MAp Journal
Mathematics & Applications
UIN IMAM BONJOL PADANG
Web : ejournal.uinib.ac.id/jurnal/index.php/MAp
Email : map_journal@uinib.ac.id

DETEKSI *OUTLIERS* DAN ANALISIS INTERVENSI DALAM MODEL ARMA

OUTLIERS DETECTION AND INTERVENTION ANALYSIS IN ARMA MODEL

Umi Yuliatin [1§]

[1]Politeknik Energi dan Mineral Akamigas [umi.yuliatin@esdm.go.id]
[§]*Corresponding Author*

Received Jun 18th 2022; Accepted Jun 30th 2022; Published Jun 30th 2022;

Abstrak

Adanya kehadiran outliers dalam analisisi runtun waktu mengaburkan estimasi parameter model yang diberikan. Selain itu outlier juga memberi dampak besaran eror yang lebih tinggi. Dalam analisis time series *Additive outliers* (AO) dan *innovational outliers* (IO) diperkenalkan sebagai usaha dalam memodelkan outliers. Usaha ini diberikan untuk menangani obserbasi yang tidak mengharmoniskan pola data sehingga membantu untuk dibentuknya model runtun waktu yang sehat terutama dalam proses ARMA. Estimator *linier square error* (LSE) digunakan untuk mengestimasi besarnya penyimpangan dari model dasarnya. Prosedur iterative dipaparkan sebagai salah satu prosedur untuk mendeteksi kedua model outliers ini. Diperkenalkan juga analisis intervensi yang digunakan untuk mengakomodasi kejadian luar sebagai variabel eksogen dalam proses ARMA. Kemudian kombinasi analisis ourliers-intervensi ini bisa digunakan sebagai kesatuan analisis untuk menangani data yang jauh dari pusat. Sebagai simulasi data dalam kasus ini adalah data PDRB D.I Yogyakarta dalam bidang pertambangan dan penggalian. Dalam analisis ini ditunjukkan deteksi outlier didalam model memberikan jumlahan kuadrat eror yang lebih kecil dibandingkan dengan model tanpa deteksi outlier sedemikian sehingga diperoleh model yang lebih baik.

Kata Kunci: *Additive outliers (AO) , innovational outliers (IO),* analisis intervensi.

Abstract

Outliers in time series analysis modeled in several types. Additive outliers and Innovational outliers (IO) is introduced to maintain the data waves. Outliers is not only a nuisance , but modelling of this extreme data help good model especially for ARMA model. Ordinary least square error is used to estimate the deviation from the basic model. Iterative procedure is introduced to detect both of outliers. In other hand intervention analysis is used to accommodate of outside factor as exogen variable in ARMA process. The combination of outliers-intervention analysis can be used as one analysis to repair data far from center. For simulation is PDRB data showed outliers detection in the model give smaller in error specification compared without detection.

Keywords*:* Additive outliers (AO) , innovational outliers (IO), intervention analysis

1. Pendahuluan

Analisis runtun waktu adalah masalah fundamental di berbagai bidang penelitian seperti keuangan, sains, demografi, industri, dan lain-lain [1]. Dengan berbagai sisi kebutuhan, seperti deskriptif, pemodelan sampai peramalan nilai masa depan. Analisis yang bisa berbasis pada waktu maupun ruang ini pun kini telah dikembangkan didalamnya banyak metode yang disesuaikan dengan kebutuhan dan perilaku data, univariat ataupun multivariate, deterministik ataupun stokastik, linier ataupun non-linier, dan lain sebagainya.

Seperti juga yang terjadi pada jenis data kebanyakan, bahwa di dalam time series juga terdapat nilai yang menjauh dari pusat, dalam hal ini adalah model dasarnya. Data ini disebut outliers. Outlier sendiri merupakan penambahan nilai error yang besar dari data observasi berpengaruh secara tak berulang atau pun berulang sebagai variabel luar yang mengintervensi pemodelan. Sebagai contoh intervensi itu dapat berupa aksi mogok, bencana alam, pandemi, kebijakan politik, perang, krisis ekonomi dan sebagainya. Sedangkan yang lain, mungkin saja terjadi kesalahan pada input data sebelum analisis. Walaupun bisa juga terjadi nilai yang bernilai ekstrem ini tanpa tendensi apapun dan sulit untuk dimengerti terjadi.

Dalam model ARMA dan ARIMA, outlier akan memberikan efek yang besar pada saat peramalan angka di masa depan. Sehingga model Box-Jenkin ini bisa memberikan angka prediksi yang jauh dari yang seharusnya kita prediksi dan jauh dari data yang sebenarnya [2]. Sedemikian sehingga outlier dalam model ARMA perlu ditangani untuk memperoleh angka prediksi yang lebih akurat.

2. Landasan Teori

2.1 Proses AR (*Autoregressive*)

Autoregresi adalah suatu bentuk regresi yang menghubungkan nilai-nilai sebelumnya dengan diri sendiri pada selang waktu (*time lag)* yang bermacam-macam. Suatu model autoregresi akan menyatakan suatu ramalan sebagai fungsi dari nilai-nilai sebelumnya. Model ini pertama kali diperkenalkan oleh Yule pada 1926 [3]. Mengacu pada referensi [4], model time series autoregresi berorde-p dengan data $X_t, t \in \mathbb{Z}$ memiliki bentuk :

$$X_t = \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \ldots + \emptyset_p X_{t-p} + a_t$$

atau ditulis dalam bentuk

$$\emptyset(B) X_t = a_t$$

dimana

$$\emptyset(B) = (1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p),$$

dengan $\emptyset_1, \emptyset_2, \ldots, \emptyset_p$ merupakan parameter-parameter AR bilangan real dan B adalah operator backshift. Sedangkan $a_t \sim WN(0, \sigma^2)$

2.2 Proses MA (Moving average)

Model runtun waktu moving average berorde-q pertama kali diperkanlakan oleh Slutzky [5] dan dibahas oleh Wie [4] , ditulis MA(q) adalah $X_t, t \in \mathbb{Z}$ yang memiliki persamaan dengan bentuk :

$$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \ldots + \theta_q a_{t-q}$$

atau ditulis $X_t = \theta(B) a_t$, dimana a_t mengikuti proses *white noise* dengan mean

$E(a_t) = 0$ dan variansi $var(a_t) = \sigma^2$dengan $\gamma_k = \begin{cases} \sigma_\varepsilon^2, k = 0 \\ 0, k \neq 0 \end{cases}$..

2.3 Proses ARMA

Metode yang memperhitungkan efek autoregressive dan juga moving average (ARMA) dengan model [6] :

$$\left(1 - \emptyset_1 B - \ldots - \emptyset_p B^p\right)(X_t - \mu)$$
$$= \left(1 - \theta_1 B - \ldots - \theta_q B^q\right) a_t$$

dimana $a_t, t \in \mathbb{Z}$ adalah *white noise* dengan mean μ dan variansi σ^2 dan $\emptyset_1, \emptyset_2, \ldots, \emptyset_p, \theta_1, \theta_2, \ldots, \theta_q \in \Re$. Sedangkan p adalah order autoregressive dan q adalah order moving average

Dengan mengunakan operator lag proses ARMA(p,q) dapat ditulis menjadi :

$$\emptyset(B) X_t = \theta(B) a_t$$
$$X_t = \frac{\emptyset(B)}{\theta(B)} a_t$$
$$X_t = \psi(B) a_t$$
$$X_t = \sum_{j=0}^{\infty} \psi\, a_{t-1}$$

Suatu proses ARMA yang stationer bisa dituliskan sebagai proses MA(∞) sedangkan suatu proses ARMA yang invertible bisa dituliskan sebagai proses AR(∞) [6].

2.4 Kausalitas dalam proses ARMA

Misalkan {Xt} adalah suatu proses ARMA(p,q) dengan polinomial D(z) dan C(z) tidak memiliki akar-akar yang sama. Maka {Xt} akan bersifat kausal jika dan hanya jika $D(Z) \neq 0$ untuk $|z| \leq 1$. Dengan kata lain polinomial D(z) yakni polinomial dari autoregressi tidak memiliki akar-akar dalam unit lingkaran, yakni jika $z_i, i = 1,2, \ldots, r$ adalah akar-akar berbeda dari D(z) maka $|z| > 1$.

Jika X_t bersifat kasual maka kondisi $\sum_{j=0}^{\infty} \theta_j^2 < \infty$ akan dipenuhi, yakni X_t akan stationer. Pada kasus kausal, penyelesaian untuk X_t dapat dituliskan sebagai

$$X_t = \frac{C(B)}{D(B)} a_t = \sum_{j=0}^{\infty} \theta_j B^j a_t = \sum_{j=0}^{\infty} \theta_j a_{t-j}$$

dengan $a_t \sim WN(0, \sigma^2)$. Jika polinomial $D(z) \neq 0$ untuk $|z| = 1$ (yakni akar polinomial dari D(z) memiliki nilai mutlak $\neq 1$), maka terdapat penyelesaian yang bersifat *"steady state"* untuk X_t . penyelesaian yang diperoleh tidak terlalu bersifat stationer, dan hanya bersifat stationer apabila $\sum_{j=-\infty}^{\infty} |C_j| < \infty$ [7].

2.5 Proses ARMA yang Invertible

Suatu proses ARMA(p,q) didefinisikan dengan $D(B) X_t = C(B) a_t$ disebut invertible jika terdapat barisan konstanta $\{h_j\}$ sedemikian sehingga

$$\sum_{j=0}^{\infty} h_j^2 < \infty, \text{ dan } a_t = \sum_{j=0}^{\infty} h_j X_{t-j}, t \in z$$

Diberikan $\{X_t\}$ suatu proses ARMA(p,q) dengan polinomial $D(z)$ dan $C(z)$ tidak memiliki akar-akar yang sama. Maka $\{X_t\}$ invertible jika dan hanya jika $C(z) \neq 0$ untuk $|z| \leq 1$. Dengan kata lain akar-akar berbeda dari polinomial $C(z), z_1, z_2, \ldots, z_k$ akan memiliki sifat $|z_i| > 1, i = 1,2, \ldots, k$.

Berdasarkan definisi dapat ditujukan bahwa proses MA*(q)*, $q < \infty$ selalu bersifat kausal. Sedangkan AR*(p)*, $p < \infty$ selalu bersifat invertibel, sedangkan untuk proses ARMA*(p,q)* bergantung kepada akar-akar dari polinomial-polinomialnya [7].

3 Hasil Dan Pembahasan

3.1 *Additive Outliers* dan *Innovational Outliers*

Diberikan proses stationer Z_t, adalah data observasi runtun waktu dan X_t adalah data runtun waktu bebas *outliers.* Diasumsikan bahwa $\{X_t\}$ mengikuti model umum ARMA (p,q). modelnya : $\emptyset(B)X_t = \theta(B)a_t$

dimana :

$$X_t = \frac{\theta(B)}{\emptyset(B)}a_t \text{ ...... 1)}$$

$\emptyset B = 1 - \emptyset_1 B - \ldots - \emptyset_p B^p$, dan

$\theta(B) = 1 - \ldots - \theta_p B^p$ adalah operator stasioner dan invertible

$\{a_t\}$ *while noise i,I,d* $N(0, \sigma_a^2)$

Additive outliers (AO) didefinisikan sebagai

$$Z_t = \begin{cases} X_t, & t \neq T \\ X_t + \omega, & t = T \end{cases}$$

$$= X_t + \omega I_t^{(T)}$$

$$= \frac{\theta(B)}{\emptyset(B)}a_t + \omega I_t^{(T)} \text{........ 2)}$$

Dimana

$I_t^{(T)} = \begin{cases} 1, t = T \\ 0, t \neq T \end{cases}$: mengindikasikan *outliers* pada waktu T, sedangkan *ω* adalah efek *Additive outliers*

Innovational outliers (IO) didefinisikan :

$$Z_t = X_t + \frac{\theta(B)}{\emptyset(B)}\omega I_t^{(T)}$$

$$= \frac{\theta(B)}{\emptyset(B)}(a_t + \omega I_t^{(T)}) \text{ ..... 3)}$$

Oleh sebab itu, sebuah *additive outliers* efeknya hanya pada observasi ke T,sedangkan *innovational outliers* ke semua observasi $Z_r, Z_{r+1}, \ldots$ sesudah waktu T yang digambarkan dengan $\frac{\theta(B)}{\emptyset(B)}$.

Secara umum, data runtun waktu yang terdiri dari *k* tipe *outliers* dapat mengikuti model umum:

$$Z_t = \sum_{j=1}^{k} \frac{\omega_j v_j}{(B) I_t^{(r_j)}} + X_t,$$

dimana :

$$X_t = \frac{\theta(B)}{\emptyset(B)}a_t,$$

$v_j(B) = 1$ untuk AO dan $v_j = \frac{\theta(B)}{\emptyset(B)}$ untuk IO pada waktu $t = T_j$

3.2 Estimasi Efek *Outliers* Dimana ARMA Parameter dan σ_a^2 Diketahui

Untuk mengestimasi *ω* dari AO dan IO, dipaparkan kasus sederhana ketika T dan semua parameter ARMA dari proses X_t diketahui. begitu juga dengan σ_a^2 sebagai variansi dari proses *white noise* X_t diketahui.

Didefinisikan:

$$\pi(B) = \frac{\emptyset(B)}{\theta(B)} = (1 - \pi_1 B - \pi_2 B^2 - \ldots)$$

$$e_t = \pi(B)Z_t \qquad \text{.... 4)}$$

Sehingga AO:

$$Z_t = \frac{\theta(B)}{\emptyset(B)}a_t + \omega I_{tt}^{(r)}$$

$$\pi(B)Z_t = \pi B\left(\frac{\theta(B)}{\emptyset(B)}a_t + \omega I_t^{(r)}\right)$$

$$\pi(B)Z_t = a_t + \omega\pi(B)I_t^{(r)}$$

$$e_t = \omega\pi(B)I_t^{(r)} + a \quad \text{.....5)}$$

Estimator LSE dari *ω* untuk model AO diperoleh dengan meminimumkan jumlah kuadrat eror dari 5)

$$\sum_{t=1}^{n} a_1^2 = \sum_{t=1}^{n}\left(e_t - \omega\pi(B)I_t^{(r)}\right)^2$$

diperoleh

$$\hat{\omega} = \frac{\sum_{t=1}^{n}\pi^2(B)I_t^{(T)}e_t}{\sum_{t=1}^{n}\pi^2(B)I_t^{(T)}}$$

dengan $I_t^{(T)} = \begin{cases} 1, untuk\ t = T \\ 0, untuk\ t \neq T \end{cases}$

Selanjutnya

$$\widehat{\omega}_{AT} = \frac{e_r - \sum_{j=1}^{n-\tau} \pi_j e_{\tau+j}}{\sum_{j=0}^{n-r} \pi_1^2}$$

$= \frac{\pi^*(F)e_\tau}{\tau^2}$ …6)

dimana

$\pi^*(F) = (1 - \pi_1 F - \pi_2 F^2 - \ldots - \pi_{n-r} F^{n-\tau})$, $\tau^2 = \sum_{j=0}^{n-\tau} \pi_t^2$, untuk j=0, maka $\pi_0^2 = 1$. Sedangkan F adalah operator forward shif, $Fe_t = e_{t+1}$.

Besar variansi estimator efek AO :

$$var\ (\widehat{\omega}_{AT}) = var\ \left(\frac{\pi^*(F)e_\tau}{\tau^2}\right)$$
$$= \frac{1}{\tau^4} var\ [\pi^*(F)a_\tau]$$
$$= \frac{\sigma_a^2}{\tau^2}$$

Dan innovational outliers (IO) menjadi :

$$Z_t = \frac{\theta(B)}{\emptyset(B)}(a_t + \omega I_t^{(\tau)}$$

Dengan mengalikan kedua ruas dengan $\pi(B)Z_t$ maka diperoleh :

$e_t = \omega I_t^{(\tau)} + a_t$ ….7) .

Kemudian $\widehat{\omega}_{rr}$ yang merupakan estimator untuk model IO diperoleh dengan metode LSE (*least square error*)

$$e_t = \widehat{\omega} I_t^{(\tau)} + a_t$$
$$a_t = e_t - \widehat{\omega} I_t^{(\tau)}$$
$$\sum_{i=1}^{n} a_i^2 = \sum_{t=1}^{n}\left(e_t - \widehat{\omega} I_t^{(\tau)}\right)^2 \ \ldots 8)$$

Dengan mengambil turunan pertama dari 8) diperoleh :

$$\widehat{\omega} = e_t$$

Dengan

$$var(\widehat{\omega}_{rr}) = \sigma_a^2$$

Jadi, estimator terbaik efek *outliers* IO diwaktu T adalah residual e_t pada waktu kejadian, sedangkan estimasi terbaik efek AO adalah kombinasi linier $e_r, e_{r+1}, \ldots, e_n$ dengan pembobotan berdasarkan struktur proses runtun yang bebas *outliers.*

Estimasi untuk $\widehat{\omega}_{AT}$ dapat ditulis sebagai :

$\widehat{\omega}_{AT} = \frac{\pi^*(F)x(B)z_t}{r^2}$ …. 9)

Untuk $\frac{1}{r^2} \leq 1$, variansi $\widehat{\omega}_{AT}$ akan sama besarnya dengan $\widehat{\omega}_{rr}$ dan di beberapa kasus ini akan lebih kecil dari σ_a^2. Untuk contoh, ketika proses X_t mengikuti orde pertama model moving average (MAI), yaitu $X_t = (1 - \theta B)a_t$, maka $var(\widehat{\omega})$ akan mendekati $\sigma_a(1 - \theta^2)$ jika *outliers* tidak terjadi di dekat akhir data. Singkatnya :

$$var(\widehat{\omega}_{AT}) \leq var(\widehat{\omega}_{rr}) = \sigma_a^2$$

Bentuk uji hipotesis untuk *Additional Outliers* (AO):

1. H_0 : Zt bukan AO atau IO ($\omega_{AT} = 0$)
   H_1 : Zt adalah AO ($\omega_{AT} \neq 0$)
2. Tingkat signifikan α
3. Statistic uji H_1 vs H_0 $\lambda_{1,r} = r\widehat{\omega}_{AT}/\sigma_a$
4. Dengan daerah kritis untuk *innovational Outliers* (IO):

Bentuk uji hipotesis untuk *Innovational Outliers* (IO) :

1. H_0 :Z_t bukan IO $\omega_{rr} = 0$)
   H_1 : Z_t adalah IO $\omega_{rr} \neq 0$)
2. Diambil ditingkat signifikan $\alpha = 0.05$
3. Statistic uji H_1 vs H_0 $\lambda_{2,r} = r\widehat{\omega}_{AT}/\sigma_a$
4. Dengan daerah kritis H_0 ditolak jika $P\left(\alpha_{hitung} \geq \lambda_{2,r}\right) < \alpha$

   Dibawah hipotesis H_0, $\lambda_{1,r}$ dan $\lambda_{2,r}$ berdistribusi N(0,1)

3.3 Analisis Intervensi

Pada data anomaly sebagaimana outlier, sering diketahui terlebih dahulu factor luar yang menyebabnya kejadian tak biasa tersebut. Faktor luar tersebut mendahului menjadi sebab terjadinya data anomaly. Prosdur parameterik dan non parametrik dengan syarat-syarat tertentu sulit untuk dijalankan. Berikut adalah strategi dalam meperlakukan data anomali tersebut:

1. Kerangkakan sebuah model untuk perubahan itu yang menggambarkan apa kira-kira yang terjadi dengan diberikannya pengetahuan itu dan intervensi yang diketahui
2. Kerjakan dengan tepat analisis data berdasarkan model tersebut
3. Jika cek diagnostic menunjukkan kekurangan dalam model, buatlah inferensi yang tepat. Jika perubahan tidak tertanggulangi, buatlah modifikasi model yang tepat, dan ulangi analisis

Anggap data … Z_{t-1}, Z_t, Z_{t+1},… adalah runtun data dengan interval waktu tertentu. kita memiliki model umum :

$$Z_t = f(I_t) + N_t \quad \ldots 10)$$

dengan $f(I_t)$ efek deterministic waktu, t , dan efek eksogen variabel I, dan fakta-fakta intervensi. dan N_t : proses stokastik dengan variansi atau *noise.*

Didefinisikan :

$$Z_t^* = Z_t - N_t = f(I_t),$$

Efek dari variabel eksogen bentuk $f(I_t)$ dapat dipresentasikan dengan model dinamik dengan bentuk :

$$f(\delta, \omega, I, t) = \sum_{j=1}^{k} Z_t^* = \sum_{j=1}^{k} \left\{\frac{\omega_j(B)}{\delta_j(B)}\right\} I_{tj} \quad ..11)$$

dengan Z_{tj}^* adalah fungsi transfer I_{tj} dan polinomial B : $\omega_j(B) = \omega_{0j} - \omega_{1j}B - \ldots - \omega_j B^{sj}$

$$\delta_j(B) = \delta_{oj} - \delta_{1j}B - \ldots - \delta_{r_j j}B^{rj}$$

dengan derajat r_j dan s_j.

Umumnya, masing-masing I_{rj} adalah faktor eksogen runtun waktu dengan nilai 0 dan 1 yang mendenotasikan kejadian intervensi.

Untuk ilustrasii, anggap variabel eksogen k=l, modelnya adalah

$$Z_t = f(\delta, \omega, I, t) + N_t$$
$$= \left(\frac{\omega_j(B)}{\delta_j(B)}\right) I_t + \left(\frac{\theta(B)}{\emptyset(B)}\right) a_t$$

Ada dua jenis fungsi untuk input dalam fungsi transfer dalam analisis inversi, yaitu :

Intervensi fungsi *step :*

$$S_t^{(T)} = \begin{cases} 0, t < T \\ 1, t \geq T \end{cases} \qquad \text{dan,}$$

Intervensi fungsi *pulse :*

$$P_t^{(T)} = \begin{cases} 1, t = T \\ 0, t \neq 0 \end{cases}$$

Fungsi *pulse* dapat diturunkan dari fungsi *step* $S_t^{(T)}$

$$P_t^{(T)} = S_t^{(T)} - S_{t-1}^{(T)}$$
$$= (1 - B)S_t^{(T)}$$

Oleh karena itu model intervensi dapat ditunjukkan dengan fungsi *step* atau fungsi *pulse.* Bentuk penggunaan biasanya didasarkan interprestasi yang sesuai.

Tujuan utama model intervensi adalah untuk mengukur efek intervensi. Sehingga, mengenai variabel intervensi I_{jt}, runtun waktu bebas intervensi disebut *noise series* dan dinotasikan dengan N_t, dan modelnya disebut dengan *noise model.*

Noise model biasanya diidentifikasi menggunakan prosedur identifikasi model univariat dari runtun Z_t sebelum data intervensi

$\{Z_t, t < T\}$. Jika cek diagnostic model menampakkan bahwa ini adalah bukan model yang kurang, maka dapat dibuat inferensi terhadap intervensi. Sebaliknya, jika inferensi terhadap intervensi ternyata tidak tepat, maka modifikasi model bisa dilakukan untuk mendapatkan model terbaik. Estimasi dan cek diagnostic selanjutnya diulangi.

Fungsi Z_t^* mempresentasikan efek additional dan bentuk intervensi dari *noise.* Kenyatannya, ketika N_t tidak stasioner, perubahan besar dapat terjadi dalam data yang datar tanpa intervensi. Memperbaiki model dapat membuat hal ini mungkin untuk membedakan apa yang dapat dan tidak dapat dijelaskan oleh *noise.*

Sedangkan untuk prosedur umum yang dalam intervensi analisis sendiri adalah :

1. Spesifikasi model Z_t menggunakan $\{X_t, \ldots, X_{r-1}\}$ yaitu data sebelum terjadinya intervensi
2. Gunakan model yang dibangun dari Z_t untuk memorediksi Z_t untuk $t \geq T$. Misal didapat $\hat{Z}_t$ untuk $t \geq T$
3. Hitung $X_t - \hat{Z}_t$ untuk $t \geq T$, untuk memperoleh $\omega(B)$ dan $\delta(B)$
4. Gunakan join estimation untuk semua data
5. Cek model tersebut atas ketidakcukupan model (cek diagnostic)
6. Bisa juga dibangun dengan model untuk intervensi yang multiple, dan kemudian cek kembali join estimasi yang dibangun dari model akhir tersebut.

Setelah *outliers* teridentifikasi, kemudian dimasukkan ke dalam model yang sesuai. Analisis dilanjutkan dengan data yang telah dikalkulasi ulang menggunakan model yang telah dirancang tersebut. Pendekatan yang baik ini bisa untuk mencari penyebab dari diteksinnya *outliers,* dan untuk pendekatan yang lebih halus bisa menggunakan model kombinasi yang telah dibahas sebelumnya.

Pendekatan ini tidak hanya untuk estimasi parameter, tapi juga untuk model *checkin*g dan *forecasting.* Dalam mencari penyebab sebuah *outliers,* salah satunya bisa dengan mencari penyebab adanya gangguan tersebut. Sebagai contoh beberapa *outliers* bisa menjadi variabel intervensi yang penting karena disebabkan oleh perubahan kebijakan. Padahal analist sebenarnya tidak familiar dengan itu sehingga terabaikan pada tahapan persiapan analisis.

Maka disamping menggunakan data yang telah disesuaikan, yaitu dengan menarik efek *outliers*, analist dapat memasukkan informasi ini ke dalam model dengan menggunakan variabel intervensi atau fungsi respon seperti yang telah didiskusikan sebelumnya.

Bentuk kombinasi dari intervensi-*outliers* model biasanya lebih berguna dalam *forecasting* dan control dari pada model univariat yang telah disesuaikan *outliers*nya.

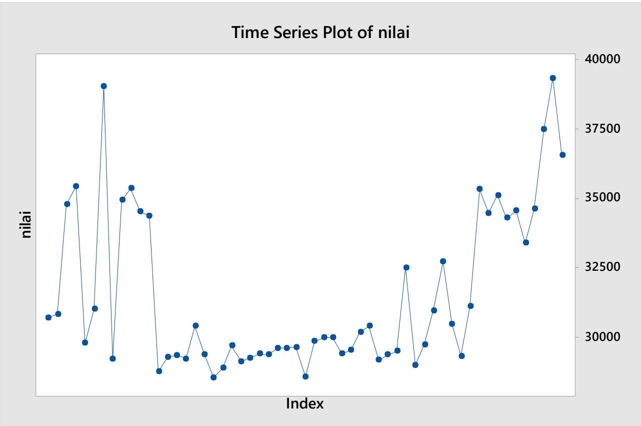
Prosedur deteksi *outliers* ini berdasarkan asumsi, bahwa model utama untuk runtun yang bebas *outliers* diketahui atau dapat diidentifikasi. Bagaimanapun dalam prakteknya, model utama biasanya tidak diketahui dan harus diidentifikasi dengan mengecek ACF dan PACF-nya.

Maka harus ditegaskan bahwa prosedur deteksi ini bekerja baik ketika efek *outliers* ini sifatnya moderat dan tidak akan mengalihkan dan mengaburkan pola pokok pada statistic sample dari data yang bebas *outliers.*

Dalam kasus yang lebih serius, adanya *outliers* bisa membuat identifikasi model tidak memungkinkan. Karena bisa jadi sebuah runtun yang *white noise* didalamnya terdapat *outliers* sebagai data empiris. Maka, ketika menemui fenomena *white noise* dalam menghitung sample statistic seperti ACF, satu yang harus diperiksa apakah disana terdapat *outliers* yang jelas-jelas dalam data. Lebih mudah memplot data terlebih dahulu untuk melihat apakah didalamnya terdapat *outliers.* Sehingga memudahkan dalam analisisnya.

3.4 Simulasi

Sebagai simulasi adalah data PDRB (PRoduk Domestik Regional Bruto) bidang Pertambgangan dan Penggalian Daerah Yogyakarta pada periodeo 1995 sampai 2009. Data quarter tahun diambil pada tiap bulan Februari, mei, agustus dan November. Berikut adalah grafik pergerakan runtun waktu dari data tersebut.



Gambar 1. Plot data Simulasi

Saat dilakukan didifferensingkan satu kali terindikasi terdapat efek musiman. Setelah dilakukan penyesuaian musiman dan deteksi ACF dan PACF, maka model terbaik ada pada ARIMA (1,0,0)(1,1,0)^4 yang aman dalam parameter dan juga check diagnostiknya telah memenuhi analisis runtun waktu. Diagnostik checking pada residual juga menunjukkan ACF dan PACF tidak menunjukkan autokorelasi dan heterogenitas residual. Estimasi model tersebut yaitu

$$(1-0.732B)(1+0.4537B)^4(1-B)^4\log(X_t)=a_t$$

Dengan $a_t \sim WN(0,\sigma^2)$.

Deteksi outlier ini menunjukkan adanya outlier jenis AO (*Additional Outlier)* sebesar 0.135 yang terjadi pada quarter ke-3 tahun 1996. Cara yang sama dilakukan oleh [8] pada data wisatawan domestik Yogyakarta. Perbandingan spesifikasi model sebelum dan sesudah deteksi outlier ditunjukkan oleh table sebagai berikut :

Tabel 1. Perbandingan sebelum dan setelah deteksi outlier

Kriteria/Model	**Dengan deteksi outlier**	**Tanpa deteksi outlier**
Stasionarr R-Square	0,579	0,475
R-Square	0,712	0,599
RMSE	1614,838	1885,551
MAE	1215173	1335,894
Normalized BIC	14,999	15,234

Dari table diatas tampak stasionary R-square dan R-Square menunjukkan angka yang lebih besar dengan deteksi outlier. Begitu jugan dengan RMSE, MAE dan Normalized BIS menunjukkan angka yang lebih kecil dengan dilakukannya deteksi outlier. Dengan demikian model ARIMA diatas akan lebih baik dengan adanya deteksi outlier yang dimodelkan.

Jika ourlier jenis AO ini merupakan sebuah data yang muncul dengan diketahui penyebabnya, maka analisis dapat dilakukan dengan analisis intervensi terlebih dahulu dan dilanjutkan dengan deteksi outlier. Sebab dalam hal ini tidak ada factor luar yang yang diasumsikan dapat mempengaruhi data maka analisis intervensi tidak dilakukan.

4 Kesimpulan Dan Saran

Dengan demikian dapat disimpulkan bahwa deteksi outliers dalam data runtun waktu diperlukan untuk mereduksi efek data yang tidak konsisten untuk memperoleh spesifikasi model ARMA yang lebih baik dari pada model yang dibangun sebelumnya. Karena outliers dapat mengaburkan nilai estimator ARMA dan selanjutnya akan mengacukan nilai peramalan di masa depan. Hal ini ditunjukkan pada data simulai yang memberikan nilai stasionary R-Square dan R-Square yang lebih tinggi serta spesifikasi error dan Normalized BIC yang lebih kecil. Outlier jenis AO pada data simulasi terdeteksi pada quarter ke 3 tahun 1996. Sebab dalam data ini tidak terdapat intervensi luar yang tidak diketahui maka analisis intervensi tidak dilakukan pada data.

Daftar Pustaka

[1] S. Makridakis, S. Wheelwright C, and V. E. McGee, *Metode dan Aplikasi Peramalan Edisi Kedua. Terjemahan dari Forecasting: Methods and Applications*. 1999.

[2] A. S. Ahmar *et al.*, “Modeling Data Containing Outliers using ARIMA Additive Outlier (ARIMA-AO),” *J. Phys. Conf. Ser.*, vol. 954, no. 1, 2018, doi: 10.1088/1742-6596/954/1/012010.

[3] G. U. Yule, “Why do we Sometimes get Nonsense-Correlations between Time-Series?--A Study in Sampling and the Nature of Time-Series,” *J. R. Stat. Soc.*, vol. 89, no. 1, 1926, doi: 10.2307/2341482.

[4] W. W. S. Wie, *Time Series Analysis: Univariate and Multivariate Methods Second Edition*, vol. SFB 373, no. Chapter 5. 2006.

[5] E. Slutzky, “The Summation of Random Causes as the Source of Cyclic Processes,” *Econometrica*, vol. 5, no. 2, 1937, doi: 10.2307/1907241.

[6] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: Forecasting and control: Fourth edition*. 2013. doi: 10.1002/9781118619193.

[7] D. Rosadi, “Pengantar Analisis Runtun Waktu,” *Diktat Kuliah, Progr. Stud. Stat. Fak. ...*, 2006.

[8] L. Budiarti, Tarno, and B. Warsito, “Analisis Intervensi dan Deteksi Outlier pada Data Wisatawan Domestik (Studi Kasus di Daerah Istimewa Yogyakarta),” *J. Gaussian*, vol. 2, no. 1, 2013.