

STUDI PENGGUNAAN DATA *GOOGLE TRENDS*:
KASUS PERAMALAN TINGKAT PENGANGGURAN USIA MUDA
A Study on the Use of Google Trends Data: The Case of Youth Unemployment
Forecasting

Siti Magfirotul Laelia & Dominicus Savio Priyarsono¹
 Fakultas Ekonomi dan Manajemen, Institut Pertanian Bogor

Abstract

Google Trends data is an unbiased sample of search data produced by the Google search machine. This data represents the number of searches for a specific keyword using the search machine. The availability of this free of charge data creates various opportunities of utilization for forecasting economic variables. On the other hand, data produced by standard methods (using surveys) cannot be issued very frequently because of financial constraints and the limited mobility during COVID-19 pandemic. This article reports a methodological study on the use of Google Trends data. The objective of this study is to show that the use of Google Trends data can improve the quality of forecasting. To achieve the objective, youth unemployment data are utilized and prepared for analysis of Autoregressive Integrated Moving Average. The quality of forecasting utilizing additional Google Trends data is compared with that without involving Google Trends data. It is concluded that the quality of the forecasting using the first method is better than that of the second one. This conclusion implies that utilizing Google Trends data can potentially improve the quality of forecasting and partially solve the problem of data scarcity resulted from financial constraints and the COVID-19 pandemic.

Key words: *Google Trends data, forecasting method, youth unemployment*

JEL: B23, J01

Abstrak

Data *Google Trends* adalah sampel takbias dari data pencarian yang dihasilkan oleh mesin pencari Google. Data ini menggambarkan jumlah pencarian untuk kata kunci tertentu melalui mesin pencari itu. Ketersediaan data gratis ini memunculkan berbagai peluang pemanfaatan antara lain peramalan variabel-variabel ekonomi. Di pihak lain, data yang dihasilkan dari metode standar (hasil survei) tidak bisa tersedia dengan frekuensi sangat tinggi karena ada kendala biaya dan keterbatasan mobilitas akibat pandemi COVID-19. Artikel ini melaporkan hasil studi metodologi penggunaan data tersebut. Tujuan studi ini adalah menunjukkan bahwa penggunaan data *Google Trends* dapat meningkatkan kualitas peramalan. Untuk itu dipilih data tingkat pengangguran usia muda yang dipersiapkan untuk analisis *Autoregressive Integrated Moving Average*. Selanjutnya, dilakukan perbandingan antara kualitas peramalan dengan menyertakan data *Google Trends* dan kualitas peramalan *tanpa* menyertakan data *Google Trends*. Dapat disimpulkan bahwa kualitas peramalan dengan metode yang pertama lebih baik daripada kualitas peramalan dengan metode yang kedua. Dengan simpulan itu dapat disarankan bahwa peramalan dengan menggunakan data *Google Trends* berpotensi meningkatkan kualitas peramalan dan dapat mengatasi sebagian persoalan kelangkaan data akibat adanya kendala biaya dan pandemi COVID-19.

Kata kunci: *data Google Trends, metode peramalan, tingkat pengangguran usia muda*

JEL: B23, J01

1. PENDAHULUAN

Penggunaan internet telah mempermudah masyarakat warganet dalam memperoleh berbagai informasi. Kini dapat diandaikan bahwa sebelum mengambil suatu keputusan untuk

¹ *Corresponding author:* D. S. Priyarsono, Departemen Ilmu Ekonomi, Fakultas Ekonomi dan Manajemen, Institut Pertanian Bogor, Kampus IPB University, Darmaga, Bogor 16680. Email: priyarsono@apps.ipb.ac.id, priyarsono@yahoo.com.

melakukan tindakan tertentu kebanyakan orang mempertimbangkan dulu informasi yang diperolehnya dari internet. Tak bisa dipungkiri, dalam fenomena itu peranan Google sangat menonjol.

Perilaku pencarian informasi di internet meninggalkan jejak digital yang terekam dalam sistem data yang disebut sebagai *big data*. Data *Google Trends* merupakan salah satu jenis *big data* yang menggambarkan jumlah pencarian untuk kata kunci tertentu melalui mesin pencari Google. Dalam beberapa literatur, data ini kemudian disebut sebagai *Google Econometrics* (Fondeur dan Karame, 2013), *Google Index* (Nagao *et al.*, 2019), atau *Search Volume Index* (González-Fernández dan González-Velasco, 2018). Data jenis ini kemudian menjadi bahan yang sangat informatif bagi riset untuk menelusuri dan meramalkan berbagai aktivitas (khususnya aktivitas ekonomi) secara *real time* (Woloszko, 2020; Bantis, Clements, dan Urquhart, 2021).

Pengangguran merupakan salah satu isu penting dalam perekonomian. Dalam kondisi yang umum terjadi, tingkat pengangguran usia muda (15-24 tahun) lebih tinggi dari tingkat pengangguran usia dewasa (25-60+ tahun). Rata-rata tingkat pengangguran usia muda Indonesia tahun 1997-2004 mencapai hampir dua kali lipat tingkat pengangguran usia dewasa (Rahman 2008). Pengangguran usia muda merupakan tantangan dalam perekonomian yang dihadapi negara maju sekaligus negara berkembang di dunia. Pengangguran usia muda secara ekonomi dapat menyebabkan ketidakstabilan pasart tenaga kerja, mereduksi tingkat kesejahteraan, rendahnya pendapatan pajak, dan pemborosan dalam investasi di bidang pendidikan dan pelatihan (ILO 2015). Mengurangi tingkat pengangguran usia muda adalah salah satu agenda utama Pembangunan Berkelanjutan (SDGs), terutama poin ke-8 tentang "*Decent work and economic growth*" (ILO 2017).

Lazimnya, data pengangguran dipublikasikan secara berkala dengan jeda waktu (*time lag*) yang cukup panjang. Di Indonesia data pengangguran dipublikasikan oleh Badan Pusat Statistik secara semesteran (BPS, 2020) yaitu pada bulan Februari dan Agustus setiap tahun. Di beberapa negara lain data pengangguran dipublikasikan secara bulanan. Untuk beberapa jenis kebijakan, baik publik atau pun swasta, ketersediaan data pengangguran secara tepat waktu atau bahkan secara *real time* sangat penting. Oleh karena itu, sering kali keharusan meramalkan data pengangguran menjadi tidak terhindarkan.

Beberapa penelitian terdahulu telah memanfaatkan data *Google Trends* untuk peramalan pengangguran usia muda. Fondeur dan Karame (2013) dan Naccarato *et al.* (2017) melakukan peramalan tingkat pengangguran usia muda berturut-turut untuk Prancis dan Italia. Dari kedua penelitian tersebut, hasil peramalan yang memanfaatkan data *Google Trends* menunjukkan galat peramalan (*forecast error*) yang lebih rendah dibanding model utama yang hanya menggunakan data pengangguran masa lalu. Dengan pendekatan yang serupa, D'Amuri dan Marcucci (2017) serta Nagao *et al.* (2019) juga mengkaji peramalan tingkat pengangguran di Amerika Serikat. Hasil kajian untuk kasus Spanyol telah dilaporkan oleh Vicente *et al.* (2015). Riset lain yang lebih mutakhir dengan topik yang agak berbeda namun masih terkait dengan persoalan pengangguran dilakukan oleh Aaronson *et al.* (2021). Peramalan tingkat pengangguran menggunakan data *Google Trends* diharapkan dapat memfasilitasi penyediaan data yang lebih cepat, sambil menunggu data yang asli dipublikasikan. Hal ini sesuai dengan karakteristik data *Google Trends* yang tersedia secara *real-time* sehingga mampu memberi gambaran fenomena ekonomi dan sosial saat itu juga. Selain itu, data ini dapat diperoleh dengan mudah dan gratis.

Belajar dari keberhasilan negara-negara lain dalam pemanfaatan data *Google Trends* untuk meningkatkan kualitas peramalan data pengangguran, maka studi ini bertujuan untuk mengkaji fenomena serupa untuk kasus Indonesia. Secara sengaja dipilih kasus yang lebih khusus, yakni kasus pengangguran usia muda, dengan asumsi bahwa tingkat penggunaan internet di kalangan kaum muda lebih tinggi daripada tingkat penggunaan internet secara umum yang mencakupi juga pengguna yang berusia lebih lanjut.

Memprediksi tingkat pengangguran usia muda di Indonesia, apalagi dengan peluang terjadinya bonus demografi pada 2030 penting untuk dilakukan. Hal ini akan berguna untuk mengestimasi jumlah lapangan kerja yang harus disiapkan serta menjadi bahan pertimbangan dalam merumuskan kebijakan yang perlu diterapkan. Untuk mencapai hal tersebut, diperlukan pengetahuan mengenai kondisi dan karakteristik pengangguran usia muda di tahun-tahun sebelumnya. Selain menjadi bahan evaluasi atas kebijakan terkait pengangguran yang telah diterapkan, dapat diketahui juga pola pertumbuhannya serta menganalisis penyebabnya.

2. METODE DAN DATA

Google Trends adalah layanan yang menampilkan data deret waktu dari intensitas pencarian di mesin pencari (*search engine*) Google untuk menunjukkan sejauh mana kata kunci tertentu dicari dalam periode dan lokasi tertentu. Intensitas pencarian distandardisasi dengan cara mengubah nilai maksimum selama periode tersebut ke angka 100 berdasarkan jumlah pencarian kata kunci spesifik relatif terhadap jumlah semua pencarian di lokasi yang ditentukan (Nagao *et al.* 2019). Data ini berbentuk indeks, yang di berbagai penelitian disebut sebagai *Google Econometric* (Fondeur dan Karame 2013), *Search Volume Index* atau SVI (Gonzalez-Fernandez dan Gonzalez-Velasco 2017), dan *Google Index* atau GI (Naccarato *et al.* 2017). Dalam Naccarato *et al.* (2017) model penghitungan *Google Index* dapat dituliskan seperti di bawah ini. *Google Trends* menghitung rasio jumlah pencarian *online* untuk kata kunci tertentu di wilayah geografis tertentu pada hari tertentu ($K_{s,t}$) hingga total jumlah pencarian *online* yang dilakukan melalui Google pada hari yang sama di area yang sama ($T_{s,t}$):

$$R_{s,t} = \frac{K_{s,t}}{T_{s,t}} \quad (1)$$

Untuk mempercepat perhitungan, angka $T_{s,t}$ dihitung berdasarkan bagian yang dipilih secara acak dari data pencarian. Akibatnya, angka bervariasi berdasarkan waktu pengambilan data. Rasio tersebut kemudian digunakan untuk menghitung tingkat mingguan sebagai rata-rata $R_{s,t}$ yang diambil setiap hari dalam minggu tertentu (γ):

$$R_{s,\gamma} = \frac{1}{7} \sum_{t=Minggu \ \gamma}^{Sabtu \ \gamma} R_{s,t} \quad (2)$$

Kemudian indeks $R_{s,\gamma}$ diskalakan ke 100 sebagai rasio terhadap nilai asli di minggu dimana ia mencapai level maksimum. Indeks Google (GI) untuk kata kunci spesifik K dalam minggu γ , dan di area s ditulis:

$$GI_{K,s,\gamma} = \frac{100}{\max_{\gamma}(R_{s,\gamma})} R_{s,\gamma} \quad (3)$$

yang mewakili persentase pengguna internet yang melakukan pencarian di Google untuk kata kunci tertentu di area tertentu dan dalam minggu tertentu.

Untuk alasan privasi, tidak ada nilai absolut dari komponen GI yang tersedia untuk umum. Data tersebut dikumpulkan dengan menggunakan alamat IP dan indeks hanya tersedia untuk umum jika jumlah pencarian melebihi ambang batas tertentu (tidak dideklarasikan nilainya). Kueri atau pencarian berulang dari satu alamat IP dalam waktu singkat dihilangkan atau tidak

dihitung (D'Amuri dan Marcucci, 2017). Data *Google Trends* ini tersedia secara gratis di <https://www.google.com/trends/?hl=id>. Pengguna dapat melihat data deret waktu dan mengunduhnya dalam format csv.

Pemilihan kata kunci merupakan langkah yang penting dalam penelitian ini. Perbedaan satu huruf saja akan mengarah pada hasil yang berbeda. Dalam mengekstraksi informasi dari *Google Trends*, kesalahan yang dapat terjadi salah satunya akibat pemilihan kata kunci yang tidak tepat. Hal ini dapat menyebabkan data mengalami korelasi semu atau palsu (Naccarato *et al.*, 2019). Beberapa penelitian terdahulu terkait topik ini menggunakan lebih dari satu kata kunci untuk mendapat data *Google Trends*. Untuk memperoleh informasi yang relevan, beberapa kata kunci dikurangi dengan menyortir kata yang paling relevan yang dapat dimasukkan sebagai variabel eksogen atau eksplanatori pada model deret waktu (Fondeur dan Karame, 2013). Nagao *et al.* (2019) menggunakan dua kata kunci untuk penelitian di negara Amerika Serikat, yaitu "*jobs*" dan "*jobs offer*". Vicente *et al.* (2015) juga menggunakan dua kata kunci untuk negara Spanyol, yaitu "*oferta de trabajo*" dan "*oferta de empleo*". Hasil penelitian Nagao *et al.* (2019) menyimpulkan penggunaan data *Google Trends* tidak begitu berpengaruh dalam meningkatkan akurasi peramalan. Adapun Fondeur dan Karame (2013) dan Naccarato *et al.* (2019) menggunakan satu kata kunci "*emploi*" dan "*offerte di lavoro*" masing-masing untuk negara Prancis dan Italia. Keduanya menunjukkan hasil peramalan yang lebih akurat dalam kasus yang melibatkan data *Google Trends*.

Dalam penelitian ini, kunci yang digunakan adalah "*lowongan kerja*". Dengan asumsi bahwa banyaknya jumlah pencarian kata "*lowongan kerja*" melalui mesin pencari Google berkorelasi positif dengan jumlah pengangguran. Beberapa pilihan kata kunci lain, seperti "*pekerjaan*" dan "*loker*" dianggap akan memberikan hasil yang bias karena keduanya mengarah ke definisi, bukan menggambarkan kegiatan mencari lapangan kerja.

Metode analisis yang digunakan dalam penelitian ini adalah ARIMA dan ARIMAX. ARIMA merupakan kombinasi antara model autoregresif (*autoregressive* atau AR) dan rata-rata bergerak (*moving average* atau MA), yang juga disebut sebagai model Box-Jenkins. Menurut Hanke *et al.* (2003) dalam Rahmawati (2006) model autoregresif merupakan fungsi dari nilai-nilai waktu lampau dalam suatu deret waktu, sehingga peramalannya bergantung pada nilai periode terdahulu. Sedangkan peramalan model rata-rata bergerak bergantung pada sisaan atau galat periode-periode sebelumnya. Dalam model ARIMA di penelitian ini, tingkat pengangguran muda dinyatakan sebagai fungsi dari nilai dan sisaan di periode sebelumnya. Model ARIMAX adalah pengembangan dari ARIMA dengan menyertakan variabel bebas lain di dalam persamaannya. Dalam hal ini, variabel *Search Volume Index* (SVI) dari *Google Trends* disertakan sebagai variabel bebas. Bentuk umum model ARIMA dan ARIMAX dalam penelitian ini adalah sebagai berikut:

$$YUR_t = \alpha_1 YUR_{t-1} + \alpha_2 YUR_{t-2} + \dots + \alpha_T YUR_{t-T} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_T e_{t-T} \quad (4)$$

$$YUR_t = \beta_1 YUR_{t-1} + \beta_2 YUR_{t-2} + \dots + \beta_T YUR_{t-T} + \varphi_1 SVI_t + v_t - \phi_1 v_{t-1} - \phi_2 v_{t-2} - \dots - \phi_T v_{t-T} \quad (5)$$

Keterangan:

YUR_t = Tingkat pengangguran usia muda

SVI_t = Indeks pencarian "*lowongan kerja*" melalui Google

$\alpha, \beta, \varphi, \theta, \phi$ = Koefisien

$t, t - 1, \dots$ = Periode data

e, v = Komponen *error*

3. PEMBAHASAN

a. Meramal dengan metode ARIMA dan ARIMAX

Bagian ini membahas hasil peramalan tingkat pengangguran muda di Indonesia dengan menggunakan metode *Autoregressive Integrated Moving Average* (ARIMA) dan *Autoregressive Integrated Moving Average Explanatory* (ARIMAX). Peramalan dilakukan dua kali, model yang pertama hanya menggunakan data tingkat pengangguran usia muda tahun 2005 sampai 2019 (ARIMA), sedangkan yang kedua (ARIMAX) dilakukan dengan menambahkan data *Search Volume Index* atau SVI pada model yang pertama. Hasil dari kedua peramalan kemudian dibandingkan untuk melihat apakah penggunaan data SVI dapat meningkatkan ketepatan peramalan dibanding hanya menggunakan data masa lalu untuk kasus tingkat pengangguran usia muda. Adapun *software* yang digunakan dalam analisis ini yaitu *E-views* 10.

Sebagai langkah pertama, dilakukan identifikasi kestasioneran data. Data tingkat pengangguran muda yang digunakan merupakan data semesteran (semi tahunan) dari tahun 2005 sampai 2019. Data terindikasi memiliki unsur *trend* menurun selama selang waktu penelitian. Pada tingkat level, data tingkat pengangguran muda terbukti stasioner dengan nilai *t-statistic* yang lebih besar dibanding *critical value* pada $\alpha = 5$ persen dan 10 persen (Tabel 1). Namun, penggunaan data tersebut pada level tidak menghasilkan kombinasi ordo maksimum AR dan MA yang signifikan sehingga dilakukan *differencing* satu kali. Dilakukan lagi tes secara formal dengan menggunakan *Augmented Dickey Fuller* (ADF) *test*. Data tingkat pengangguran muda terbukti stasioner dari uji ADF yang dilakukan pada tingkat *difference* (Tabel 2). Nilai *t-statistic* uji ADF pada data tersebut lebih besar dibandingkan dengan *critical value* yang dipakai dalam penelitian ini yaitu pada $\alpha = 5$ persen. Demikian juga dengan *critical value* pada level 1 persen dan 10 persen.

Data *Search Volume Index* (SVI) berupa data bulanan sehingga data dirata-ratakan menjadi semi tahunan untuk menyamakan periodenya dengan data tingkat pengangguran. Pada data ini juga dilakukan uji ADF untuk melihat kestasionerannya. Hasil menunjukkan bahwa data SVI terbukti stasioner pada tingkat level dengan nilai *t-statistic* pada $\alpha = 5$ persen lebih besar dibandingkan *critical value*-nya, begitu juga pada $\alpha = 10$ persen (Tabel 3). Dapat disimpulkan bahwa kedua data dapat dianalisis menggunakan ARIMA.

Tabel 1 Hasil uji ADF data tingkat pengangguran muda pada level

	t-Statistic	Prob.
Augmented Dickey-Fuller test statistic	-4.238153	0.0130
Test critical values:		
1% level	-4.356068	
5% level	-3.595026	
10% level	-3.233456	

Sumber: Laelia (2020)

Tabel 2 Hasil uji ADF data tingkat pengangguran muda pada *difference*

	t-Statistic	Prob.
Augmented Dickey-Fuller test statistic	-3.686790	0.0007
Test critical values: 1% level	-2.660720	
5% level	-1.955020	
10% level	-1.609070	

Sumber: Laelia (2020)

Tabel 3 Hasil uji ADF data SVI

	t-Statistic	Prob.
Augmented Dickey-Fuller test statistic	-2.149778	0.0331
Test critical values: 1% level	-2.669359	
5% level	-1.956406	
10% level	-1.608495	

Sumber: Laelia (2020)

Selanjutnya dilakukan estimasi parameter. Pada tahap ini dilakukan identifikasi signifikansi parameter dari model yang akan digunakan. Namun sebelumnya, ditentukan terlebih dulu ordo maksimal AR(p) dan MA(q) dengan melihat plot *autocorrelation* (ACF) dan *partial autocorrelation* (PACF).

Untuk model ARIMA, plot ACF dan PACF disajikan pada Lampiran 1. Dari plot ACF, diketahui terdapat beberapa autokorelasi parsial yang signifikan yaitu dari *lag* pertama hingga delapan, artinya ordo maksimal MA(q) sebanyak delapan. Sedangkan ordo maksimal AR(p) adalah satu dan tiga, digambarkan oleh plot PACF yang hanya signifikan pada *lag* pertama dan ketiga. Ordo yang digunakan pada derajat *differencing* atau derajat perbedaan adalah satu. Kombinasi model yang diperoleh dari plot ACF dan PACF yang ditentukan secara tentatif ada sebanyak dua puluh enam, yakni di antaranya adalah model ARIMA (0,1,1), ARIMA (0,1,2), ARIMA (0,1,3), ARIMA (0,1,4), ARIMA (0,1,5), ARIMA (0,1,6), ARIMA (0,1,7), ARIMA (0,1,8), ARIMA (1,1,0), ARIMA (1,1,1), ARIMA (1,1,2), ARIMA (1,1,3), ARIMA (1,1,4), ARIMA (1,1,5), ARIMA (1,1,6), ARIMA (1,1,7), ARIMA (1,1,8), ARIMA (3,1,0), ARIMA (3,1,1), ARIMA (3,1,2), ARIMA (3,1,3), ARIMA (3,1,4), ARIMA (3,1,5), ARIMA (3,1,6), ARIMA (3,1,7), dan ARIMA (3,1,8).

Pada tiap-tiap kombinasi model dilakukan estimasi dengan metode *Least Square* untuk memeriksa signifikansi parameter. Hanya terdapat enam model yang memiliki parameter signifikan. Model-model tersebut adalah ARIMA (0,1,3), ARIMA (0,1,4), ARIMA (0,1,5), ARIMA (1,1,0), ARIMA (1,1,1), dan ARIMA (3,1,0).

Untuk model ARIMAX disertakan data *Search Volume Index* atau data pencarian di Google atas kata kunci "*lowongan pekerjaan*" sebagai variabel independen. Plot ACF dan PACF model kedua ini disajikan pada Lampiran 1. Sama seperti pada model pertama, ordo maksimum untuk

AR(p) adalah 1 dan 3 sementara ordo maksimum MA(q) sebesar 8. Begitu juga dengan kombinasi model tentatifnya, terdapat dua puluh enam model.

Setelah dilakukan analisis regresi pada kedua model tersebut, diperoleh empat model dengan parameter yang signifikan. Model tersebut yaitu ARIMAX (1,1,0), ARIMAX (1,1,2), ARIMAX (3,1,0), dan ARIMAX (3,1,6).

b. Menentukan model ARIMA dan ARIMAX yang terbaik

Untuk memperoleh model terbaik yang dapat digunakan untuk proses peramalan dilakukan evaluasi berdasarkan tiga kriteria (Hanke dan Wichern, 2005). Pertama, koefisien di dalam model harus signifikan secara statistik. Artinya *p-value* yang dimiliki model harus lebih kecil dari taraf nyata yang digunakan dalam penelitian, yaitu sebesar 5 persen. Kedua, model harus *white noise* yaitu memiliki residual yang acak, dilihat dari nilai probabilitas Q statistik yang harus lebih besar dari $\alpha = 5$ persen. Ketiga, seleksi model didasarkan pada prinsip *parsimony*, yaitu memilih model dengan parameter paling efisien (jumlah sesedikit mungkin).

Untuk model ARIMA, dari enam model yang telah terbukti memiliki parameter signifikan dilakukan uji pada residualnya untuk melihat apakah residual sudah bersifat *white noise*. Kriteria dalam uji ini yaitu nilai probabilitas dari Q statistik harus lebih dari $\alpha = 5$ persen. Pada Lampiran 2 dapat dilihat bahwa hanya satu model yang memenuhi kondisi *white noise*. Model tersebut adalah ARIMA (3,1,0) dengan probabilitas sebesar 0.0001. Pada ARIMA (1,1,0) dan ARIMA (1,1,1) hampir semua *lag* tidak signifikan kecuali *lag* ketiga yang kurang dari 0.05. Sementara model lainnya memiliki beberapa *lag* yang signifikan sehingga juga tidak memenuhi kondisi *white noise*. Dengan demikian, model peramalan yang didapat adalah ARIMA (3,1,0). Model matematis ARIMA (3,1,0) dapat disajikan sebagai berikut:

$$\Delta YUR_t = \alpha_1 \Delta YUR_{t-3} + e_t \quad (6)$$

Karena $\Delta YUR_t = YUR_t - YUR_{t-1}$, maka persamaan (6) dapat disubstitusikan ke dalam variabel asli YUR_t seperti berikut:

$$YUR_t = YUR_{t-1} + \alpha_1 YUR_{t-3} - \alpha_1 YUR_{t-4} + e_t \quad (7)$$

Untuk model ARIMAX, sebelumnya telah diperoleh empat kombinasi model untuk peramalan 2 yang memiliki koefisien signifikan. Setelah dilakukan uji pada residualnya hanya satu model yang terbukti memenuhi kondisi *white noise*. Model ARIMAX (1,1,0) dan ARIMAX (1,1,2) memiliki beberapa nilai probabilitas Q statistik kurang dari 0,05, begitu juga dengan ARIMAX (3,1,6) yang pada *lag* ketiga terbukti signifikan sehingga ketiga model tersebut tidak memenuhi kriteria model terbaik. Seperti model ARIMA, pada model yang menambahkan variabel SVI ini kombinasi p, d, dan q yang diperoleh juga sama, yaitu ARIMAX (3,1,0).

Model matematis ARIMAX (3,1,0) dapat ditulis sebagai berikut:

$$\Delta YUR_t = \alpha_2 \Delta YUR_{t-3} + \beta_2 SVI_t + e_t \quad (8)$$

Persamaan (4) juga diubah ke dalam variabel asli YUR_t berdasarkan persamaan $\Delta YUR_t = YUR_t - YUR_{t-1}$, maka persamaan (32) dapat dituliskan sebagai berikut:

$$YUR_t = YUR_{t-1} + \alpha_2 YUR_{t-3} - \alpha_2 YUR_{t-4} + \beta_2 SVI_t + v_t \quad (9)$$

c. Membandingkan kualitas peramalan model ARIMA dan ARIMAX

Model ARIMA yang telah diperoleh selanjutnya digunakan untuk meramalkan tingkat pengangguran muda Indonesia. Sebagai contoh, diberikan hasil peramalan untuk periode satu tahun segera sesudah selang waktu data yang diobservasi, yaitu periode Agustus 2019 dan Februari 2020. Hasil peramalan dapat dilihat di Tabel 4. Peramalan dengan model ARIMA hanya menggunakan data tingkat pengangguran usia muda dengan senjang (*di-lag*-kan). Adapun data *Google Trends* atau SVI dengan kata kunci “lowongan kerja” digunakan sebagai variabel tambahan untuk peramalan dengan model ARIMAX. Dasar penggunaan data SVI ini yaitu sifat data yang tersedia secara *immediate* dan mampu memberi gambaran terkini tentang fenomena sosial dan ekonomi. Selain itu, data SVI dapat digunakan dengan biaya yang sangat rendah, bahkan gratis.

Tabel 4 Hasil Peramalan Tingkat Pengangguran Muda

Periode	ARIMA (%)	ARIMAX (%)
2019s2	18.55825	18.49301
2020s1	15.97678	15.40150

Sumber: Laelia (2020)

Dari Tabel 4 dapat dilihat bahwa peramalan yang menggunakan data tingkat pengangguran muda dan SVI (ARIMAX) memiliki keluaran yang lebih rendah. Namun bila dibandingkan dengan peramalan yang hanya menggunakan data tingkat pengangguran muda (ARIMA), margin hasil peramalan hanya berbeda pada angka desimalnya. Keduanya sama-sama berada pada angka 18 dan 15 persen berturut-turut untuk periode Agustus 2019 dan Februari 2020.

Setiap metode peramalan tidak dapat dijamin akan meramalkan kondisi di masa depan dengan tepat. Dengan kata lain, kesalahan prediksi atau *forecast error* dalam peramalan adalah kejadian yang lazim. Semakin kecil tingkat kesalahan yang diperoleh dari metode peramalan tertentu merupakan indikasi bahwa hasil peramalan semakin baik (mendekati tepat). Dalam penelitian ini untuk menilai metode peramalan yang lebih baik, digunakan indikator *mean squared error* (MSE), *mean absolute error* (MAE), dan *mean absolute percentage error* (MAPE) berdasarkan keseluruhan data yang digunakan dalam peramalan seperti yang disajikan pada Tabel 5.

Tabel 5 Perbandingan MSE, MAE dan MAPE

Indikator	ARIMA	ARIMAX
Mean Squared Error (MSE)	94.51	4.15
Mean Absolute Error (MAE)	9.28	1.73
Mean Absolute Percent Error (MAPE)	48.32	8.42

Sumber: Laelia (2020)

MSE menggambarkan perbedaan antara nilai aktual dan prediksi yang didapat dengan mengkuadratkan masing-masing perbedaan dari kedua nilai tersebut. Oleh karena itu, nilai MSE cenderung besar. Formula MSE sebagai berikut:

$$MSE = \frac{1}{n} \sum_{t=1}^n (Z_t - \hat{Z}_t)^2 \quad (10)$$

MAE mengukur besarnya rata-rata kesalahan dalam serangkaian prediksi, tanpa mempertimbangkan arahnya. Rumus menghitung MAE:

$$MAE = \frac{1}{n} \sum_{t=1}^n |Z_t - \hat{Z}_t| \quad (11)$$

MAPE merupakan rata-rata dari persentase kesalahan absolut. MAPE tidak dibatasi dengan skala tertentu dan mudah ditafsirkan. Namun nilai yang dihasilkan tidak terbatas, atau tidak terdefinisi jika nilai aktual sebesar nol. Jika nilai aktual sangat kecil (kurang dari satu), MAPE menghasilkan persentase kesalahan yang sangat besar (Kim dan Kim 2016). Formula MAPE dapat ditulis seperti:

$$MAE = \frac{100\%}{n} \sum_{t=1}^n \frac{Z_t - \hat{Z}_t}{Z_t} \quad (12)$$

dimana

Z_t = Nilai inflasi aktual

\hat{Z}_t = Nilai inflasi peramalan

MAE dan MAPE digunakan untuk menunjukkan seberapa besar kesalahan dari hasil peramalan, sementara MSE mewakili standar deviasi perbedaan antara nilai ramalan dan nilai aktual yang diamati. Semakin kecil nilai dari ketiga indikator tersebut, maka nilai hasil ramalan makin akurat. Dari Tabel 5 dapat dilihat bahwa metode ARIMAX menghasilkan nilai MSE, MAE dan MAPE yang lebih kecil. Dengan demikian dapat disimpulkan bahwa metode ARIMAX lebih baik dalam meramalkan tingkat pengangguran muda bila dibandingkan dengan metode ARIMA. Penggunaan data SVI atau data pencarian lowongan pekerjaan melalui Google terbukti meningkatkan akurasi peramalan tingkat pengangguran muda di Indonesia.

Selain itu, model ARIMAX yang menambahkan data SVI memiliki nilai *R-squared* sebesar 67.6 persen yaitu lebih tinggi dibanding model ARIMA yang nilainya 56.7 persen. Demikian juga dengan *adjusted R-squared*, model ARIMAX memiliki nilai yang lebih tinggi dibanding model ARIMA yang masing-masing sebesar 64.5 dan 54.9 persen. Hal ini menunjukkan bahwa keragaman pada tingkat pengangguran muda dapat dijelaskan dengan lebih baik oleh variabel SVI bersama dengan data deret waktu pengangguran usia muda dibandingkan dengan yang dapat dijelaskan oleh semata-mata hanya data deret waktu tingkat pengangguran muda saja. Hal ini juga menunjukkan bahwa data SVI dapat menggambarkan fenomena tingkat pengangguran usia muda di Indonesia. Kata kunci yang digunakan untuk mendapatkan data sudah tepat. Kata "lowongan kerja" lebih tepat dibanding kata-kata lainnya, dan dianggap dapat mencerminkan kondisi pengangguran usia muda.

4. SIMPULAN

Berdasarkan temuan tersebut dapat disimpulkan bahwa *search volume index* (SVI) yang menggunakan kata kunci "lowongan kerja" ternyata cukup baik untuk digunakan sebagai alat untuk mengestimasi tingkat pengangguran usia muda. Data SVI baru mulai tersedia pada tahun 2004. Di pihak lain, data tingkat pengangguran di Indonesia hanya tersedia secara semesteran.

Kendala-kendala demikian menyebabkan jumlah titik pengamatan yang digunakan analisis perbandingan antara model ARIMA dan ARIMAX cukup terbatas. Oleh karena itu dapat disarankan bahwa untuk penelitian mendatang hal ini perlu memperoleh perhatian. Terkait dengan hal tersebut, hasil penelitian Choi dan Varian (2009a, 2009b), Borup dan Schutte (2020), serta Medeiros dan Pires (2021) dapat menjadi rujukan yang baik.

















































Saran lain yang dapat disampaikan adalah penggabungan antara teknik-teknik peramalan dengan teknik-teknik yang lebih baru dengan pendekatan *machine learning* (Babii, Ghysels, dan Striaukas, 2021). atau *deep learning* (Petropoulos, 2021) atau pemanfaatan *big data* pada umumnya (Jun, Yoo, dan Choi, 2018). Pendekatan-pendekatan ini diperkirakan akan terus berkembang seiring dengan kemajuan sains data. Pengujian keefektifan metode-metode baru itu dari sudut pandang statistika juga merupakan wilayah baru bagi penelitian-penelitian di masa mendatang (Mavragani, Ochoa, dan Tsagarakis, 2018).

DAFTAR PUSTAKA

- Aaronson, D., Brave, S. A., Butters, R. A., Fogarty, M., Sacks, D. W., Seo, B. (2021). Forecasting Unemployment Insurance Claims in Realtime with Google Trends. *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2021.04.001>.
- Babii, A., Ghysels, E., Striaukas, J. (2021). Machine Learning Time Series Regressions with an Application to Nowcasting. *Journal of Business & Economic Statistics*. <https://doi.org/10.1080/07350015.2021.1899933>. Terbit daring 21 April 2021.
- Badan Pusat Statistik. (2020). Keadaan Angkatan Kerja di Indonesia 2005-2019. Jakarta.
- Bantis, E., Clements, M., Urquhart, A. (2021). *Forecasting GDP Growth Rates Using Google Trends in the United States and Brazil*. ICMA Centre, Henley Business School, University of Reading.
- Borup, D., Schutte, E. C. M. (2020). In Search of a Job: Forecasting Employment Growth Using Google Trends. *Journal of Business & Economic Statistics*. 40:1, 186-200.
- Choi, H. & Varian, H. (2009a). Predicting Initial Claims for Unemployment Benefits. Google Inc.
- Choi, H. & Varian, H. (2009b). Predicting the Present with Google Trends. Google Inc.
- D'Amuri, F. & Marcucci, J. (2017). The Predictive Power of Google Searches in Forecasting US Unemployment. *International Journal of Forecasting*. 33: 801-816.
- Fondeur, Y. & Karamé, F. (2013). Can Google Data Help Predict French Youth Unemployment? *Economic Modelling*. 30: 117-125.
- González-Fernández, M. & González-Velasco, C. (2018). Can Google Econometrics Predict Unemployment? Evidence from Spain. *Economics Letters*. 170: 42-45.
- Hanke, J. E. & Wichern, D. W. (2005). *Business Forecasting*, Eighth Edition. Pearson Education: New Jersey.
- International Labour Organization. (2015). *Toward Solutions for Youth Employment*. https://www.ilo.org/wcmsp5/groups/public/---ed_emp/documents/publication/wcms_413826.pdf
- International Labour Organization. (2017). *Global Employment Trends For Youth 2017*. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_598669.pdf
- Jun, S. P., Yoo, H. S., Choi, S. (2018). Ten Years of Research Change Using Google Trends: From the Perspective of Big Data Utilizations and Applications. *Technological Forecasting & Social Change*. 130:69-87. <https://doi.org/10.1016/j.techfore.2017.11.009>.

- Laelia, S. M. (2020). Dinamika dan Peramalan Tingkat Pengangguran Usia Muda di Indonesia dengan Menggunakan Data *Google Trends*. [Skripsi]. Bogor (ID): Institut Pertanian Bogor.
- Mavragani, A., Ochoa, G., Tsagarakis, K. P. (2018). Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. *Journal of Medical Internet Research*. 20:11. <https://doi.org/10.2196/jmir.9366>.
- Medeiros M. C., Pires, H. F. (2021). *The Proper Use of Google Trends in Forecasting Models*. Pontifical Catholic University of Rio de Janeiro. <https://arxiv.org/pdf/2104.03065.pdf>.
- Naccarato, A., Falorsi, S., Loriga, S., Pierini, A. (2018). Combining Official and Google Trends Data to Forecast the Italian Youth Unemployment Rate. *Technological Forecasting & Social Change*. 130: 114-122.
- Nagao, S., Takeda, F., Tanaka, R. (2019). Nowcasting of the U.S. Unemployment Rate Using Google Trends. *Finance Research Letters*. 30: 103-109.
- Petropoulos, A., Siakoulis, V., Stavroulakis, E., Lazaris, P., Vlachogiannakis, N. (2021). Employing Google Trends and Deep Learning in Forecasting Financial Market Turbulence. *Journal of Behavioral Finance*. <https://doi.org/10.1080/15427560.2021.1913160>. Terbit daring 25 Mei 2021.
- Rahman, A. (2008). Analisis Eksistensi Persistensi Pengangguran di Indonesia. [Skripsi]. Bogor (ID): Institut Pertanian Bogor.
- Vicente, M. R., López-Menéndez, A. J., Pérez, R. (2015). Forecasting Unemployment with Internet Search Data: Does it Help to Improve Predictions When Job Destruction is Skyrocketing? *Technological Forecasting & Social Change*. 92:132-139.
- Woloszko, N. (2020). Tracking Activity in Real Time with Google Trends. *OECD Working Papers*. www.oecd.org/eo/workingpapers.

Date: 04/22/20 Time: 18:25
 Sample: 2005S1 2019S1
 Included observations: 28

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.706	-0.706	15.489	0.000
		2	0.651	0.304	29.168	0.000
		3	-0.722	-0.413	46.692	0.000
		4	0.567	-0.143	57.952	0.000
		5	-0.509	0.046	67.421	0.000
		6	0.458	-0.163	75.413	0.000
		7	-0.439	-0.152	83.121	0.000
		8	0.373	-0.044	88.974	0.000
		9	-0.323	-0.044	93.597	0.000
		10	0.326	-0.019	98.546	0.000
		11	-0.259	0.036	101.85	0.000
		12	0.231	-0.016	104.65	0.000
		13	-0.279	-0.181	108.99	0.000
		14	0.254	0.064	112.86	0.000
		15	-0.275	-0.146	117.74	0.000
		16	0.259	-0.144	122.43	0.000
		17	-0.299	-0.133	129.24	0.000
		18	0.310	-0.050	137.30	0.000
		19	-0.234	0.038	142.44	0.000
		20	0.250	-0.132	149.00	0.000
		21	-0.253	-0.067	156.69	0.000
		22	0.170	-0.172	160.71	0.000
		23	-0.167	-0.097	165.41	0.000
		24	0.183	-0.011	172.42	0.000

Lampiran 2. Hasil Uji Parameter dan Residual ARIMA

a) Hasil Uji Model ARIMA (0,1,3)

Dependent Variable: D(YUR)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 04/21/20 Time: 18:57
 Sample: 2005S2 2018S1
 Included observations: 26
 Convergence achieved after 14 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(3)	-0.659148	0.243577	-2.706122	0.0123
SIGMASQ	6.563356	3.110156	2.110298	0.0454

R-squared	0.404256	Mean dependent var	-0.473240
Adjusted R-squared	0.379434	S.D. dependent var	3.384932
S.E. of regression	2.666515	Akaike info criterion	4.938955
Sum squared resid	170.6473	Schwarz criterion	5.035732
Log likelihood	-62.20642	Hannan-Quinn criter.	4.966824
Durbin-Watson stat	2.484270		

Inverted MA Roots			
	.87	-.44+.75i	-.44-.75i

Date: 04/21/20 Time: 18:57
 Sample: 2005S1 2018S1
 Included observations: 26
 Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.603	-0.603	10.586	
		2 0.463	0.156	17.086	0.000
		3 -0.493	-0.263	24.775	0.000
		4 0.448	0.084	31.415	0.000
		5 -0.419	-0.097	37.514	0.000
		6 0.316	-0.094	41.153	0.000
		7 -0.338	-0.107	45.543	0.000
		8 0.255	-0.125	48.164	0.000
		9 -0.237	-0.063	50.569	0.000
		10 0.256	0.028	53.556	0.000
		11 -0.195	0.013	55.410	0.000
		12 0.139	-0.082	56.418	0.000
		13 -0.202	-0.161	58.693	0.000
		14 0.214	0.004	61.478	0.000
		15 -0.284	-0.245	66.799	0.000
		16 0.197	-0.159	69.626	0.000
		17 -0.210	-0.131	73.196	0.000
		18 0.295	0.025	81.124	0.000
		19 -0.217	0.014	86.032	0.000
		20 0.187	-0.136	90.288	0.000
		21 -0.127	0.028	92.645	0.000
		22 0.119	-0.124	95.238	0.000
		23 -0.136	-0.122	99.703	0.000
		24 0.181	0.080	111.68	0.000

b) Hasil Uji Model ARIMA (0,1,4)

Dependent Variable: D(YUR)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 04/21/20 Time: 18:56
 Sample: 2005S2 2018S1
 Included observations: 26
 Convergence achieved after 13 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(4)	0.614962	0.295778	2.079136	0.0485
SIGMASQ	7.167810	2.817496	2.544035	0.0178
R-squared	0.349391	Mean dependent var		-0.473240
Adjusted R-squared	0.322282	S.D. dependent var		3.384932
S.E. of regression	2.786598	Akaike info criterion		5.034299
Sum squared resid	186.3631	Schwarz criterion		5.131075
Log likelihood	-63.44588	Hannan-Quinn criter.		5.062167
Durbin-Watson stat	2.574998			
Inverted MA Roots	.63+.63i	.63+.63i	-.63+.63i	-.63+.63i

Date: 04/21/20 Time: 18:56
 Sample: 2005S1 2018S1
 Included observations: 26
 Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.366	-0.366	3.9001	
		2 0.402	0.309	8.7971	0.003
		3 -0.535	-0.409	17.858	0.000
		4 0.141	-0.244	18.517	0.000
		5 -0.400	-0.235	24.072	0.000
		6 0.196	-0.188	25.467	0.000
		7 -0.213	-0.255	27.202	0.000
		8 0.304	-0.113	30.949	0.000
		9 -0.067	-0.055	31.139	0.000
		10 0.246	-0.056	33.889	0.000
		11 -0.101	0.006	34.385	0.000
		12 0.096	0.037	34.867	0.000
		13 -0.231	-0.083	37.865	0.000
		14 0.064	0.030	38.113	0.000
		15 -0.320	-0.267	44.908	0.000
		16 0.219	-0.058	48.388	0.000
		17 -0.112	0.009	49.411	0.000
		18 0.301	-0.064	57.642	0.000
		19 -0.100	-0.067	58.688	0.000
		20 0.098	-0.223	59.863	0.000
		21 -0.143	-0.099	62.852	0.000
		22 0.050	-0.107	63.309	0.000
		23 -0.115	-0.107	66.532	0.000
		24 0.130	0.069	72.713	0.000

c) Hasil Uji Model ARIMA (0,1,5)

Dependent Variable: D(YUR)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 04/21/20 Time: 18:55
 Sample: 2005S2 2018S1
 Included observations: 26
 Convergence achieved after 11 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(5)	-0.541271	0.253462	-2.135513	0.0431
SIGMASQ	8.112145	3.174523	2.555390	0.0174

R-squared	0.263676	Mean dependent var	-0.473240
Adjusted R-squared	0.232995	S.D. dependent var	3.384932
S.E. of regression	2.964483	Akaike info criterion	5.151652
Sum squared resid	210.9158	Schwarz criterion	5.248429
Log likelihood	-64.97148	Hannan-Quinn criter.	5.179520
Durbin-Watson stat	2.719500		

Inverted MA Roots				
.88	.27-.84i	.27+.84i	-.72-.52i	-.72+.52i

Date: 04/21/20 Time: 18:56
 Sample: 2005S1 2018S1
 Included observations: 26
 Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.516	-0.516	7.7668	
		2 0.295	0.039	10.411	0.001
		3 -0.602	-0.594	21.864	0.000
		4 0.428	-0.185	27.927	0.000
		5 -0.120	0.116	28.425	0.000
		6 0.225	-0.090	30.268	0.000
		7 -0.303	-0.122	33.795	0.000
		8 0.069	-0.081	33.988	0.000
		9 -0.065	-0.124	34.170	0.000
		10 0.206	-0.034	36.094	0.000
		11 -0.107	-0.061	36.650	0.000
		12 0.096	0.121	37.128	0.000
		13 -0.231	-0.097	40.115	0.000
		14 0.200	-0.001	42.533	0.000
		15 -0.227	-0.215	45.947	0.000
		16 0.268	-0.136	51.180	0.000
		17 -0.234	-0.093	55.614	0.000
		18 0.193	-0.181	58.999	0.000
		19 -0.162	0.015	61.721	0.000
		20 0.166	-0.003	65.071	0.000
		21 -0.104	-0.028	66.648	0.000
		22 0.052	-0.079	67.139	0.000
		23 -0.089	-0.131	69.081	0.000
		24 0.102	0.014	72.863	0.000

d) Hasil Uji Model ARIMA (1,1,0)

Dependent Variable: D(YUR)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 04/21/20 Time: 18:45
 Sample: 2005S2 2018S1
 Included observations: 26
 Convergence achieved after 28 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.709589	0.186043	-3.814115	0.0008
SIGMASQ	5.876699	2.084973	2.818597	0.0095

R-squared	0.466583	Mean dependent var	-0.473240
Adjusted R-squared	0.444357	S.D. dependent var	3.384932
S.E. of regression	2.523177	Akaike info criterion	4.789649
Sum squared resid	152.7942	Schwarz criterion	4.886426
Log likelihood	-60.26544	Hannan-Quinn criter.	4.817518
Durbin-Watson stat	1.286032		

Inverted AR Roots	-0.71
-------------------	-------

Date: 04/21/20 Time: 18:45
 Sample: 2005S1 2018S1
 Included observations: 26
 Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.227	0.227	1.5052	
		2	-0.066	-0.125	1.6390	0.200
		3	-0.418	-0.399	7.1749	0.028
		4	0.025	0.256	7.1956	0.066
		5	-0.107	-0.287	7.5951	0.108
		6	-0.017	-0.112	7.6053	0.179
		7	-0.178	-0.037	8.8186	0.184
		8	0.072	-0.047	9.0281	0.251
		9	0.074	0.040	9.2654	0.320
		10	0.186	0.055	10.834	0.287
		11	0.004	-0.006	10.835	0.371
		12	-0.020	-0.029	10.855	0.455
		13	-0.146	-0.070	12.047	0.442
		14	-0.083	-0.080	12.464	0.490
		15	-0.190	-0.199	14.853	0.388
		16	-0.010	0.035	14.860	0.462
		17	-0.002	-0.077	14.860	0.535
		18	0.180	0.026	17.806	0.401
		19	0.054	0.009	18.112	0.448
		20	0.063	-0.073	18.599	0.483
		21	-0.118	-0.078	20.640	0.419
		22	-0.089	-0.137	22.081	0.395
		23	-0.044	0.083	22.541	0.428
		24	0.091	0.005	25.541	0.323

e) Hasil Uji Model ARIMA (1,1,1)

Dependent Variable: D(YUR)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 04/21/20 Time: 18:44
 Sample: 2005S2 2018S1
 Included observations: 26
 Convergence achieved after 34 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.981020	0.040243	-24.37752	0.0000
MA(1)	0.679985	0.196752	3.456054	0.0021
SIGMASQ	4.110908	1.773879	2.317468	0.0297

R-squared	0.626860	Mean dependent var	-0.473240
Adjusted R-squared	0.594414	S.D. dependent var	3.384932
S.E. of regression	2.155716	Akaike info criterion	4.547739
Sum squared resid	106.8836	Schwarz criterion	4.692904
Log likelihood	-56.12061	Hannan-Quinn criter.	4.589542
Durbin-Watson stat	1.680566		

Inverted AR Roots	-.98
Inverted MA Roots	-.68

Date: 04/21/20 Time: 18:44
 Sample: 2005S1 2018S1
 Included observations: 26
 Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.058	0.058	0.0976	
		2	-0.110	-0.113	0.4625	
		3	-0.356	-0.348	4.4701	0.034
		4	0.034	0.061	4.5081	0.105
		5	0.044	-0.034	4.5767	0.206
		6	-0.105	-0.256	4.9779	0.290
		7	-0.051	-0.003	5.0769	0.407
		8	-0.062	-0.106	5.2299	0.515
		9	0.176	0.048	6.5493	0.477
		10	0.096	0.080	6.9719	0.540
		11	0.064	0.028	7.1732	0.619
		12	-0.150	-0.085	8.3424	0.595
		13	-0.031	0.046	8.3965	0.677
		14	-0.079	-0.110	8.7761	0.722
		15	-0.067	-0.143	9.0757	0.767
		16	-0.074	-0.060	9.4776	0.799
		17	0.027	-0.036	9.5347	0.848
		18	0.138	0.017	11.273	0.792
		19	0.056	-0.005	11.603	0.824
		20	0.025	-0.016	11.681	0.863
		21	-0.094	-0.058	12.957	0.841
		22	-0.128	-0.172	15.926	0.721
		23	0.005	0.024	15.932	0.773
		24	0.073	0.035	17.851	0.715

f) Hasil Uji Model ARIMA (3,1,0)

Dependent Variable: D(YUR)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 04/21/20 Time: 18:35
 Sample: 2005S2 2018S1
 Included observations: 26
 Convergence achieved after 29 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(3)	-0.782282	0.165574	-4.724656	0.0001
SIGMASQ	4.766501	1.850918	2.575209	0.0166
R-squared	0.567354	Mean dependent var		-0.473240
Adjusted R-squared	0.549327	S.D. dependent var		3.384932
S.E. of regression	2.272380	Akaike info criterion		4.662566
Sum squared resid	123.9290	Schwarz criterion		4.759342
Log likelihood	-58.61335	Hannan-Quinn criter.		4.690434
Durbin-Watson stat	2.071461			
Inverted AR Roots	.46+.80i	.46-.80i		-.92

Date: 04/21/20 Time: 18:35
 Sample: 2005S1 2018S1
 Included observations: 26
 Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.270	-0.270	2.1274	
		2 0.202	0.140	3.3702	0.066
		3 -0.166	-0.089	4.2419	0.120
		4 0.174	0.099	5.2462	0.155
		5 -0.143	-0.054	5.9566	0.202
		6 -0.204	-0.330	7.4681	0.188
		7 -0.046	-0.130	7.5504	0.273
		8 0.031	0.044	7.5902	0.370
		9 0.041	0.066	7.6632	0.467
		10 0.053	0.145	7.7915	0.555
		11 -0.042	-0.068	7.8768	0.641
		12 -0.024	-0.236	7.9066	0.722
		13 -0.114	-0.260	8.6324	0.734
		14 0.065	0.014	8.8865	0.781
		15 -0.201	-0.064	11.564	0.641
		16 0.051	0.056	11.750	0.698
		17 -0.102	-0.062	12.585	0.703
		18 0.153	-0.110	14.726	0.615
		19 -0.055	-0.131	15.035	0.660
		20 0.073	-0.042	15.680	0.679
		21 0.028	0.067	15.791	0.730
		22 0.005	0.016	15.796	0.781
		23 -0.036	-0.089	16.103	0.811
		24 0.095	0.051	19.420	0.677

Lampiran 3. Hasil Uji Parameter dan Residual ARIMAX

a) Hasil Uji Model ARIMAX (1,1,0)

Dependent Variable: D(YUR)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 04/21/20 Time: 18:23
 Sample: 2005S2 2018S1
 Included observations: 26
 Convergence achieved after 11 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SVI	-0.011084	0.004556	-2.432845	0.0232
AR(1)	-0.780007	0.145583	-5.357815	0.0000
SIGMASQ	4.773383	1.964227	2.430159	0.0233
R-squared	0.566729	Mean dependent var		-0.473240
Adjusted R-squared	0.529053	S.D. dependent var		3.384932
S.E. of regression	2.322929	Akaike info criterion		4.667761
Sum squared resid	124.1080	Schwarz criterion		4.812926
Log likelihood	-57.68089	Hannan-Quinn criter.		4.709563
Durbin-Watson stat	1.356303			
Inverted AR Roots	-0.78			

Date: 04/21/20 Time: 18:23
 Sample: 2005S1 2018S1
 Included observations: 26
 Q-statistic probabilities adjusted for 1 ARMA term and 1 dynamic regressor

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*	
		1	0.274	0.274	2.1842	
		2	-0.200	-0.297	3.3921	0.066
		3	-0.479	-0.387	10.643	0.005
		4	-0.067	0.171	10.791	0.013
		5	-0.092	-0.383	11.081	0.026
		6	-0.073	-0.211	11.275	0.046
		7	-0.181	-0.203	12.528	0.051
		8	0.080	-0.151	12.785	0.078
		9	0.176	-0.085	14.116	0.079
		10	0.266	0.006	17.330	0.044
		11	0.046	-0.062	17.431	0.065
		12	-0.052	-0.051	17.573	0.092
		13	-0.136	-0.019	18.610	0.098
		14	-0.102	-0.145	19.247	0.116
		15	-0.159	-0.175	20.923	0.104
		16	-0.039	-0.049	21.035	0.136
		17	0.046	-0.123	21.209	0.171
		18	0.195	-0.016	24.669	0.102
		19	0.119	-0.050	26.143	0.097
		20	0.066	-0.054	26.667	0.113
		21	-0.109	-0.158	28.406	0.100
		22	-0.113	-0.162	30.743	0.078
		23	-0.032	0.020	30.992	0.096
		24	0.062	-0.077	32.378	0.093

*Probabilities may not be valid for this equation specification.

b) Hasil Uji Model ARIMAX (1,1,2)

Dependent Variable: D(YUR)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 04/21/20 Time: 18:21
 Sample: 2005S2 2018S1
 Included observations: 26
 Convergence achieved after 27 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SVI	-0.009970	0.001108	-8.996226	0.0000
AR(1)	-1.000000	0.456711	-2.189568	0.0395
MA(2)	-0.999785	0.000152	-6592.473	0.0000
SIGMASQ	3.442061	1.214350	2.834489	0.0096
R-squared	0.687570	Mean dependent var	-0.473240	
Adjusted R-squared	0.644966	S.D. dependent var	3.384932	
S.E. of regression	2.016901	Akaike info criterion	4.608516	
Sum squared resid	89.49359	Schwarz criterion	4.802069	
Log likelihood	-55.91070	Hannan-Quinn criter.	4.664252	
Durbin-Watson stat	1.073387			
Inverted AR Roots	-1.00			
Inverted MA Roots	1.00	-1.00		

Date: 04/21/20 Time: 18:22
 Sample: 2005S1 2018S1
 Included observations: 26
 Q-statistic probabilities adjusted for 2 ARMA terms and 1 dynamic regressor

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*
1	0.415	0.415	5.0161		
2	0.241	0.083	6.7719		
3	-0.159	-0.346	7.5674	0.006	
4	-0.022	0.186	7.5838	0.023	
5	-0.157	-0.135	8.4374	0.038	
6	-0.160	-0.213	9.3687	0.053	
7	-0.125	0.161	9.9643	0.076	
8	-0.047	-0.058	10.053	0.122	
9	0.123	0.115	10.703	0.152	
10	0.058	-0.028	10.857	0.210	
11	0.099	-0.018	11.336	0.253	
12	-0.083	-0.093	11.693	0.306	
13	-0.097	-0.122	12.224	0.347	
14	-0.239	-0.102	15.700	0.205	
15	-0.262	-0.181	20.226	0.090	
16	-0.183	0.069	22.673	0.066	
17	-0.090	-0.037	23.325	0.077	
18	0.057	-0.016	23.618	0.098	
19	0.031	-0.009	23.717	0.127	
20	0.067	-0.107	24.264	0.147	
21	-0.017	-0.042	24.303	0.185	
22	-0.017	-0.078	24.354	0.227	
23	-0.007	0.095	24.367	0.276	
24	0.063	0.109	25.815	0.260	

*Probabilities may not be valid for this equation specification.

c) Hasil Uji Model ARIMAX (3,1,0)

Dependent Variable: D(YUR)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 04/21/20 Time: 18:13
 Sample: 2005S2 2018S1
 Included observations: 26
 Convergence achieved after 11 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SVI	-0.011094	0.004444	-2.496471	0.0202
AR(3)	-0.847915	0.146931	-5.770843	0.0000
SIGMASQ	3.566329	1.403257	2.541465	0.0182
R-squared	0.676291	Mean dependent var		-0.473240
Adjusted R-squared	0.648142	S.D. dependent var		3.384932
S.E. of regression	2.007860	Akaike info criterion		4.486636
Sum squared resid	92.72455	Schwarz criterion		4.631801
Log likelihood	-55.32626	Hannan-Quinn criter.		4.528438
Durbin-Watson stat	2.533602			
Inverted AR Roots	.47+.82i	.47-.82i		-.95

Date: 04/21/20 Time: 18:13
 Sample: 2005S1 2018S1
 Included observations: 26
 Q-statistic probabilities adjusted for 1 ARMA term and 1 dynamic regressor

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*
		1 -0.346	-0.346	3.4820	
		2 0.092	-0.032	3.7369	0.053
		3 -0.149	-0.145	4.4399	0.109
		4 0.131	0.040	5.0041	0.171
		5 -0.105	-0.052	5.3854	0.250
		6 -0.285	-0.409	8.3357	0.139
		7 0.031	-0.259	8.3720	0.212
		8 0.052	-0.085	8.4828	0.292
		9 0.078	-0.010	8.7418	0.365
		10 0.069	0.162	8.9599	0.441
		11 0.013	0.054	8.9680	0.535
		12 -0.006	-0.189	8.9701	0.625
		13 -0.074	-0.276	9.2733	0.679
		14 0.033	-0.103	9.3412	0.747
		15 -0.148	-0.115	10.796	0.702
		16 0.035	0.063	10.885	0.761
		17 -0.061	-0.021	11.191	0.798
		18 0.118	-0.134	12.459	0.772
		19 -0.037	-0.261	12.599	0.815
		20 0.068	-0.200	13.161	0.830
		21 0.055	-0.008	13.603	0.850
		22 -0.039	0.049	13.874	0.875
		23 -0.030	-0.036	14.088	0.898
		24 0.077	0.026	16.222	0.846

*Probabilities may not be valid for this equation specification.

d) Hasil Uji Model ARIMAX (3,1,6)

Dependent Variable: D(YUR)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 04/21/20 Time: 18:25
 Sample: 2005S2 2018S1
 Included observations: 26
 Convergence achieved after 13 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SVI	-0.009657	0.001952	-4.947551	0.0001
AR(3)	-0.999999	7.33E-05	-13650.14	0.0000
MA(6)	-0.997836	0.002395	-416.6050	0.0000
SIGMASQ	1.841987	1.018691	1.808190	0.0843
R-squared	0.832806	Mean dependent var		-0.473240
Adjusted R-squared	0.810007	S.D. dependent var		3.384932
S.E. of regression	1.475430	Akaike info criterion		4.290906
Sum squared resid	47.89165	Schwarz criterion		4.484459
Log likelihood	-51.78178	Hannan-Quinn criter.		4.346642
Durbin-Watson stat	2.570943			
Inverted AR Roots	.50+.87i	.50-.87i	-1.00	
Inverted MA Roots	1.00	.50+.87i	.50-.87i	-.50+.87i
		-1.00		

Date: 04/21/20 Time: 18:49
 Sample: 2005S1 2018S1
 Included observations: 26
 L-statistic probabilities adjusted for 2 ARMA terms

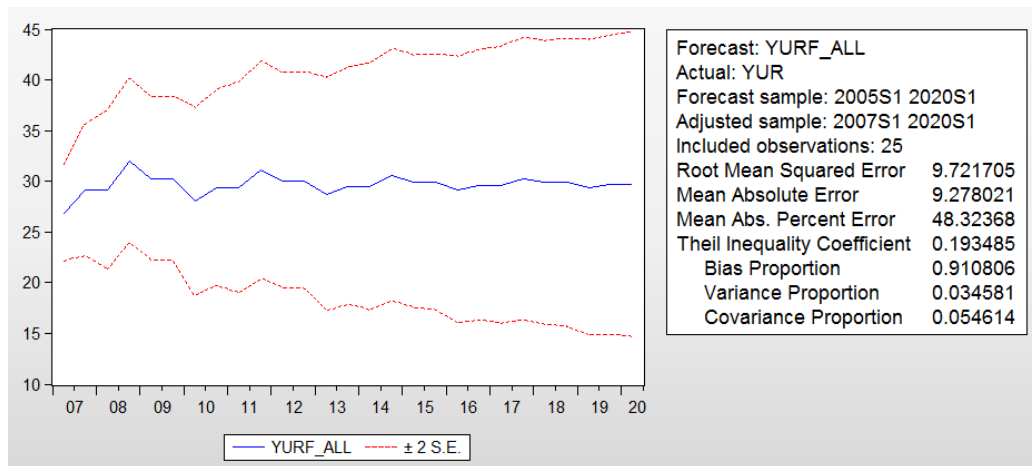
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.281	-0.281	2.3025	
		2	0.205	0.137	3.5816	
		3	-0.163	-0.083	4.4235	0.035
		4	0.172	0.097	5.3990	0.067
		5	-0.151	-0.063	6.1935	0.103
		6	-0.160	-0.286	7.1216	0.130
		7	-0.046	-0.109	7.2015	0.206
		8	0.019	0.016	7.2157	0.301
		9	0.035	0.057	7.2678	0.402
		10	0.068	0.147	7.4766	0.486
		11	-0.050	-0.066	7.5995	0.575
		12	-0.074	-0.258	7.8846	0.640
		13	-0.107	-0.262	8.5299	0.665
		14	0.075	0.020	8.8692	0.714
		15	-0.191	-0.063	11.282	0.587
		16	0.036	0.037	11.376	0.656
		17	-0.089	-0.076	12.020	0.677
		18	0.165	-0.102	14.500	0.562
		19	-0.068	-0.138	14.986	0.597
		20	0.078	-0.039	15.731	0.611
		21	0.025	0.089	15.824	0.669
		22	0.013	0.033	15.853	0.726
		23	-0.050	-0.108	16.462	0.743
		24	0.117	0.030	21.414	0.495

Lampiran 4. Tingkat Pengangguran Usia Muda di Indonesia

Waktu	Data Asli	Dugaan ARIMA	Dugaan ARIMAX
Semester 1, 2007	25.43	26.88	25.08
Semester 2, 2007	25.14	29.12	25.91
Semester 1, 2008	21.09	29.10	24.26
Semester 2, 2008	23.31	32.00	27.31
Semester 1, 2009	22.44	30.26	24.94
Semester 2, 2009	22.19	30.27	24.59
Semester 1, 2010	19.93	28.00	20.48
Semester 2, 2010	21.45	29.36	21.13
Semester 1, 2011	23.36	29.36	20.20
Semester 2, 2011	22.81	31.13	22.70
Semester 1, 2012	19.10	30.06	21.30
Semester 2, 2012	19.62	30.07	21.34
Semester 1, 2013	17.49	28.68	18.55
Semester 2, 2013	21.58	29.52	19.09
Semester 1, 2014	17.05	29.51	18.39
Semester 2, 2014	22.20	30.60	20.14
Semester 1, 2015	18.28	29.94	19.05
Semester 2, 2015	22.59	29.95	19.03
Semester 1, 2016	17.85	29.10	16.94
Semester 2, 2016	19.45	29.61	17.28
Semester 1, 2017	15.29	29.61	16.77
Semester 2, 2017	20.44	30.27	18.03
Semester 1, 2018	16.38	29.87	17.29
Semester 2, 2018	19.68	29.87	17.33
Semester 1, 2019	15.38	29.35	15.90

Gambar 1. Perbandingan Tebaran Titik Data Asli, Titik Dugaan Model ARIMA, dan Titik Dugaan Model ARIMAX: Kasus Tingkat Pengangguran Usia Muda 2005-2019

a) Model ARIMA



b) Model ARIMAX

