PROPERTY DAN PERDAGANGAN SEBAGAI SEKTOR DOMINAN PADA DATA BURSA SAHAM

DENGAN Principal Component Analysis (PCA)

Oleh :

Hanna A Parhusip, Deva Widyananto[1] dan Bernadeta Desinova Kr[2]

Program Studi Statistika Matematika

Fakultas Sains dan Matematika (FSM)

Universitas Kristen Satya Wacana (UKSW) (www.uksw.edu)

[1]mahasiswa S1, matematika – FSM – UKSW

[2]mahasiswa S1, matematika – FSM – UKSW

Masalah

Data dengan variabel banyak dapat membingungkan dalam formulasi masalah
pada suatu model matematika
Oleh karena itu diperlukan cara memilih variabel yang dominan untuk mereduksi banyaknya variabel.

METODE : PCA (*Principal Component Analysis)*

Pada makalah ini dibahas tentang cara memilih sektor - sektor saham domina yang ada di Indonesia. Data diambil dari Bursa Efek Jakarta (BEJ) dan menggunakan data pada bulan Januari 2008 sampai dengan Januari 2010.

Tabel 3. Data Saham Januari 2008 – Januari 2010

Sumber : http //: bankindonesia/statistic/moneter/semil1

Tahun	Akhir Periode	Pertanian	Industri Dasar	Aneka Industri	Barang Konsumsi	Keuangan	Pertambangan	Properti	Perdagangan
	2006	1218.45	147.1	284.12	392.46	206.57	933.33	122.92	275.08
	2007	2754.76	238.05	477.35	436.04	260.57	3270.09	251.82	392.24
2008	September	1489.57	162.93	326.15	381.36	203.37	1833.24	142.42	261.33
	Oktober	738.17	112.18	199.97	321.92	151.79	1095.87	101.35	158.76
	November	803.89	114.45	215.82	320.9	150.9	897.51	105.63	137.78
	Desember	918.77	134.99	214.94	326.84	176.33	877.68	103.49	148.33
2009	Januari	969.43	126.39	246.57	337.85	161.24	922.16	96.03	147.6
	Februari	1046.64	124.08	220.41	346.16	145.95	963.89	96.56	147.9
	Maret	1094.59	134.66	287.9	352.8	172.71	1045.31	100.54	161.37
	April	1333.25	151.15	316.67	381.32	215.73	1444.46	112.32	185.56
	Mei	1576.52	182.05	362.72	433.73	227.65	1818.96	130.99	205.21
	Juni	1527	192.93	416.21	495.73	243.66	1848.54	144.79	217.84
	Juli	1659.55	222.8	504.6	591.2	272.79	2144.91	159398	250.01
	Agustus	1797.12	229.12	538.05	559.18	280.46	2140.43	157.96	259.85
	September	1784.21	238.46	584.96	597.63	300.7	2238.59	162.29	277.4
	Oktober IV	1823.1	256.83	572.19	591.66	288.32	2231.36	159.22	272.33
	November I	1733.43	260.78	533.45	595.51	293.81	2068.46	156.21	256.74
	November II	1760.16	259.53	267.64	608.09	296.37	2129.23	152.96	256.38
	November III	1786.82	263.74	578.61	631.32	298.2	2230.38	153.95	264.14
	November IV	1745.19	252.88	261.22	625.73	293.48	2129.87	143.64	248.47
	Desember I	1838.58	260.48	271.65	661.59	301.75	2248.63	147.95	255.4
	Desember II	1831.67	264.47	587.64	651.07	300.32	2211.08	146.86	260.62
	Desember III	1785.5	268.77	598.88	654.62	296.4	2152.68	144.28	267.77

Principal Component Analysis

Secara aljabar PCA merupakan suatu kombinasi linear khusus untuk *p* variabel random $X_1, \ldots, Xp$. Secara geometri, kombinasi linear menyatakan pemilihan sistem koordinat baru yang diperoleh dari merotasi sistem mula-mula $X_1, \ldots, Xp$ sebagai sumbu-sumbu koordinat. Sumbu koordinat yang baru sangat tergantung dari matriks kovariansi. (atau matrik korelasi). Matriks kovariansi pada makalah ini disimbolkan dan haruslah positif tegas (*positive definite*). Istilah ini dijelaskan pada Definisi 1. $\sum$

Definisi 1: (Peressini,1988)

Misalkan $\Sigma = (\sigma_{ij})$ sebuah matriks simetri $n \times n$ maka matriks $\Sigma = (\sigma_{ij})$ positif tegas (*definite positive*) jika dan hanya jika semua nilai eigennya positif . Untuk negatif tegas didefinisikan secara analog.

Teorema 1.

(halaman 358, Johnson and Wichern, 2007)

Sebutlah matriks ***X*** = $[X_1, \ldots, X_p]$ adalah matriks yang vektor-vektor kolomnya adalah vektor random (dianggap berdistribusi normal) yang mempunyai matriks kovariansi (simetris dan positif tegas Σ (*positive definite*)) dengan nilai eigen $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ dan sebutlah vektor eigen yang bersesuaian untuk setiap $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ adalah $\vec{e}_1, \ldots, \vec{e}_p$ yang saling ortogonal. Komponen prinsip ke-*i* adalah

$$Y_i = \vec{e}_i^{\,T} X = e_{1i} X_1 + e_{21} X_2 + \ldots + e_{pi} X_p, \quad i = 1,2,\ldots,p \qquad (1.a)$$

Dengan pemilihan ini

Var $(Y_i) = \vec{e}_i^{\,T} \Sigma \vec{e}_i = \lambda_i$, i =1,2,...,p (1.b)

Cov (Y_i, Y_k) = $\vec{e}_i^{\,T} \Sigma \vec{e}_k = 0$, $\vec{e}_i^{\,T} \Sigma \vec{e}_k = 0$. (1.c)

Bukti **:** (halaman 358, Johnson and Wichern, 2002) .

METODE PENELITIAN

- Data : data saham Januari 2008 – 2010. Data setiap vektor-vektor kolom dianggap sebagai variabel random yang berdistribusi normal.
- Menyusun matriks kovariansi
- Menghitung nilai eigen dan vektor eigen Vektor eigen sebagai penyusun koefisien pada komponen prinsip.
- Menyatakan variabel prinsip sebagai kombinasi linear dari variabel mula-mula
- Menentukan koefisisen korelasi antara komponen prinsip dan variabel mula mula dengan persamaan

$$Y_i = \vec{e}_i^{\,T} X = e_{1i} X_1 + e_{21} X_2 + \ldots + e_{pi} X_p$$

4. ANALISA DAN PEMBAHASAN

Tabel 4. Matriks Kovariansi (S)

S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
0.0251	0.0292	0.0285	0.0229	0.0250	0.0287	0.0010	0.0227
0.0292	0.0471	0.0445	0.0420	0.0402	0.0354	0.0011	0.0268
0.0285	0.0445	0.0644	0.0397	0.0382	0.0346	0.0047	0.0280
0.0229	0.0420	0.0397	0.0398	0.0363	0.0286	0.0034	0.0213
0.0250	0.0402	0.0382	0.0363	0.0356	0.0303	0.0024	0.0232
0.0287	0.0354	0.0346	0.0286	0.0303	0.0346	0.0031	0.0256
0.0010	0.0011	0.0047	0.0034	0.0024	0.0031	0.0357	0.0008
0.0227	0.0268	0.0280	0.0213	0.0232	0.0256	0.0008	0.0238

Tampak bahwa matriks kovariansi *S* adalah matriks simetris. Kita dapat mencari nilai eigen sebagaimana ditunjukkan pada Bab 2 dan dengan menggunakan bantuan MATLAB maka nilai eigen adalah

$$\lambda = \begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5 & \lambda_6 & \lambda_7 & \lambda_8 \end{bmatrix}$$
$$= \begin{bmatrix} 0.2312 & 0.0358 & 0.0191 & 0.0154 & 0.0029 & 0.0008 & 0.0006 & 0.0003 \end{bmatrix}$$

Sedangkan vektor eigen untuk tiap nilai eigen diperoleh berturut-turut ditunjukkan tiap kolom pada Tabel 5 dan dapat ditunjukkan bahwa vektor eigen tersebut saling ortonormal.

Tabel 5. Nilai Vektor eigen untuk data Tabel 3

$\vec{u}_1$	$\vec{u}_2$	$\vec{u}_3$	$\vec{u}_4$	$\vec{u}_5$	$\vec{u}_6$	$\vec{u}_7$	$\vec{u}_8$
0.4043	0.6699	0.1841	-0.1714	-0.2206	0.4298	-0.0566	0.2965
-0.6881	0.3835	-0.2473	-0.0466	0.3426	-0.0002	-0.0703	0.4402
0.0070	0.0086	0.0249	-0.0667	-0.5731	-0.6643	0.0763	0.4683
0.5947	-0.0871	-0.3955	0.1113	0.5286	-0.2029	0.0152	0.3861
-0.0354	-0.2525	0.8242	0.0760	0.3240	0.0055	-0.0212	0.3800
-0.0487	-0.5485	-0.2184	-0.5459	-0.1750	0.4399	-0.0096	0.3562
-0.0402	0.0456	0.0044	0.0324	0.0360	0.1035	0.9910	0.0336
-0.0635	-0.1726	-0.1445	0.8043	-0.2987	0.3585	-0.0563	0.2785

Oleh karena itu komponen prinsip adalah

$$Y_1 = 0.4043X_1 - 0.6881X_2 + 0.0070X_3 + 0.5947X_4 - 0.0354X_5 - 0.0487X_6 - 0.0402X_7 - 0.0635X_8$$
$$Y_2 = 0.6699X_1 + 0.3835X_2 + 0.0086X_3 - 0.0871X_4 - 0.2525X_5 - 0.5485X_6 + 0.0456X_7 - 0.1726X_8$$
$$Y_3 = 0.1841X_1 - 0.2473X_2 + 0.0249X_3 - 0.3955X_4 + 0.8242X_5 - 0.2184X_6 + 0.0044X_7 - 0.1445X_8$$
$$Y_4 = -0.1714X_1 - 0.0466X_2 - 0.0667X_3 + 0.1113X_4 + 0.0760X_5 - 0.5459X_6 - 0.0324X_7 + 0.8043X_8$$
$$Y_5 = -0.2206X_1 + 0.3426X_2 - 0.5731X_3 + 0.5286X_4 + 0.3240X_5 - 0.1750X_6 + 0.0360X_7 - 0.2987X_8$$
$$Y_6 = 0.4298X_1 - 0.0002X_2 - 0.6643X_3 - 0.2029X_4 + 0.0055X_5 + 0.4399X_6 + 0.1035X_7 + 0.3585X_8$$
$$Y_7 = -0.0566X_1 - 0.0703X_2 + 0.0763X_3 + 0.0152X_4 - 0.0212X_5 - 0.0096X_6 + 0.9910X_7 - 0.0563X_8$$
$$Y_8 = 0.2965X_1 + 0.4402X_2 + 0.4683X_3 + 0.3861X_4 + 0.3800X_5 + 0.3562X_6 + 0.0336X_7 + 0.2785X_8$$

Dapat ditunjukkan bahwa Y_i dan Y_j saling bebas linear , *i,j*=1,..,8. Oleh karena itu sebagaimana disebutkan pada Bab 2 diperoleh bahwa Cov(Y_i , Y_j) = 0.

Untuk selanjutnya korelasi antara antara komponen prinsip pertama dan variabel mula-mula berturut-turut adalah

$$\rho_{Y_1,X_1} = \frac{e_{11}\sqrt{\lambda_1}}{\sqrt{s_{11}}} = 0.0442; \qquad \rho_{Y_1,X_2} = \frac{e_{21}\sqrt{\lambda_1}}{\sqrt{s_{22}}} = 0.0419; \qquad \rho_{Y_1,X_3} = \frac{e_{31}\sqrt{\lambda_1}}{\sqrt{s_{33}}} = 0.0027$$

$$\rho_{Y_1,X_4} = \frac{e_{41}\sqrt{\lambda_1}}{\sqrt{s_{44}}} = 0.0301; \qquad \rho_{Y_1,X_5} = \frac{e_{51}\sqrt{\lambda_1}}{\sqrt{s_{55}}} = 0.2132; \qquad \rho_{Y_1,X_6} = \frac{e_{21}\sqrt{\lambda_1}}{\sqrt{s_{66}}} = 0.3266$$

$$\rho_{Y_1,X_7} = \frac{e_{21}\sqrt{\lambda_1}}{\sqrt{s_{77}}} = 0.9931; \qquad \rho_{Y_1,X_8} = \frac{e_{21}\sqrt{\lambda_1}}{\sqrt{s_{88}}} = 0.8682$$

Karena nilai korelasi variabel X_7 (properti) dan X_8 (perdagangan) dekat dengan 1, maka variabel X_7 dan X_8 sebagai variabel yang paling berpengaruh terhadap nilai saham.

Hasil PCA dengan standarisasi variabel

Hasil PCA dengan standarisasi variabel diharapkan memberikan kesimpulan yang sama tentang variabel yang dianggap dominan. Dengan menggunakan persamaan (5c) data distandarisasi. Diperoleh nilai eigen

$$\lambda^T = [\lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \lambda_4 \quad \lambda_5 \quad \lambda_6 \quad \lambda_7 \quad \lambda_8]$$

$$= [6.0037 \quad 1.0036 \quad 0.5486 \quad 0.2895 \quad 0.1072 \quad 0.0222 \quad 0.0179 \quad 0.0072].$$

Sehingga vektor eigen untuk nilai eigen terbesar adalah

$$\vec{e}_1^{\,T} = [-0.3793 \quad -0.3964 \quad -0.3482 \quad -0.3727 \quad -0.3854 \quad -0.0324 \quad -0.3664].$$

Komponen prinsip dengan variabel standard untuk nilai eigen terbesar adalah

$$Y_1 = -0.3793Z_1 - 0.3964Z_2 - 0.3482Z_3 - 0.3727Z_4 - 0.3937Z_5 - 0.3854Z_6$$
$$- 0.0324Z_7 - 0.3664Z_8 \quad .$$

Kita dapat menyusun korelasi antara Y_1 dengan Z_i, i =1,...,8 menggunakan persamaan (5h) diperoleh

ρ_{Y_1,Z_1} = -0.1548, ρ_{Y_1,Z_2} = -0.1618, ρ_{Y_1,Z_3} = -0.1421, ρ_{Y_1,Z_4} = -0.1521, ρ_{Y_1,Z_5} = -0.1607, ρ_{Y_1,Z_6} = -0.1573 , ρ_{Y_1,Z_7} = -0.0132 dan ρ_{Y_1,Z_8} = -0.1496. Kita menyimpulkan bahwa semua variabel mempunyai makna yang sama atau tidak ada yang dominan.

Dari hasil ini disimpulkan bahwa menggunakan standarisasi variabel tidak dapat menunjukkan dominasi salah satu variabel. Hal ini juga ditunjukkan pada literatur (Johnson and Wichern, 2002) . Pemilihan variabel dengan standarisasi dilakukan bila data mempunyai perbedaan yang sangat signifikan ataupun dengan satuan yang berbeda. Hal ini belum diselidiki lebih lanjut.

Kesimpulan dan Saran

Pada makalah ini ini telah ditunjukkan analisa variabel dengan menggunakan *Principal Component Analysis* untuk 8 variabel mengenai sektor – sektor yang memiliki nilai sahan cukup tinggi. Variabel tersebut adalah pertanian, industri dasar, aneka industri, barang konsumsi, keuangan, pertambangan, properti, dan perdagangan. Seluruh variabel diuji untuk mendapatkan variabel yang dominan. Diperoleh bahwa **properti dan perdagangan** adalah variabel yang dominan yang ditunjukkan dengan variansi terbesar melalui *Principal Component Analysis*.