

Домашняя работа

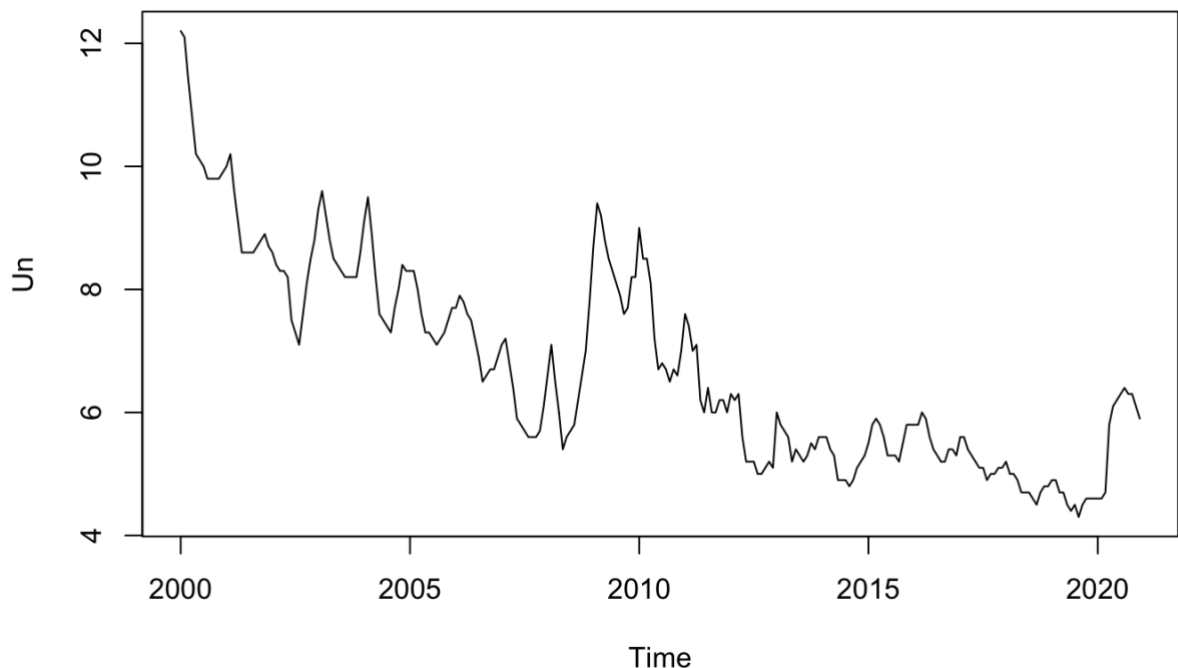
Эконометрика

Весь процесс написания кода производился в программе R и к файлу приложен документы html и Rmd формата с кодом необходимым для повторения полученных результатов.

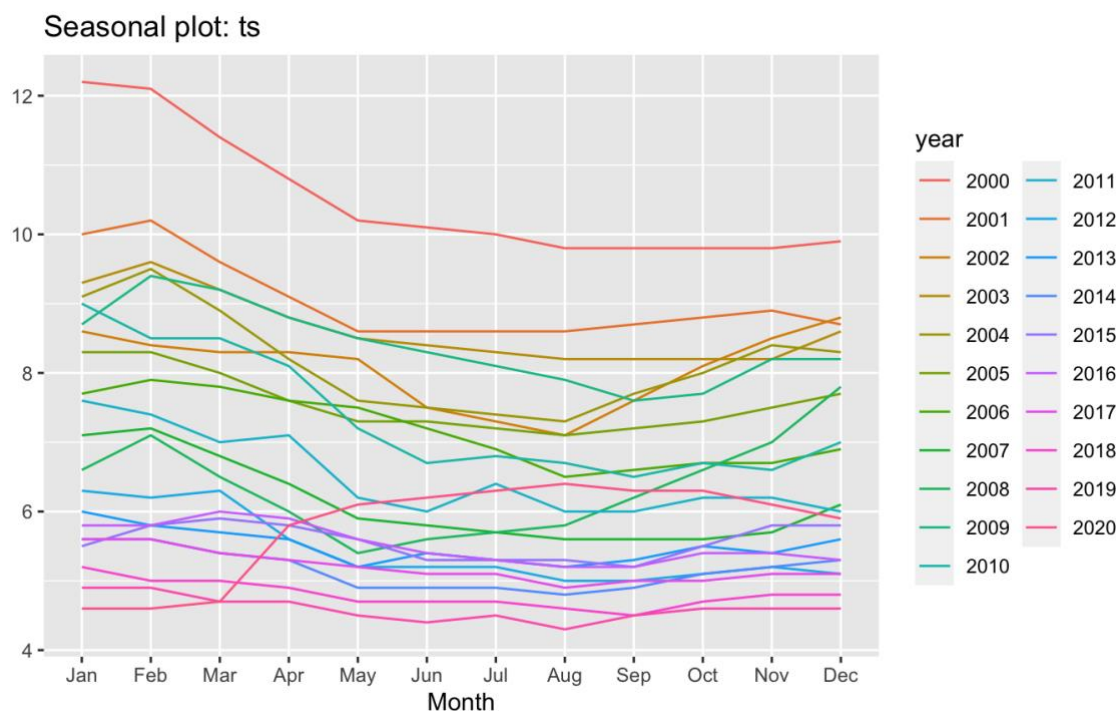
Часть 1

В качестве используемого временного ряда мы взяли месячный ряд уровня безработицы в РФ с сайта <http://sophist.hse.ru/hse/nindex.shtml>

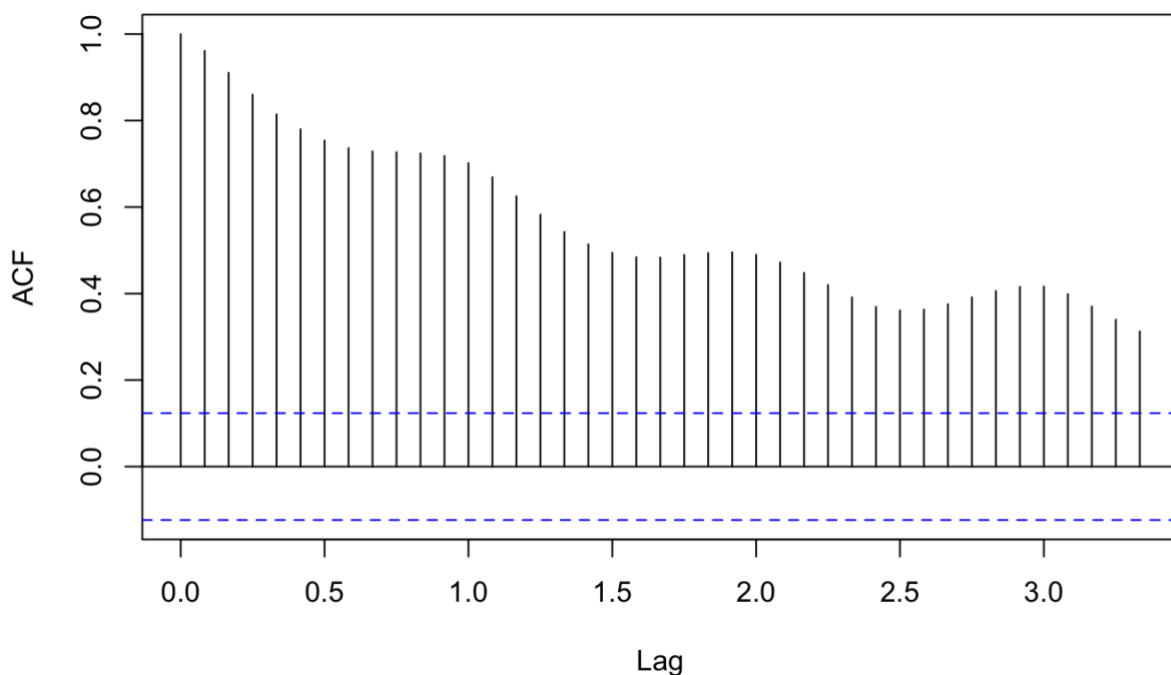
В нашем ряду мы имеем 252 наблюдения с января 2000 года до декабря 2020 года. Сначала мы построили график нашего временного ряда.



На графике можно увидеть явный нисходящий тренд, который связан со стабилизацией российской экономики после 90-х годов. Более того, в России с годами заработная плата стала намного более показательной величиной для анализа рынка труда, так как работодатели предпочитают не увольнять сотрудников, а уменьшать заработную плату. Также в динамике можно увидеть несколько резких скачков: первый из них – в 2008 году, когда в стране произошел кризис на фоне финансового кризиса в США. Второй – в 2020 году, когда из-за коронавируса часть предприятий сферы услуг закрылась и многие работники данной отрасли потеряли работу. На графике присутствует некая цикличность, но для более точного анализа мы построили график с разделениями по годам.

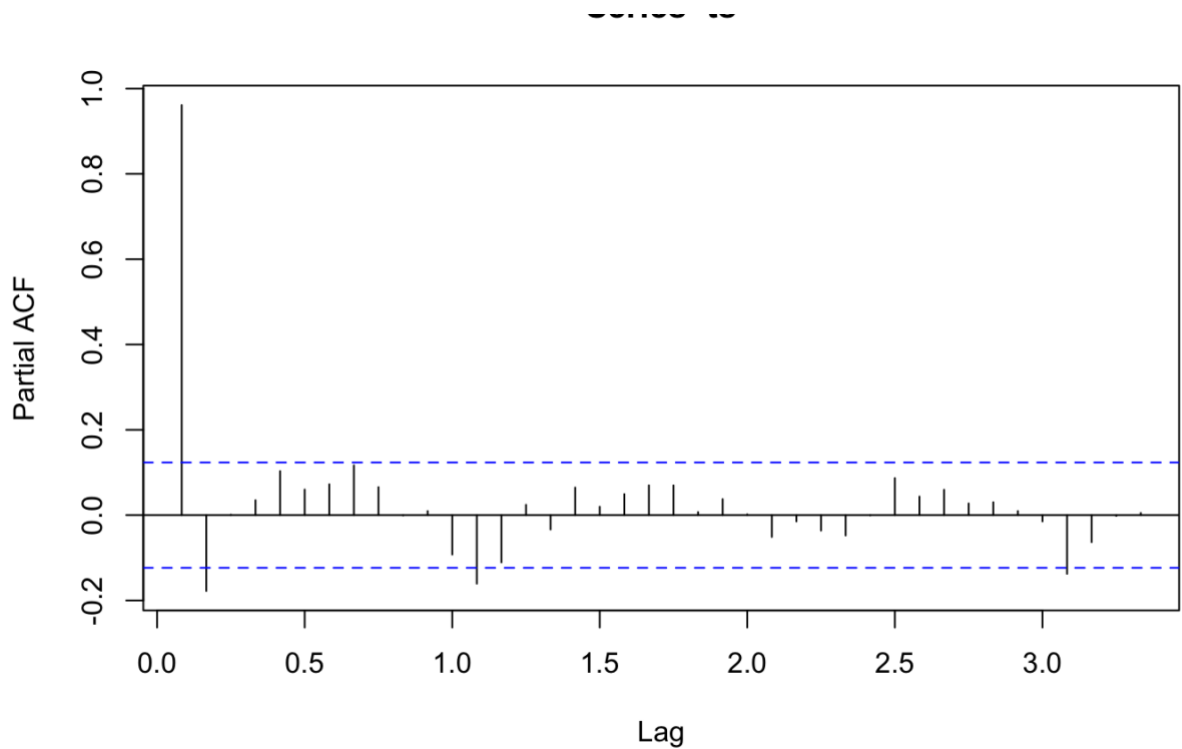


На данном графике явной сезонности действительно не наблюдается. Можно сказать, что в начале двухтысячных в данных была особенность “провисания” ряда во второй половине года, однако, это скорее характеризует тренда ряда, а не его сезонность.



На графике ACF можно увидеть постепенное понижение значений корреляции с увеличением лага, что говорит о наличии явного тренда в ряду. Более того, на PACF графике первый лаг достигает значения 1, что также

указывает на предопределенность каждого последующего значения ряда. Однако сезонность ряда также подтверждается с помощью графиков ACF и PACF, что не было очевидно на предыдущих графиках. В данном случае на обоих графиках значения лагов 12,24 и т.д. имеют большие значения относительно общей динамики лагов. А на PACF графике эти скачки являются значимыми (в первую очередь для лага 12), что еще раз подтверждает наличие сезонности.

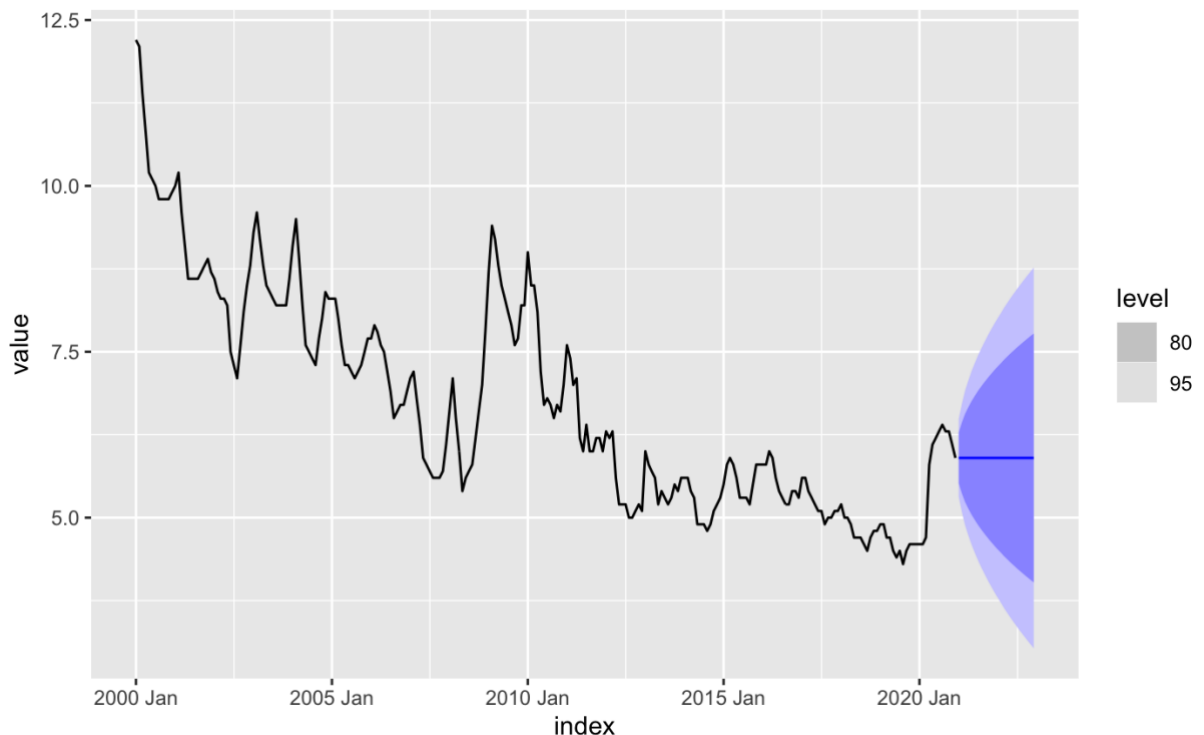


После этого мы построили 6 моделей на промежутке 2000-2019 и сравнили значение показателя MAE для значений 2020 года.

Модель	MAE
Наивная модель	1.175
Наивная сезонная модель	1.275
ETS (AAA)	1.745783
ETS (авто.)	1.473252
SARIMA (1,1,1)(1,0,0)(12)	1.398274
SARIMA (auto)	1.56997

Оказалось, что лучше всего себя на данном промежутке ведет обычная наивная модель. Это скорее всего связано с тем, что остальные модели выявляли наличие тренда в ряду и строили прогноз с учетом данного тренда, а безработица в 2020 году повела себя иначе из-за экзогенного шока. Именно поэтому модель, которая просто повторяет значения декабря 2019 года оказывается наилучшей.

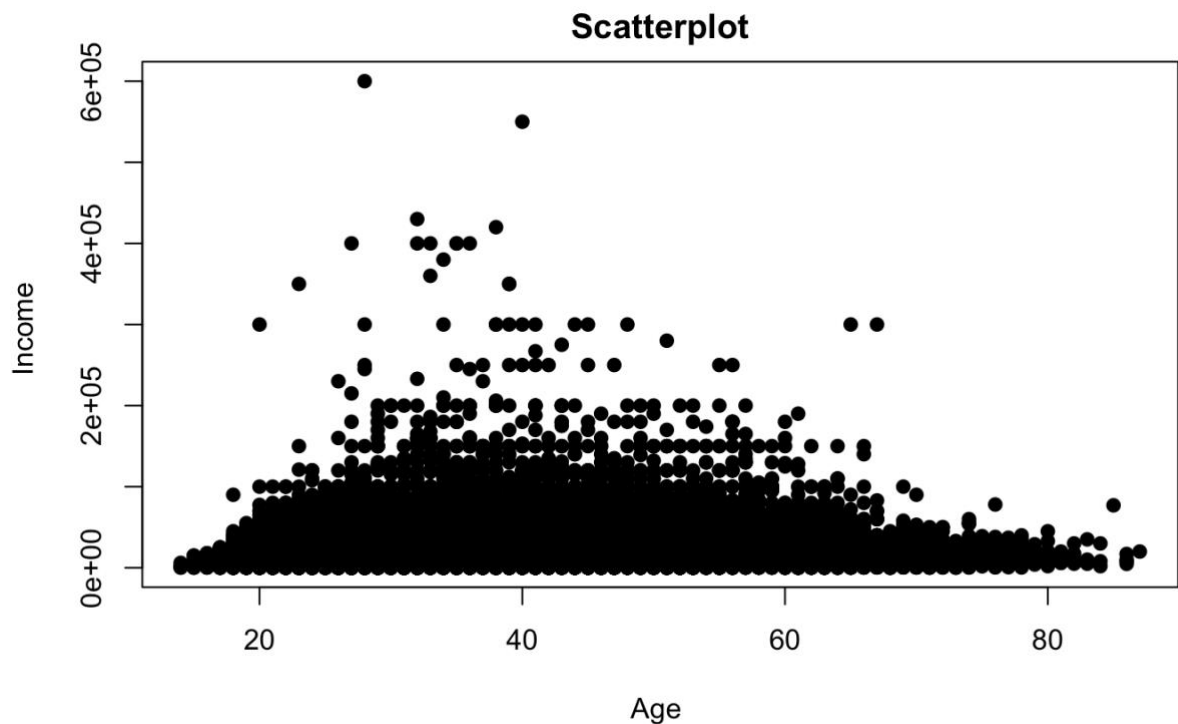
С помощью этой модели был построен прогноз и на 2021 год. При этом все будущие значения ряда оказываются равны 5,9.



Часть 2

Для изучения панельных данных, была использована выборка RLMS с 2000 до 2019 года. Всего в данных имеется 133645 наблюдений, при этом сами данные являются несбалансированными из-за проведенной очистки пропущенных значений.

В данной работе нас интересует вопрос: “Влияет ли возраст на размер заработной платы?”



По диаграмме рассеивания можно предположить, что возраст действительно влияет на доход индивида, однако данное влияние скорее всего нелинейно. Так, от 20 до 45 лет люди зарабатывают больше всего, а разброс показателей также наибольший. До и после этого промежутка среднее значение данного показателя сильно уменьшается, что может указывать на наличие квадратичной составляющей в модели.

Для оценивания зависимости на панельных данных, нами были оценены Pooled, FE и RE модели.

Pooled и FE модели сравниваются друг с другом с помощью F теста на значимость детерминированных индивидуальных эффектов.

F test for individual effects

data: J10 ~ AGE

F = 12.421, df1 = 26923, df2 = 106720, p-value < 2.2e-16

alternative hypothesis: significant effects

В данном случае p-value признается очень близким к нулю, что указывает на значимость индивидуальных эффектов.

Pooled и RE модели сравниваются друг с другом с помощью теста Бройша-Пагана, нулевая гипотеза в котором заключается в равенстве дисперсии случайных эффектов нулю.

Lagrange Multiplier Test - (Breusch-Pagan) for unbalanced panels

```
data: J10 ~ AGE
chisq = 119891, df = 1, p-value < 2.2e-16
alternative hypothesis: significant effects
```

В данном случае p-value признается очень близким к нулю, что указывает на значимость индивидуальных эффектов.

RE и FE модели сравниваются с помощью теста Хаусмана. При принятии нулевой гипотезы обе модели признаются состоятельными, но только модель RE признается эффективной. В то же время при принятии альтернативной гипотезы, состоятельной является только модель FE.

Hausman Test

```
data: J10 ~ AGE
chisq = 32829, df = 1, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

В данном случае p-value признается очень близким к нулю, что указывает на несостоятельность RE модели.

Так как по результатам тестирования наилучшей была признана модель FE, именно ее нужно использовать для ответа на поставленный вопрос.

Oneway (individual) effect Within Model

Call:

```
plm(formula = J10 ~ AGE, data = Panel, model = "within")
```

Unbalanced Panel: n = 26924, T = 1-20, N = 133645

Residuals:

Min.	1st Qu.	Median	3rd Qu.	Max.
-222223.6	-2785.6	0.0	2427.6	390176.7

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t)
AGE	1529.3019	7.4361	205.66	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 1.314e+13

Residual Sum of Squares: 9.4108e+12

R-Squared: 0.28383

Adj. R-Squared: 0.10316

F-statistic: 42295.8 on 1 and 106720 DF, p-value: < 2.22e-16

Показатель R² является не особо большим, хотя модель признается адекватной в целом, а независимый регрессор признается значимым, что определяет данную модель как правильно специфицированную. По результатам оценивания можно сказать, что увеличение возраста индивида на 1 год увеличивает в среднем доход индивида на 1529 рублей.