Republic of North Macedonia
Ministry of Finance

Descriptive and Predictive analytics with customs data: Case of Republic of North Macedonia

Jordan Simonov M.Sc
Zoran Gligorov M.Sc

Skopje, 22-24 October 2019

Outline

1. Types of Big Data Analytics :
- Descriptive analytics
- Predictive analytics

2.Descriptive analytics:
- MFN tariff rates
- Preferential tariff rates

3. Predicitive analytics:
- Seasonal Naïves
- Exponential smoothing
- Autoregressive integrated moving average-ARIMA

4. Conclusions

1. Types of Big Data Analytics

Big data is defined as collections of datasets whose volume, velocity or variety is so large that it is difficult to store, manage, process and analyze using traditional databases and data processing tools (Prajapati,2013)

Big data definitions (evolution) :

- **3Vs (Volume, Velocity and Variety)**
- **4Vs (Volume, Velocity, Variety and Veracity)**
- **5Vs (Volume, Velocity, Variety, Veracity and Value)**

Types of Big Data Analytics

2. Descriptive analytics

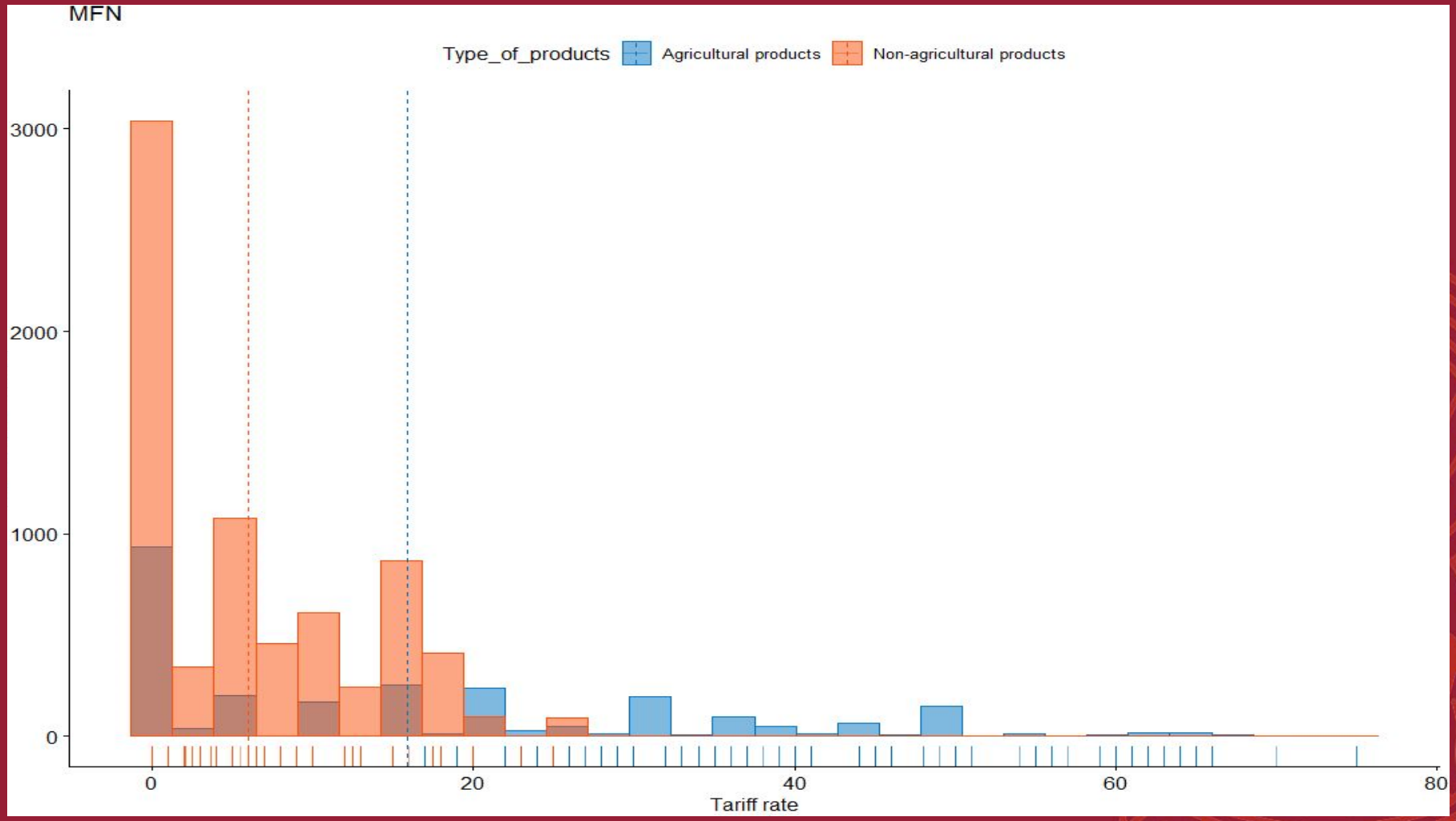
Simple average MFN tariff rate 8,6%

Descriptive statistics of MFN tariff rates

Description	Agricultural products	Non-agricultural products
Simple average	15,9	6
Standard deviation	17,4	6,6
Median	10	5
Trimmed mean	13,3	5,2
Median absolute deviation	14,8	7,4
Min	0	0
Max	75	25
Range	75	25
Skew	1	0,8
Kurtosis	-0,1	-0,5
Standard error	0,3	0,1
Var	304,3	43,28

2. Descriptive analytics

Histogram of MFN tariff rates by HS division



2. Descriptive analytics

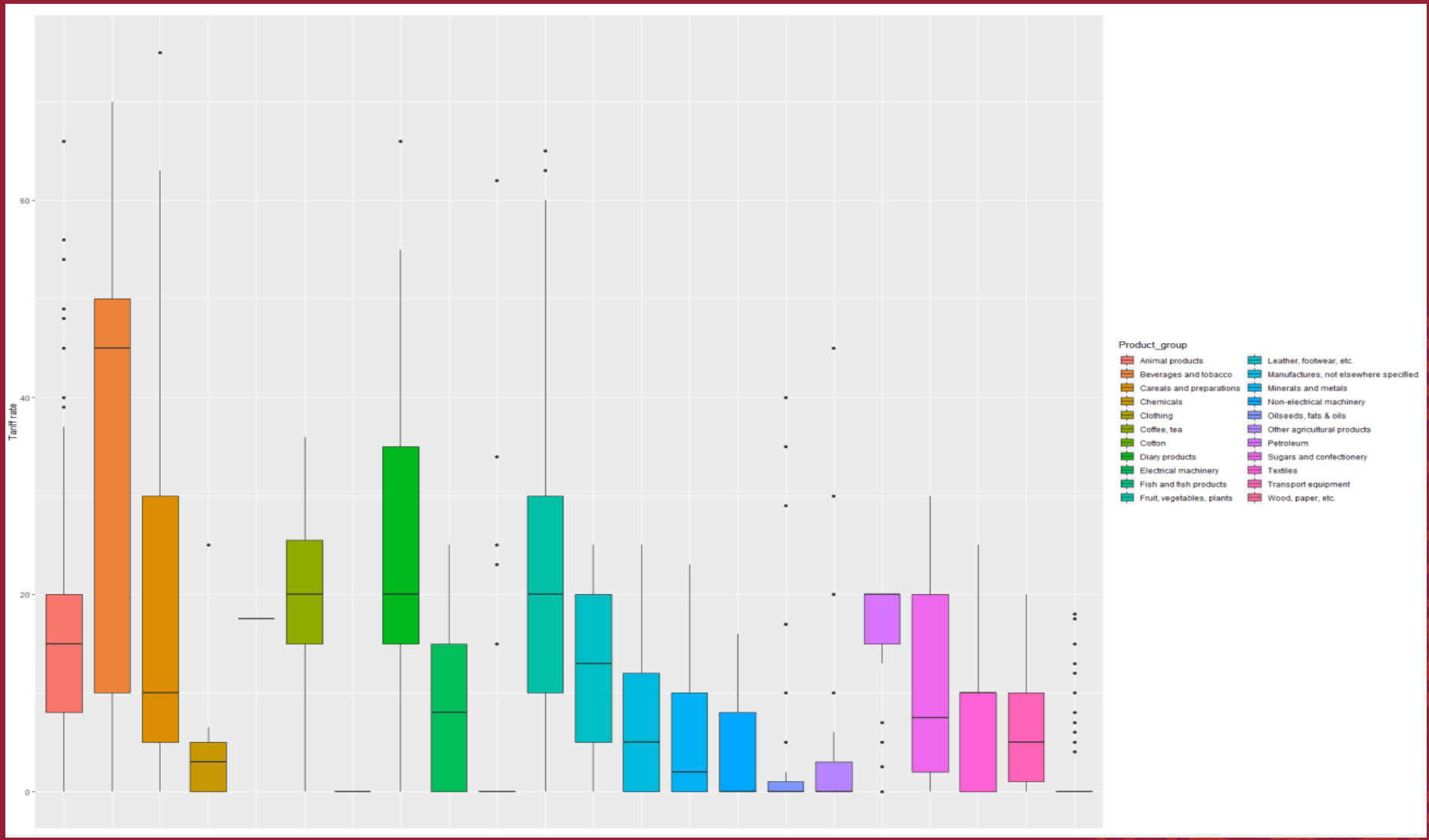
Descriptive statistics of MFN tariff rates

Description	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Animal products	359	15,7	11,0	15,0	14,8	7,4	0,0	66,0	66,0	1,1	2,4	0,6
Dairy products	153	21,9	13,5	20,0	20,9	14,8	0,0	66,0	66,0	0,9	1,2	1,1
Fruit, vegetables, plants	570	21,8	15,2	20,0	20,8	14,8	0,0	65,0	65,0	0,6	0,0	0,6
Coffee, tea	47	18,6	9,6	20,0	19,0	7,4	0,0	36,0	36,0	-0,4	-0,6	1,4
Cereals & preparations	237	19,5	16,3	10,0	18,1	14,8	0,0	75,0	75,0	0,8	0,3	1,1
Oilseeds, fats & oils	175	2,6	7,0	0,0	0,9	0,0	0,0	40,0	40,0	3,9	16,1	0,5
Sugars and confectionery	44	10,8	10,4	7,5	9,9	11,1	0,0	30,0	30,0	0,4	-1,4	1,6
Beverages and tobacco	356	34,0	22,0	45,0	34,9	14,8	0,0	70,0	70,0	-0,6	-1,3	1,2
Cotton	6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Other agricultural products	261	2,8	6,7	0,0	1,1	0,0	0,0	45,0	45,0	4,0	18,6	0,4
Fish and fish products	528	1,9	7,6	0,0	0,0	0,0	0,0	62,0	62,0	5,6	36,8	0,3
Minerals & metals	1465	5,1	6,4	2,0	4,4	3,0	0,0	23,0	23,0	0,9	-0,8	0,2
Petroleum	49	15,2	6,5	20,0	16,1	0,0	0,0	20,0	20,0	-1,2	-0,1	0,9
Chemicals	1380	3,0	3,3	3,0	2,7	4,5	0,0	25,0	25,0	2,3	13,6	0,1
Wood, paper, etc.	473	2,7	5,1	0,0	1,6	0,0	0,0	18,0	18,0	1,6	1,0	0,2
Textiles	838	8,1	5,5	10,0	7,9	3,7	0,0	25,0	25,0	-0,2	-0,7	0,2
Clothing	341	17,5	0,0	17,5	17,5	0,0	17,5	17,5	0,0	0,0	0,0	0,0
Leather, footwear, etc.	270	12,7	8,8	13,0	12,8	11,9	0,0	25,0	25,0	0,1	-1,3	0,5
Non-electrical machinery	915	4,6	5,8	0,0	3,8	0,0	0,0	16,0	16,0	0,8	-0,9	0,2
Electrical machinery	480	7,6	6,5	8,0	7,6	10,4	0,0	25,0	25,0	0,1	-1,5	0,3
Transport equipment	286	6,4	5,5	5,0	6,0	7,4	0,0	20,0	20,0	0,6	-0,8	0,3
Manufactures, n.e.s.	621	6,4	6,5	5,0	5,8	7,4	0,0	25,0	25,0	0,7	-0,3	0,3

Definition of groups used from World Tariff Profiles 2018

2. Descriptive analytics

Box-plot of MFN tariff rates



2. Descriptive analytics

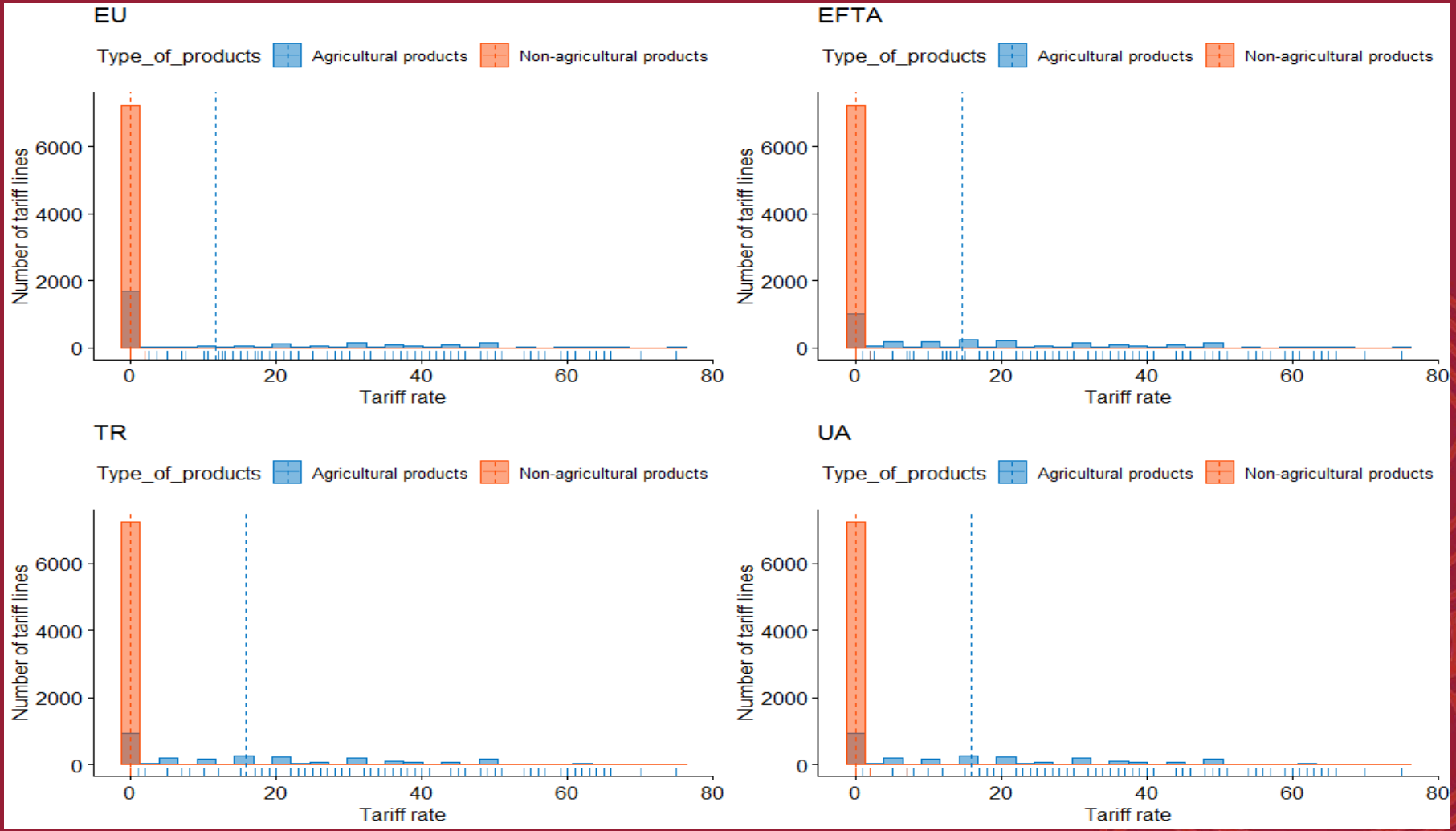
Simple averages preferential tariff rates: EU-3,1%,EFTA -3,9%,Turkey 4,2% and Ukraine 4,2%

Descriptive statistics of preferential tariff rates

	EU		EFTA		TR		UA	
Description	Agricultural products	Non-agricultural products	Agricultural products	Non-agricultural products	Agricultural products	Non-agricultural products	Agricultural products	Non-agricultural products
Simple average	11,6	0	14,7	0	15,9	0	15,9	0
Standard deviation	18,1	0	17,2	0	17,5	0	17,5	0
Median	0	0	10	0	10	0	10	0
Trimmed mean	8,1	0	11,9	0	13,3	0	13,3	0
Median absolute deviation	0	0	14,8	0	14,8	0	14,8	0
Min	0	0	0	0	0	0	0	0
Max	75	0	75	0	75	0	75	0
Range	75	0	75	0	75	0	75	0
Skew	1,3	0	1,1	0	1	0	1	0
Kurtosis	0,4	0	0,2	0	0	0	-0,1	0
Standard error	0,4	0	0,3	0	0,3	0	0,3	0
Var	327,5	0	296,6	0	304,6	0	304,6	0

2. Descriptive analytics

Histogram of preferential tariff rates by HS division



2. Descriptive analytics

Descriptive statistics of preferential tariffs with EU

Description	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Animal products	359	6,76	13,12	0	3,77	0	0	66	66	1,94	3,14	0,69
Dairy products	153	16,6	17,21	17,5	14,84	25,95	0	66	66	0,71	-0,21	1,39
Fruit, vegetables, plants	570	16,35	18,21	10	14,02	14,83	0	65	65	0,7	-0,62	0,76
Coffee, tea	47	4,95	6,73	0	4,29	0	0	18	18	0,64	-1,5	0,98
Cereals & preparations	237	13,08	17,42	0	10,51	0	0	75	75	1,16	0,73	1,13
Oilseeds, fats & oils	175	1,33	6,56	0	0	0	0	40	40	5,13	25,21	0,5
Sugars and confectionery	44	6,99	8,9	0	5,76	0	0	30	30	1	-0,12	1,34
Beverages and tobacco	356	32,3	22,7	44	32,85	16,31	0	70	70	-0,48	-1,44	1,2
Cotton	6	0	0	0	0	0	0	0	0	0	0	0
Other agricultural products	261	0,52	4,81	0	0	0	0	45	45	9,1	81,25	0,3
Fish and fish products	528	1,2	5,22	0	0	0	0	43	43	5,81	39,02	0,23
Minerals & metals	1465	0	0	0	0	0	0	0	0	0	0	0
Petroleum	49	0	0	0	0	0	0	0	0	0	0	0
Chemicals	1380	0	0	0	0	0	0	0	0	0	0	0
Wood, paper, etc.	473	0	0	0	0	0	0	0	0	0	0	0
Textiles	838	0	0	0	0	0	0	0	0	0	0	0
Clothing	341	0	0	0	0	0	0	0	0	0	0	0
Leather, footwear, etc.	270	0	0	0	0	0	0	0	0	0	0	0
Non-electrical machinery	915	0	0	0	0	0	0	0	0	0	0	0
Electrical machinery	480	0	0	0	0	0	0	0	0	0	0	0
Transport equipment	286	0	0	0	0	0	0	0	0	0	0	0
Manufactures, n.e.s.	621	0	0	0	0	0	0	0	0	0	0	0

2. Descriptive analytics

Descriptive statistics of preferential tariffs with EFTA

Description	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Animal products	359	15.68	11	15	14.82	7.41	0	66	66	1.1	2.37	0.58
Dairy products	153	21.16	13.91	20	19.94	17.79	0	66	66	0.96	0.93	1.12
Fruit, vegetables, plants	570	21.17	15.36	20	20.05	14.83	0	65	65	0.58	-0.01	0.64
Coffee, tea	47	9.1	8.3	12	8.45	8.9	0	30	30	0.36	-0.56	1.21
Cereals & preparations	237	15.93	15.73	10	14.05	14.83	0	75	75	1.21	1.39	1.02
Oilseeds, fats & oils	175	2.43	7	0	0.67	0	0	40	40	3.98	16.53	0.53
Sugars and confectionery	44	7.99	8.35	5	6.99	7.41	0	30	30	0.99	0.05	1.26
Beverages and tobacco	356	33.43	21.92	44	34.21	18.53	0	70	70	-0.53	-1.31	1.16
Cotton	6	0	0	0	0	0	0	0	0	0	0	0
Other agricultural products	261	2.56	6.74	0	0.82	0	0	45	45	4.01	18.83	0.42
Fish and fish products	528	0.06	0.92	0	0	0	0	15	15	16.11	258.01	0.04
Minerals & metals	1465	0	0	0	0	0	0	0	0	0	0	0
Petroleum	49	0	0	0	0	0	0	0	0	0	0	0
Chemicals	1380	0	0	0	0	0	0	0	0	0	0	0
Wood, paper, etc.	473	0	0	0	0	0	0	0	0	0	0	0
Textiles	838	0	0	0	0	0	0	0	0	0	0	0
Clothing	341	0	0	0	0	0	0	0	0	0	0	0
Leather, footwear, etc.	270	0	0	0	0	0	0	0	0	0	0	0
Non-electrical machinery	915	0	0	0	0	0	0	0	0	0	0	0
Electrical machinery	480	0	0	0	0	0	0	0	0	0	0	0
Transport equipment	286	0	0	0	0	0	0	0	0	0	0	0
Manufactures, n.e.s.	621	0	0	0	0	0	0	0	0	0	0	0

2. Descriptive analytics

Descriptive statistics of preferential tariffs with TR

Description	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Animal products	359	15.69	10.99	15	14.83	7.41	0	66	66	1.11	2.39	0.58
Dairy products	153	21.99	13.57	20	20.98	17.79	0	66	66	0.89	1.06	1.1
Fruit, vegetables, plants	570	21.77	15.18	20	20.79	14.83	0	65	65	0.56	-0.01	0.64
Coffee, tea	47	18.55	9.63	20	19	7.41	0	36	36	-0.39	-0.63	1.41
Cereals & preparations	237	19.51	16.33	10	18.12	14.83	0	75	75	0.76	0.27	1.06
Oilseeds, fats & oils	175	2.43	7	0	0.67	0	0	40	40	3.98	16.53	0.53
Sugars and confectionery	44	10.82	10.4	7.5	9.89	11.12	0	30	30	0.4	-1.39	1.57
Beverages and tobacco	356	34	21.97	45	34.92	14.83	0	70	70	-0.59	-1.27	1.16
Cotton	6	0	0	0	0	0	0	0	0	0	0	0
Other agricultural products	261	2.55	6.74	0	0.81	0	0	45	45	4.01	18.82	0.42
Fish and fish products	528	1.85	7.63	0	0	0	0	62	62	5.61	36.75	0.33
Minerals & metals	1465	0	0	0	0	0	0	0	0	0	0	0
Petroleum	49	0	0	0	0	0	0	0	0	0	0	0
Chemicals	1380	0	0	0	0	0	0	0	0	0	0	0
Wood, paper, etc.	473	0	0	0	0	0	0	0	0	0	0	0
Textiles	838	0	0	0	0	0	0	0	0	0	0	0
Clothing	341	0	0	0	0	0	0	0	0	0	0	0
Leather, footwear, etc.	270	0	0	0	0	0	0	0	0	0	0	0
Non-electrical machinery	915	0	0	0	0	0	0	0	0	0	0	0
Electrical machinery	480	0	0	0	0	0	0	0	0	0	0	0
Transport equipment	286	0	0	0	0	0	0	0	0	0	0	0
Manufactures, n.e.s.	621	0	0	0	0	0	0	0	0	0	0	0

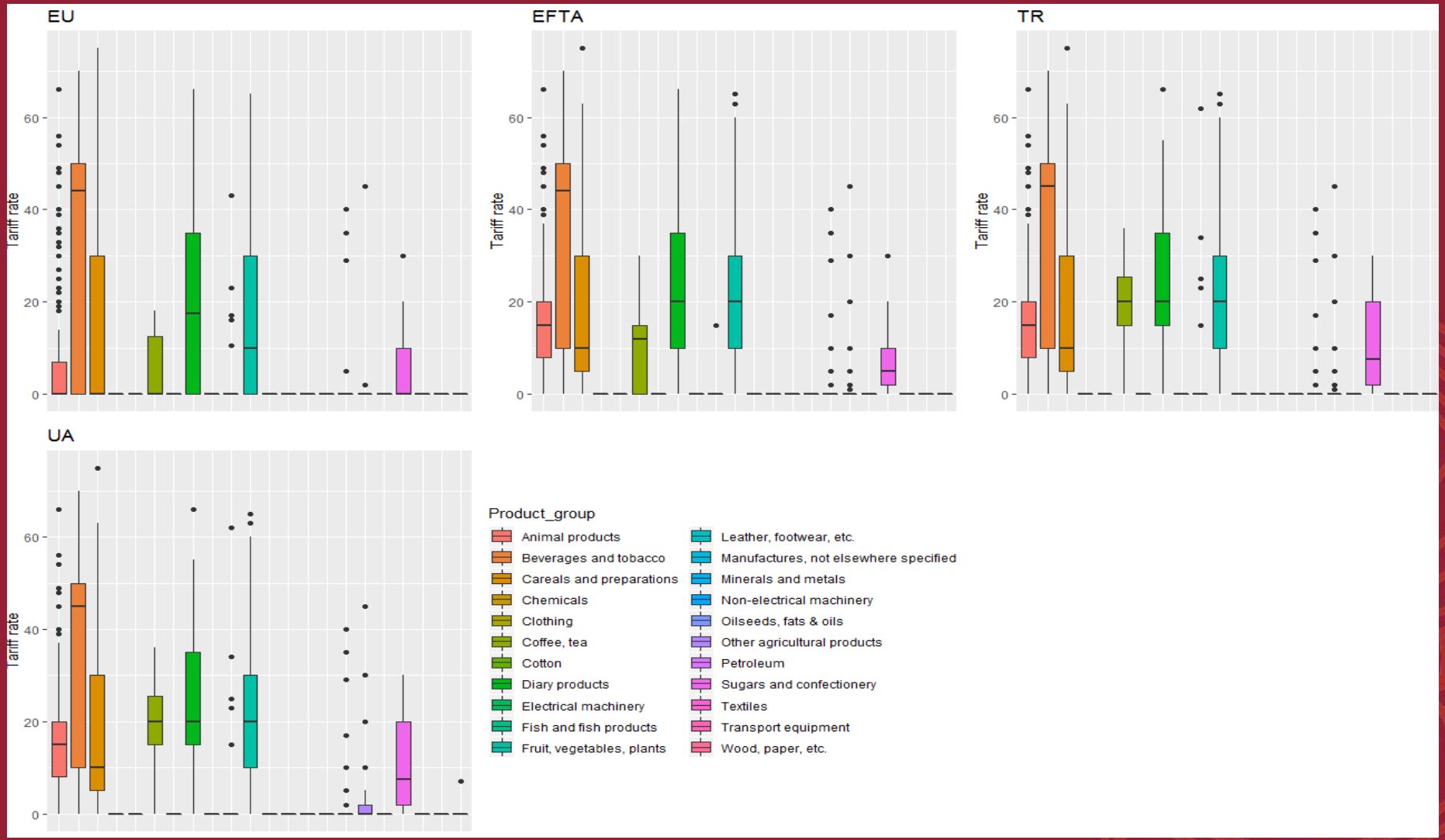
2. Descriptive analytics

Descriptive statistics of preferential tariffs with UA

Description	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Animal products	359	15.69	10.99	15	14.83	7.41	0	66	66	1.11	2.39	0.58
Dairy products	153	21.99	13.57	20	20.98	17.79	0	66	66	0.89	1.06	1.1
Fruit, vegetables, plants	570	21.77	15.18	20	20.79	14.83	0	65	65	0.56	-0.01	0.64
Coffee, tea	47	18.55	9.63	20	19	7.41	0	36	36	-0.39	-0.63	1.41
Cereals & preparations	237	19.51	16.33	10	18.12	14.83	0	75	75	0.76	0.27	1.06
Oilseeds, fats & oils	175	2.43	7	0	0.67	0	0	40	40	3.98	16.53	0.53
Sugars and confectionery	44	10.82	10.4	7.5	9.89	11.12	0	30	30	0.4	-1.39	1.57
Beverages and tobacco	356	34	21.97	45	34.92	14.83	0	70	70	-0.59	-1.27	1.16
Cotton	6	0	0	0	0	0	0	0	0	0	0	0
Other agricultural products	261	2.59	6.73	0	0.86	0	0	45	45	4.01	18.85	0.42
Fish and fish products	528	1.85	7.63	0	0	0	0	62	62	5.61	36.75	0.33
Minerals & metals	1465	0	0	0	0	0	0	0	0	0	0	0
Petroleum	49	0	0	0	0	0	0	0	0	0	0	0
Chemicals	1380	0	0	0	0	0	0	0	0	0	0	0
Wood, paper, etc.	473	0.01	0.32	0	0	0	0	7	7	21.61	466.01	0.01
Textiles	838	0	0	0	0	0	0	0	0	0	0	0
Clothing	341	0	0	0	0	0	0	0	0	0	0	0
Leather, footwear, etc.	270	0	0	0	0	0	0	0	0	0	0	0
Non-electrical machinery	915	0	0	0	0	0	0	0	0	0	0	0
Electrical machinery	480	0	0	0	0	0	0	0	0	0	0	0
Transport equipment	286	0	0	0	0	0	0	0	0	0	0	0
Manufactures, n.e.s.	621	0	0	0	0	0	0	0	0	0	0	0

2. Descriptive analytics

Box-plots of preferential tariff rates

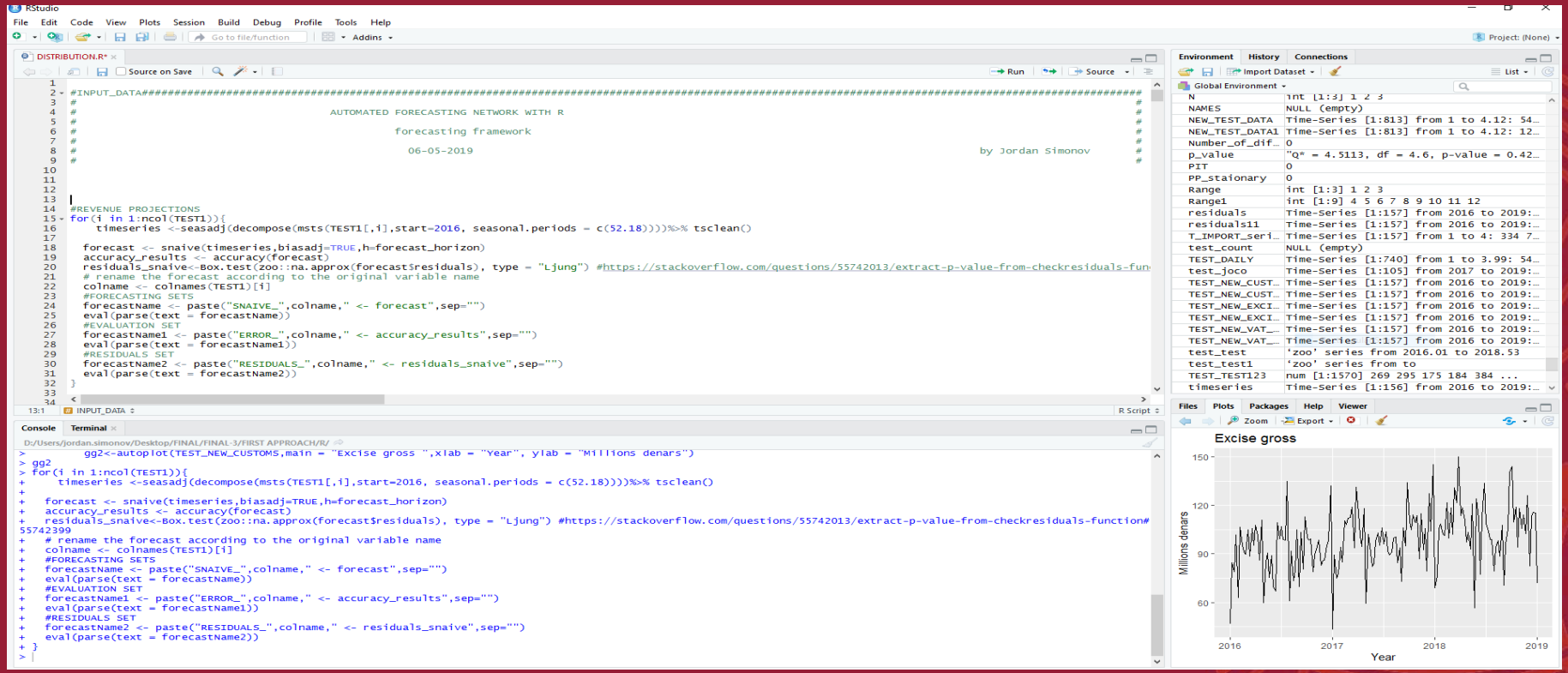


3. Predictive analytics

Automated forecasting with data-driven methods :

- Seasonal Naïves;
- Exponential smoothing and
- ARIMA.

Automated Forecasting Network with *R* and *R Studio*



3. Predictive analytics

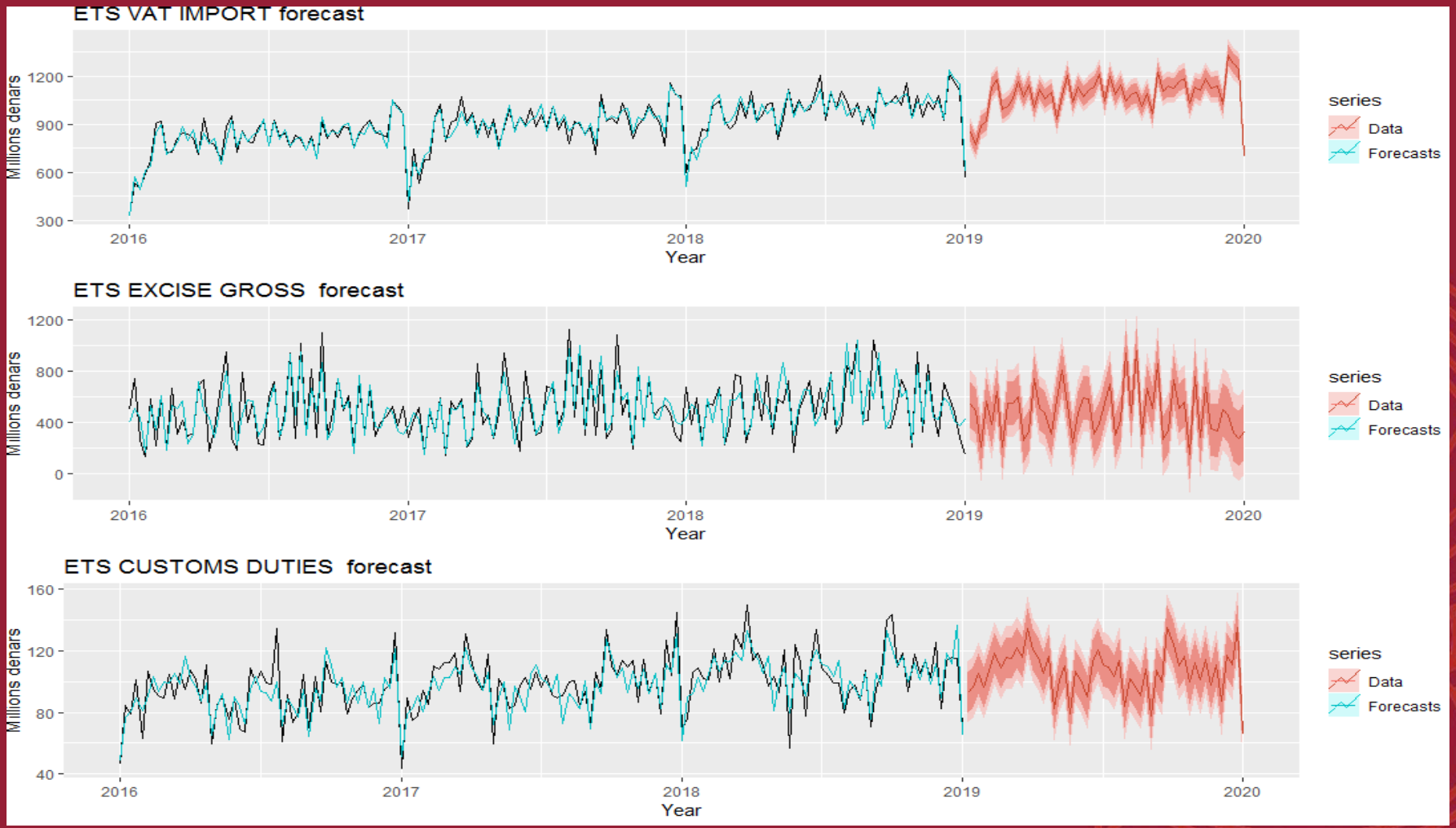
Weekly revenue collection (VAT import, Gross Excise and Customs duties)

3. Predictive analytics

Snaive forecast

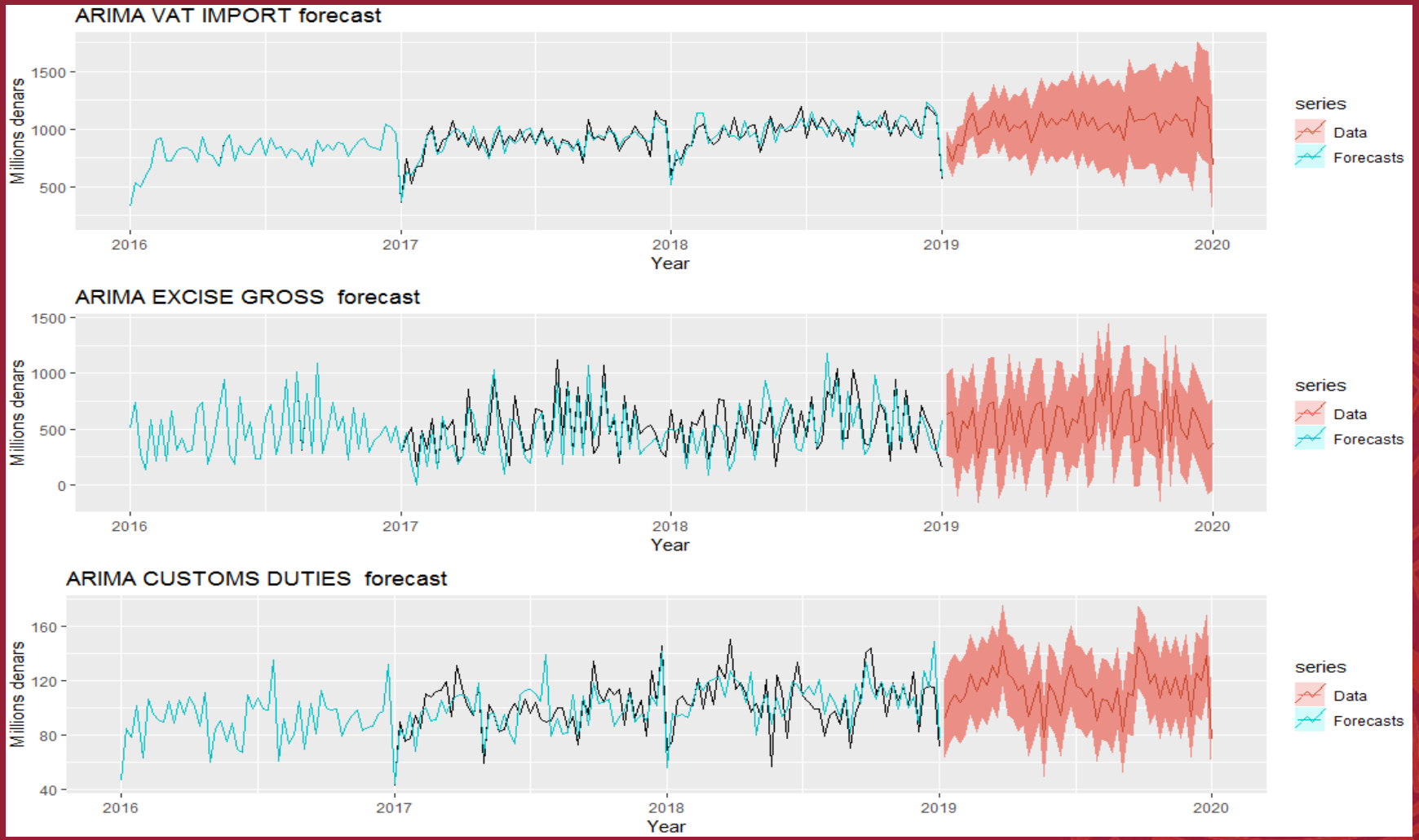
3. Predictive analytics

Exponential smoothing forecast



3. Predictive analytics

ARIMA forecast



3. Predictive analytics

Residuals diagnostics and Evaluations forecast accuracy

Model	Type of time series	Q	df	p.value	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Snaive	VAT_IMPORT	21.302	31	0.9037	90,87	115,25	94,97	9,6	10,11	1	-0,01
Snaive	EXCISE_GROSS	23.53	31	0.8292	25,71	213,25	160,27	-5,06	36,69	1	-0,33
Snaive	CUS	30.662	31	0.4833	7,37	17,97	14,55	5,86	14,17	1	-0,23
ETS (M,A,N)	VAT_IMPORT	38.559	27	0.06945	1,06	39,49	31,86	-0,06	3,67	0,34	0
ETS (M,A,N)	EXCISE_GROSS	50.514	27	0.003968	-1,64	129,73	98,3	-8,16	23,01	0,61	-0,34
ETS (A,N,N)	CUS	62.875	29	0.000267	1,37	10,06	8,29	0,34	8,63	0,57	-0,27
ARIMA(1,1,2)(1,1,1)$_{[52]}$	VAT_IMPORT	54.199	26	1.835e-09	5,54	58,00	35,73	0,33	3,84	0,33	-0,03
ARIMA(5,1,0)(0,1,0)$_{[52]}$	EXCISE_GROSS	30.172	26	0.8704	3,00	159,24	98,50	-6,22	23,72	0,68	-0,02
ARIMA(1,1,1)(0,1,1)$_{[52]}$	CUS	27.567	28	0.4864	0,24	11,79	7,62	-0,66	7,7	0,52	0,01

4. Conclusions

- By using the descriptive analysis, we completely determined the nominal customs duties in the Republic of North Macedonia, according to both MFN-tariff rates and the preferential tariff rates. At the same time, we hereby concluded that the average tariff rate calculated as simple average is not the best indicator for the level of tariff rates protection, and therefore the measurement of the level of nominal customs duties should not be based solely on this measure, but for the purpose of its correct determination, it is rather necessary to take into account the other measures of the descriptive statistics as well (measures of the central tendency, measures of dispersion, asymmetry and roundness) that should be complementarily used with graphical visualization (Box-plot, Histogram etc.).
- By using the predictive analytics, we created an automated forecasting network in which we made more projections of univariate times series (VAT, Excise and Customs duties),whereby we selected the models such as: ETS (M, A, N), ARIMA (5,1,0) (0,1,0) [52] and ARIMA (1,1,1) (0,1,1) [52] respectively, based on the slightest error (RMSE), as well as the residual tests. We concluded that this approach provided us with fast and flexible manner of projecting the revenues, which is especially useful when simultaneously forecasting multiple time series, whereby the main focus is placed on evaluating the models and their errors, rather than the modeling itself.
- Finally, we can conclude that from the aspect of Big data analytics, the Ministry of Finance has already adapted its working methods and practices, so it is quite ready to welcome the new digital era that has just begun with CDEPS. New completely paperless work environment will provide much more data in a digital form and therefore a much greater opportunity to perform analyzes for the purpose of conducting sound fiscal policy.

Thank you for attention !

E-mail: jordan.simonov@finance.gov.mk
zoran.gligorov@finance.gov.mk