilah yang merupakan model regresi ganda terbaik.

Berdasarkan regresi ganda terbaik ini, selanjutnya diestimasi parameter model regresi ganda yang bersesuaian. Hasilnya adalah $\hat{\beta}_0 = 5.8995$, $\hat{\beta}_1 = 3.0213$, $\hat{\beta}_2 = 5.0453$ dan $\hat{\sigma}^2 = 0.9368$. 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. Prediksi untuk nilai y_{20} jika x_1 = 7 dan x_2 = 2 adalah 37.1391 dan selang prediksi yang bersesuaian adalah (36.5309, 37.7162).

Data Riil

Tabel 3 menunjukkan jumlah uang yang beredar (Y) dan faktor-faktor yang mempengaruhi, yaitu aktiva luar negeri bersih (X_1), tagihan bersih kepada pemerintah pusat (X_2), tagihan kepada lembaga dan BUMN pusat berupa kredit (X_3), dan tagihan kepada perusahaan swasta dan perorangan dalam bentuk kredit (X_4) pada bulan Juli 2007 sampai dengan April 2008 dalam milyar rupiah (www.bi.go.id).

Tabel 3 : Uang yang beredar dan faktor-faktor yang mempengaruhinya

Bulan	Y	X_1	X_2	X_3	X_4
Juli 2007	144179	498496	444352	38911	826194
Agustus 2007	149194	498091	443878	40264	846472
September 2007	160327	519360	439649	40281	866979
Oktober 2007	156955	517566	437701	46136	884017
November 2007	161272	518424	447846	44920	908339
Desember 2007	183419	524703	497478	51038	944074
Januari 2008	166950	529580	446397	43221	937040
Februari 2008	165633	543467	433322	40313	955010
Maret 2008	164995	549049	375976	44748	984424
April 2008	171049	544746	371557	45675	1009072

Data pada Tabel 3 dicocokkan terhadap regresi ganda. Algoritma bootstrap digunakan untuk mendapatkan estimator orde, parameter model regresi dan variansi σ^2. Estimasi orde, dilakukan dengan melihat nilai statistik C_k untuk tiga model.

Tabel 4 : Nilai statistik C_k

Variabel Terikat	Variabel Bebas	Statistik C_k
Y	X_1	1.7714×10^8
Y	X_2	3.2492×10^8
Y	X_3	1.3029×10^8
Y	X_4	1.3118×10^8
Y	X_1, X_2	6.8238×10^7
Y	X_1, X_3	6.8440×10^7
Y	X_1, X_4	1.4807×10^8
Y	X_2, X_3	1.2984×10^8
Y	X_2, X_4	2.8434×10^7

Y	X_3, X_4	7.1298×10^7
Y	X_1, X_2, X_3	9.2289×10^7
Y	X_1, X_2, X_4	4.2453×10^7
Y	X_1, X_3, X_4	7.7687×10^7
Y	X_2, X_3, X_4	2.3552×10^7
Y	X_1, X_2, X_3, X_4	3.9092×10^7

Dari Tabel 4 terlihat bahwa nilai statistik C_k terkecil dicapai oleh persamaan regresi ganda ke-14. Dengan demikian, regresi ganda ke-14 inilah yang merupakan model regresi ganda terbaik.

Berdasarkan regresi ganda terbaik ini, selanjutnya diestimasi parameter model regresi ganda yang bersesuaian. Hasilnya adalah $\hat{\beta}_0 = -92898.48$, $\hat{\beta}_1 = 0.17$, $\hat{\beta}_2 = 0.52$, $\hat{\beta}_3 = 0.18$ dan $\hat{\sigma}^2 = 7.61 \times 10^6$. Prediksi untuk nilai y_{10} jika x_2 = 375976, x_3 = 44748 dan x_4 = 984424 adalah 169552.29 dan (167191.75, 173326.86).

3.4 KESIMPULAN

Uraian di atas menunjukkan bahwa betapa sederhananya algoritma bootstrap dapat digunakan untuk menghasilkan taksiran parameter dalam model regresi ganda dan menentukan selang prediksi untuk variabel terikat dalam model regresi ganda apabila gangguan stokhastiknya berdistribusi sembarang. Dari hasil simulasi menunjukkan bahwa algoritma bootstrap dapat menaksir parameter dan menentukan selang prediksi dengan baik.

Sebagai implementasi metode bootstrap, diambil data jumlah uang yang beredar (Y) dan faktor-faktor yang mempengaruhi, yaitu aktiva luar negeri bersih (X_1), tagihan bersih kepada pemerintah pusat (X_2), tagihan kepada lembaga dan BUMN pusat berupa kredit (X_3), dan tagihan kepada perusahaan swasta dan perorangan dalam bentuk kredit (X_4) pada bulan Juli 2007 sampai dengan April 2008 (dalam milyar rupiah). Dengan menggunakan metode bootstrap diperoleh model matematik yaitu

$$Y = -92898.48 + 0.17X_2 + 0.52X_3 + 0.18X_4$$

Model matematik ini sangat bermanfaat bagi pengambilan keputusan, misalnya untuk memprediksikan nilai atau menghitung selang prediksi variabel Y di masa mendatang.

Meskipun dalam artikel ini hanya dibahas model regresi ganda, tetapi algoritma bootstrap dapat diterapkan juga pada model-model ekonometrik linear yang lainnya (Suparman, 1997). 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).

BAB 4

METODE BOOTSTRAP DALAM INFERENSI MODEL REGRESI POLINOMIAL

4.1 RUMUSAN MASALAH

Model regresi polinomial ini merupakan salah satu model regresi yang sering digunakan untuk menyatakan hubungan antara sejumlah pasangan data. Hubungan antara biaya iklan dengan penjualan, hubungan antara banyaknya curah hujan dengan jumlah kotoran udara, dan hubungan antara laju inflasi dan indeks harga konsumen, hubungan antara merokok dan penyakit kanker paru-paru, hubungan temperatur dengan kuat tekan beton merupakan beberapa contoh data riil yang dapat dinyatakan dengan model regresi polinomial.

Misalkan y_t adalah variabel tak bebas, x_t variabel yang menjelaskan, u_t adalah gangguan yang stokhastik, dan t menyatakan pengamatan yang ke-t, maka model regresi polinomial orde p bisa ditulis sebagai (Devoire, 2008; Black, 2009)

$$y_t = a_0 + a_1 x_t + a_2 x_t^2 + \cdots + a_p x_t^p + u_t \qquad (1)$$

untuk t = 1, 2, 3, …, n. Di mana p adalah orde, $a = (a_0, a_1, \cdots, a_m)^t$ adalah vektor koefisien dan u_t adalah barisan gangguan stokhastik dengan mean 0 dan variansi σ^2. Di sini, u_t diasumsikan berdistribusi sembarang dan tidak diketahui.

Apabila model regresi polinomial dicocokkan terhadap data riil, umumnya orde p dan parameter model regresi polinomial tidak diketahui. Sehingga permasalahan penelitian ini adalah bagaimana mengestimasi orde p, vektor koefisien a dan variansi gangguan stokhastik σ^2. Dan tujuan penelitian ini adalah berdasarkan data (y_t, x_t) (t = 1, 2, …, n), akan ditemukan penaksir untuk orde p, vektor koefisien a dan variansi gangguan stokhastik σ^2.

4.2 METODE PENELITIAN

Teori-teori yang digunakan untuk menaksir harga p, a dan σ^2 adalah metode kuadrat terkecil, statistic C_p dan algoritma bootstrap.

METODE KUADRAT TERKECIL

Meskipun model (1) memiliki variabel x yang nonlinier, tetapi koefisien-koefisiennya berbentuk linier. Oleh karena itu model tersebut merupakan model linier. Dengan demikian, parameter model regresi polinomial dapat ditaksir dengan menggunakan metode kuadrat terkecil. Persamaan (1) adalah bentuk ringkas untuk sekumpulan n persamaan simultan berikut :

$$\begin{aligned} Y_1 &= a_0 + a_1 X_1 + a_2 X_1^2 + \cdots + a_m X_1^p + u_1 \\ Y_2 &= a_0 + a_1 X_2 + a_2 X_2^2 + \cdots + a_m X_2^p + u_2 \\ &\vdots \\ Y_n &= a_0 + a_1 X_n + a_2 X_n^2 + \cdots + a_p X_n^p + u_n \end{aligned} \tag{2}$$

Dalam bentuk matriks, persamaan (2) menjadi

$$Y = Xa + u \tag{3}$$

di mana

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_1 & X_1^2 & \cdots & X_1^p \\ 1 & X_2 & X_2^2 & \cdots & X_2^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_n & X_n^2 & \cdots & X_n^p \end{bmatrix}, a = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix}, dan\ u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}.$$

Untuk mendapatkan taksiran kuadrat terkecil dari a, mari kita mula-mula menuliskan regresi polinomial orde p sampel :

$$\mathrm{Y}_\mathrm{i} = \hat{a}_0 + \hat{a}_1 X_i + \hat{a}_2 X_i^2 + \cdots + \hat{a}_p X_i^p + e_i \tag{4}$$

untuk i = 1, 2, 3, ..., n, yang dapat ditulis secara ringkas dalam notasi matriks sebagai :

$$Y = X\hat{a} + e \tag{5}$$

di mana

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_1 & X_1^2 & \cdots & X_1^p \\ 1 & X_2 & X_2^2 & \cdots & X_2^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_n & X_n^2 & \cdots & X_n^p \end{bmatrix}, \hat{a} = \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_p \end{bmatrix}, dan\ e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}.$$

Di sini, $\hat{a}$ adalah suatu vektor kolom dari penaksir kuadrat terkecil koefisien regresi polinomial dan e adalah suatu vektor kolom dari n residual.

Menurut metode kuadrat terkecil, penaksir kuadrat terkecil diperoleh dengan meminimumkan

$$\sum_{i=1}^{n} e_i^2 \quad = \sum_{i=1}^{n} (y_i - \hat{a}_0 - \hat{a}_1 X_i - \hat{a}_2 X_i^2 - \cdots - \hat{a}_p X_i^p)^2 \tag{6}$$

Ini dicapai dengan menurunkan (6) secara parsial terhadap $\hat{a}_0$, $\hat{a}_1$, …, $\hat{a}_p$ dan menyamakan hasil yang diperoleh dengan nol. Proses ini menghasilkan p+1 persamaan simultan dalam p+1 peubah yang tidak diketahui.

$$\begin{aligned}
n\hat{a}_0 + \hat{a}_1 \sum_{i=1}^{n} X_i + \hat{a}_2 \sum_{i=1}^{n} X_i^2 + \cdots + \hat{a}_o \sum_{i=1}^{n} X_i^p \quad &= \sum_{i=1}^{n} y_i \\
\hat{a}_0 \sum_{i=1}^{n} X_i + \hat{a}_1 \sum_{i=1}^{n} X_i^2 + \hat{a}_2 \sum_{i=1}^{n} X_i^3 + \cdots + \hat{a}_p \sum_{i=1}^{n} X_i^{p+1} \quad &= \sum_{i=1}^{n} x_i y_i \\
\hat{a}_0 \sum_{i=1}^{n} X_i^2 + \hat{a}_1 \sum_{i=1}^{n} X_i^3 + \hat{a}_2 \sum_{i=1}^{n} X_i^4 + \cdots + \hat{a}_p \sum_{i=1}^{n} X_i^{p+2} \quad &= \sum_{i=1}^{n} x_i^2 y_i \qquad (7) \\
\cdots\cdots\cdots \\
\hat{a}_0 \sum_{i=1}^{n} X_i^p + \hat{a}_1 \sum_{i=1}^{n} X_i^{p+1} + \hat{a}_2 \sum_{i=1}^{n} X_i^{p+2} + \cdots + \hat{a}_p \sum_{i=1}^{n} X_i^{2p} \quad &= \sum_{i=1}^{n} x_i^p y_i
\end{aligned}$$

Dalam bentuk matriks, persamaan (7) dapat disajikan sebagai :

$$(X'X)\hat{a} \quad = X'Y \tag{8}$$

di mana

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_1 & X_1^2 & \cdots & X_1^p \\ 1 & X_2 & X_2^2 & \cdots & X_2^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_n & X_n^2 & \cdots & X_n^p \end{bmatrix}, \ \hat{a} = \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_p \end{bmatrix},$$

dan

$$X'X = \begin{bmatrix} n & \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 & \cdots & \sum_{i=1}^{n} X_i^p \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 & \sum_{i=1}^{n} X_i^3 & \cdots & \sum_{i=1}^{n} X_i^{p+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} X_l^p & \sum_{i=1}^{n} X_i^{p+1} & \sum_{i=1}^{n} X_i^{p+2} & \cdots & \sum_{i=1}^{n} x_i^{2p} \end{bmatrix}$$

Kalau invers dari *(X'X)* ada, katakan $(X'X)^{-1}$, maka dengan mengalikan di muka kedua sisi dari (8) dengan invers ini, kita memperoleh

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

atau

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

Penaksir kuadrat terkecil untuk $a = (a_0, a_1, \cdots, a_p)^t$ adalah

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

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

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ 1 & x_2 & x_2^2 & & x_2^p \\ 1 & x_3 & x_3^2 & & x_3^p \\ \vdots & & & \ddots & \\ 1 & x_n & x_n^2 & & x_n^p \end{pmatrix}$$

Sedangkan penaksir kuadrat terkecil untuk σ^2 adalah (Gujarati, 2003) :

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

STATISTIK C_P

Untuk menaksir orde p digunakan kriteria statistik Cp. Orde terbaik dipilih adalah orde yang memiliki nilai C_p terkecil. Nilai C_p dihitung dengan menggunakan persamaan berikut (Elfron and Tibshirani, 1993) :

$$C_p = \frac{\sum_{i=1}^{n}(y_i - \hat{a}_0 - \hat{a}_1 x_i - \cdots - \hat{a}_p x_i^p)^2}{n} - \frac{2p\hat{\sigma}^2}{n}$$

ALGORITMA BOOTSTRAP

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 regresi polinomial. Dalam metode klasik para peneliti mendasarkan pada asumsi bahwa gangguan stokhastik dalam model regresi polinomial 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 peneliti 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.

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 $x_1^b, x_2^b, \cdots, x_n^b$. Sampel bootstrap $x_1^b, x_2^b, \cdots, x_n^b$ adalah sampel random berukuran n yang diambil dengan pengembalian dari populasi x_1, x_2, ... ,x_n. Anggota data bootstrap $x_1^b, x_2^b, \cdots, x_n^b$ 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, ... ,x_n ~ F adalah sampel random berukuran n dari distribusi F yang tidak diketahui	• Sampel bootstrap $x_1^b, x_2^b, \cdots, x_n^b$ ~ $\hat{F}$ adalah sampel buatan berukuran n dari distribusi $\hat{F}$.
• $\theta = \theta(F)$ adalah nilai riil dari suatu parameter yang menjadi perhatian.	• $\theta = \theta(\hat{F})$
• Jika T adalah statistik untuk θ maka $\hat{\theta} = T(x_1, x_2, \cdots, x_n)$.	• $\hat{\theta} = T(x_1^b, x_2^b, \cdots, x_n^b)$.

Langkah-langkah komputasi untuk 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 $\hat{u}_t$ dengan menggunakan persamaan $o_t = y_t - a_0 - \sum_{i=1}^{p} \hat{a}_i x_t^i$.

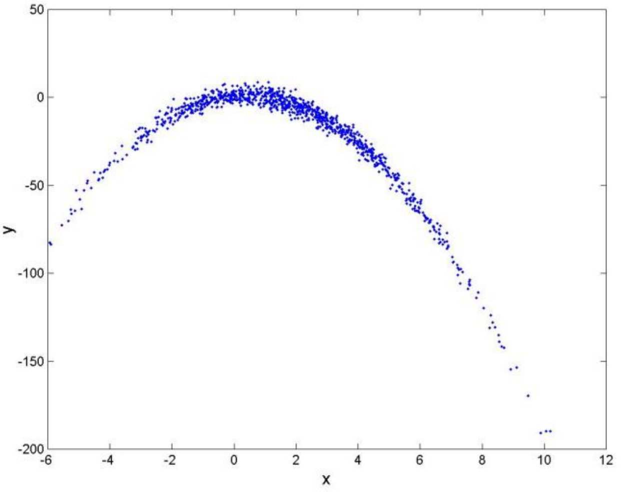
- Untuk b = 1, 2, ..., B :
    - Resampling $\hat{u}_t^{(b)}$.
    - Hitunglah $y_t^{(b)}$ dengan menggunakan persamaan $y_t^{(b)} = a_0 + \sum_{i=1}^{p} \hat{a}_i x_t^i + \hat{u}_t^{(b)}$.
    - Hitunglah $\hat{\beta}^{(b)}$, $\hat{\sigma}^{2(b)}$ dan $C_p^{(b)}$.
    - Hitunglah $\hat{y}_{t+1}^{(b)}$.
- Hitunglah $\hat{\beta}_{boot}$, $\hat{\sigma}^2_{boot}$, $C_{(p)(boot)}$ dan $\hat{y}_{(t+1)(boot)}$.
- Hitunglah interval kepercayaan $100(1-\alpha)\%$ untuk $\hat{y}_{t+1}$.

4.3 HASIL DAN PEMBAHASAN

Sebagai ilustrasi, algoritma bootstrap diterapkan untuk menginferensi parameter data sintesis dan data riil. Data sintesis (Law and Kelton, 2000) ditempuh untuk mengkonfirmasi kinerja dari algoritma bootstrap apakah dapat bekerja dengan baik. Algoritma dibuat dalam bahasa pemrograman MATLAB (Hanselman and Littlefield, 1997). Sedangkan data riil diberikan untuk memberikan contoh penerapan penelitian dalam memecahkan permasalahan dalam kehidupan sehari-hari. Di sini resampling dilakukan sebanyak B = 2000 dan α = 0.05. Nilai α ini menyatakan probalilitas kesalahan jenis I.

Data Sintesis

Gambar 1 menunjukkan 1000 data sintesis regresi polinoial orde 2. Nilai x ditentukan sedangkan nilai y dibuat dengan menggunakan persamaan (1) di atas. Nilai koefisien regresi dan variansi gangguan stokhastik adalah a_0 = 0.7160, a_1 = 1.5986, a_2 = -2.0647 dan $\sigma^2 = 9$.



Gambar 1. Data sintesis dari regresi polinomial orde 2

Berdasarkan data sintesis tersebut, nilai statistik $\hat{C}_p$ dihitung untuk p = 1, 2, ..., 7. Hasilnya disajikan dalam Tabel 1 berikut.

Tabel 1. Nilai statistik $\hat{C}_p$ untuk p =1,2, ...,7

p	$\hat{C}_p$
1	504.1765
2	9.5483
3	9.5725
4	9.5717
5	9.5923
6	9.5951
7	9.5819

Dari Tabel 1, nampak bahwa nilai $\hat{C}_p$ terkecil dicapai pada p = 2. Dengan demikian, model regresi polinoial terbaik adalah orde 2.

Selanjutnya mengestimasi koefisien regresi polinomial orde 2 dan variansi σ^2 yang bersesuaian dengan menggunakan algoritma bootstrap. Hasilnya adalah $\hat{a}_0 = 0.8409$, $\hat{a}_1 = 1.5618$, $\hat{a}_2 = -2.0578$ dan $\hat{\sigma}^2 = 9.5103$. Apabila nilai parameter dan nilai estimatornya baik untuk koefisen regresi maupun variansi dibandingkan, maka terlihat bahwa algoritma bootstrap dapat bekerja dengan "baik" pada data sintesis.

Data Riil

Indeks Tendensi Bisnis dan Indeks Tendensi Konsumen

Tabel 2 menunjukkan data indeks tendensi bisnis (y) dan indeks tendensi konsumen (x) di Indonesia untuk periode triwulan dari Januari 2000 sampai dengan Desember 2009

Tabel 2. Indeks Tendensi Bisnis (ITB) dan Indeks tendensi Konsumen (ITK)

Tahun	Triwulan	ITB	ITK
2000	II	122.50	113.29
	III	117.44	108.04
	IV	116.06	114.23
2001	I	107.73	110.52
	II	111.75	104.10
	III	105.36	119.21
	IV	101.03	125.19
2002	I	100.03	113.75
	II	113.38	116.65
	III	108.77	119.96
	IV	102.37	120.28
2003	I	95.78	105.87
	II	105.16	117.28
	III	111.41	114.17
	IV	114.13	121.73

Tabel 2. Indeks Tendensi Bisnis (ITB) dan Indeks tendensi Konsumen (ITK) (lanjutan)

Tahun	Triwulan	ITB	ITK
2004	I	104.35	115.20
	II	113.74	112.30
	III	114.12	120.22
	IV	115.03	109.96
2005	I	98.93	96.72
	II	106.31	98.68

	III	105.7	93.20
	IV	98.45	94.43
2006	I	95.12	96.01
	II	108.5	109.77
	III	108.72	109.16
	IV	107.43	106.96
2007	I	100.19	106.93
	II	110.96	105.78
	III	112.58	109.48
	IV	112.25	106.10
2008	I	104.41	95.01
	II	111.72	93.84
	III	111.12	102.78
	IV	102.19	100.93
2009	I	96.91	102.15
	II	110.43	106.42
	III	112.86	107.79
	IV	108.45	108.76

Berdasarkan data riil tersebut, nilai statistik $\hat{C}_p$ dihitung untuk p = 1, 2, ..., 10. Hasilnya disajikan dalam Tabel 3 berikut.

Tabel 3. Nilai statistik $\hat{C}_p$ untuk p =1,2, ...,10

p	$\hat{C}_p$
1	39.9206
2	38.2947
3	40.0926
4	42.2993
5	45.1948
6	48.9944
7	55.0465
8	65.1955
9	83.1599
10	113.9575

Dari Tabel 3, nampak bahwa nilai $\hat{C}_p$ terkecil dicapai pada p = 2. Dengan demikian, model regresi polinoial terbaik adalah orde 2.

Selanjutnya mengestimasi koefisien regresi polinomial orde 2 dan variansi σ^2 yang bersesuaian dengan menggunakan algoritma bootstrap. Hasilnya adalah $\hat{a}$=(-182.9973, 5.1756, -0.0229)' dan $\hat{\sigma}^2$ = 34.6350.

Model yang diperoleh dapat digunakan untuk memprediksi ITB apabila ITK diketahui. Sebagai contoh untuk ITK = 108.76, nilai ramalan untuk ITB disajikan pada Tabel 4

Tabel 4. Prediksi Titik dan Interval Bootstrap untuk y

$\hat{y}$	Interval Kepercayaan 95% untuk $\hat{y}$
108.7309	(106.4970 , 110.9655)

Laju Inflasi dan Indeks Harga Konsumen

Tabel 5 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 5. 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

Tabel 5. Laju Inflasi dan Indeks Harga Konsumen (Sumber : http://www.bps.go.id)

Berdasarkan data riil tersebut, nilai statistik $\hat{C}_p$ dihitung untuk p = 1, 2, ..., 10. Hasilnya disajikan dalam tabel berikut. Hasilnya disajikan dalam Tabel 6 berikut.

Tabel 6. Nilai statistik $\hat{C}_p$ untuk p =1,2, ...,10

p	$\hat{C}_p$
1	0.1189
2	0.1284
3	0.1421
4	0.1579
5	0.1755
6	0.1991
7	0.2225
8	0.2491
9	0.2832
10	0.3246

Dari Tabel 6, nampak bahwa nilai $\hat{C}_p$ terkecil dicapai pada p = 1. Dengan demikian, model regresi polinoial terbaik adalah orde 1 atau regresi linier. Selanjutnya mengestimasi koefisien regresi polinomial orde 1 dan variansi σ^2 yang bersesuaian dengan menggunakan algoritma bootstrap. Hasilnya adalah $\hat{a}$=(-3.6966, 0.0286)' dan $\hat{\sigma}^2$=0.1082.

Model yang diperoleh digunakan untuk memprediksi laju inflasi apabila IHK = 153,53. Hasilnya disajikan pada Tabel 7

Tabel 7. Prediksi Titik dan Interval Bootstrap untuk y

$\hat{y}$	Interval Kepercayaan 95% untuk $\hat{y}$
0.7003	(0.4586 , 0.9469)

4.4 KESIMPULAN

Uraian di atas menunjukkan bahwa betapa sederhananya algoritma bootstrap dapat digunakan untuk menghasilkan taksiran parameter/koefisien hubungan dalam model 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 : indeks tendensi bisnis vs indeks tendensi konsumen dan laju inflasi vs indeks harga konsumsi. Untuk data indeks tendensi bisnis vs indeks tendensi konsumen, model pada persamaan (1) dapat diselesaikan dengan menggunakan prosedur algoritma bootstrap dan menghasilkan

$$\hat{y}_t = -182.9973 + 5.1756 x_t - 0.0029 x_t^2$$

Untuk data indeks tendensi bisnis vs indeks tendensi konsumen, model pada persamaan (1) dapat diselesaikan dengan menggunakan prosedur algoritma bootstrap dan menghasilkan

$$\hat{y}_t = -3.6966 + 0.0286 x_t$$

Taksiran ini sangat bermanfaat bagi pengambilan keputusan, misalnya untuk memprediksi indeks tendensi bisnis.

Meskipun dalam artikel ini hanya dibahas model regresi polinomial, tetapi algoritma bootstrap dapat diterapkan juga pada model-model ekonometrik yang lainnya (Suparman, 2007). 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).

PROGRAM KOMPUTER

1. Listing instruksi dalam MATLAB untuk membuat data sistesis

```
clear all
clc

x=normrnd(2,3,1000,1);
p=2; b=normrnd(0,1,p+1,1);
sigma = 3;

n = length(x);
mx = zeros(n,p+1);
for j = 1:p+1,
mx(:,j) = x.^(j-1);
end;
y=(b'*mx')'+normrnd(0,sigma,n,1);

plot(x,y,'.'); xlabel('x','FontSize',14); ylabel('y','FontSize',14);
```

2. Listing instruksi dalam MATLAB untuk memghitung estimator dari a dan σ^2

```
function [beta_topi,var_topi] = lse(y,x,p);

n = length(x);
mx = zeros(n,p+1);
for j = 1:p+1,
mx(:,j) = x.^(j-1);
end;
beta_topi = pinv(mx'*mx)*mx'*y;
var_topi=sum((y-(beta_topi'*mx')').^2)/(n-p-1);
```

3. Listing instruksi dalam MATLAB untuk memghitung estimator bootstrap

```
function [bb,y_boot,ba,beta_boot,var_boot,cp_boot] = bootsrap(y,x,p);

[beta_topi_0,var_topi] = lse(y,x,p);
n=length(x);
mx = zeros(n,p+1);
for j = 1:p+1,
mx(:,j) = x.^(j-1);
end;
galat=(y-(beta_topi_0'*mx')');

B=2000; alpa=0.05;
matriks_beta=zeros(p+1,B);
matriks_var=zeros(1,B);
matriks_y=zeros(1,B);
matriks_cp=zeros(1,B);
for b=1:B,
galat_bintang=galat(unidrnd(n,n,1));
mx = zeros(n,p+1);
for j = 1:p+1,
   mx(:,j) = x.^(j-1);
end;
y_bintang=(beta_topi_0'*mx')'+galat_bintang;
[beta_topi,var_topi] = lse(y_bintang,x,p);
matriks_beta(:,b)=beta_topi;
```

```
matriks_var(:,b)=var_topi;
matriks_y(:,b)=(beta_topi'*mx(n,:)')';
matriks_cp(:,b)=sum(galat_bintang.^2)/n+2*p*var_topi/n;
end;
beta_boot=mean(matriks_beta,2);
var_boot=mean(matriks_var,2);
y_boot=mean(matriks_y,2);
cp_boot=mean(matriks_cp,2);

y_terurut=sort(matriks_y(:,:));
bb=y_terurut(B*alpa);
ba=y_terurut(B*(1-alpa));
```

BAB 5

METODE BOOTSTRAP DALAM PENGUJIAN HIPOTESIS MENGENAI DUA MEAN POPULASI

5.1 RUMUSAN MASALAH

Dalam kegiatan penelitian ilmiah, banyak perhatian dicurahkan untuk menjawab pertanyaan tentang kebenaran atau kesalahan hipotesis mengenai suatu parameter populasi. Apakah suatu metode pembelajaran baru akan lebih efektif dibandingkan dengan metode pembelajaran konvensional ? Apakah lama belajar berpengaruh terhadap hasil belajar ? Apakah motivasi belajar berhubungan dengan prestasi belajar ?

Jika parameter populasi dinotasikan dengan θ maka pengujian hipotesis mengenai θ akan dirumuskan menggunakan istilah hipotesis nol H_0 dan hipotesis alternatif H_1. Jika Ω merupakan ruang parameter, maka hipotesis nol berkaitan dengan Ω_0 yaitu himpunan bagian dari Ω. Hipotesis alternatif berkaitan dengan komplemennya, $\Omega - \Omega_0$. Dalam kasus hipotesis sederhana dengan $\Omega = \{\theta_0, \theta_1\}$, himpunan Ω_0 dan himpunan komplemennya mempunyai satu anggota saja, $\Omega_0 = \{\theta_0\}$ dan $\Omega - \Omega_0 = \{\theta_1\}$, di mana $\theta_0 \neq \theta_1$.

Pengujian hipotesis akan mengacu pada proses untuk memutuskan kebenaran atau kesalahan hipotesis tersebut berdasarkan data sampel yang diambil dari populasi. Dengan kata lain, penerimaan atau penolakan hipotesis nol didasarkan pada data sampel. Statistik penguji yang sesuai dengan hipotesis akan membagi daerah di bawah kurva distribusi samplingnya menjadi dua daerah, yaitu daerah kritis dan daerah penerimaan. Jika nilai statistik penguji dari data sampel berada di daerah kritis, maka H_0 akan ditolak. Sebaliknya, jika nilai statistik penguji dari data sampel tidak berada di daerah kritis, maka H_0 akan diterima.

Sebagian besar prosedur pengujian hipotesis statistik yang telah dikembangkan sejauh ini dibangun dengan asumsi bahwa populasi didistribusikan menurut distribusi normal. Padahal kenyataannya seringkali asumsi nornalitas tidak dipenuhi. Tujuan artikel ini adalah untuk mengembangkan pengujian hipotesis mengenai dua mean populasi yang tidak dipenuhinya asumsi normalitas dan homogenitas.

5.2 METODE PENELITIAN

Penelitian dimulai dengan mengkaji berbagai pustaka terkait pengujian hipotesis secara parametrik mengenai dua mean populasi, metode bootstrap dan pengujian hipotesis mengenai dua mean populasi dengan metode bootstrap. Berdasarkan teori yang dihasilkan dari berbagai kajian pustaka tersebut, selanjutnya dibuat program komputasinya dengan menggunakan MATLAB. Program komputer digunakan untuk menguji hipotesis mengenai dua mean populasi.

Pengujian Hipotesis Secara Parametrik

Misalkan dua sampel yang saling bebas diambil dari dua populasi normal yang berbeda. Misalkan $z = (z_1, z_2. \cdots, z_n)$ merupakan suatu sampel random berukuran n yang dimbil dari populasi 1 dengan mean μ_1 dan variansi σ_1^2. Juga misalkan $y = (y_1, y_2. \cdots, y_m)$ merupakan suatu sampel random berukuran m yang diambil dari populasi 2 dengan mean μ_2 dan variansi σ_2^2. Jika dipenuhi dua asumsi, yaitu kedua populasi berdistribusi normal dan kedua variansi sama $\sigma_1^2 = \sigma_2^2 = \sigma^2$ tidak diketahui, maka untuk menguji hipotesis

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

digunakan statistik penguji :

$$t_0 = \frac{\bar{z} - \bar{y}}{s\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

di mana

$$s = \frac{s_1^2(n-1) + s_2^2(m-1)}{n+m-2}$$

Pada taraf signifikansi α, hipotesis H_0 ditolak jika $t_0 > t_{\alpha/2}(n+m-2)$ atau $t_0 < -t_{\alpha/2}(n+m-2)$.

Jika tidak ada asumsi bahwa variansi kedua populasi adalah sama, maka uji hipoitesis didasarkan pada

$$t(x) = \frac{\bar{z} - \bar{y}}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}}$$

Namun karena distribusi sampling t(x) tidak lagi berdistribusi t, maka beberapa pendekatan diusulkan. Dalam literatur, ini dikenal sebagai masalah Behrens-Fisher (Efron and Tibshirani, 1993).

Bootstrap

Bootstrap diperkenalkan oleh Efron. Metode bootstrap adalah metode berbasis komputer untuk mengestimasi suatu distribusi dengan menggunakan sampel bootstrap. Secara formal, misalnya x = (x_1, ..., x_n) merupakan sampel random yang diambil dari populasi dengan distribusi F. Pertimbangkan statistik t(x). Salah satu tujuan utama dalam inferensi statistik adalah untuk menentukan distribusi sampling dari t(x). Jika Fn menunjukkan distribusi empiris dari x yang dimbil dengan probabilitas 1/n pada setiap x_1, ..., x_n, maka versi bootstrap dari t(x) diberikan oleh $t(x^{*b})$ di mana x^{*b} = (x_1^{*b}, x_2^{*b}, ... , x_n^{*b}) adalah sampel bootstrap ke b (b =1, 2, ..., B) yang diambil dengan pengembalian dari x = (x_1, ..., x_n) (Efron and Tibshirani, 1993).

Uji Hipotesis Bootstrap

Dua sampel random $z = (z_1, z_2. \cdots, z_n)$ dan $y = (y_1, y_2. \cdots, y_m)$ yang saling independen diambil dari dua populasi, yang mungkin berbeda. Misalkan populasi 1 mempunyai mean μ_1 dan variansi σ_1^2. Sedangkan populasi 2 mempunyai mean μ_2 dan variansi σ_2^2 . berdasarkan sampel z dan y, akan diuji hipotesis bahwa tidak ada perbedaan mean antara populasi 1 dan populasi 2. Sehingga

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Asumsi variansi sama merupakan asumsi yang sangat penting untuk uji t karena menyederhanakan bentuk distribusi sampling yang dihasilkan. Dalam mempertimbangkan pengujian hipotesis dengan menggunakan metode bootstrap tidak ada alasan kuat untuk menganggap bahwa variansi kedua populasi sama. Oleh karena itu di sini tidak diasumsikan bahwa variansi kedua populasi adalah sama. Sehingga uji hipotesis didasarkan pada statistik penguji

$$t(x) = \frac{\bar{z} - \bar{y}}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}}$$

Kemudian dihitung nilai statistik penguji untuk tiap sampel bootstrap (b = 1,2, …, B)

$$t(x^{*b}) = \frac{\bar{z}^{*b} - \bar{y}^{*b}}{\sqrt{\frac{s_1^{2*b}}{n} + \frac{s_2^{2*b}}{m}}} \quad .$$

Dan nilai dari *Achieved Significance Level* (ASL) bootstrap diperoleh dengan rumus berikut :

$$\hat{ASL}_{boot} = \frac{\#\left\{\left|t(x^{*b})\right| \geq \left|t(x)\right|\right\}}{B}$$

Pada taraf signifikansi α, jika $\hat{ASL}_{boot} < \alpha$ maka H_0 ditolak.

Prosedur pengujian ini disajikan sebagai berikut :

1. Misalkan $\hat{F}$ menempatkan probabilitas yang sama pada titik $\tilde{z}_i = z_i - \bar{z} + \bar{x}$ untuk i = 1,2, …, n dan $\hat{G}$ menempatkan probabilitas yang sama pada titik $\tilde{y}_i = y_i - \bar{y} + \bar{x}$ untuk i=1,2, …m, di mana $\bar{z}$ dan $\bar{y}$ adalah mean masing-masing sampel dan $\bar{x}$ adalah mean dari sampel gabungan.
2. Bentuk B himpunan data bootstrap (z^{*b}, y^{*b}) di mana z^{*b} adalah sampel dengan penggantian dari $\tilde{z}_1, \tilde{z}_2, \cdots, \tilde{z}_n$ dan y^{*b} adalah sampel dengan penggantian dari $\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_m$.
3. Mengevaluasi pada setiap set data
$$t(x^{*b}) = \frac{\bar{z}^{*b} - \bar{y}^{*b}}{\sqrt{\frac{s_1^{2*b}}{n} + \frac{s_2^{2*b}}{m}}}$$
4. Menaksir nilai ASL_{boot} dengan menggunakan persamaan
$$\hat{ASL}_{boot} = \frac{\#\left\{\left|t(x^{*b})\right| \geq \left|t(x)\right|\right\}}{B}$$

Berikut merupakan listing program yang ditulis dalam instruksi bahasa pemrograman MATLAB untuk pengujian hipotesis mengenai mean dua populasi :

```
clear all
clc

alpha = 0.05
z=normrnd(2,2,n,1);
y=normrnd(4,3,m,1);
B=1000

n=length(z); zbar=mean(z);
m=length(y); ybar=mean(y);
xbar=((n*zbar)+(m*ybar))/(n+m);
ztilda=z-zbar+xbar; ytilda=y-ybar+xbar;

tobs=(mean(z)-mean(y))
    /(sqrt(var(z)/n+var(y)/m));

zboot=zeros(n,B); yboot=zeros(m,B);
tboot=zeros(1,B);
nt=0;
for i=1:B,
   b1=randi(n,n,1); b2=randi(m,m,1);
   zboot(:,i)=ztilda(b1);
   yboot(:,i)=ytilda(b2);
   tboot(i)=(mean(zboot(:,i))
         -mean(yboot(:,i)))
         /sqrt(var(zboot(:,i))
         /n+var(yboot(:,i))/m);
   if abs(tboot(i))>=abs(tobs)
      nt=nt+1;
   else
      nt=nt;
   end;
end;
aslboot=nt/B

if aslboot<=alpha
   disp('Tolak hipotesis nol')
else
   disp('Terima hipotesis nol')
end;
```

5.3 HASIL DAN PEMBAHASAN

Sebagai ilustrasi, kita akan menerapkan metode bootstrap untuk menguji hipotesis mengenai dua mean populasi pada data sintesis (studi simulasi) dan data riil (studi kasus). Studi simulasi (Law and Kelton, 2000) ditempuh untuk mengkonfirmasi kinerja dari pendekatan yang diusulkan apakah dapat bekerja dengan baik. Sedangkan studi kasus diberikan untuk memberikan contoh penerapan penelitian dalam memecahkan permasalahan dalam kehidupan sehari-hari. Komputasi ditulis dalam bahasa pemrograman MATLAB (Hanselman and Littlefield, 1997).

Data Sintesis

Tabel 1 menunjukkan data simulasi sampel 1 diambil dari populasi berdistribusi normal dengan mean 0 dan variansi 16. Sedangkan sampel 2 diambil dari populasi berdistribusi normal dengan mean 5 dan variansi 25.

Tabel 1 : Data sintesis

z	y
-1.6665	1.5244
-0.9786	12.7383
-2.3251	7.2338
4.7958	-0.3053
-7.6532	4.5957
-0.8577	6.2931
1.1777	18.9566
-2.3565	3.7741
-1.3786	6.3701
2.9073	-0.1921
7.855	5.2916
2.9077	5.3109
-3.8819	1.8309
2.351	-1.7485
5.983	1.4346
-4.4519	
3.3374	
4.3556	
8.6611	
1.1849	

3.8737
0.651
6.0478
4.6373
-4.4509

Kedua sampel digunakan untuk menguji hipotesis

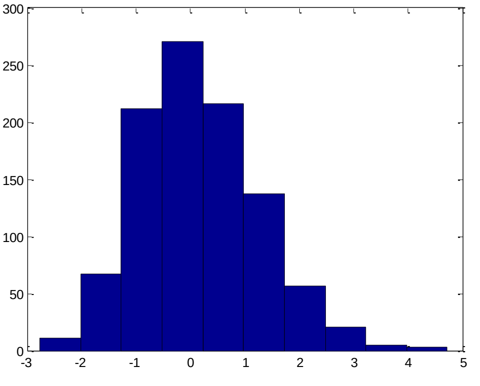
$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Tabel 2 menyajikan hasil pengujian hipotesis untuk nilai $\alpha = 0.05$ dan berbagai nilai B.

Tabel 2 : Hasil pengujian hipotesis mengenai dua mean populasi sintesis

B	$\hat{ASL}_{boot}$	Kesimpulan
1000	0.0440	Tolak H_0
5000	0.0436	Tolak H_0
10000	0.0417	Tolak H_0
20000	0.0428	Tolak H_0



T

Gambar 1 : Nilai statistik penguji untuk 1000 sampel bootstrap.

Jadi terdapat perbedaan antara mean populasi 1 dengan mean populasi 2. Dari data simulasi terlihat bahwa uji hipotesis bootstrap memberikan keputusan yang benar.

Data Riil

Tabel 3 menunjukkan data riil (Rokhmah, 2012). Sampel 1 menyatakan nilai tes hasil belajar matematika kelas yang menggunakan strategi pembelajaran aktif *Index Card Match* (ICM). Sedangkan sampel 2 menyatakan nilai tes hasil belajar matematika kelas yang menggunakan strategi pembelajaran aktif *Team Quiz* (TQ).

Kedua sampel digunakan untuk menguji hipotesis

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Tabel 4 menyajikan hasil pengujian hipotesis untuk nilai $\alpha = 0.05$ dan berbagai nilai B.

Tabel 4 : Hasil pengujian hipotesis mengenai dua mean populasi riil

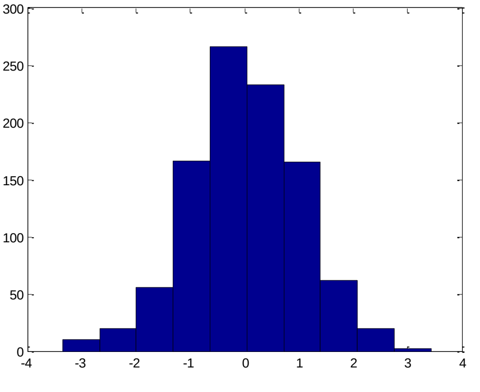
B	$\hat{ASL}_{boot}$	Kesimpulan
1000	0.0040	Tolak H_0
5000	0.0048	Tolak H_0
10000	0.0059	Tolak H_0
20000	0.0060	Tolak H_0

Tabel 3 : Nilai tes hasil belajar matematika dengan strategi ICM dan TQ

ICM	TQ
62	50
44	62
50	25
56	62
81	69
62	44
31	44
50	31

56	62
75	81
69	69
88	75
75	75
75	50
81	50
50	56
81	44
75	31
69	25
75	50
69	56
75	50
81	62
50	75
62	62
75	75
69	38
56	31
	69

Sedangkan hasil nilai statistik penguji untuk 1000 sampel bootstrap disajikan dalam Gambar 2.



Gambar 2 : Nilai statistik penguji untuk 1000 sampel bootstrap.

Jadi terdapat perbedaan antara mean nilai tes hasil belajar matematika kelas yang menggunakan strategi pembelajaran aktif ICM dengan nilai tes hasil belajar matematika kelas yang menggunakan strategi pembelajaran aktif TQ.

5.4 KESIMPULAN

Dalam artikel ini dikembangkan uji hipotesis mengenai dua mean populasi yang tidak memerlukan prasarat normalitas dan homogenitas. Distribusi sampling dari statistik uji bootstrap tidak berdistribusi t.

Metode bootstrap dapat menguji kesamaan dua mean populasi baik jika kedua variansinya diketahui sama, maupun jika mungkin kedua variansinya tidak sama. Dalam kasus prasarat normalitas dan homogenitas dipenuhi sehingga memungkinkan digunakan uji t, maka metode bootstrap ini akan memberikan suatu alternatif untuk pengujian hipotesis mengenai dua mean populasi.

DAFTAR PUSTAKA

Allison, P.D. (1999) Multiple Regression : A Primer, Pine Forge Press, California.

Bain, L.J. and Engelhardt, M., 1992, *Introduction to Probability and mathematical statistics*, California : Duxbury Press.

Baltagi, B.H., 2008, "*Econometric"*, Berlin : Springer-Verlag.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C., 1994, "*Time Series Analysis : Forecasting and Control"*, Prentice Hall, New Jersey.

Brockwell, P.J. and Davis, R.A., 1991, *Times Series : Theory and Methods*, Springer, New York.

Black, K., 2009, "Business Statistics : Contempory Decision Making", John Wiley & Sons, USA.

Devoire, J.L., 2008, "Probability and Statistics for Engineering and The Sciences", Thomson Higher Education, USA

Dupont, W.D. (2002) Statistical Modeling for Biometrical Resercher : A Simple Introduction to The Analysis of Complex Data, Cambridge University Press, United Kingdom.

Efron, B. and Tibshirani, R., 1993, "*An Introduction to the Bootstrap"*, New York : Chapman & Hall.

Freund, R.J. and Wilson, W.J. (1998) Regression Analysis Statistical Modeling of a Response Variable, Academic Press, United States of America.

Gentle, J.E. (2002) Elements of Computational Statistics, Springer-Verlag, New York.

Greene, W.H., 2003, "*Econometric Analysis"*, New Delhi : Pearson Education.

Gujarati, D., 1978, "*Ekonometrika Dasar"*, Erlangga, Jakarta.

Hanselman, D and Littlefield, B., 1997, "*Matlab : Bahasa Komputasi Teknis"*, Yogyakarta : Andi.

Heij, C, Paul de Boer, Frenser, P.H, Kloek, T and Dijk, H.K.V. (2004) Econometric Methods with Applications in Business and Economics, Oxford University Press, New York.

Johnston, J., 1972, "*Econometric Methods"*, McGraw-Hill, New York

Law, A.M. and Kelton, W.D., 2000, "*Simulation Modeling and Analysis"*, Singapore : McGraw-Hill.

Lee, S and Kim, S.H., 2006, "A Lag Effect of IT Investment on Firm Perfomance" , *Information Resources Management Journal*, **19** (1), 43-70.

Leeflang, P.S.H., Wittink, D.R., Wedel, M., and Naert, P.A., 2000, "*Building Models for Marketing Decisions*", Netherlands : Kluwer Academic.

Mackinnon, J.G., 2006, "Bootstrap Methods in Econometrics", The economic record, Volume 82.

Munandar, A., Fajriyah, R dan Suparman, 2008, Bootstrap dengan S-plus dalam Uji Hipotesis Mean Satu Sampel. *Jurnal Eksakta*, 10(1) : 36-46.

Rokhmah, N. 2012, *Efektivitas Penggunaan Strategi Pembelajaran Aktif Index Card Match dan Team Quiz Terhadap Hasil Belajar Matematika Siswa Kelas VII Semester Genap SMP Muhammadiyah 13 Wonosegoro Kabupaten Boyolali Tahun Pelajaran 2011/2012*, Yogyakarta : Skripsi Pendidikan Matematika UAD.

Soetharaman, P.B., 2004, "Modeling Multiple Sources of State Dependence in Random Utility Models : A Distribution Lag Approach", *Marketing Science Journal*, **23** (3), 263.

Sen, A and Srivastawa, M. (1990) Regression Analysis Theory, Methods and Applications, Springer-Verlag, New York.

Suparman (1997) Algoritma Bootstrap dalam Inferensi Model Ekonometrik dan Aplikasinya untuk Memprediksikan Beberapa Indikator Ekonomi, *Jurnal Teknologi Informasi dan Bisnis*, Vol. 8, No. 2, hal. 125-134.

Suparman (2007) Algoritma Bootstrap dalam Inferensi Model Ekonometrik dan Aplikasinya untuk Memprediksikan Beberapa Indikator Ekonomi, Jurnal Teknologi Informasi dan Bisnis, Vol. 8, No. 2, hal 125-149.

Suparman, 2012, *Statistika Matematika*, Yogyakarta : FMIPA UAD Press.

Walpole, R.E dan Myers, R.H., 1995, Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan, Bandung : ITB Press.

Wooldridge, J.M., 2009, "*Introduction Econometric : A Modern Approach*", United States of America : Cengage Learning.

www.bi.go.id.

www.bps.go.id/
www.finance.yahoo.com.

Tentang Penulis

Dr. Suparman, M.Si., DEA dilahirkan di Bantul Yogyakarta. Menyelesaikan S1 Pendidikan Matematika di Universitas Lampung tahun 1992. S2 Matematika di Universitas Gadjah Mada Yogyakarta diselesaikan tahun 1997. Menyelesaikan S2 Matematika Terapan di Universitas Toulouse III Perancis tahun 2000. S3 Matematika Terapan di Universitas Toulouse III Perancis diselesaikan tahun 2003.

Menjabat Dekan Fakultas Sains dan Teknologi Universitas Teknologi Yogyakarta dari 2006 s/d 2010. Menjabat Wakil Rektor Bidang Kemahasiswaan Universitas Teknologi Yogyakarta dari 5 Juli 2010 s/d 31 Januari 2011. Saat ini, sebagai Dosen Tetap Universitas Ahmad Dahlan Yogyakarta.

Metode Bootstrap dan Aplikasinya

Buku ini disusun berdasarkan penelitian dan pengajaran yang penulis lakukan selama lima tahun. Di samping itu, penulis juga telah mengkaji berbagai literatur dan hasil penelitian. Buku ini ditulis sebagai buku referensi untuk para peneliti, para pengajar di Perguruan Tinggi dan para mahasiswa dalam memahami Metode Bootstrap. Dalam buku ini disertakan beberapa aplikasinya sehingga membuat buku ini cocok juga untuk para praktisi yang ingin memahami Metode Bootstrap dan permasalahan yang bisa diselesaikan dengan metode ini.

Dalam buku ini dibahas berbagai aplikasi, yaitu :

- ☑ Inferensi dalam Model ekonometrik
- ☑ Inferensi dalam Model Lag yang Didistribusikan
- ☑ Inferensi dalam Model Regresi Ganda
- ☑ Inferensi dalam Model Regresi Polinomial
- ☑ Uji Hipotesis Mengenai Dua Mean Populasi

Dr. Suparman, M.Si., DEA

ISBN : 978-602-18282-3-6