

Министерство науки и образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего образования
Национальный исследовательский университет "Высшая школа экономики"
Факультет экономических наук

Домашняя работа по теме «Стационарность и автокорреляция. Трендовые модели»

По дисциплине «Временные ряды и их практическое применение»

Выполнила:
Студентка группы БЭК187
Андреичева Полина

Москва – 2021 г.

- 1) Описать постановку задачи, гипотезы для проверки, графики динамики; в первом пункте, когда будете писать задачу исследования, нужно упомянуть и внешние факторы (хоть в модели пока и они не участвуют), а также проанализировать графики по ним (т.е. не только по зависимой переменной)

Данные представляют собой количество нанятых работников в тысячах человек в четырех отраслях в Висконсине. Выбранными отраслями являются торговля, продовольственная отрасль, машиностроение и металлургия. Зависимой переменной выбрана занятость в тысячах человек в торговой отрасли.

Исследовательской задачей является выявление тренда и сезонности зависимой переменной, а также построение нескольких моделей и прогнозирование по ним. Занятость является хорошим объектом для исследования, так как обладает трендом и сезонностью. Внешние факторы так же могут влиять на зависимую переменную, так как возможно перетекание рабочей силы.

Можно выдвинуть гипотезы, что данные будут обладать сезонностью и детерминированным трендом, а также будут хорошо предсказываться.

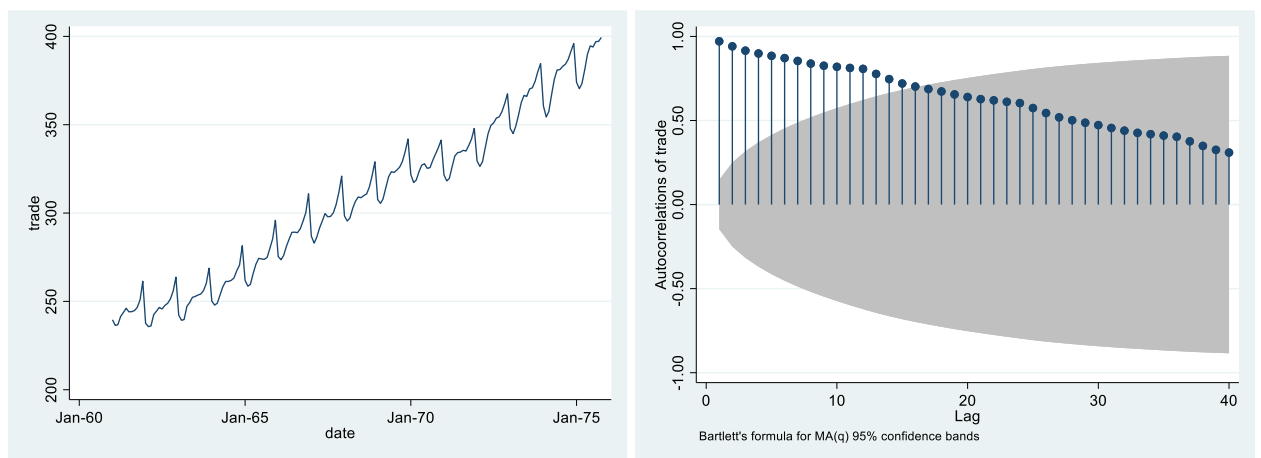


Рисунок 1. График временного ряда занятости в торговой отрасли и его автокорреляционная функция.

По графику заметно наличие возрастающего тренда и сезонности с периодичностью примерно год. Временной ряд не выглядит стационарным, так как имеется тренд, то есть со временем меняется математическое ожидание. В то же время не заметно увеличение дисперсии, а также сезонность не меняется. Автокорреляционная функция так же указывает на нестационарность и на наличие тренда.

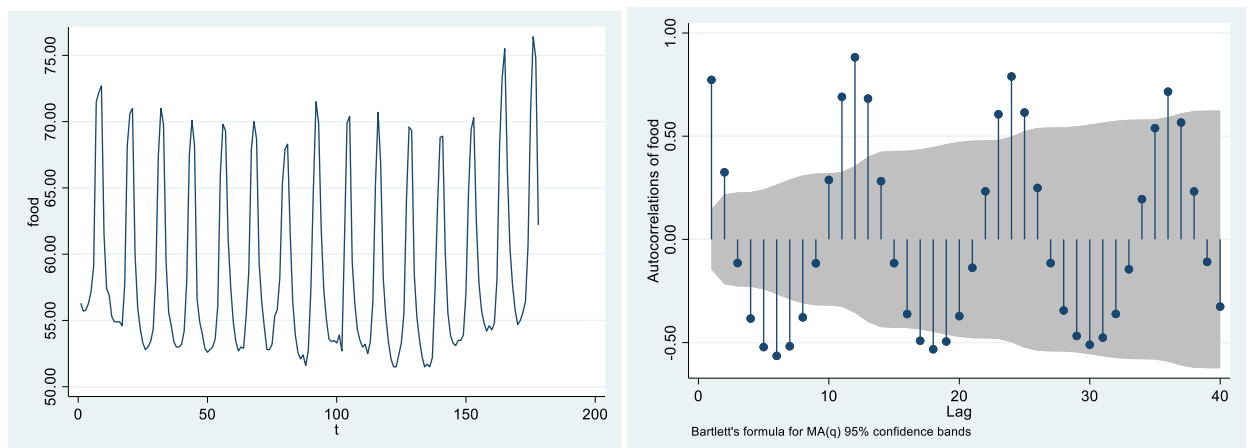


Рисунок 2. График временного ряда занятости в пищевой промышленности.

По графикам видна сильные колебания в связи с сезонностью. Математическое ожидание и дисперсия не меняется, судя по графику временного ряда, однако автокорреляционная функция медленно входит в нулевой доверительный интервал. Временной ряд возможно стационарен.

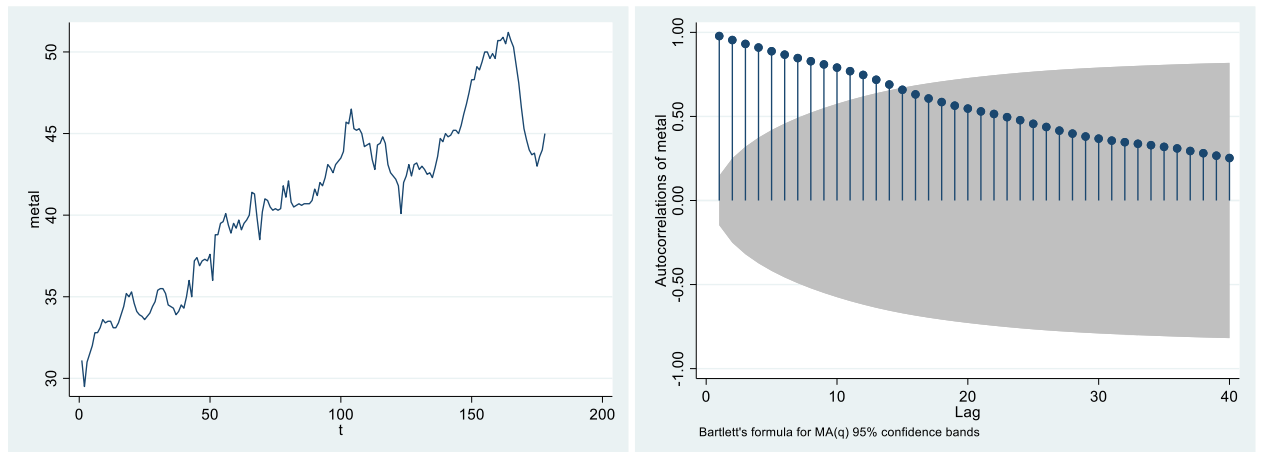


Рисунок 3. График временного ряда занятости в металлургической промышленности.

Временной ряд имеет возрастающий тренд, математическое ожидание меняется со временем, дисперсия так же визуальнo меняется, это говорит о нестационарности ряда. Автокорреляционная функция это подтверждает. Сезонность не заметна в ряду.

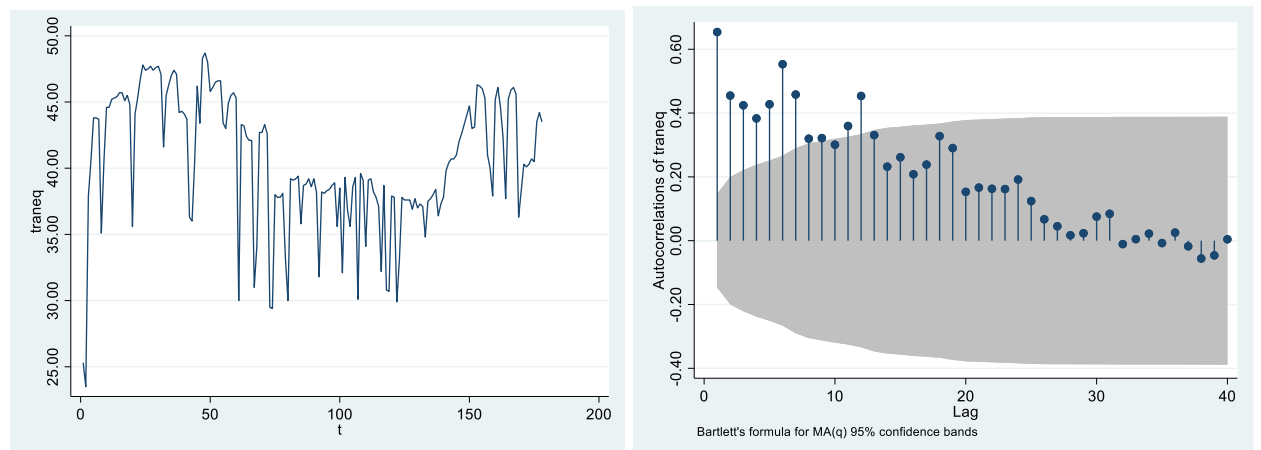


Рисунок 4. График временного ряда занятости в машиностроительной отрасли.

Ряд, возможно, имеет полиномиальный тренд, сезонность не заметна. Математическое ожидание меняется, про дисперсию сложно судить визуальнo. Автокорреляционная функция не очень быстро входит в нулевой доверительный интервал, но сложно говорить о нестационарности.

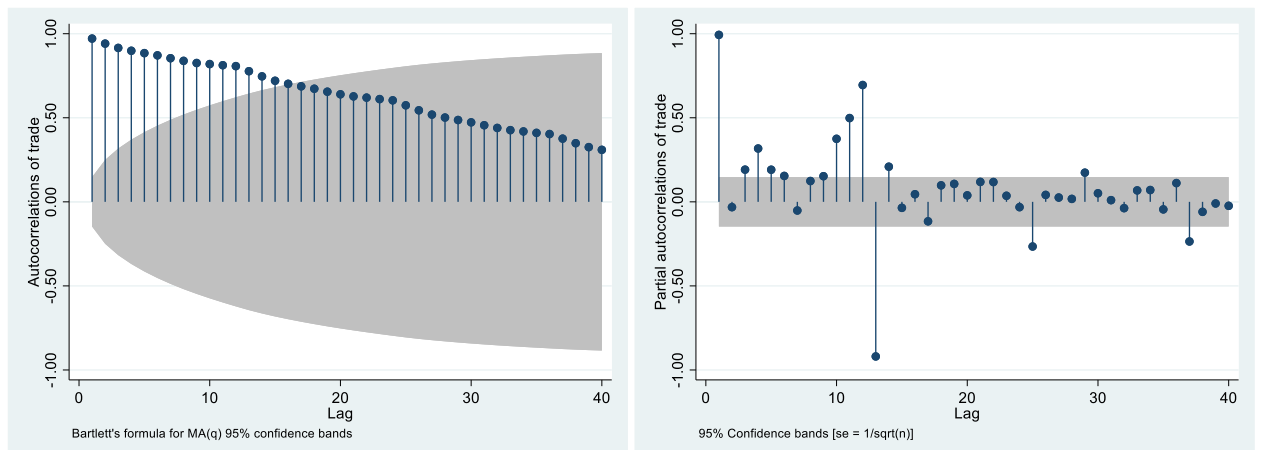


Рисунок 5. Автокорреляционная и частная автокорреляционная функция зависимой переменной.

По автокорреляционной функции видно, что временной ряд имеет тренд, а также не является стационарным, так как автокорреляционная функция очень медленно входит в доверительный интервал. Частная автокорреляционная функция так же говорит о нестационарности ряда, так как значения выходят из нулевого доверительного интервала.

- 2) проанализировать наличие детерминированного тренда, сезонности простого периода, построить модели, оценивающие эти компоненты, скорректировать в случае необходимости.

Визуально по графику временного ряда и по автокорреляционной функции можно сделать вывод о наличии тренда. Тренд напоминает простую линейную функцию, зависящую от переменной времени в данном случае – t . Для устранения тренда возьмем первую разность.

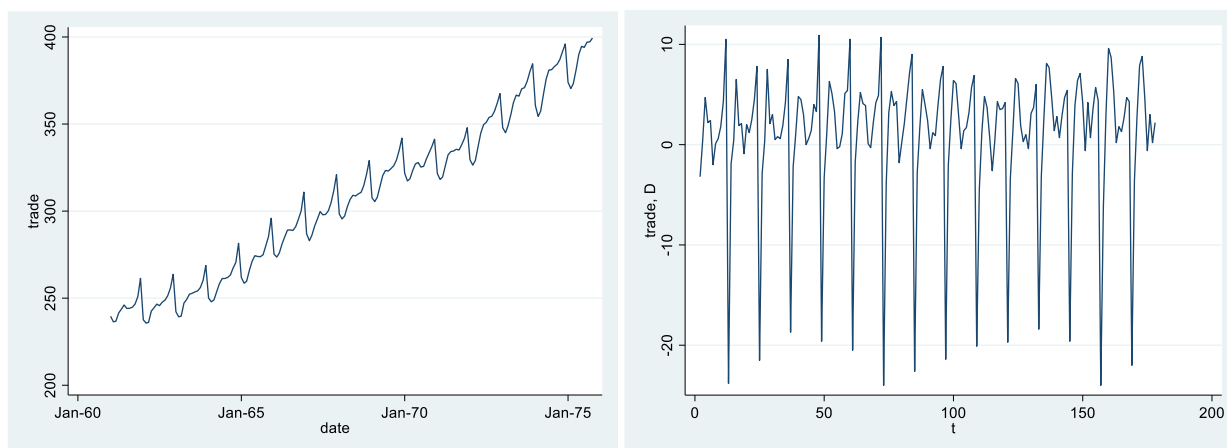


Рисунок 7. График временного ряда до и после взятия первой разности.

После взятия второй разности тренд устранен, но сезонность с возможным интервалом год стала более явной. Однако графики автокорреляционной и частной автокорреляционной функции выглядят более стационарными при взятии разности второго порядка, поэтому далее будем работать с ней.

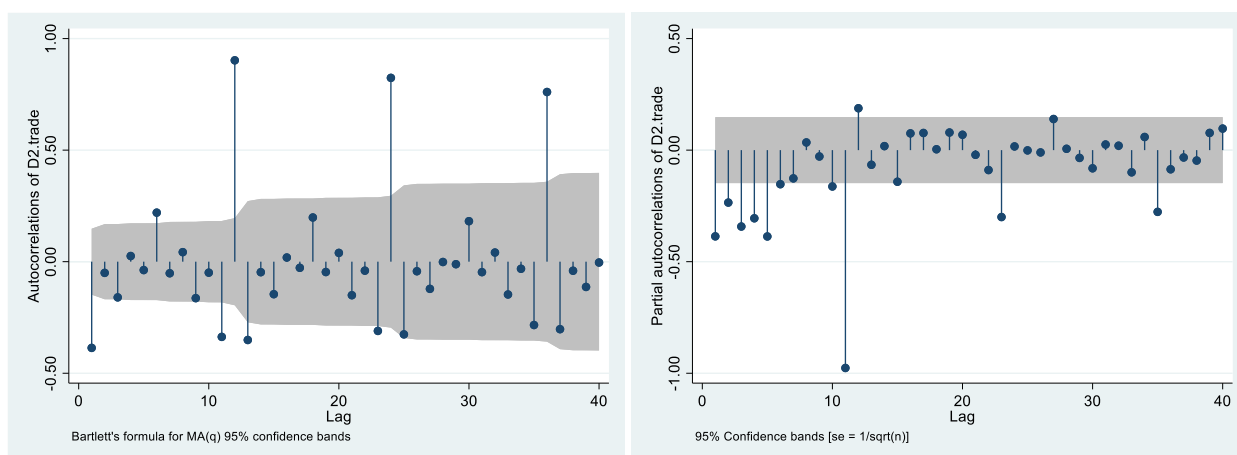


Рисунок 8. График автокорреляционной и частной автокорреляционной функции после взятия второй разности.

После взятия второй разности ряд все еще не стационарный судя по коррелограмме. Далее необходимо избавиться от сезонности. Сезонность можно устранить с помощью сезонного оператора. Визуально по данным можно заметить годовую сезонность, и так как данные

месячные установим сезонный оператор с параметром 12. Ниже приведен график временного ряда с устранением годовой сезонности и с второй разностью.

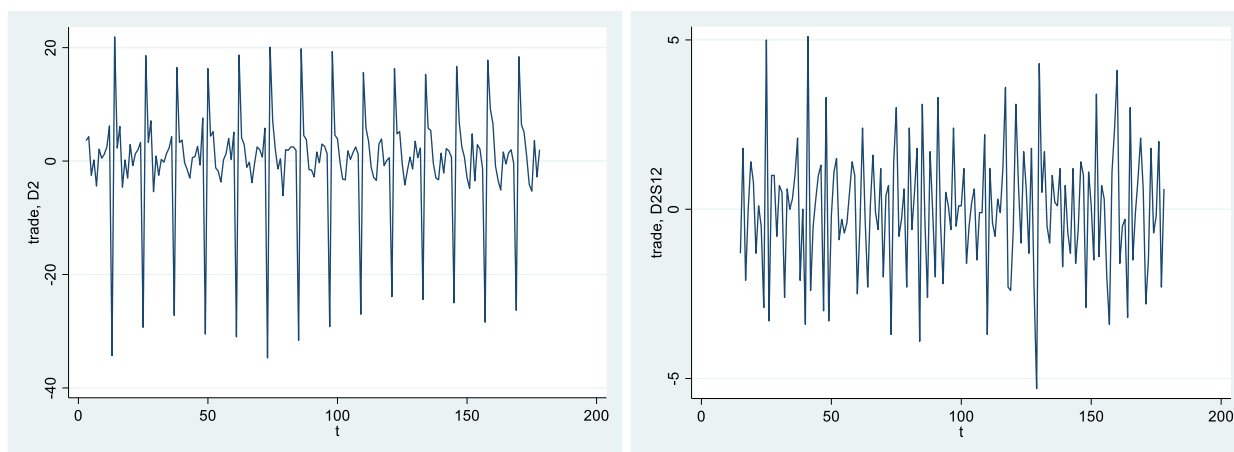


Рисунок 9. График временного ряда до и после устранения годовой сезонности.

Визуально удалось удалить тренд и сезонность. Так же нужно проверить автокорреляционную и частную автокорреляционную функцию на стационарность и автокорреляцию.

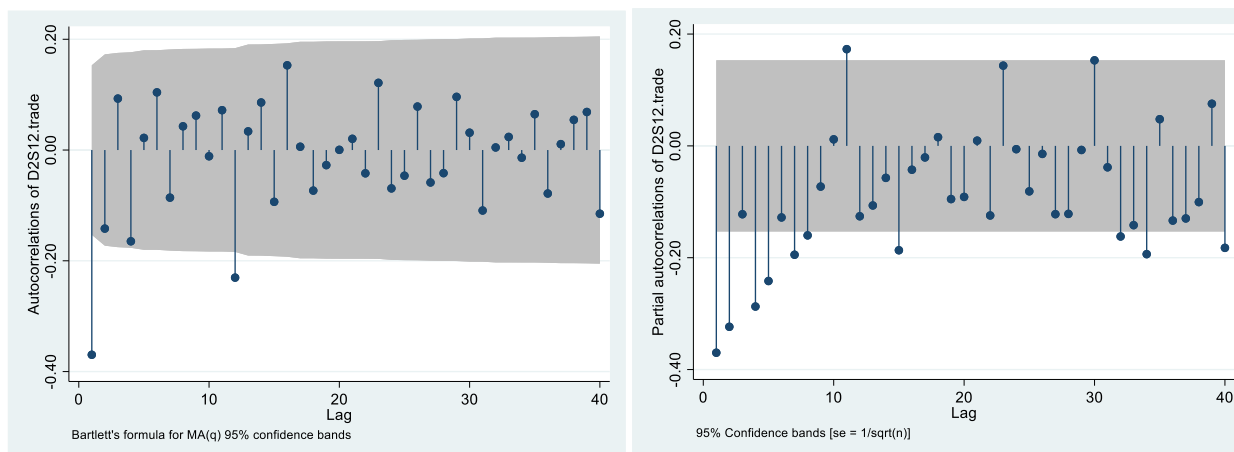


Рисунок 10. График автокорреляционной и частной автокорреляционной функции после взятия второй разности и устранения сезонности.

Автокорреляционная функция имеет пик на первом лаге, а далее входит в 95% нулевой доверительный интервал, что является характеристикой стационарного временного ряда. Частная автокорреляционная функция медленно входит в доверительный интервал, а значительно выходит из него только на первых четырех лагах. Визуально нельзя исключить того, что временной ряд после взятия второй разности и устранения сезонности является стационарным. Следовательно, требуется проведение тестов на стационарность.

На данном этапе исследования удалось обнаружить наличие тренда и сезонности во временном ряду. После их устранения данные выглядят стационарно и с низкой автокорреляцией. Таким образом, можно перейти к построению модели после определения линейного тренда, соответствующего переменной времени – t , а также годовой сезонности.

4) На основе анализа тренда и сезонности строится общая модель тренда и сезонности, корректируется в случае необходимости, на основе нее строится прогноз.

Для построения модели построим регрессию основной переменной на переменные, определяющие тренд и сезонность. В модели выделяется линейный возрастающий тренд,

который можно задать с помощью временной переменной $t = 1, \dots, 178$. Сезонную компоненту определим с помощью создания новой переменной, которая является разницей исходного временного ряда, избавленного от тренда, и построенного программой Stata временного ряда, избавленного от годовой сезонности и тренда. Таким образом создается вторая независимая переменная для регрессии, определяющая сезонность.

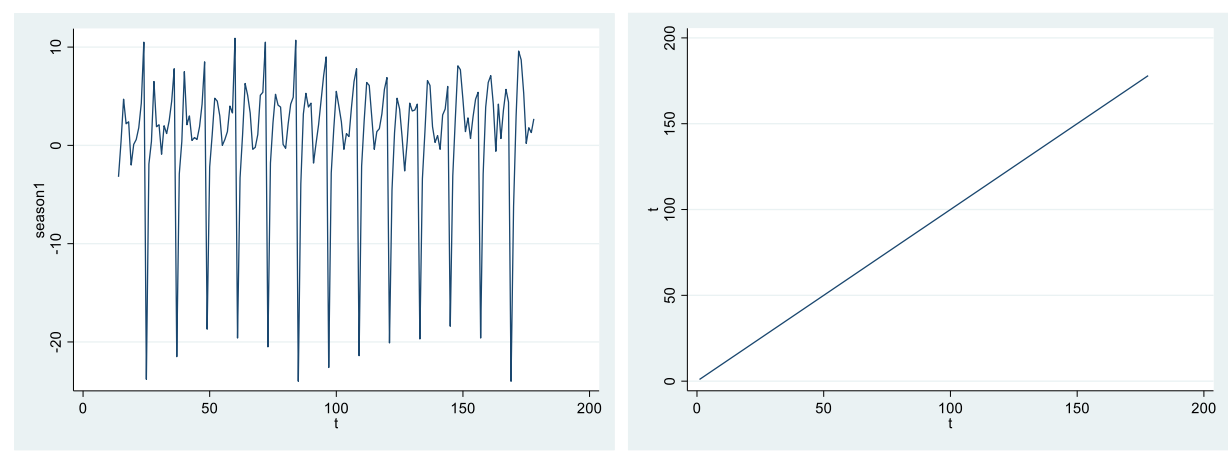


Рисунок 11. График сезонной и трендовой компонент.

На представленные выше переменные проводится регрессия, результаты которой ниже.

Source	SS	df	MS	Number of obs	=	166
Model	332718.502	2	166359.251	F(2, 163)	=	9292.01
Residual	2918.26545	163	17.903469	Prob > F	=	0.0000
				R-squared	=	0.9913
				Adj R-squared	=	0.9912
Total	335636.767	165	2034.16222	Root MSE	=	4.2312

trade	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
t	.1889099	.0348388	5.42	0.000	.1201164	.2577034
season	.8405878	.0391345	21.48	0.000	.7633118	.9178638
_cons	40.43418	8.577215	4.71	0.000	23.4974	57.37096

Рисунок 12. Регрессия зависимой переменной на тренд и сезонность.

Переменные тренда и сезонности являются значимыми в данной модели. Сама модель имеет скорректированный коэффициент R-квадрат равный 99%, что означает, модель хорошо соответствует исходным данным. График построенной модели и исходного временного ряда представлен ниже.

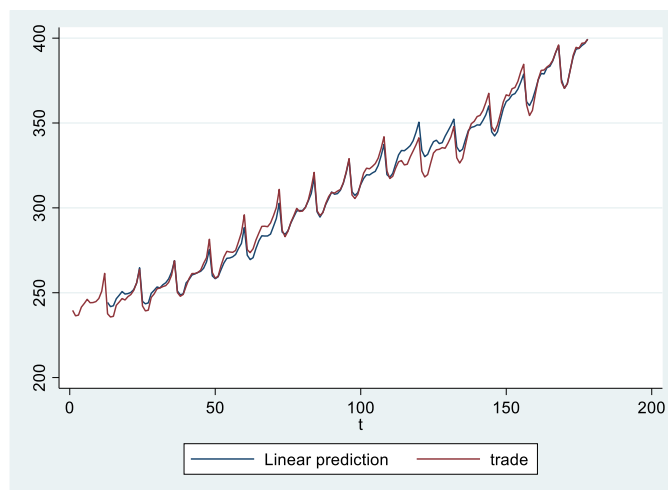


Рисунок 13. График исходной переменной и построенной с помощью регрессии.

5) проанализировать стационарность временного ряда соответствующими тестами, в случае необходимости перейти к конечным разностям или логарифмам.

В тесте Дики-Фуллера основная гипотеза говорит о том, что ряд не стационарен. Для подтверждения стационарности теста тестовая статистика должна быть меньше критического значения на определенном уровне значимости.

```
. dfuller e, regress
```

Dickey-Fuller test for unit root Number of obs = 165

Interpolated Dickey-Fuller				
Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value	
Z(t)	-2.943	-3.488	-2.886	-2.576

MacKinnon approximate p-value for Z(t) = 0.0406

D.e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
e						
L1.	-.0924893	.0314255	-2.94	0.004	-.154543	-.0304357
_cons	.0407529	.1321609	0.31	0.758	-.2202152	.3017211

Рисунок 14. Результаты теста Дики-Фуллера.

Согласно результатам теста, ряд является стационарным на уровне значимости 5%, так как p-value теста равно 0,04.

В тесте KPSS основная гипотеза говорит о том, что ряд стационарен. Тестовая статистика должна быть меньше критической на выбранном уровне значимости, чтобы основная гипотеза не была отвергнута.

```
. kpss e , notrend auto
```

KPSS test for e

Automatic bandwidth selection (maxlag) = 9
Autocovariances weighted by Bartlett kernel

Critical values for H_0 : e is level stationary

10%: 0.347 5% : 0.463 2.5%: 0.574 1% : 0.739

Lag order	Test statistic
9	.122

Рисунок 15. Результаты теста KPSS.

По результатам данного теста ряд является стационарным на уровне значимости 1%. Так как тест KPSS является более мощным, можно утверждать, что ряд стационарен и продолжить работать с ним без взятия разностей.

6) проанализировать наличие и причины автокорреляции в исходном ряду и в остатках от модели из п4 разными тестами, подтвердить коррелограммой.

Тест серий

Основная гипотеза говорит о том, что ряд не имеет автокорреляцию. Альтернативная гипотеза: имеется автокорреляция первого порядка.

```
. runtest e, threshold(0)
N(e <= 0) = 73
N(e > 0) = 93
obs = 166
N(runs) = 24
z = -9.289999999999999
Prob>|z| = 0
```

Рисунок 16. Результаты теста серий.

P-value равно нулю, следовательно изначальный временной ряд не имеет автокорреляции, согласно критерию серий.

Тест Дарбина – Уотсона

Основная гипотеза говорит о том, что ряд не имеет автокорреляцию. Проведем тест на уровне значимости 1%.

```
. estat dwatson
```

```
Durbin-Watson d-statistic( 3, 166) = .1696211
```

Рисунок 17. Тестовая статистика Дарбина-Уотсона.

В данном случае тестовая статистика входит в интервал [1,665;2,335], в котором основная гипотеза не отвергается. Следовательно ряд стационарен, согласно данному тесту.

Оба теста говорят об отсутствии автокорреляции, посмотрим на коррелограммы.

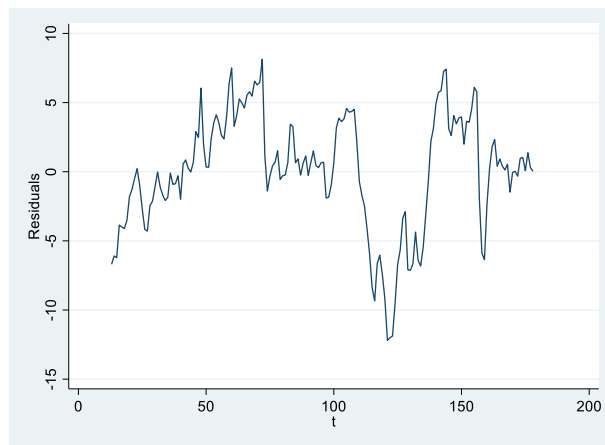


Рисунок 18. Временной ряд, очищенный от тренда и сезонности.

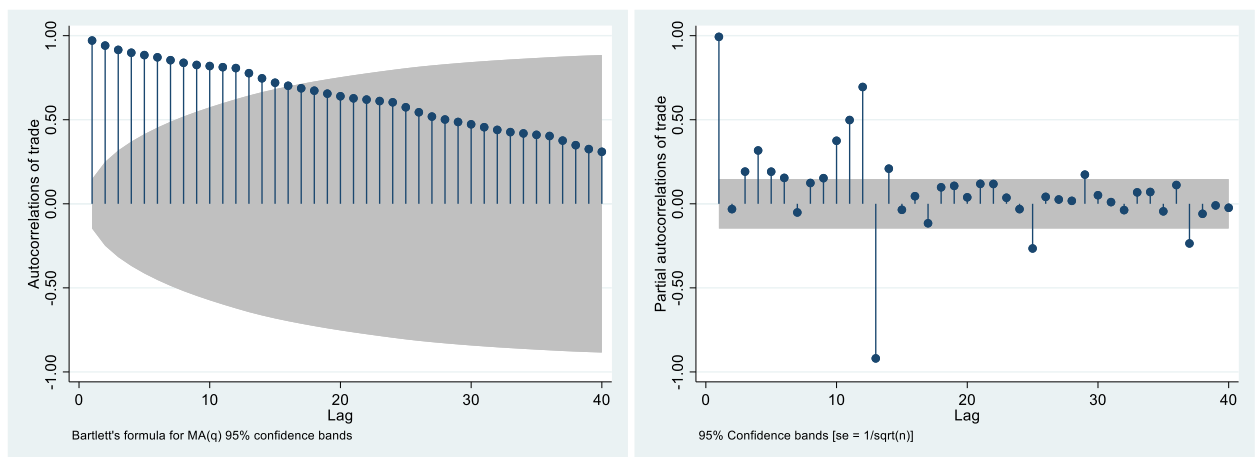


Рисунок 19. Автокорреляционная и частная автокорреляционная функция зависимой переменной.

В исходном временном ряду видна небольшая автокорреляция, обе функции входят в доверительный интервал только после примерно 14-го лага. Возможно положительная автокорреляция.

Встроенная в программу Stata команда `wntestq` является тестом Льюнга-Бокса. Основная гипотеза теста говорит об отсутствии автокорреляции. Другой тест ниже `wntestb` использует другую статистику, но основная гипотеза остается такой же.

```
. wntestq e
```

```
Portmanteau test for white noise
```

```
Portmanteau (Q) statistic = 762.9434
Prob > chi2(40)           = 0.0000
```

```
.
```

```
. wntestb e, table
```

```
Cumulative periodogram white-noise test
```

```
Bartlett's (B) statistic = 6.5061
Prob > B                 = 0.0000
```

Рисунок 20. Результаты тестов на отсутствие автокорреляции остатков.

Тесты говорят об отсутствии автокорреляции на уровне значимости 1%. Посмотрим на графики автокорреляционных функций.

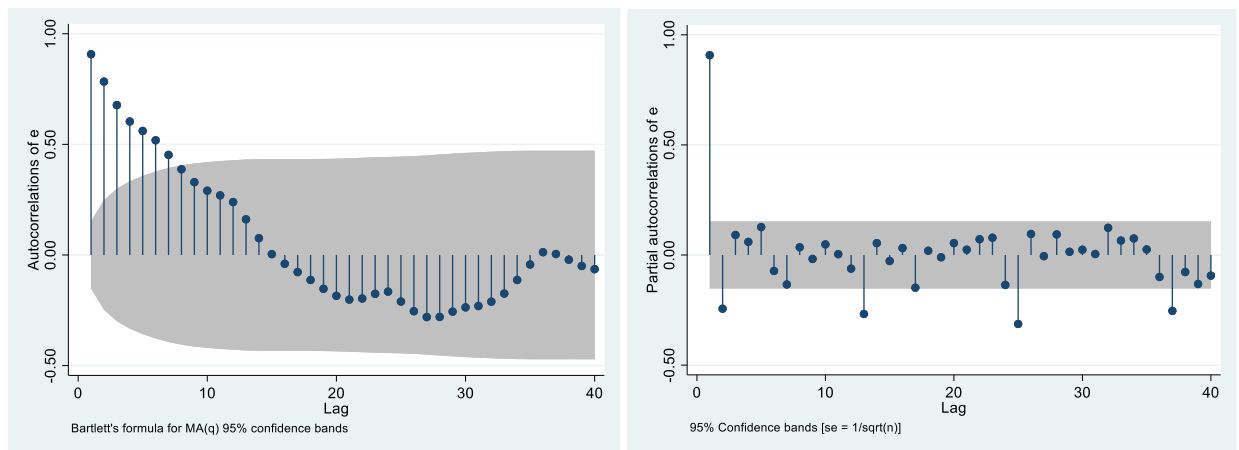


Рисунок 21. Автокорреляционная и частная автокорреляционная функция временного ряда, очищенного от тренда и сезонности.

Получившийся ряд визуально показывает меньшую автокорреляцию и более стационарен.

7) Строится точечный прогноз по модели из п4, приводится алгоритм расчетов, график прогноза.

Точечный прогноз будет строится по модели обычной регрессии (OLS) в gretl. Сначала необходимо выбрать данную модель регрессии, чтобы повторить модель, построенную в программе Stata. Далее выбрать те же независимые переменные и нужную зависимую, эти действия представлены ниже.

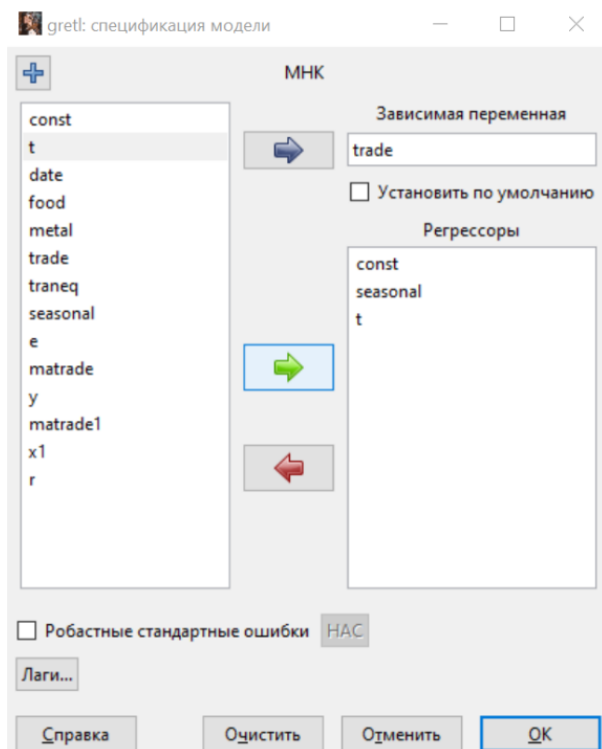


Рисунок 22. Ввод данных в программу gretl.

После этого программа построит требуемую модель, идентичную по коэффициентам с построенной в Пункте 4.

Модель 2: МНК, использованы наблюдения 1962:01-1975:10 (T = 166)
Зависимая переменная: trade

	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	40,4342	8,57722	4,714	5,18e-06	***
seasonal	0,840588	0,0391345	21,48	2,11e-049	***
t	0,188910	0,0348388	5,422	2,09e-07	***
Среднее зав. перемен	312,1042	Ст. откл. зав. перемен	45,10169		
Сумма кв. остатков	2918,265	Ст. ошибка модели	4,231249		
R-квадрат	0,991305	Испр. R-квадрат	0,991199		
F(2, 163)	9292,012	P-значение (F)	1,1e-168		
Лог. правдоподобие	-473,4846	Крит. Акаике	952,9692		
Крит. Шварца	962,3052	Крит. Хеннана-Куинна	956,7588		
Параметр rho	0,907510	Стат. Дарбина-Вотсона	0,169621		

Рисунок 23. Модель регрессии, построенная программой gretl.

После этого программа gretl может построить прогноз по каждому наблюдению с выделением 95% интервала.

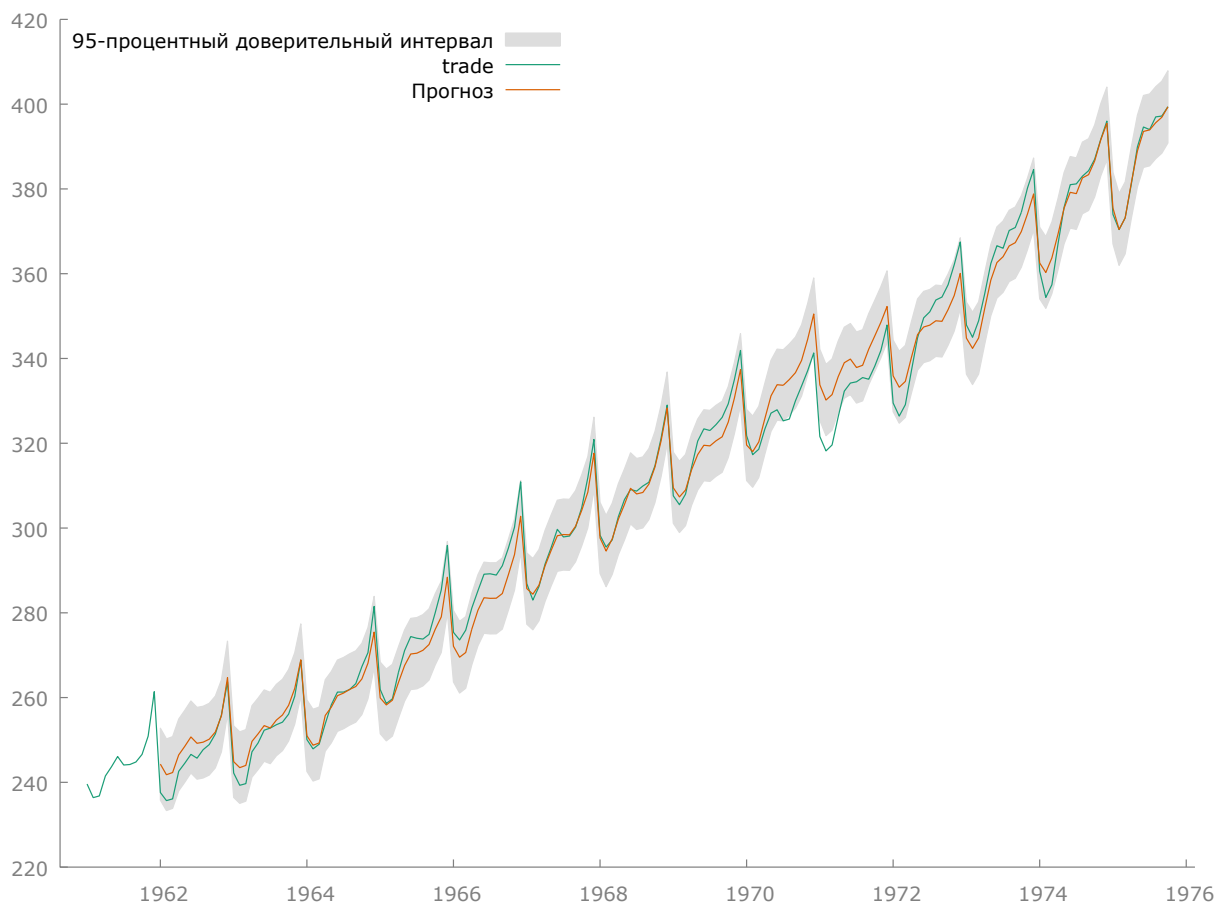


Рисунок 24. Прогноз модели регрессии, построенной программой gretl, с выделением 95% интервала.

Выдели точечные прогнозы на 6 последних месяцев.

trade Предсказание Ст. ошибка 95% доверительный интервал

1974:11	391,7	391,6	4,32	383,0 –	400,1
1974:12	396,0	395,5	4,34	386,9 –	404,0
1975:01	374,0	375,5	4,28	367,0 –	383,9
1975:02	370,4	370,5	4,30	362,0 –	378,9
1975:03	373,2	373,2	4,29	364,7 –	381,6
1975:04	381,1	381,4	4,28	373,0 –	389,9
1975:05	389,9	388,9	4,28	380,5 –	397,4
1975:06	394,6	393,6	4,30	385,1 –	402,1
1975:07	394,0	393,9	4,30	385,4 –	402,4
1975:08	397,0	395,6	4,30	387,1 –	404,1
1975:09	397,2	396,9	4,30	388,4 –	405,4
1975:10	399,4	399,4	4,31	390,9 –	407,9

Рисунок 25. Прогноз модели регрессии на полгода, построенной программой gretl, с выделением 95% интервала.

В последнем месяце прогноз не отличается от исходных данных ошибка прогноза равно 0.

8) Анализ коррелограммы, построение модели АРИМА и прогнозирование по ней.

Построим модель ARIMA с помощью встроенных в программу Stata функций очищения от тренда и стационарности. Как выяснено в пункте 2, лучше взять вторую разность и годовую сезонность.

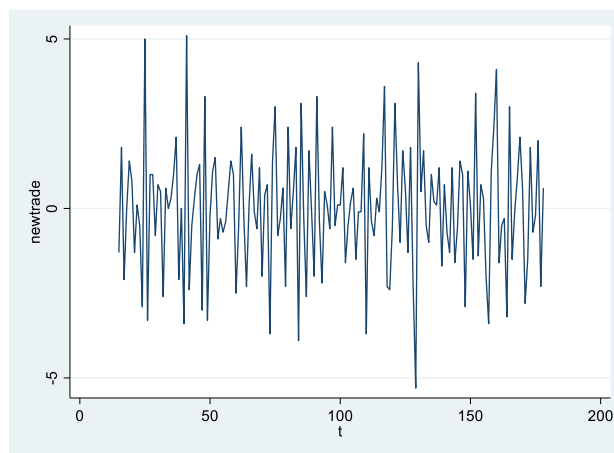


Рисунок 26. Временной ряд, очищенный от тренда и сезонности.

Визуально ряд не имеет тренда или сезонности, математическое ожидание и дисперсия неизменны.

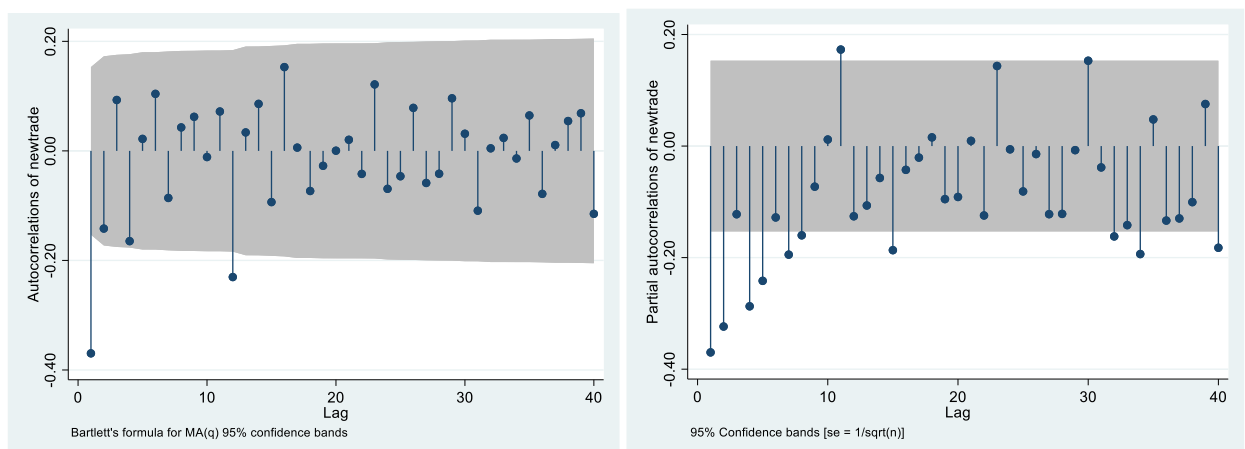


Рисунок 27. Автокорреляционная и частная автокорреляционная функция временного ряда, очищенного от тренда и сезонности.

По автокорреляционной функции видно, что AR-параметр модели ARIMA равен одному, так как функция имеет пик на первом лаге. По частной автокорреляционной функции заметно, что параметр MA модели скорее всего равен 5, 6 или 7. Выберем параметр 5 для построения модели.

Прогнозирование будет строится в программе gretl, алгоритм аналогичен Пункту 6. Выберем модель ARIMA(1,0,5) для построения прогноза на данных с двойной разностью.

gretl: спецификация модели

ARIMA

Зависимая переменная: newtrade_01

☒ Установить по умолчанию

Регрессоры

Несезонные компоненты: AR 1 I 0 MA 5

Перечислить: ☐ ☐

Сезонные компоненты: AR 0 I 0 MA 0

☐ Включить константу ☐ Показать отчет об итерациях

☒ Оценивать матрицу ковариаций с помощью Гессияна

☒ Разность для независимых переменных

☐ Использовать X-12-ARIMA

Точный метод максимального правдоподобия

Параметры

Справка Очистить Отменить OK

Рисунок 28. Ввод данных в программу gretl.

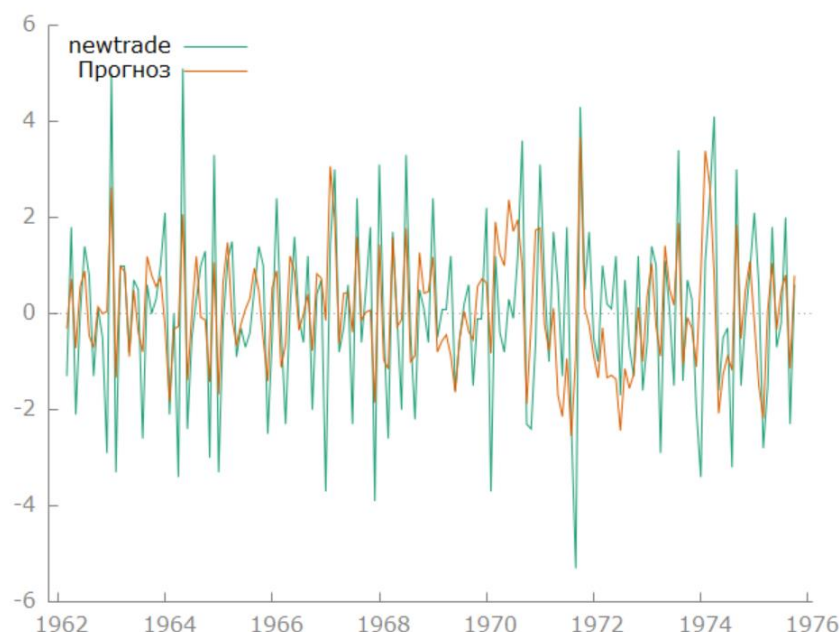


Рисунок 29. Данные без сезонности и с второй разницей и построенная ARIMA(1,2,5).

	newtrade	Предсказание
1975:04	-1,500000	0,063320
1975:05	1,800000	1,048131
1975:06	-0,700000	-0,328883
1975:07	-0,200000	0,465643
1975:08	2,000000	0,805865
1975:09	-2,300000	-1,142131
1975:10	0,600000	0,790148

Рисунок 30. Прогнозы по ARIMA(1,2,5) на полгода.

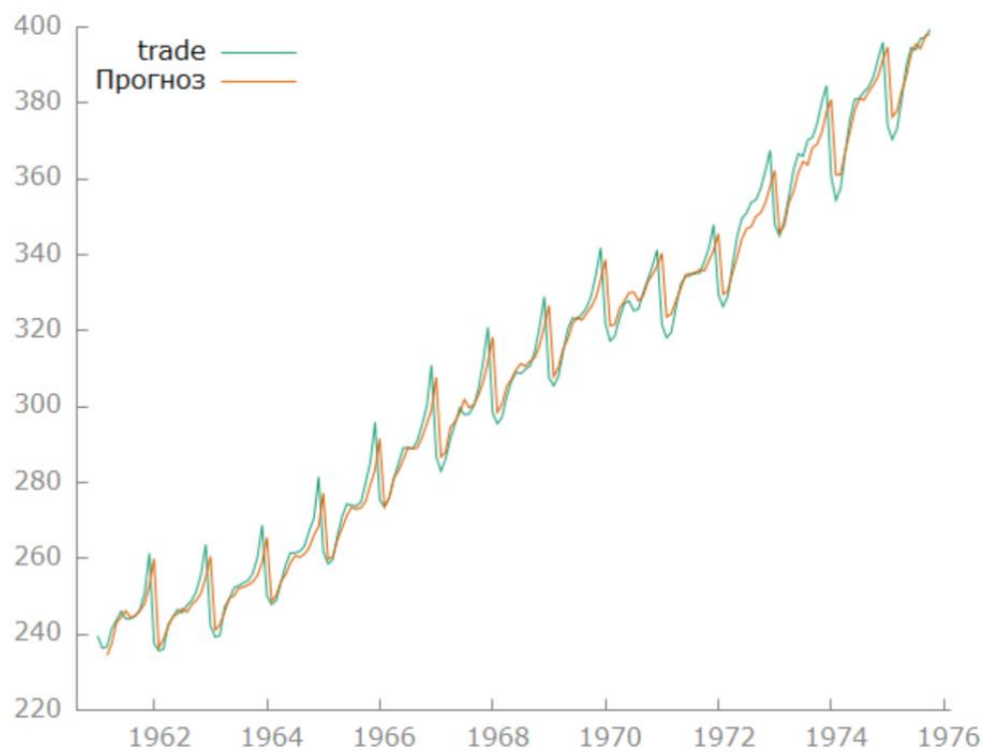


Рисунок 31. Исходные данные и модель ARIMA(1,2,5)

	trade	Предсказание
1975:04	381,1	383,1
1975:05	389,9	387,0
1975:06	394,6	393,0
1975:07	394,0	395,5
1975:08	397,0	394,3
1975:09	397,2	397,7
1975:10	399,4	398,2

Рисунок 32. Исходные данные и предсказания по модели ARIMA(1,2,5).

Средняя ошибка (ME) равна -0,065804.