

**Perluakah *Cross Validation* dilakukan?
Perbandingan antara *Mean Square Prediction Error* dan *Mean Square Error* sebagai
Penaksir Harapan Kuadrat Kekeliruan Model**

Yusep Suparman
(yusep.suparman@ unpad.ac.id)
Universitas Padjadjaran

Abstrak

Seleksi model merupakan tahapan terakhir dari suatu analisis regresi. Salah satu pendekatan yang sering dipergunakan adalah *cross-validation*. Ukuran *mean square prediction error* (MSPE) dianggap sebagai ukuran yang lebih baik untuk mengevaluasi tingkat prediktibilitas dibandingkan *mean square error* (MSE) yang diperoleh tanpa harus melalui *cross-validation*. Dalam penelitian ini penulis membandingkan kinerja MSPE dengan MSE sebagai penaksir harapan kuadrat kekeliruan model melalui sebuah simulasi Monte Carlo. Penulis menemukan bahwa kedua statistik tersebut merupakan penaksir yang bias untuk nilai harapan kuadrat kekeliruan model. Namun demikian, keduanya merupakan penaksir yang konsisten. MSE bersifat *underestimate* sementara MSPE bersifat *overestimate*. Bias taksiran MSPE lebih besar dari MSE. Selain itu, standar error MSPE juga lebih besar dibanding dengan MSE. Untuk ukuran sampel yang kecil, MSPE mempunyai tingkat kesetabilan yang rendah, hal ini ditunjukkan oleh nilai standar error yang sangat besar. Dengan demikian dapat disimpulkan bahwa MSE merupakan ukuran yang lebih baik dalam melakukan evaluasi prediktibilitas model dari pada MSPE.

Kata kunci: *cross-validation*, *mean square prediction error*, *mean square error*, regresi multipel, simulasi Monte-Carlo.

1. Pendahuluan

Dalam analisis regresi salah satu langkah penting yang harus dilakukan adalah seleksi model. Pendekatan yang paling sering dipergunakan dalam seleksi model adalah pembagian data. Dalam pendekatan ini data dibagi menjadi dua bagian. Bagian yang pertama disebut *construction set*, dipergunakan untuk membangun model. Bagian yang kedua disebut *validation set*, model regresi yang dibangun berdasarkan *construction set* dievaluasi kesesuaiannya terhadap *validation set*. Prosedur ini biasa disebut *cross-validation* (Weisberg, 2005).

Salah satu hal yang dievaluasi dalam *cross-validation* adalah prediktibilitas model. Ketika sebuah model regresi dibangun dengan data tertentu, tidak dapat dihindari lagi bahwa model yang terpilih adalah yang terbentuk, karena modelnya

memang cocok dengan data, setidaknya untuk sebagian besar data. Hal ini dapat mengakibatkan tingkat kesalahan yang dihitung berdasarkan *construction set* (*apparent error*) lebih rendah dari tingkat kesalahan dari dalam prediksi di luar data *construction set* (*error rate*) (Kutner dkk., 2005, h. 370). Dengan kata lain *error rate* merupakan ukuran yang lebih baik dibandingkan dengan *apparent error*.

Yang perlu dicermati dalam hal ini apakah rendahnya *apparent error* dibanding dengan *error rate* merupakan variasi sampel yang terjadi perkasus atau merupakan suatu kecenderungan umum. Untuk itu perlu dilakukan suatu evaluasi terhadap perilaku *apparent error* dan *error rate* kemudian dibandingkan kecenderungan umum yang terjadi di antaranya. Selain dari itu tentu harus diperhatikan pula sifat-sifat lain keduanya sebagai statistik yang baik.

Di dalam penelitian ini penulis bermaksud mempelajari perilaku *apparent error* yang diwakili oleh *Means Square Prediction Error* (MSPE) dan *error rate* yang diwakili oleh *Means Square Error* (MSE) untuk kasus model regresi linier multipel tanpa pelanggaran asumsi Gaus-Markov. Dari penelitian ini diharapkan dapat diperoleh informasi tentang kecenderungan umum perilaku MSPE dan MSE dari sebuah model regresi linier multipel. Selanjutnya informasi dapat dipergunakan sebagai bahan evaluasi dari penggunaan MSPE sebagai ukuran prediktabilitas model dan penggunaan *cross-validation* dalam validasi suatu model regresi.

2. Simulasi

Dalam suatu kasus di mana analisis matematis dipandang jauh lebih sulit untuk dilakukan, simulasi merupakan alternatif yang dapat ditempuh untuk memecahkan kasus tersebut. Demikian pula halnya dalam membandingkan kinerja MSPE dan MSE, penulis melakukan sebuah simulasi Monte-Carlo. Simulasi ini ditujukan untuk mengevaluasi sifat-sifat MSPE dan MSE sebagai penakasir dari harapan kuadrat kekeliruan model yang merupakan suatu ukuran prediktabilitas model. Untuk itu dalam simulasi ini dihitung rata-rata taksiran MSPE dan MSE, bias keduanya terhadap rata-rata kuadrat kekeliruan dan standard error keduanya.

Berikut ini adalah ketentuan-ketentuan yang diterapkan dalam simulasi yang dilakukan.

- a. Populasi dari data akan mengikuti model regresi sebagai berikut.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

dengan

$$\beta_0 = 5, \beta_1 = 3, \text{ dan } \beta_2 = 5, \text{ kemudian } \varepsilon \sim N(0,4), X_1 \sim N(8,9), \text{ dan}$$

$X_2 \sim N(6,2)$ saling independen satu dengan lainnya.

- b. Untuk memenuhi kriteria *cross-validasi* yang baik, ukuran *construction set* dan *validation set* akan sama (Roecker, 1991). Dengan demikian ukuran *construction* dan *validation set* setengah dari ukuran sampel yang dipergunakan. Ukuran sampel (n) yang dipergunakan adalah 10, 20, 40, 80, 160, 320, 640 dan 1280.
- c. Setiap ukuran sampel diulangi sebanyak 5000 kali (banyak replikasi).
- d. Dalam setiap replikasi MSE dirumuskan sebagai

$$MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}, \quad (2.1)$$

sedangkan MSPE dirumuskan sebagai

$$MSE = \frac{\sum_{i=1}^{n^*} (Y_i - \hat{Y}_i)^2}{n^*} \quad (\text{Kutner dkk, 2005}). \quad (2.2)$$

Dengan $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$ dan n^* merupakan ukuran *validation set*.

- e. Selanjutnya dihitung rata-rata, bias dan standar error dari MSE dan MSPE untuk masing-masing ukuran sampel.

3. Hasil dan Pembahasan

Dalam mengevaluasi suatu statistik, terdapat dua konsep yang harus diperhatikan yaitu efisiensi dan bias (Mendenhall dkk, 1986). Suatu statistik akan lebih disukai dibanding dengan statistik lainnya jika statistik tersebut lebih efisien. Efisiensi ini ditunjukkan oleh nilai standar error. Jadi statistik yang mempunyai

standar error yang lebih kecil akan lebih disukai dibandingkan statistik lainnya yang mempunyai standar error lebih besar. Sedangkan mengenai bias, statistik yang lebih disukai adalah yang tak bias. Namun demikian, ketakbiasan ini bisa diperlonggar dengan sifat kekonsistenan, yaitu untuk ukuran sampel besar statistik tersebut mempunyai bias yang dapat diabaikan. Di antara statistik yang konsisten, tentunya statistik dengan bias terkecil yang lebih disukai untuk dipergunakan. Berdasarkan kedua konsep inilah kita akan mengevaluasi hasil simulasi.

Dari taksiran hasil rata-rata MSPE dan MSE, seperti yang ditampilkan dalam Tabel 1 dan Gambar 1, dapat diketahui bahwa MSPE merupakan penaksir yang over estimate untuk rata-rata kuadrat galat. MSE sebaliknya, merupakan penaksir yang underestimate. Klaim yang selama ini menyatakan bahwa MSE cenderung untuk lebih kecil adalah benar. Walaupun demikian baik MSPE maupun MSE merupakan penaksir yang secara asimtotis konvergen terhadap rata-rata kuadrat galat. Dengan kata lain baik MSPE maupun MSE merupakan penaksir yang konsisten untuk rata-rata kuadrat galat.

Tabel 1. Taksiran, Bias, dan Standar Error MSPE dan MSE

n	Taksiran		Bias		Standar error	
	MSE	MSPE	MSE	MSPE	MSE	MSPE
10	1.593	14.339	1.203	10.339	1.598	62.160
20	3.386	5.955	0.614	1.955	1.162	3.627
40	3.709	4.722	0.291	0.722	0.856	1.637
80	3.855	4.314	0.145	0.314	0.622	1.006
160	3.931	4.145	0.069	0.145	0.444	0.668
320	3.967	4.082	0.033	0.082	0.319	0.465
640	3.980	4.042	0.020	0.042	0.221	0.323
1280	3.992	4.013	0.008	0.013	0.156	0.229

Selanjutnya dari nilai bias taksiran dapat kita ketahui bahwa MSPE mempunyai bias taksiran lebih besar dari MSE (Lihat gambar 2). Pada setiap ukuran sampel, bias taksiran MSPE selalu lebih lebih besar dibandingkan dengan bias MSE. Semakin besar ukuran sampel semakin kecil perbandingan antara bias MSE dan MSPE. Pada ukuran sampel 10, bias MSPE 8,6 kali bias MSE. Sedangkan

pada ukuran sampel 1280, bias MSPE 1,6 kali bias MSE. Dari perbandingan nilai bias, kita dapat pastikan bahwa MSE merupakan ukuran yang lebih disukai karena mempunyai nilai bias yang selalu lebih kecil.

Terakhir, kita lihat perbandingan antara taksiran standar error MSPE dan MSE. Sama seperti halnya bias, taksiran standar error MSPE selalu lebih besar dari taksiran standar error MSE. Dari Gambar 3, perbedaan yang mencolok terjadi pada ukuran sampel di bawah 20. Pada ukuran sampel 10, taksiran standar error MSPE 38,9 kali lebih besar dari taksiran standar error MSE. Sedangkan pada ukuran sampel terbesar, yaitu 1280, taksiran standar error MSPE 1,5 kali lebih besar taksiran standar error MSE. Dari sini, kita dapat mengetahui bahwa MSPE merupakan ukuran yang lebih tidak reliabel untuk mengevaluasi kecocokan suatu model regresi dibandingkan dengan MSE, terutama untuk ukuran sampel kecil.

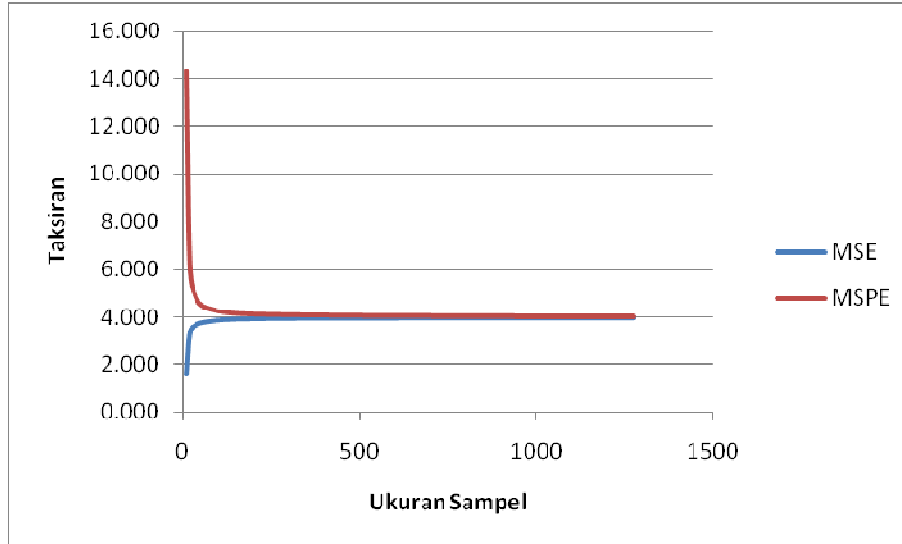
4. Kesimpulan

Baik MSPE maupun MSE merupakan penaksir yang konsisten untuk harapan kuadrat kekeliruan model. Namun demikian MSE mempunyai bias yang lebih kecil dibanding MSPE. Selain itu MSE lebih efisien dibanding MSPE. Dengan begitu, MSE merupakan penaksir ukuran prediktabilitas model yang lebih baik dibanding MSPE. Khusus untuk ukuran sampel di bawah 20, MSPE tidak dianjurkan untuk dipergunakan sebagai ukuran prediktabilitas model menimbang standar errornya yang jauh lebih besar dibanding standar error MSE.

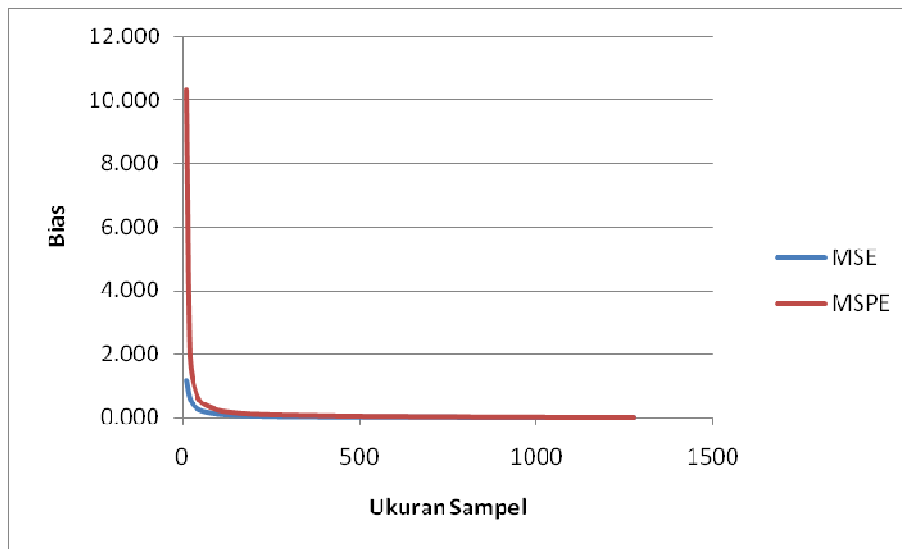
5. Referensi

- Kutner, M.H., Nachtsheim, C.J., Neter, J. dan Li, W. (2005) *Applied Linear Statistical Model*, McGraw-Hill.
- Mendenhall, W., Scheaffer, R.L., dan Wackerly, D.D. (1986) *Mathematical Statistics with Applications*, ed. ke-3, Duxbury, Boston.
- Roecker, E.B. (1991) Prediction Error and Its Estimations for Subset-Selected Models, *Technometrics*, **33**(4), hal 468.

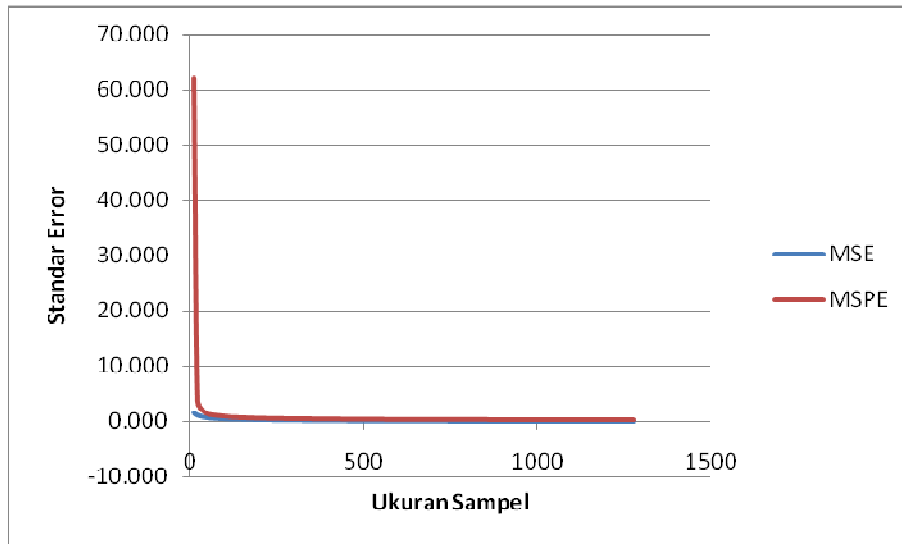
Weisberg, S. (2005) *Applied Linear Regression*, ed. Ke-3, John Wiley and Sons, New Jersey.



Gambar 1. Taksiran MSPE dan MSE



Gambar 2. Bias MSPE dan MSE



Gambar 3. Standar Error MSPE dan MSE