UJI LINEARITAS BERDASARKAN ESTIMASI MEAN DAN VARIANSI BERSYARAT UNTUK PROSES RUNTUN WAKTU

Supriyanto

Program Studi Matematika, Fakultas Sains dan Tekink
Universitas Jenderal Soedirman, Purwokerto

Herni Utami

Jurusan Matematika, Fakultas MIPA
Universitas Gajah Mada, Yogyakarta

ABSTRACT. *This study aims to examine the benefits of testing linearity in the case of live test data. Bootstrap procedure is used to form the estimators of the statistics. Hypothetical form is used to follow the linear model. And compare the value of criticism from the distribution of this value with the test statistics that have been calculated based on the observed time series data existing. This procedure starts with a model determines autoregression to the data. By using the Akaike information criterion, order estimation obtained from the autoregression models.*

Keywords. *lag, kernel method, heteroscedastic procedure, autoregresi linear autoregression, Akaike information.*

I. PENDAHULUAN

Misalkan $\{X_t, t \geq 0\}$ adalah proses stasioner yang mempunyai momen sampai order 4. Uji linearitas akan didasarkan pada mean bersyarat dan variansi bersyarat :

$$M_k(x) = E\left[X_t \middle| X_{t-k} = x\right]$$

$$V_k(x) = \mathrm{var}\left[X_t \middle| X_{t-k} = x\right]$$

k adalah lag, yang dapat bernilai negatif dan bervariasi. Mean dan variansi bersyarat di atas, diestimasi dengan metode kernel, dari n observasi, yaitu $\{X_1, X_2, ..., X_n\}$. Adapun estimasi mean bersyarat sebagai berikut :

$$\hat{M}_k(x) = \frac{(n-k)\sum_{t=k+1}^{n} X_t K_h(x - X_{t-k})}{n^{-1}\sum_{t=1}^{n} K_h(x - X_t)} = \frac{\hat{a}_k(x)}{\hat{p}(x)}$$

dimana $K_h(x) = h^{-1}K(h^{-1}x)$, dengan K adalah fungsi kernel dan h adalah bandwidth. Dalam hal ini digunakan $h = n^{-1/5}{}_{SD}(X)$, yang mana untuk variansi

Diterima 15 Februari 2009, disetujui untuk dipublikasikan 21 April 2009.

dan bias kuadrat dari estimasi kernel ordinal, dalam hal ini baik untuk kasus normal. Estimator $\hat{p}(x)$ adalah estimasi dari densitas probabilitas marginal p(x).

Demikian juga variansi bersyarat, diestimasikan dengan

$$\hat{V}_k(x) = \frac{(n-k)^{-1}\sum_{t=k-1}^{n} X_t^{\,2} K_h(x - X_{t-k})}{n^{-1}\sum_{t=1}^{n} K_h(x - X_t)} - \hat{M}_k^{\,2}(x)$$

Untuk pembuktian konsistensi secara lemah dan konsistensi secara kuat dari estimasi-estimasi di atas dapat dilihat Robinson(1983), Journal of Time Series Vol. 4 No. 3.

Selanjutnya diperkenalkan indek nonlinearitas :

$$L(M_k) = E\left[\{M_k(X_{t-k}) - \rho_k X_{t-k}\}^2\right] = \int \{M_k(x) - \rho_k x\}^2 p(x)dx$$

Untuk mencegah dependensi pada skala, digunakan $\{X_t'\} = \left[\{X_t - E(X_t)\}/_{SD}(X_t)\right]$ dimana $E(X_t)$ dan $_{SD}(X_t)$ adalah mean dan standar deviasi dari $\{X_t\}$ yang diestimasikan dengan $\overline{X} = n^{-1}\sum_t X_t$ dan $s = \{(n-1)^{-1}\sum_t (X_t - \overline{X})2\}^{1/2}$.

Nonlinearitas dideskripsikan dalam variansi bersyarat dengan memperkenalkan

$$L'(V_k) = E\left[\{V_k(X_{t-k}) - \sigma_k^{\,2}\}^2\right]$$

dimana $\sigma_k^{\,2} = E\{(X_t - \rho_k X_{t-k})^2\} = (1 - \rho_k^{\,2})\mathrm{var}(X_t)$. Namun untuk konsistensi dengan prosedur heteroscedastik untuk pemodelan variansi bersyarat digunakan

$$L(V_k) = E\left[\{V_k(e_{t-k}) - \sigma_e^{\,2}\}^2\right]$$

dengan $\sigma_e^{\,2} = \mathrm{var}(e_t)$ dan $\{e_t\}$ adalah proses residual dari autoregresi linear yang terbaik, yaitu $e_t = X_t - a_1X_{t-1} - \ldots - a_pX_{t-p}$ dengan p order yang ditentukan dengan kriteria informasi akaike, dan a_i diestimasikan dengan kuadrat terkecil.

II. ESTIMASI FUNGSIONAL

Fungsional $L(M_k)$ diestimasikan dengan

$$\hat{L}(M_k) = \frac{1}{n}\sum_{t=1}^{n}\left\{\hat{M}_k(X_t) - \hat{\rho}_k X_t\right\}^2 w(X_t)$$

dengan $\hat{\rho}_k = \sum_t \left(X_t - \overline{X}\right)\left(X_{t-k} - \overline{X}\right) / \sum_t \left(X_t - \overline{X}\right)^2$

Fungsi w adalah fungsi pemberat, dimaksudkan untuk mencegah munculnya kuantitas bernilai ekstrim. Sedang estimasi untuk $L(V_k)$ adalah

$$\hat{L}(V_k) = \frac{1}{n-p}\sum_{t=1+\hat{p}}^{n}\left\{\hat{V}_k(\hat{e}_t) - \hat{\sigma}_{\hat{e}}^{\,2}\right\}^2 w(\hat{e}_t)$$

dan $\hat{e}_t = X_t - \sum_{i=1}^{\hat{p}} \hat{a}_i X_{t-i}$

Pada uji linearitas, H_0 akan ditolak jika nilai $\hat{L}(M_k)$ dan $\hat{L}(V_k)$ terlalu besar.

III. SIFAT-SIFAT ASIMTOTIS

Untuk $h = o(n^{-1/5})$ dan $h = O(n^{-1/5})$, kita dapat memperoleh asimtotik normalitas untuk $\hat{L}(M_k)$ dan $\hat{L}(V_k)$ dengan tambahan syarat bahwa $\{X_t\}$variabel random identik dan independen. Untuk variabel random dependen $\{X_t\}$ kami menduga bahwa asimtotik normalitas akan terus memenuhi asimtotik lemah. Jika $h \propto n^{-r}$, r < 1/5 asimtotik normal dapat dibuktikan untuk kasus dependen, dengan menggunakan teorema limit pusat mixing strong. Terhadap hipotesis null H_0' : $M_k(x) = \rho_k X$, k = 1,2,... hal ini benar untuk model Gaussian dan model-model autoregresi order pertama pada umumnya, tetapi tidak untuk semua model-model autoregresi moving average. Mean distribusi asimtotik sebagai berikut :

$$E\left\{\hat{L}(M_k)\right\} \propto (nh)^{-1}\int K^2(z)dz\int V_k(x)w(x)dx + h^4\left\{\int K^2(z)dz\right\}^2 \rho_k^{\,2}\int p^{-1}(x)\left\{p'(x)\right\}^2 w(x)dx$$

dengan p' adalah derivatif dari p(30

x). Kasus selanjutnya adalah asimtotik variansi dari $\hat{L}(M_k)$, didefinisikan sebagai berikut :

$$\operatorname{var}\left\{\hat{L}(M_k)\right\} = 6(n^2h)^{-1}\operatorname{var}(X_t)\int G^2(z)dz\int w(x)dx + o\left\{(n^2h)^{-1}\right\},$$

$$G(z) = \int K(y)K(z+y)dy$$

Dengan argumen yang sama, dapat dilakukan untuk $L(V_k)$. Untuk residual proses yang telah diketahui, terhadap hipotesis null bahwa $\{X_t\}$ adalah suatu model autoregresi moving average, $\{e_t\}$ adalah suatu deret variabel independen, dan untuk $h = O(n^{-1/5})$, maka

$$E\{\hat{L}(V_k)\} \infty (nh)^{-1} \int K^2(z)dz \int \{s_k(x) - r_k^2(x)\} w(x)dx,$$

$$\mathrm{var}\{\hat{L}(V_k)\} \infty 2(n^2 h)^{-1} \int G^2(z)dz \int \{s_k^2(x) - r_k^2(x)\} w(x)dx.$$

dengan $s_k(x) = E(e^4{}_{t+k} | e_t = x) = E(e^4{}_t)$, $r_k(x) = E(e^2{}_{t+k} | e_r = x) = \sigma^2{}_e$.

Dalam prakteknya kita menggunakan $e_t = X_t - \sum \hat{a}_i X_{t-i}$.

IV. PROSEDUR UJI BERDASARKAN BOOTSTRAP

Ide dasar adalah menggunakan boostrap untuk membentuk distribusi null dari $\hat{L}(M_k)$ dan $\hat{L}(V_k)$ terhadap hipotesis null :

H_0 : { X_t } mengikuti suatu model autoregresi linear.

Dan membandingkan nilai kritik dari distribusi ini dengan nilai uji statistik yang telah dihitung berdasarkan data runtun waktu terobservasi yang ada. Prosedur ini dimulai dengan menentukan suatu model autoregresi untuk data yang ada. Dengan menggunakan kriteria informasi Akaike, diperoleh estimasi order dari model autoregresi. Selanjutnya dihitung estimasi residual :

$$\hat{e}_t = X_t - \sum_{i=1}^{\hat{p}} \hat{a}_i X_{t-i}$$

dan diasumsikan $\{e_t\} = \left\{ X_t - \sum_{i=1}^{p} a_i X_{t-i} \right\}$ independen untuk model autoregresi order p.

Centering $\{\hat{e}_t\}$ dan bentuk replikasi bootstrap sebanyak B, sehingga diperoleh $\{\hat{e}^*{}_t\}$ sebanyak B. Untuk masing-masing $\{\hat{e}^*{}_t\}$ dihitung X_t^* dengan rumus sebagai berikut :

$$X_t^* = \sum_{i=1}^{\hat{p}} a_i X_{t-1} + e_t^*$$

Kemudian kita hitung nilai $\hat{L}(M_k)$ dan $\hat{L}(V_k)$ untuk $\{X_t^*\}$. Untuk $\hat{L}(M_k)$ kita hitung

$$\hat{L}^*(M_k) = \frac{1}{n}\sum_{t=1}^{n}\left\{\hat{M}_k^*(X_t^*) - \hat{\rho}_k X_t^*\right\}^2 w(X_t^*)$$

Dari hasil perhitungan $\hat{L}(M_k)$ untuk masing-masing $\{X_t^*\}$ diperoleh estimasi bootsrap $L(M_k)$ sebanyak B. Kemudian dari hasil tersebut dapat dihitung distribusi $\{X_t^*\}$. Daerah kritik dihitung berdasarkan distribusi tersebut. Untuk nilai kritik kita tentukan dengan menggunakan quantil ke α. Langkah terakhir, kita hitung

$$\hat{L}(M_k) = \frac{1}{n}\sum_{t=1}^{n}\left\{\hat{M}_k(X_t) - \hat{\rho}_k X_t\right\}^2 w(X_t)$$

untuk proses original $\{X_t\}$, dan hipotesis linearitas ditolak jika $\hat{L}(M_k)$ lebih besar dari nilai kritik α dari distribusi bootstrap.

Jika proses autoregresi conditional heteroscedastik type heterogenitas di dalam residual, yaitu $e_t = g(e_{t-1})\eta_t$ dimana η_t independen, maka residual $\{e_t\}$ menjadi tak berkorelasi tetapi tidak independen. Pelanggaran dari sifat independensi ini tidak mudah dideteksi dengan $L(M_k)$, dan fungsional $L(V_k)$ lebih cocok untuk statistik uji linearitas, karena ini didasarkan pada variansi bersyarat. Distribusi bootstrap dikonstruksikan dari replikasi $\{\hat{e}^*_t\}$, sedemikian sehingga

$$\hat{L}^*(V_k) = \frac{1}{n}\sum_{t}\left\{\hat{V}_k^*(e_t^*) - \hat{\sigma}^2_{k,e^*}\right\}^2 w(e_t^*)$$

dan hipotesis dari linearitas ditolak jika $\hat{L}(V_k)$ lebih besar dari nilai kritik dari distribusi bootstrap.

Karena $L(M_k)$ dan $L(V_k)$ bervariasi nilainya untuk semua k, maka untuk uji linearitas digunakan statistik uji :

$$\hat{L}_{sup}(M_k) = \sup_{j\le k}\hat{L}(M_j),$$

$$\hat{L}_{sup}(V_k) = \sup_{j\le k}\hat{L}(V_j),$$

$$\hat{L}_{ave}(M_k) = \frac{1}{k}\sum_{j=1}^{k}\hat{L}(M_j)$$

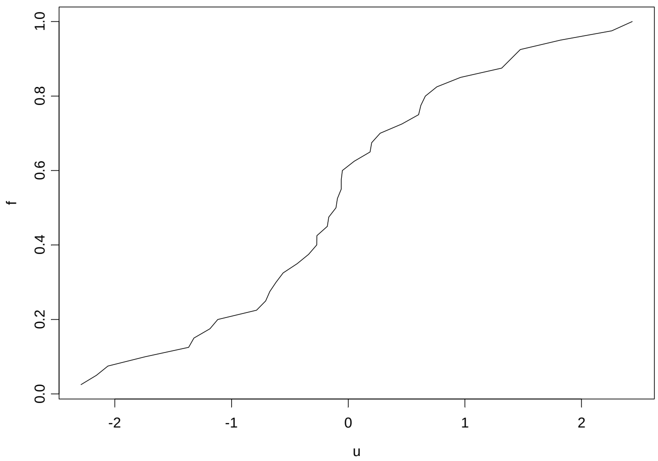
$$\hat{L}_{ave}(V_k) = \frac{1}{k}\sum_{j=1}^{k}\hat{L}(V_j)$$

V. HASIL SIMULASI

Untuk simulasi, digunakan data runtun waktu berukuran n = 100. Dari data ini akan diuji apakah data mempunyai model linear, dengan kata lain akan diuji mengenai linearitas. Untuk pengujian ini digunakan prosedur resampling. Dalam uji kali ini, hipotesisnya adalah :

H_0 : $\{X_t\}$ mengikuti suatu model autoregresi linear.

Sebelum dilakukan pengujian, kita periksa apakah data stasioner (Gambar 1). Untuk simulasi kali ini, setelah data di plot ternyata data stasioner. Dari data tersebut mempunyai model autoregresi order pertama, yaitu $X_t = 0.9129181X_{t-1} + e_t$. Dilakukan resampling $\{e_t{*}\}$ sebanyak 100 kali. Dari resampel tersebut diperoleh distribusi pendekatan bootstrap, yang dapat dilihat dari hasil program komputer. Statistik uji diperoleh 0,9151852 dan nilai kritik 1,45456. Karena statistik uji kurang dari nilai kritik, maka kita dapat menyimpulkan bahwa data mengikuti model autoregresi.



Gambar 1. Plot data hasil simulasi

VI. DAFTAR PUSTAKA

Bain, L.J., dan Engelhardt, M., *Introduction to probability and Mathematical Statistics*, Wadsworth Publishing Company, California, 1992.

Goodman, Roe, *Introduction to Stochastic Models*, Benjamin/Cummings Publishing Company, California, 1998.

Hu, X. J. dan Lawless, J.F., *Estimation of rate and Mean Functions from Truncated Recurrent Event Data*, JASA, **91**(1996), 300-310.

Hu, X. J. dan Lawless, J.F., *Estimation from Truncated Lifetime Data with Supplementary Information on Covariate and Censoring Times*, Biometrika, **83**(1996), 747-761.

Hu, X. J.,Lawless, J.F., dan Suzuki, K., *Nonparametric Estimation of a Lifetime Distribution When Censoring Times are Missing*, Technometrics, **40**(1998), 3-13.

Kalbfleisch, J.D. dan Lawless, J.F., *Estimation of Reliability in Field Performance Studies*, Technometrics, **30**(1998), 365-388.

Kalbfleisch, J.D. , Lawless, J.F., dan Robinson, J.A., *Methods for the Analysis and Prediction of Warranty Claims*, Technometrics, **33**(1991), 273-285.

Lawless, J.F., *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York, 1982.

Lawless, J.F., *Introduction to Two Classics in Reliability Theory*, Technometrics, **42**(2000), 5-6.

Supriyanto, *Estimasi Nonparametrik Distribusi Tahan Hidup : Kasus Data Tersensor dari Unit-Unit tak Terputus*, Tesis, Pascasarjana Universitas Gadjah Mada , Yogyakarta, 2001.

Shao, Jun, dan Tu, Dongsheng, *The Jacknife and Bootstrap*, Springer, New York, 1995.

Suzuki, K., *Nonparametric Estimation of Lifetime Distributions from a Record of Failures and Follow-Ups*, JASA, **80**(1985), 68-72.

Thall, P.F. dan Lachin, J.M., *Analysis of Recurrent Event : Nonparametric Methods of Random-Interval Count Data*, JASA, **83**(1988), 339-347.