

IMPUTASI *MISSING VALUES* DENGAN *EXPECTATION MAXIMIZATION WITH BOOTSTRAPPING (EMB)* DATA *TIME SERIES-CROSS SECTION (TSCS)* VOLUME PRODUKSI *CRUDE PALM OIL (CPO)*

Lasmiyati¹, Sutawanir Darwis², Budhi Handoko³

¹Mahasiswa Program Magister Statistika Terapan, FMIPA-UNPAD, Bandung

²Program Magister Statistika Terapan, FMIPA-UNPAD, Bandung

³Program Magister Statistika Terapan, FMIPA-UNPAD, Bandung

Abstrak

Pengumpulan data CPO merupakan hal yang cukup penting mengingat manfaatnya yang cukup besar sebagai bahan baku dalam berbagai industri. Akan tetapi, dalam pelaksanaannya terjadi kendala non respon unit pengamatan yang tidak dapat ditanggulangi yang berdampak pada terjadinya *missing values*. Untuk menangani *missing values*, salah satu metode yang dapat digunakan adalah imputasi melalui *Expectation Maximization with Bootstrapping (EMB)*. Metode ini dapat diterapkan pada data yang memiliki struktur gabungan *time series* dan *cross section (TSCS)* sebagaimana data CPO yang dikumpulkan perbulan untuk setiap perusahaan perkebunan. Dari hasil pengolahan menggunakan metode EMB dengan *package Amelia II* melalui software R, diperoleh nilai prediksi imputasi *missing values* yang cukup baik dan masuk akal. Hal tersebut dapat terlihat dari perbandingan densitas antara nilai rata-rata imputasi dan nilai yang terobservasi yang sangat mirip. Selain itu, dari *overimpute* yaitu plot antara distribusi rata-rata nilai imputasi dengan distribusi nilai yang terobservasi, berada di sekitar garis diagonal yang menunjukkan bahwa model imputasi sudah cukup baik.

Kata Kunci: EM, Bootstrap, EMB, Missing values, TSCS

1. Pendahuluan

Salah satu komoditas perkebunan yang cukup diminati di Indonesia adalah kelapa sawit. Prospeknya cukup menjanjikan mengingat tanaman kelapa sawit merupakan sumber minyak nabati yang digunakan di seluruh dunia, diantaranya digunakan sebagai bahan baku minyak goreng, sabun, kosmetika, margarin, dan sebagainya. Minyak yang berasal dari kelapa sawit terdiri dari dua macam, yaitu minyak yang berasal dari daging buah yang dikenal sebanyak minyak sawit kasar atau *Crude Palm Oil (CPO)* dan minyak yang berasal dari inti sawit yang dikenal sebagai minyak inti sawit atau *Palm Kernel Oil (PKO)*. Cukup penting untuk melakukan pengumpulan data komoditas kelapa sawit dalam hal ini CPO, mengingat manfaatnya cukup besar sebagai bahan baku dari berbagai industri yang menempatkan posisinya sebagai komoditas yang strategis.

Pengumpulan data CPO dilakukan melalui kuesioner bulanan yang dikirimkan kepada setiap perusahaan perkebunan kelapa sawit yang berada pada wilayah negara kesatuan Republik Indonesia. Setiap akhir tahun, Badan Pusat

Statistik (BPS) mengirimkan satu set kuesioner yang terdiri dari 12 kuesioner untuk bulan Januari sampai dengan bulan Desember tahun berikutnya pada setiap perusahaan perkebunan kelapa sawit. Perusahaan-perusahaan perkebunan kelapa sawit tersebut diminta untuk melaporkan luas dan produksi perusahaannya pada setiap bulan dengan mengirimkan kembali kuesioner bulanan pada setiap awal bulan berikutnya.

Dalam pelaksanaan lapangannya, banyak dari kuesioner bulanan tersebut tidak sampai kembali ke Badan Pusat Statistik (BPS) dengan berbagai alasan, sehingga berdampak pada non respon unit pengamatan. Non respon dari unit pengamatan yang tidak dapat ditanggulangi pada akhirnya mengakibatkan terjadinya *missing values*. *Missing values* ini harus diatasi mengingat pentingnya kebutuhan akan data kelapa sawit dalam hal ini CPO. Salah satu cara mengatasi *missing values* adalah dengan membuang observasi yang *missing* tersebut, akan tetapi cara demikian bukanlah solusi yang baik karena terkait dengan bias dan *inefisiensi* (Honaker dan King, 2010), sehingga diperlukan suatu metode yang tepat untuk mengatasi masalah *missing values*.

Tabel 1.1 Contoh Non Respon Unit Pengamatan.

Perusahaan (observasi)	Luas (Ha)	Volume Produksi CPO (Ton)						
		Jan	Feb	Mar	Apr	Mei	Jun	Jul
A	955	899.110	647.870	831.800	842.860	900.991	1.121.040	1.147.160
B	262,56	233.525	176.620	220.415	247.566	208.650	253.180	233.935
C	3.328,5	11.127.030	10.176.230	9.696.920	11.794.850	11.756.830	10.818.360	12.370.640
D	32,1			11.084	16.392	22.025	29.685	24.468
E	1.258	328.513	493.102	724.213	1.238.528	1.127.695	1.219.837	1.148.459

Sumber: BPS, diolah

Dari tabel 1.1 dapat dilihat bahwa terjadi non respon unit pengamatan pada bulan Januari dan Februari perusahaan D. Kejadian tersebut dapat terjadi pada sekian banyak perusahaan. Dengan demikian, akan terjadi *missing values* pada bulan-bulan tertentu pada sekian banyak perusahaan sehingga struktur data yang terbentuk adalah panel data atau gabungan *time series* dan *cross section* (TSCS) yaitu *time series* untuk runtun waktu 12 bulan serta *cross section* untuk banyaknya responden dalam hal ini adalah perusahaan perkebunan kelapa sawit. Sehingga diperlukan suatu metode untuk memprediksi nilai imputasi dalam mengatasi *missing values* yang terjadi pada data bulanan komoditas kelapa sawit dengan struktur TSCS.

2. Metode

Sebelum memilih metode imputasi, pertama kali harus diidentifikasi mekanisme *missing data* seperti yang didefinisikan oleh Rubin (1976). Terdapat tiga mekanisme *missing data* yaitu:

1) *Missing Completely at Random* (MCAR)

Mekanisme *missing data* MCAR terjadi jika peluang *missing* identik untuk seluruh unit data, mengindikasikan bahwa pola *missing* independen

dengan observasi. Sebagai contoh, jika *missing data* ditentukan oleh sebuah koin, maka mekanisme *missing data*-nya adalah MCAR.

2) *Missing at Random* (MAR)

Mekanisme *missing data* disebut MAR jika *missing data* tergantung pada variabel lain bukan pada variabel itu sendiri.

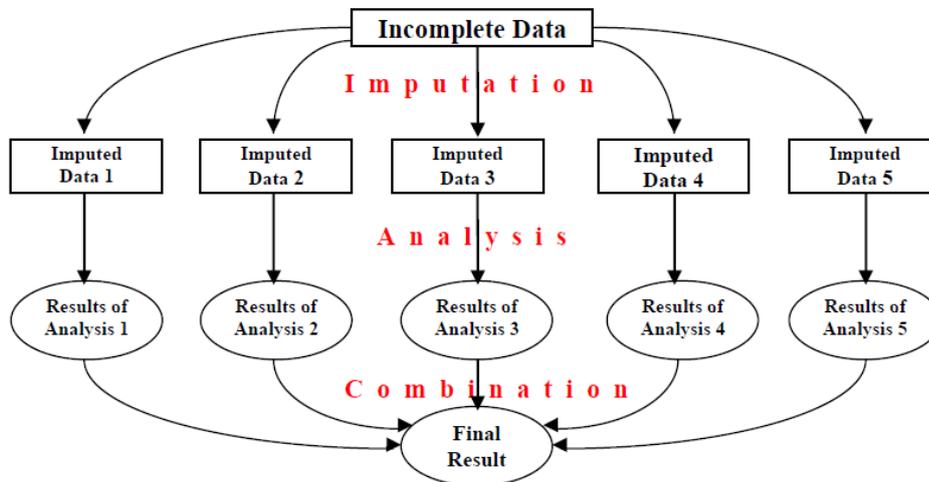
3) *Missing Nonignorable at Random* (MNAR)

Ketika mekanisme *missing data* MNAR, peluang terjadinya *missing* bergantung pada *missing data* itu sendiri, sehingga tidak dapat diprediksi karena tidak ada informasi tentang *missing values* tersebut.

Dalam penelitian ini, mekanisme *missing data* mengikuti MAR, karena *missing data* yang terjadi pada volume produksi CPO dapat diprediksi dari luas tanaman menghasilkan komoditas kelapa sawit.

Salah satu pendekatan dalam prosedur berbasis imputasi adalah *single imputation*. Dalam *single imputation*, *missing values* diisi dengan *predictive values* seperti *mean imputation*, *cold deck imputation*, *hot deck imputation*, dan *regression imputation* (Rubin, 2002 dalam Takahashi dan Ito, 2012). *Single imputation* lebih efisien bila dibandingkan dengan *listwise deletion* dengan tidak adanya unit yang dikorbankan, karena penggunaan sampel lengkap tersebut mencegah hilangnya kekuatan yang disebabkan oleh berkurangnya ukuran sampel. Jika data observasi berisi informasi yang bermanfaat untuk memprediksi *missing values*, prosedur estimasi dapat memanfaatkan informasi ini dan menjamin tingkat ketelitian yang tinggi (Schafer dan Graham, 2002). Masalah umum dalam *single imputation* adalah menempatkan kembali *missing values* yang tidak diketahui dengan nilai tunggal dan kemudian memperlakukannya sebagaimana nilai tersebut merupakan *true values* (Little dan Rubin, 1987). *Missing data* atau data yang tidak terobservasi tidak dapat dianggap sebagai *true values*, sehingga metode *single imputation* kurang cocok karena menghasilkan varians yang underestimated dibanding varians yang sebenarnya (Grittner dkk, 2011).

Dengan keterbatasan *single imputation*, maka selanjutnya dikembangkan metode *multiple imputation*. Metode *multiple imputation* berperan memberikan estimasi yang lebih valid secara statistik dengan mewujudkan ketidakpastian dengan memperluas asumsi konfidens interval, sehingga imputasi tersebut benar (Kenward dan Carpenter, 2007 dalam Grittner dkk, 2011). *Multiple imputation* memiliki sejumlah manfaat sebagai suatu pendekatan *missing data*, dengan mengisi *missing values* beberapa kali, akhirnya menghasilkan beberapa kali set data yang lengkap. *Missing values* diimputasi berdasarkan nilai observasi individu dan hubungan data observasi partisipan lain, mengasumsikan variabel observasi termasuk dalam model imputasi (Schafer dan Graham, 2002). Karena menciptakan beberapa kali prediksi untuk masing-masing *missing values*, analisis *multiple imputation* akan menghitung ketidakpastian dalam imputasi dan menghasilkan standar error yang akurat (Azur dkk, 2011).



Gambar 2.1 Langkah-Langkah *Multiple Imputation*
 Sumber: Takahashi, 2014

Contoh metode *multiple imputation* adalah MICE dan EM yang dapat dijelaskan sebagai berikut:

1) MICE

MICE dikenal juga dengan “*fully conditional specification*” atau “*sequential regression multiple imputation*” yang menciptakan *multiple imputation* untuk menghitung ketidakpastian secara statistik dalam imputasi (Azur dkk, 2011). Dalam metode ini, imputasi *missing data multivariate* dilakukan per variabel atau dapat diartikan bahwa model imputasi dilakukan secara terpisah untuk masing-masing variabel dengan *missing values* dengan memperlakukan variabel lain sebagai prediktor (Grittner dkk, 2011).

2) EM

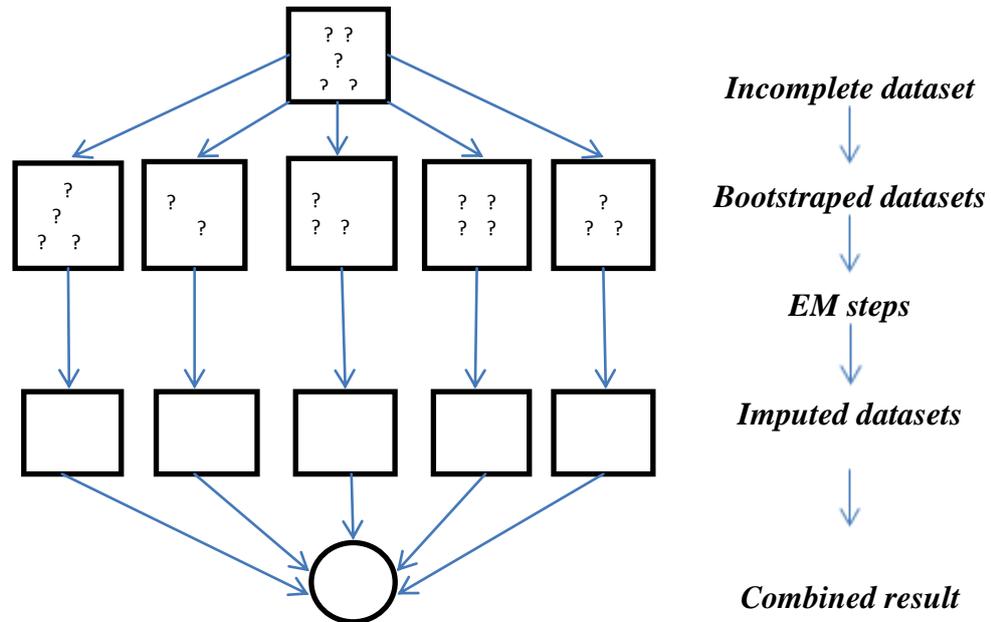
Estimasi maksimum likelihood rata-rata dan matrik kovarian dari sampel *multivariate normal* yang tidak lengkap mengasumsikan data MAR (Little dan Rubin, 1987). Maksimum likelihood menggunakan seluruh data yang tersedia, baik yang lengkap maupun yang tidak lengkap untuk mendapatkan nilai estimasi parameter yang mempunyai peluang tertinggi dalam menghasilkan data sampel.

Dalam TSCS sumber informasi yang bernilai digunakan untuk meningkatkan nilai imputasi yaitu melalui *prior bayesian*. *Prior* ini biasanya digunakan untuk analisis bayesian sebagai distribusi parameter dalam model yang mengasumsikan pengetahuan hubungan antar variabel atau distribusi marginalnya. Strategi untuk memasukkan informasi nilai awal ke dalam algoritma EM adalah dengan memasukkan informasi awal pada langkah E dan menggunakannya untuk mempengaruhi langkah M secara tidak langsung melalui pengaruhnya pada langkah E.

Untuk mengatasi *missing values* pada data TSCS dengan ukuran data yang besar sekaligus seringkali diikuti dengan jumlah variabel yang banyak, dapat dilakukan dengan *Expectation Maximization with Bootstrapping* (EMB). Dengan algoritma EMB, imputasi yang dihasilkan lebih akurat untuk data dengan struktur TSCS serta dapat memasukkan *prior information* dalam estimasinya (Honaker

dan King, 2013). Algoritma EMB juga memasukkan *bootstrap* untuk melakukan *re-sampling* set data. Pengolahan dilakukan menggunakan *package* Amelia II dengan software R.

Prosedur EMB



Gambar 2.2 *Multiple Imputation* dengan menggunakan algoritma EMB
 Sumber: Honaker dan King, 2013

Prosedur EMB dapat dijelaskan sebagai berikut:

1. *Incomplete Dataset* / Set Data yang Tidak Lengkap

Dimisalkan ada suatu matrik M dengan elemen matrik M adalah 1 ketika *missing* dan 0 ketika observasinya ada atau terobservasi. $X \equiv D$. i adalah baris ke i (unit) dan x_{ij} adalah elemen ke j (variabel) pada baris tersebut. Kemudian menciptakan $X^{obs} = X * (1 - M)$, yaitu dengan menggantikan nilai yang *missing* dengan nol, asumsikan kolom pertama dari X adalah konstan, sehingga tidak ada baris yang sepenuhnya *missing*, elemen konstan pada kolom pertama disebut x_{i0} .

2. *Bootstrapped Dataset* / Set Data *Bootstrap*

Jadi, dari suatu set data yang tidak lengkap berukuran n dilakukan *bootstrap*, yaitu penarikan subsampel berukuran n dengan pengembalian (*with replacement*) sebanyak m kali sehingga diperoleh m set data yang tidak lengkap berukuran n .

3. *EM Steps* / Langkah-Langkah EM

Langkah-langkah EM terdiri dari langkah E dan langkah M sebagai berikut:

a. Langkah E

$$E[x_{ij}] = x_i^{obs} \theta \{1 - M_i\}_j^t$$

$$E[x_{ij} x_{ik}] = E[x_{ij}] E[x_{ik}] + \theta \{1 - M_i\}_{jk}^t$$

$$\hat{x}_i^{t+1} = x_i^{obs} + M_i * (x_i^{obs} \theta \{1 - M_i\}^t)$$

Algoritma EM memasukkan informasi *single prior* dalam langkah E untuk mengikat informasi tersebut dalam estimasi parameter. Selain itu, EM dapat juga memasukkan *multiple priors*, karena *prior* dapat terjadi untuk *multiple* observasi dan *multiple missing* elemen dalam observasi yang sama.

b. Langkah M

$$Q^{t+1} = \sum_{i=1}^n (\hat{x}_i^{t+1'} \hat{x}_i^{t+1} + \Sigma_{i|x_i^{obs}}^{t+1}) = \hat{X}^{t+1'} \hat{X}^{t+1} + \sum_{i=1}^n (\Sigma_{i|x_i^{obs}}^{t+1})$$

4. *Imputed Datasets*

Melalui *running* algoritma EM pada m set data yang tidak lengkap diperoleh m set data yang lengkap dengan nilai imputasi yang beragam atau berbeda pada setiap set data untuk setiap elemen observasi yang *missing*.

5. *Combined Results*

Ada beberapa metode yang dapat digunakan untuk menggabungkan m set data lengkap hasil imputasi. Dalam penelitian ini digunakan metode yang dikembangkan oleh Rubin dan Schenker (1986) dalam Miyama dan Managi (2014). Kombinasi estimasi dari parameter $\bar{\theta}_m$ secara sederhana merupakan rata-rata aritmatika dari m estimasi.

$$\bar{\theta}_m = \frac{1}{m} \sum_{i=1}^m \hat{\theta}_i$$

Diagnosa Hasil Imputasi dapat dilakukan dengan perbandingan densitas dan *overimpute*. Perbandingan densitas dapat dilihat melalui plot garis antara densitas rata-rata imputasi dari m set data dengan distribusi dari nilai yang terobservasi (Honaker dkk, 2013). *Overimputing* memperlakukan nilai yang terobservasi seolah-olah nilai tersebut *missing* dengan membangkitkan beberapa ratus nilai imputasi. *Overimputing* dapat dilihat melalui *scatterplot* antara distribusi rata-rata nilai imputasi dengan distribusi nilai yang terobservasi (Honaker dkk, 2013).

3. Hasil dan Pembahasan

Data yang digunakan adalah data bulanan volume produksi CPO dan luas tanaman menghasilkan komoditas kelapa sawit pada tahun 2013. Data tersebut diperoleh dari Survei Perusahaan Perkebunan Bulanan Komoditas Kelapa Sawit. Terdapat 226 perusahaan perkebunan besar swasta yang tersebar di beberapa propinsi di Pulau Sumatera, Kalimantan, dan sebagian Sulawesi. Imputasi *missing values* dilakukan pada variabel volume produksi CPO dengan variabel prediktor luas tanaman menghasilkan komoditas kelapa sawit.

Langkah-langkah yang dilakukan dalam prosedur EMB melalui *package* Amelia II melalui software R adalah sebagai berikut:

1. Dibentuk suatu set data yang tidak lengkap dari 226 perusahaan perkebunan besar swasta dalam kurun waktu 12 bulan pada tahun 2013. Sehingga struktur data yang terbentuk adalah gabungan data *time series* dan *cross section* (TSCS). Dari eksplorasi data, diketahui terdapat 883 observasi *missing* dari 2712 total observasi atau *missing values* yang terjadi sekitar 32,5 persen dari total observasi

2. Selanjutnya dilakukan pengambilan sampel *bootstrap*. Dalam hal ini, diambil sampel sebesar 2712 dengan pengembalian dari set data berukuran 2712 sebanyak 5 kali. Sehingga diperoleh 5 set data yang tidak lengkap berukuran 2712.
3. Kemudian pada setiap set data dilakukan *running* algoritma EM untuk mendapatkan prediksi nilai imputasi observasi yang *missing*. *Prior information* yang yang dalam hal ini adalah μ dan Σ dimasukkan ke dalam langkah E agar secara tidak langsung mempengaruhi langkah M di mana pada langkah tersebut parameter diperbarui. Misalkan terdapat *missing value* pada observasi perusahaan b bulan ke 8, maka *prior information* diambil dari μ dan Σ dari observasi perusahaan b dengan *series* yang terobservasi (bulan 1-7 dan bulan 9-12) yaitu $\mu = 236.447$ dan $\Sigma = 32.914$. *Prior Information* dapat dimasukkan pada setiap observasi yang *missing*.
4. Dari *running* algoritma EM yang dilakukan pada ke 5 set data *bootstrap*, diperoleh 5 set data lengkap dengan prediksinilai imputasi yang berbeda-beda untuk setiap *missing values*. Hal tersebut menunjukkan bahwa multiple imputation mewujudkan ketidaktentuan dalam imputasi.

Tabel 3.1 Tabel Hasil Prediksi Imputasi (Bold) Data Volume Produksi CPO 2013.

bulan	id	luas	Imputasi				
			1	2	3	4	5
			produksi	produksi	produksi	produksi	produksi
1	a	955	899.110	899.110	899.110	899.110	899.110
2	a	955	647.870	647.870	647.870	647.870	647.870
.
6	b	262.56	253.180	253.180	253.180	253.180	253.180
7	b	262.56	233.935	233.935	233.935	233.935	233.935
8	b	262.56	207.624,1788	191.480,8655	258.322,8343	255.809,7853	215.487,9494
9	b	262.56	300.720	300.720	300.720	300.720	300.720
10	b	262.56	227.050	227.050	227.050	227.050	227.050
.
.
11	c	3.328.52	10.780.270	10.780.270	10.780.270	10.780.270	10.780.270
12	c	3.328.52	12.969.590	12.969.590	12.969.590	12.969.590	12.969.590
1	d	32.1	24.939,00727	9.592,154241	43.576,53241	59.750,85075	44.483,98966
2	d	32.1	47.794,78882	44.389,87651	13.848,52242	40.105,59559	51.031,77079
3	d	32.1	11.084	11.084	11.084	11.084	11.084
4	d	32.1	16.392	16.392	16.392	16.392	16.392
5	d	32.1	22.025	22.025	22.025	22.025	22.025
.

Sumber: Hasil Pengolahan Data Survei Perusahaan Perkebunan Komoditas Kelapa Sawit.

5. Berikutnya dilakukan kombinasi dari kelima hasil imputasi untuk memperoleh prediksi nilai imputasi tunggal dengan menghitung rata-rata dari kelima hasil prediksi nilai imputasi *missing values*.

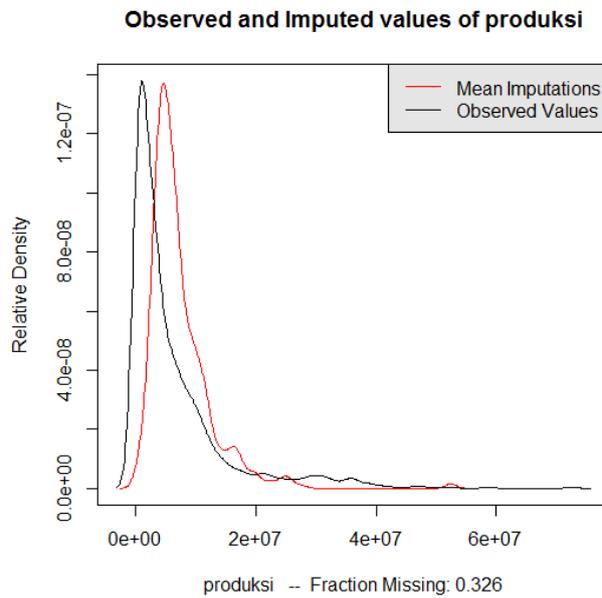
Tabel 3.2 Hasil Kombinasi Rata-Rata 5 Prediksi Nilai Imputasi *Missing Value*(Bold) Data Volume Produksi CPO 2013

Bulan	Id	Luas	Rata-rata Produksi
1	a	955	899.110
2	a	955	647.870
.	.	.	.
6	b	262.56	253.180
7	b	262.56	233.935
8	b	262.56	225.745,1
9	b	262.56	300.720
10	b	262.56	227.050
.	.	.	.
.	.	.	.
11	c	3.328.52	10.780.270
12	c	3.328.52	12.969.590
1	d	32.1	36.468,51
2	d	32.1	39.434,11
3	d	32.1	11.084
4	d	32.1	16.392
5	d	32.1	22.025
.	.	.	.

Sumber: Hasil Pengolahan Data Survei Perusahaan Perkebunan Komoditas Kelapa Sawit.

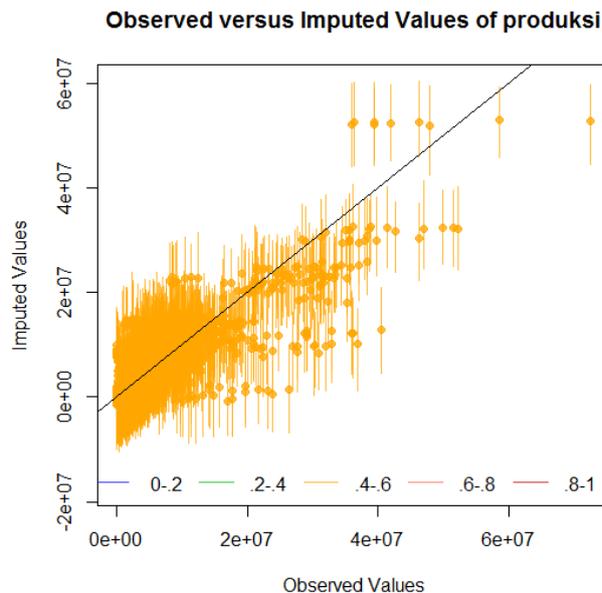
Dari rata-rata prediksi imputasi *missing values* pada bulan Agustus perusahaan b dan bulan Januari serta Februari perusahaan d, dapat dilihat bahwa nilai rata-rata prediksi imputasi yang diperoleh berada dalam selisih nilai yang tidak terlalu jauh bila dibandingkan dengan nilai yang terobservasi pada bulan-bulan yang lain untuk perusahaan yang sama.

Selanjutnya dilakukan diagnosa hasil imputasi yang dilakukan dengan perbandingan densitas dan *overimpute*. Dari plot garis antara densitas rata-rata imputasi dari m set data dengan distribusi dari nilai yang terobservasi (Honaker dkk, 2013) dapat dilihat bahwa densitas rata-rata imputasi dengan distribusi nilai yang terobservasi hampir mirip. Dengan demikian dapat dikatakan bahwa model imputasi sudah masuk akal.



Gambar 3.1 Plot Densitas antara Rata-Rata Imputasi dengan Distribusi Nilai yang Terobservasi

Berikutnya, dari *overimpute* atau *scatterplot* antara distribusi rata-rata nilai imputasi dengan distribusi nilai yang terobservasi (Honaker dkk, 2013) dapat dilihat bahwa titik-titik berada di sekitar garis diagonal. Sehingga, dapat disimpulkan bahwa model imputasi yang terbentuk sudah cukup baik.



Gambar 3.2 *Scatter Plot* antara Distribusi Rata-Rata Nilai Imputasi dengan Distribusi Nilai yang Terobservasi

4. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dapat diambil kesimpulan bahwa EMB dapat digunakan sebagai metode alternatif untuk memperoleh nilai taksiran imputasi.

Daftar Pustaka

- [1] Allison, P. 2008. *Missing Data*. www.ssc.upenn.edu/~allison
- [2] Azur, M., Stuart, E., Frangakis, C., dan Leaf, P. 2011. *Multiple Imputation by Chained Equations: What is it and how does it work?*. *International Journal Methods Psychiatric Res* 20(1): 40-49.
- [3] Baraldi, A., dan Enders, C. 2009. *An Introduction to Modern Missing Data Analysis*. *Journal of School Psychology* 48:5-37.
- [4] Brownstone, D., dan Valleta, R. 2001. *The Bootstrap and Multiple Imputations: Harnessing Increased Computed Power for Improved Statistical tests*. *Journal of Economic Perspectives* 15(4): 129-141.
- [5] Buuren, S., dan Groothuis, K. 2011. *Mice: Multiple Imputation by Chained Equations in R*. *Journal of Statistical Software: Volume 45, issue 3*.
- [6] Demster, A. P., Laird, N. M., dan Rubin, D. B. 1977. *Maximum Likelihood from Incomplete Data via the EM Algorithm*. *Journal of the Royal Statistical Society* 39(1):1-38.
- [7] Durrant, G. 2005. *Imputation Methods for Handling Item Nonresponse in the Social Sciences: A Methodological Review*. Research Paper: University of Southampton.
- [8] Efron, B. 1979. *A Bootstrap Methods: Another Look A the Jackknife*. *The Annals of Statistics* 7(1):1-26.
- [9] Efron, B. 1987. *Better Bootstrap Confidence Interval*. *Journal of the American Statistical Association* 82(397):171-185.
- [10] Efron, B. 1992. *Missing Data, Imputation, and The Bootstrap*. Technical Report No.397: Stanford University, California.
- [11] Eichler. 2003. *Introduction to The Bootstrap*. <http://galton.uchicago.edu>.
- [12] Grittner, U., Gmel, G., Ripatti, S., Bloomfield, K., dan Wicki, M. 2011. *Missing Value Imputation in Longitudinal Measures of Alcohol Consumption*. *International Journal Methods Psychiatric Res* 20(1):50-61.
- [13] Hogg, R., McKean, J., dan Craig, A. 2013. *Introduction to Mathematical Statistics(seventh edition)*. www.pearsonhighered.com.
- [14] Honaker, J., dan King, G. 2010. *What to Do about Missing Values in Time-Series Cross-Section Data*. *American Journal of Political Science* 54(2): 561-581.
- [15] Honaker, J., King, G., dan Blackwell, M. 2013. *AMELIA II: A Program of Missing Data*. <http://gking.harvard.edu/amelia>.
- [16] Hsiao, C. 2003. *Analysis of Panel Data*. New York: CambridgeUniversity Press.
- [17] Jerez, J.M., dan Molina, I. (2010). *Missing Data Imputation Using Statistical And Machine Learning Methods In A Real Breast Cancer Problem*. *Artificial Intelligence in Medicine*, Vol. 50:105-115.

- [18] King, G., Tomz, M., dan Wittenberg, J. 2000. *Making the Most of Statistical Analyses: Improving Interpretation and Presentation*. . American Journal of Political Science 44(2): 341-355.
- [19] Lahiri, P. 2003. *On the Impact of Bootstrap in Survey Sampling and Small-Area Estimation*. Statistical Science 18(2): 199-210.
- [20] Little, R. J. A., dan Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc, New York.
- [21] Miyama, E., dan Managi, S. 2014. *The Environmental Cuznets Curve: An Application of Multiple Imputation*.
- [22] Rubin, D. 1976. *Inference and Missing Data*. Biometrika 63(3): 581-592.
- [23] Rubin, D. 1987. *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Inc.
- [24] Schafer, J., dan Graham, J. 2002. *Missing Data: Our View of the State of the Art*. Psychological Methods 7(2):147-177.
- [25] Singh, Kesar., and Xie, M. *Bootstrap: A Statistical Method*. Rutgers University.
- [26] Takahashi, M., dan Ito, T. 2012. *Multiple Imputation of Turnover in Edinet Data: Toward the Improvement of Imputation for Economic Census*. Working Paper: United Nations Economic Commission for Europe, Conference of European Statisticians.
- [27] Takahashi, M. 2014. *An Assessment of Automatic Editing via the Contamination Model and Multiple Imputation*. Working paper: United Nations Economic Commission for Europe, Conference of European Statisticians.
- [28] Yuan, Y. 2011. *Multiple Imputation using SAS Software*. Journal of Statistical Software 45(6).