

კორონავირუსის (COVID-19-ის) გავრცელების პროგნოზირების პრობლემა მსოფლიოსთვის

აკაკი გაბელაია

ფიზიკა-მათემატიკურ მეცნიერებათა დოქტორი,
ახალი უმაღლესი სასწავლებლის მონვეული პროფესორი,
საქართველოს ტექნიკური უნივერსიტეტის პროფესორი

აბსტრაქტი

როგორც ცნობილია, უკანასკნელ პერიოდში კორონავირუსის COVID-19-ის გავრცელება სასიცოცხლო მნიშვნელობის პრობლემად იქცა მთელი მსოფლიოსთვის. საკმარისია იმის თქმაც, რომ იგი მსოფლიო ჯანდაცვის ორგანიზაციამ, 2020 წლის 11 მარტს, პანდემიად გამოაცხადა. აქედან გამომდინარე, გასაგებია თუ რა მნიშვნელობა შეიძლება ენიჭებოდეს მისი გავრცელების პროგნოზირების პრობლემას. მოცემული ნაშრომი ეძღვნება ამ მხრივ ჩვენ მიერ ჩატარებული კვლევის გამოცდილებას და შედეგებს.

გარკვეულობისათვის უნდა შევნიშნოთ, რომ ჩვენ მიერ პროგნოზირების თვალსაზრისით განხილული იყო კორონავირუსის გავრცელების ისეთი ძირითადი მაჩვენებლები, როგორცაა ინფიცირების საერთო შემთხვევათა რაოდენობა (მსოფლიოში) მიმდინარე მომენტისათვის (total cases) და აქტიური შემთხვევების რაოდენობა მიმდინარე მომენტისათვის (active cases).

ამ მაჩვენებლების პროგნოზული შეფასებები ჩვენ მიერ ნაპოვნი იყო 13 თებერვლიდან დაწყებული. ამასთან, იმის გათვალისწინებით, რომ ამ პერიოდიდან დაწყებული ვირუსი (რომელიც თავდაპირველად ჩინეთში გავრცელდა) მოედო მთელ მსოფლიოს, ანუ დრამატულად შეიცვალა მისი გავრცელების არეალი და, გარდა ამისა, ჩვენ მიერ პროგნოზირებისთვის გამოყენებულ მოდელთა უმრავლესობა ატარებდა მოკლევადინ ხასიათს, აუცილებელი გახდა ჩვენი პროგნოზული შეფასებების პერიოდული კორექტირება. ამასთან, უნდა შევნიშნოთ, რომ თავდაპირველად ჩვენ მიერ ნაპოვნი პროგნოზული შეფასებები (რომელიც ძირითადად ჩინეთში ვირუსის გავრცელების მონაცემებს ეყრდნობოდა) აღმოჩნდა საკმაოდ ზუსტი. კერძოდ, ჩვენი პროგნოზის თანახმად, მარტის ბოლოსთვის ვირუსით ინფიცირებულთა საერთო რაოდენობა არ უნდა გასცდენოდა 85000-ს, რეალურად კი ამ მაჩვენებელმა ჩინეთში (რომელმაც ეს ვირუსი ფაქტობრივად უკვე დამარცხა!), 21 მარტის მდგომარეობით, 81008 შეადგინა.

საკვანძო სიტყვები: კორონავირუსი, პროგნოზირება, პროგნოზირების მოდელები, პროგნოზული შეფასებების სიზუსტე.

როგორც ცნობილია, უკანასკნელ პერიოდში კორონავირუსის COVID-19-ის გავრცელება სასიცოცხლო მნიშვნელობის პრობლემად იქცა მთელი მსოფლიოსთვის. საკმარისია იმის თქმაც, რომ იგი მსოფლიო ჯანდაცვის ორგანიზაციამ 2020 წლის 11 მარტს, პანდემიად გამოაცხადა. აქედან გამომდინარე, გასაგებია თუ რა მნიშვნელობა შეიძლება ენიჭებოდეს მისი გავრცელების პროგნოზირების პრობლემას. ქვემოთ აღწერილი იქნება ამ მხრივ ჩვენ მიერ ჩატარებული კვლევის შედეგები და გამოცდილება.

გარკვეულობისთვის უნდა შევნიშნოთ, რომ ჩვენ მიერ პროგნოზირების თვალსაზრისით განხილული იყო კორონავირუსის გავრცელების ისეთი ძირითადი მაჩვენებლები, როგორცაა **ინფიცირების საერთო შემთხვევათა რაოდენობა (მსოფლიოში) მიმდინარე მომენტისათვის (total cases)**, რომელსაც ქვემოთ აღვნიშნავთ **infic** ცვლადით და **აქტიური შემთხვევების (ე.ი. ამ ვირუსით დაავადებულთა) რაოდენობა მიმდინარე მომენტისათვის (active cases)**, რომელსაც ქვემოთ აღვნიშნავთ **ac** ცვლადით. (გასაგებია, რომ ამ ორ სიდიდეს შორის სხვაობა წარმოადგენს გამოჯანმრთელებულ და გარდაცვლილ პაციენტთა ჯამურ რაოდენობას (მთელი მსოფლიოს მასშტაბით).)

თავდაპირველად ჩვენ ვცადეთ infic ცვლადის მნიშვნელობათა პროგნოზირება, 22/01-13/02 2020 პერიოდის მონაცემთა ბაზაზე (დღეების წრილში), 13/02-29/02 პერიოდისათვის. აქედან გამომდინარე, გასაგებია, რომ 13 თებერვლის მონაცემი ჩვენ მიერ გამოყენებული იყო ე.წ. ex post პროგნოზირების ანუ პროგნოზირების მოდელის შერჩევის მიზნით.

საინტერესოა შევნიშნოთ, რომ ამ (პირველ ეტაპზე) ამ მოდელთა მახასიათებლებიდან და 13 თებერვლის მონაცემიდან გამომდინარე, პროგნოზირების ყველაზე საიმედო მოდელად ჩავთვალეთ ფერჰიულსტის ლოგისტიკური ზრდის მოდელი [Verhulst, 1845 :1-45] რომლის ზოგადი სახეც ასე გამოიყურება:

$$P(t) = \frac{K_0}{(K - P_0)e^{-rt} + P_0}, \quad (1)$$

სადაც P_0 აღნიშნავს პოპულაციის (ამ შემთხვევაში ინფიცირებულთა) რიცხოვნობას საწყის მომენტში, ხოლო K პოპულაციის (ამ შემთხვევაში ინფიცირებულთა) მაქსიმალური რიცხვია.

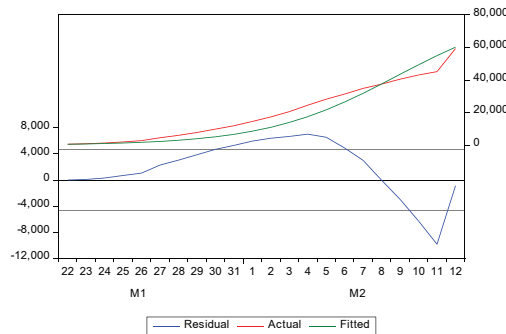
უნდა შევნიშნოთ, რომ ამ პერიოდში ვირუსი ძირითადად ჩინეთში იყო გავრცელებული და მონაცემებიც ძირითადად ჩინეთში მის გავრცელებას შეესაბამებოდა.

მოცემულ შემთხვევაში, 13 თებერვლის მონაცემის გათვალისწინებით, ჩვენ მიერ K პარამეტრის სიდიდე აღებული იყო 85000-ის ტოლად (რაც ამ შემთხვევაში საკმარისად ზუსტი გამოდგა, იმის გათვალისწინებით, რომ ამ მაჩვენებელმა 15 სექტემბრის მდგომარეობით 85214 შეადგინა! (ე.ი. პროგნოზირების მომენტიდან 7 თვის განმავლობაში პროგნოზის ცდომილება რეალური მნიშვნელობის მიმართ, მხოლოდ 0.25 % გამოვიდა!!)

ასე რომ ზრდის მაქსიმალური ტემპის, r -ის სიდიდეს, მოცემულ შემთხვევაში ვეძებდით რეგრესიის შემდეგი განტოლებიდან (უნდა შევნიშნოთ, რომ ჩვენ მიერ ყველა ძირითადი გაანგარიშება ჩატარებული იყო ცნობილი კომპიუტერული პროგრამის EViews-10-ის [გაბელაია ა., გაბელაია ლ., 2017: 433] ბაზაზე):

$$\text{INFIC}=(85000*580)/((85000-580)*\text{EXP}(-C(1)*@\text{TREND})+580).$$

ამ რეგრესიის შედეგად მიღებულმა r პარამეტრის მნიშვნელობამ შეადგინა $r=0.279355$. ამასთან, ამ რეგრესიის დეტერმინაციის კოეფიციენტი 0,93-ის რიგისა იყო, ძალზე მაღალი (59,2-ის ტოლი) t-სტატისტიკით, მხოლოდ დარბინ-უოტსონის სტატისტიკა გამოვიდა ძალზე მცირე, რაც იმაზე მეტყველებს, რომ რეგრესიას შეიძლება ახასიათებდეს სისტემატური ცდომილებები. გარდა ამისა, აღნიშნული მოდელის სიზუსტეზე გარკვეულ წარმოდგენას იძლევა მისი ცდომილების დიაგრამაც (იხ. ნახ. 1)).

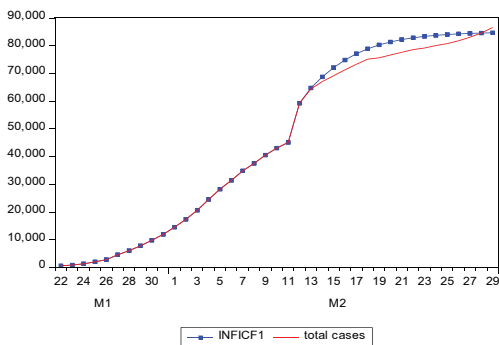


ნახ.1. ინფიცირების ლოგისტიკური მოდელის ცდომილების დიაგრამა

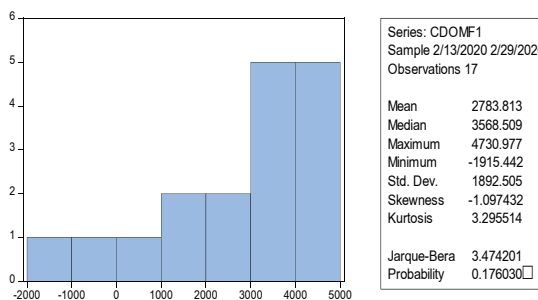
რაც შეეხება ამ მოდელიდან გამომდინარე infic სიდიდის პროგნოზულ შეფასებებს და მათ თანაფარდობას ამ სიდიდის რეალურ მნიშვნელობებთან მოცემული პერიოდისათვის, ნაჩვენებია ნახ.

2-ზე, სადაც infic და inficf1, შესაბამისად, აღნიშნავს ინფიცირებულთა რაოდენობის რეალურ სიდიდეს და მის (წინასწარ) პროგნოზულ შეფასებას.

მიღებული პროგნოზული შეფასებების ცდომილებაზე უფრო ზუსტ წარმოდგენას იძლევა მისი ცდომილების cdomf1=inficf1-infic ჰისტოგრამა და სტატისტიკური მახასიათებლები, რომლებიც ნაჩვენებია ნახ. 3-ზე.



ნახ.2. ინფიცირების პროგნოზული და რეალური მაჩვენებლების დინამიკა 13/02-29/02 2020 პერიოდისათვის



ნახ.3. ინფიცირების პროგნოზული შეფასებების ცდომილების ჰისტოგრამა და სტატისტიკური მახასიათებლები

როგორც ვხედავთ, ამ (პირველ) ეტაპზე პროგნოზირების შედეგები არც ისე ცუდია (საშუალო ცდომილებამ და სტანდარტულმა გადახრამ საპროგნოზო პერიოდისათვის, შეადგინეს, შესაბამისად 2784 და 1593 ერთეული).

შეიძლება დავრწმუნდეთ, რომ ამ პროგნოზული შეფასებების ე.წ. აპროქსიმაციის საშუალო ცდომილებამ (ცდომილების მოდულების პროცენტულმა საშუალომ რეალურ მნიშვნელობათა მიმართ), რომელიც, მოცემულ შემთხვევაში იანგარიშება ფორმულით:

$$A = \left(\frac{1}{n} \sum_{i=1}^n \left| \frac{\text{inf icf1}_i - \text{inf } \dot{c}_i}{\text{inf } \dot{c}_i} \right| \right) * 100\%$$

(სადაც n პროგნოზირების ჰორიზონტის სიგრძეს აღნიშნავს) 3.3 % შეადგინა!

უნდა შევნიშნოთ, რომ თანდათანობით, მას შემდეგ რაც ვირუსი გასცდა ერთი ქვეყნის საზღვრებს, მისი გავრცელების დინამიკა მნიშვნელოვნად დაშორდა ზემოთ გამოყენებული ლოგისტიკური მრუდის ფორმას (რომელიც, როგორც ზემოთაც აღვნიშნეთ, იძლევა შედარებით გრძელვადიანი პროგნოზირების საშუალებას!) და მოითხოვა უფრო რთული, მაგ. ARIMA-ს (ავტორეგრესიისა და მცოცავი საშუალოს ინტეგრირებული მოდელების) გამოყენება (რაც ვიზუალურადაც კარგად ჩანს ვირუსის რეალური გავრცელების გრაფიკებზე). ამასთან, ამ მოდელების მინუსი ისაა, რომ ისინი ჩვეულებრივ იძლევიან მოკლევადიან პროგნოზებს!

კერძოდ, მეორე ეტაპზე ჩვენ ვცადეთ ინფიცირებულთა საერთო რაოდენობის პროგნოზირება მარტის თვისთვის, უკვე იანვარ-თებერვლის მონაცემებიდან (უფრო ზუსტად, 22/01-29/02 2020) გამომდინარე, უკვე ზემოხსენებული ARIMA-ს ტიპის მოდელების ბაზაზე.

infic ცვლადის ტრენდულმა მოდელმა, ავტორეგრესიული წევრით, ამ პერიოდისათვის მიიღო სახე:

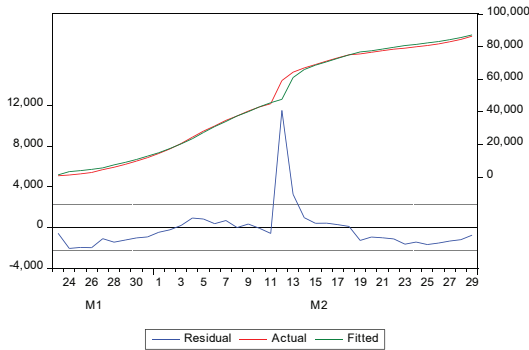
$$\text{INFIC} = 2419.82056382 * @TREND + [\text{AR}(1)=0.934518885427, \text{UNCOND}, \text{ESTSMPL}=\gg 1/23/2020 2/29/2020\gg]. (2)$$

უნდა შევნიშნოთ, რომ ამ რეგრესიის დეტერმინაციის კოეფიციენტი ძალიან მაღალი გამოვიდა (0,99-ის რიგის), პარამეტრთა t-სტატისტიკები, შესაბამისად, 6.4, 10.8 და 5.8-ის ტოლი, მხოლოდ

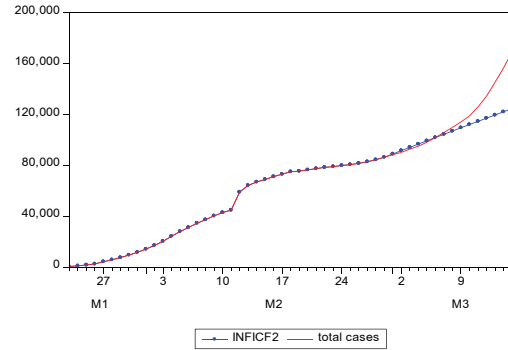
დარბინ-უოტსონის სტატისტიკა გამოვიდა ცოტა დაბალი (1.26-ის რიგის), რაც იმაზე მეტყველებს, რომ რეგრესია არაა დაზღვეული სისტემატური ცდომილებებისგან. ცხადია, აღნიშნული მოდელის სიზუსტეზე გარკვეულ წარმოდგენას იძლევა მისი ცდომილების დიაგრამაც (იხ. ნახ. 4)).

რაც შეეხება ამ მოდელიდან გამომდინარე infic სიდიდის პროგნოზულ შეფასებებს და მათ თანაფარდობას ამ სიდიდის რეალურ მნიშვნელობებთან 1-15 მარტის მონაცემებიდან გამომდინარე, ნაჩვენებია ნახ. 5-ზე, სადაც infic და inficf2, შესაბამისად, აღნიშნავს ინფიცირებულთა რაოდენობის რეალურ სიდიდეს და მის (წინასწარ) პროგნოზულ შეფასებას.

როგორც ვხედავთ, პროგნოზირების სიზუსტე სწრაფად უარესდებოდა 11 მარტიდან დაწყებული, როცა ვირუსი, ფაქტობრივად, მთელ მსოფლიოს მოედო. ბუნებრივია, რომ ეს დრამატული ცვლილებები, ვირუსის გავრცელების არეალსა და პირობებში ძალზე მნიშვნელოვნად ართულებდა მისი პროგნოზირების პრობლემას, მით უფრო იმის გათვალისწინებით, რომ ჩვენ აქ ჩვეულებრივ ვიყენებდით პროგნოზირების პასიურ მეთოდებს, რომლებიც ემყარება იმ დაშვებას, რომ მომავალი უნდა ჰგავდეს წარსულს. სწორედ ამით აიხსნება ის გარემოება, რომ ქვემოთ ჩვენ იძულებული ვიყავით გვეკეთებინა ე.წ. ოპტიმისტური და პესიმისტური პროგნოზული შეფასებები, რომლებიც ძალზე მნიშვნელოვნად განსხვავდებოდა ერთმანეთისგან.



ნახ.4. ინფიცირების (2) მოდელის ცდომილების დიაგრამა



ნახ.5. ინფიცირების პროგნოზული და რეალური მაჩვენებლების დინამიკა 1/03-15/03 2020 პერიოდისათვის

აქედან გამომდინარე, ჩვენ მიერ მიღებული იყო გადაწყვეტილება ჩვენი პროგნოზული შეფასებების კორექტირების შესახებ მიმდინარე წლის 11 მარტიდან.

რაც შეეხება ჩვენ მიერ მიღებული პროგნოზული შეფასებების სიზუსტეს **მარტის პირველი დეკადისთვის**, მისი აპროქსიმაციის საშუალო ცდომილებამ 2 % შეადგინა!

უნდა შევნიშნოთ, რომ ზემოთ მოყვანილის ანალოგიური მოდელებიდან მიღებული infic ცვლადის პროგნოზული შეფასებების აპროქსიმაციის საშუალო ცდომილებამ **მარტის მეორე დეკადისთვის 2.8 % შეადგინა!**

ყოველივე ზემოაღნიშნულიდან გამომდინარე (უკვე **მეთხე ეტაპზე!**), ჩვენ გავაკეთეთ პროგნოზული შეფასებები **მარტის მესამე დეკადისთვის, რომლისთვისაც აპროქსიმაციის საშუალო ცდომილებამ 1.75 % შეადგინა!**

გარდა ამისა, როგორც შესაბამისმა გაანგარიშებებმა გვიჩვენა, **13 თებერვალი-31 მარტი პერიოდისათვის, ჩვენი პროგნოზების აპროქსიმაციის საშუალო შეწონილმა ცდომილებამ 2.57 % შეადგინა!**

მეხუთე ეტაპზე, აპრილის პირველი ნახევრისთვის ინფიცირების პროგნოზული შეფასების აპროქსიმაციის საშუალო ცდომილებამ 2.14 % შეადგინა.

მექვსე ეტაპზე, ანუ აპრილის მეორე ნახევარში პროგნოზული შეფასებების აპროქსიმაციის საშუალო ცდომილებამ 1.7 % შეადგინა.

მეშვიდე ეტაპზე, ანუ **მაისის პირველ ნახევარში** ჩვენი პროგნოზული შეფასებების აპროქსიმაციის საშუალო ცდომილებამ 0.54 % შეადგინა.

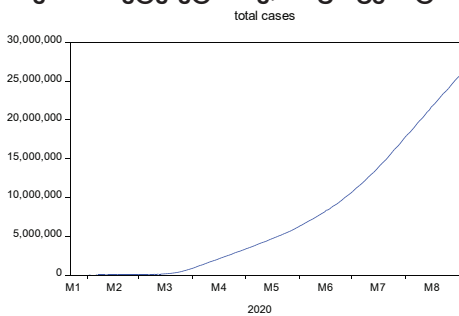
მერვე ეტაპზე, **16 მაისი-15 ივნისის პერიოდისათვის** ჩვენი პროგნოზული შეფასებების აპროქსიმაციის საშუალო ცდომილებამ 0.9 % შეადგინა.

მეცხრე ეტაპზე, ანუ **ივნისის მეორე ნახევრისთვის** ჩვენი პროგნოზული შეფასებების აპროქსიმაციის საშუალო ცდომილებამ 2.73 % შეადგინა.

მეათე ეტაპზე, **ივლისის პირველ ნახევარში** ჩვენი პროგნოზული შეფასებების აპროქსიმაციის საშუალო ცდომილებამ 2.23 % შეადგინა.

მეთერთმეტე ეტაპზე, ანუ **ივლისის მეორე ნახევარში** ჩვენი პროგნოზული შეფასებების აპროქსიმაციის საშუალო ცდომილებამ 1.2 % შეადგინა.

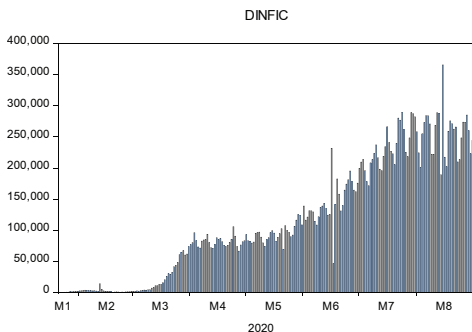
მეთორმეტე ეტაპზე, ანუ **აგვისტოში** პროგნოზული შეფასებების აპროქსიმაციის საშუალო ცდომილებამ 2.15 % შეადგინა.



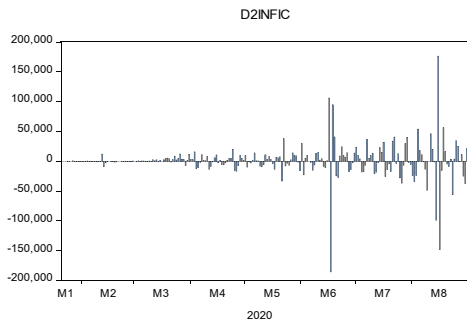
დაბოლოს, ვირუსის გავრცელების ძირითადი მახასიათებლების დინამიკა მსოფლიოსათვის (დღეების წრილში) 2020 წლის იანვარ-აგვისტოს პერიოდში ასე გამოიყურებოდა (იხ. ნახ. 6-9):

ნახ. 6. ინფიცირებულთა რაოდენობის $infc$ დინამიკა 2020 წლის იანვარ- აგვისტოს პერიოდში (დღეების წრილში)

როგორც ვხედავთ, ჩვენ მიერ გამოყენებული პროგნოზირების მოდელები (რომლებიც, როგორც ცნობილია კარგად მუშაობენ მოკლევადიან პერსპექტივაში!) მართლაც საკმარისად მაღალ სიზუსტეს აჩვენებენ მაქსიმუმ თვის პერსპექტივაში (შემდეგ მათი სიზუსტე ეცემა!).

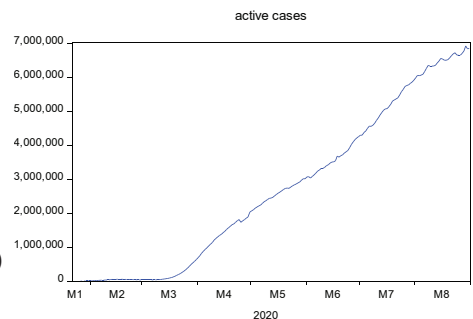


ნახ. 7. ინფიცირებულთა რაოდენობის ზრდის „სიჩქარის“ $d(infc)$ დინამიკა 2020 წლის იანვარ- აგვისტოს პერიოდში (დღეების წრილში)

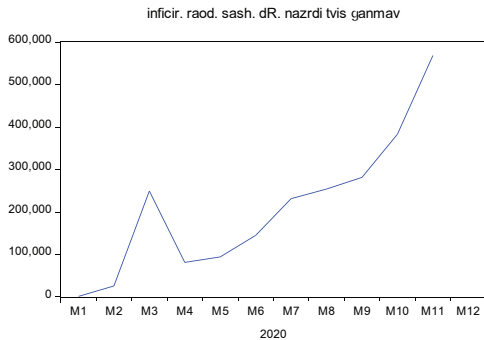


ნახ. 8. ინფიცირებულთა რაოდენობის ზრდის „აჩქარების“ $d(infc,2)$ დინამიკა 2020 წლის იანვარ- აგვისტოს პერიოდში (დღეების წრილში)

ნახ. 9. ვირუსით დაავადებულთა რაოდენობის ac დინამიკა 2020 წლის იანვარ- აგვისტოს პერიოდში (დღეების წრილში)



მეორე მხრივ, ვირუსი უახლოეს პერსპექტივაში „გაჩერებას არ აპირებს“, რაც პროგნოზირების ზემოთ მოყვანილი მეთოდების გამოყენების პერსპექტიულობას ეჭვქვეშ აყენებს. აქედან გამომდინარე (პროგნოზირების ჰორიზონტის ზრდის მიზნით!), შეიძლება აზრი ჰქონდეს ახალი მაჩვენებლის განხილვას, როგორცაა მაგ., „ინფიცირებულთა რაოდენობის საშუალო დღიური ნაზრდი თვის განმავლობაში“, რაც მოგვცემს საშუალებას გავაკეთოთ ამ მაჩვენებლის პროგნოზი რამდენიმე თვის შემცველი ჰორიზონტისათვის, რითაც ჩვენ შემდგომში ვაპირებთ დავკავდეთ! მით უფრო, რომ ცენტრალური ზღვარითი თეორემის თანახმად, ამ მაჩვენებლის განხილვა ახლოს უნდა იყოს ნორმალურთან, რაც გაამარტივებს მისთვის სარწმუნო პროგნოზული შეფასებების პოვნის ამოცანას.



კერძოდ, ამ მაჩვენებლის დინამიკა მსოფლიოსთვის 2020 წლის იანვარ-ოქტომბრის მონაცემებით ასე გამოიყურებოდა:

ნახ. 10. ინფიცირებულთა რაოდენობის საშუალო დღიური ნაზრდის დინამიკა მსოფლიოში 2020 წლის იანვარ-ოქტომბრის მონაცემებით

როგორც ვხედავთ, მოცემულ პერიოდში ეს მაჩვენებელი (მარტში ლოკალურ მაქსიმუმის თუ არ ჩავთვლით) აღმავალი ხაზით ვითარდებოდა, ე.ი. მსოფლიოში ვირუსის გავრცელებას ჰქონდა პერმანენტურად აღმავალი დინამიკა, თუმცა ამ ზრდას არ ჰქონია ისეთი ზვავისებური ზრდის ხასიათი, როგორსაც ადგილი ჰქონდა საქართველოში ბოლო ორი თვის განმავლობაში.

უნდა შევნიშნოთ, რომ რაც შეეხება ამ ვირუსისგან გამონვეულ სიკვდილიანობის მაჩვენებელს (გარდაცვლილთა პროცენტულ შეფარდებას გამოჯანმთელებულთა და გარდაცვლილთა რაოდენობების ჯამთან), იგი თავდაპირველად საკმარისად სწრაფად გაიზარდა 21 %-მდე, თუმცა შემდეგ ნელ-ნელა ჩამოვიდა 3 %-მდე (უკვე ოქტომბრის ბოლოს მონაცემებით!). ამასთან, უნდა გავითვალისწინოთ, რომ ზემოთ ჩვენ ვსაუბრობდით ამ მაჩვენებლის კუმულატიურ (დაგროვილ) მნიშვნელობაზე! რაც შეეხება მის მიმდინარე მნიშვნელობას (იქიდან გამომდინარე, რომ „კრიტიკულები“ შეადგენენ დაავადებულთა მიმდინარე რაოდენობის მხოლოდ 1 %-ს), უნდა ვიგულისხმოთ, რომ სიკვდილიანობის მიმდინარე მაჩვენებელი ამ ვირუსითვის კიდევ უფრო დაბალია. თუმცა, სამწუხაროდ, მას გავრცელების ძალზე მაღალი სიჩქარე ახასიათებს.

ლიტერატურა:

1. Verhulst, P. F., Recherches Mathématiques sur La Loi D'Accroissement de la Population, *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, 18, Art. 1, 1-45, 1845 (Mathematical Researches into the Law of Population Growth Increase).
2. გაბელაია ა., გაბელაია ლ. ეკონომეტრიკული ანალიზისა და პროგნოზირების კომპიუტერული პაკეტის EViews-ის, საფუძვლები, თბილისი, 2017, 433 გვ.