

М. Г. Коротков, А. А. Петров, М. В. Куркина

**ПРОБЛЕМЫ ИДЕНТИФИКАЦИИ ИНТЕРВАЛОВ ВРЕМЕННОГО РЯДА
ПРИ ПРОГНОЗИРОВАНИИ ДИНАМИКИ ЧИСЛА ЗАРАЖЁННЫХ COVID-19
СТАТИСТИЧЕСКИМИ МЕТОДАМИ НА ПРИМЕРЕ ЮГРЫ**

Целью работы является разработка подхода по выделению из временного ряда динамики новых случаев заражения коронавирусом в Югре числа зараженных COVID-19 весной–летом 2020 года интервала данных для статистического прогнозирования.

Ключевые слова: временной ряд, статистические методы, прогнозирование, модель.

M. G. Korotkov, A. A. Petrov, M. V. Kurkina

**PROBLEMS OF IDENTIFYING TIME SERIES INTERVALS WHEN PREDICTING
THE DYNAMICS OF THE NUMBER OF INFECTED COVID-19 BY STATISTICAL
METHODS USING THE EXAMPLE OF YUGRA**

The aim of this work is to develop an approach to isolate the data interval for statistical forecasting from the time series of dynamics of new cases of coronavirus infection in the Yugra of the number of COVID-19 infected in the spring-summer of 2020.

Key words: time series, statistical methods, forecasting, model.

Введение

При исследовании процесса распространения COVID-19 в открытых источниках доступными являются данные о:

- 1) количестве зараженных за день;
- 2) количестве выздоровевших за день;
- 3) количестве умерших за день.

Остальные данные являются производными этих трех.

Скорее всего, нельзя рассматривать данные 2, 3 в качестве временного ряда случайных величин [1], чего нельзя сказать о количестве зараженных за день, поскольку отсутствуют ежедневно меняющиеся неслучайные воздействия, влияющие на эту величину. Возможны разовые изменения, влияющие на характер (режим) процесса распространения коронавируса (введение/отмена режима самоизоляции, запрет отдельных видов деятельности, ввод масочного режима и т. п.). Реакция количества ежедневных заражений на эти изменения запаздывает, и какое-то время продолжает сказываться влияние предыдущего режима.

Все это оставляет возможность применения статистических подходов к прогнозированию [2] количества ежедневных заражений с учетом вышеизложенных замечаний.

Описание алгоритма

Рассмотрим временной ряд числа зараженных COVID-19 за сутки в ХМАО с 14.04.2020 по 31.08.2020 (открытые данные сайта coronavirus-monitor.ru) длиной 140 (рис. 1).



Рисунок 1

Поведение ряда близко к «колоколообразной» гауссовой кривой [3], поэтому имеет смысл прологарифмировать исходный ряд (рис. 2).

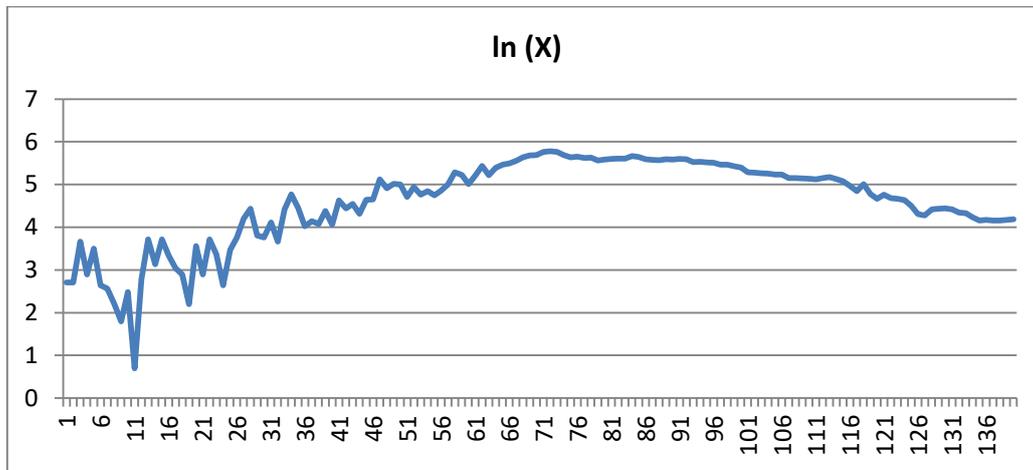


Рисунок 2

В прологарифмированном ряду наблюдается квадратичная трендовая зависимость. Применяя МНК, получим модель тренда

$$y = -0,00050 x^2 + 0,08439 x + 1,94620$$

Удалив из прологарифмированного ряда тренд, получим ряд остатков (рис. 3).

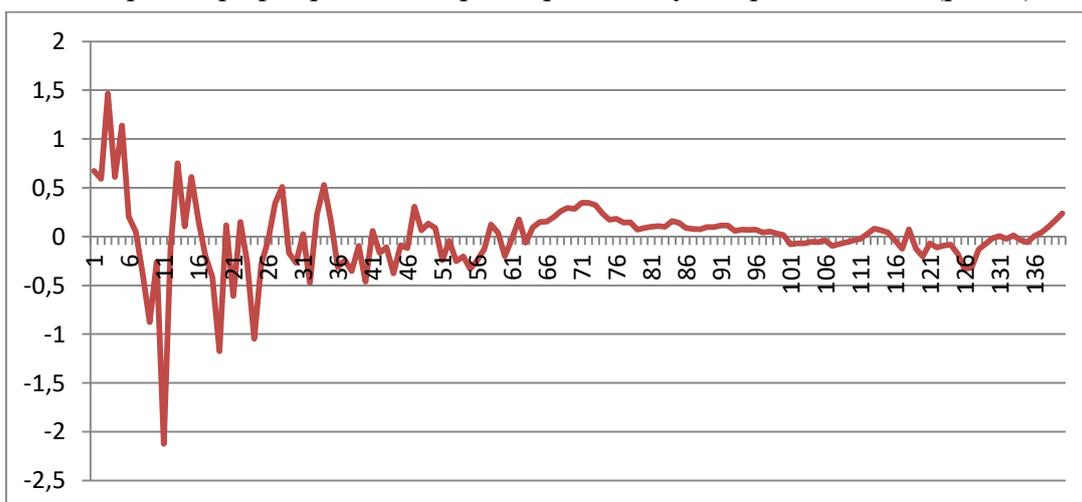


Рисунок 3

Из графика видно, что, начиная с 64 уровня исходного ряда, наблюдается изменение характера процесса. Включение в рассмотрение данных с 1 по 63 уровень (с 14.04.20 по

15.06.20) при идентификации типа и оценки параметров модели [4] для статистического прогноза может только ухудшить его качество.

Результат удаления из рассмотрения указанного интервала исходного временного ряда представлен на рис. 4.

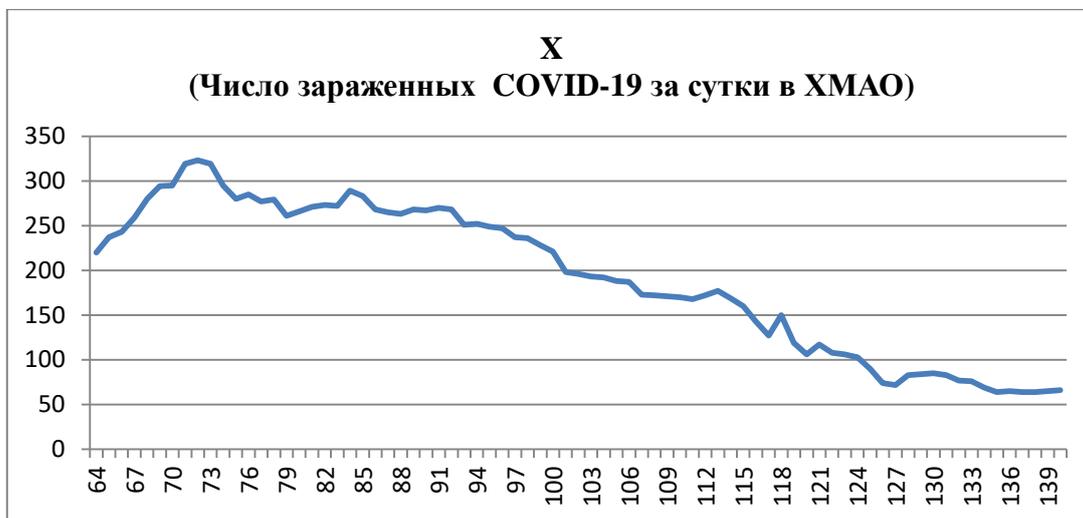


Рисунок 4

Повторив вышеизложенные шаги, получим модель тренда:

$$y = -0,00041 x^2 + 0,06706 x + 3,37319$$

для прологарифмированного исходного ряда. Удалив этот тренд из логарифмированного ряда, получим ряд остатков для логарифмированного ряда (рис. 5).

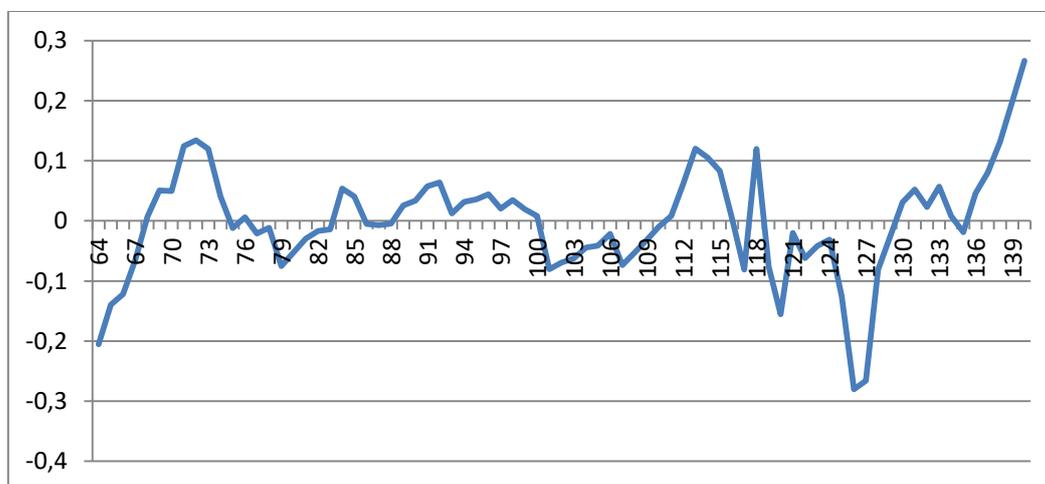


Рисунок 5

Данный ряд нестационарен [5], что препятствует переходу к следующим шагам построения модели прогноза. Для приведения ряда к стационарному применим стандартную операцию нахождения разностей ряда первого порядка [4] (рис. 6).



Рисунок 6

Как видим, полученный ряд также не является стационарным. Ряд разностей второго порядка (рис. 7) показывает, что начиная с 110 уровня (31.07.20) наблюдается существенное изменение характера процесса.

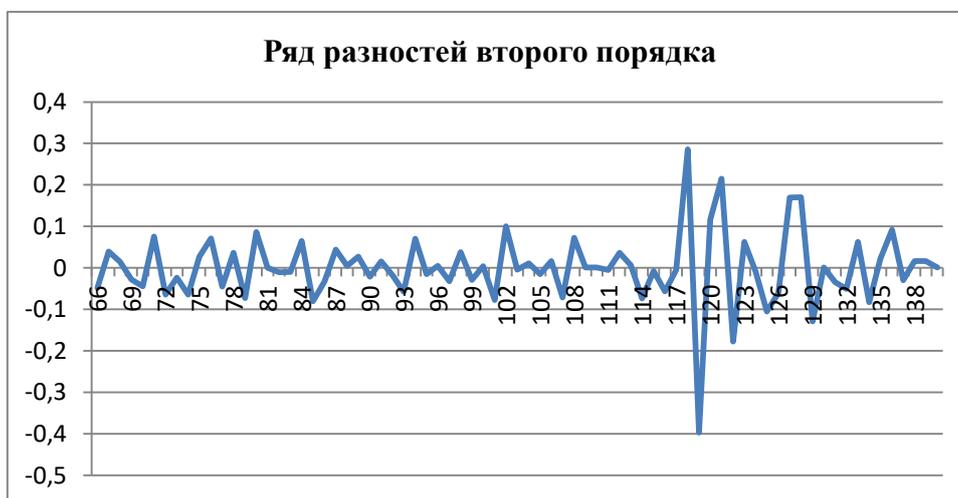


Рисунок 7

Следовательно, при построении модели статистического прогноза целесообразно использовать данные начиная с 31.07.20.

Повторив шаги, аналогичные предыдущим, получим соответствующий ряд остатков, приводящийся к стационарному взятием первых и вторых разностей. Это позволит перейти к следующим этапам построения статистической модели (идентификации типа и оценки параметров модели) и прогноза.

Обобщая вышеизложенное, перечислим этапы, позволяющие, по крайней мере, не ухудшить качество статистической модели прогноза при ее построении:

1. Визуализация исходных данных.
2. Выявление вида тренда исходя из результатов визуализации.
3. Преобразование (при необходимости) исходных данных к виду, удобному для выделения тренда.
4. Получение оценок параметров тренда на основе МНК.
5. Удаление из преобразованного ряда тренда.
6. Проверка на стационарность полученного ряда остатков (графический способ, способ применения автокорреляционной функции и ряд других).
7. Выявление временных интервалов, соответствующих принципиально различным по характеру процессам.

8. Определение интервала данных для построения прогностической статистической модели.

Выводы

При использовании полного ряда данных при статистическом прогнозировании очень часто возникает ситуация, когда учет более ранних данных вносит существенное ухудшение качества модели, используемой для прогноза.

Такого рода проблема может возникнуть и при построении иных (не статистических) моделей, используемых для прогнозирования. В этом случае подход, рассмотренный в данной работе, может также оказаться полезным.

Литература

1. Мишулина, О. А. Статистический анализ и обработка временных рядов / О. А. Мишулина. – Москва : МИФИ, 2004. – 180 с. – ISBN 5-7262-0536-7. – Текст : непосредственный.
2. Айвазян, С. А. Прикладная статистика и основы эконометрики / С. А. Айвазян, В. С. Мхитарян. – Москва : ЮНИТИ-ДАНА, 2001. – 656 с. – Текст : непосредственный.
3. Вентцель, Е. С. Теория вероятностей / Е. С. Вентцель. – 10-е издание, стереотипное. – Москва : Academia, 2005. – 576 с. – Текст : непосредственный.
4. Боровиков, В. П. Прогнозирование в системе STATISTICA в среде Windows / В. П. Боровиков, Г. И. Ивченко. – Москва : Финансы и статистика, 1999. – 384 с. – Текст : непосредственный.
5. Дуброва, Т. А. Статистические методы прогнозирования : учебное пособие для вузов / Т. А. Дуброва. – Москва : Юнити, 2003. – 106 с. – ISBN 5-238-00497-4. – Текст : непосредственный.