

Econometric Modelling of Solid Waste Generation in Jordan

Veronika Alhanaqtah

College of Business, Tafila Technical University, PO box 66110, Tafila, Jordan

* E-mail of the corresponding author: ver-grabko@yandex.ru

Abstract

The objective of this study is to develop an econometric model that might help decision-makers to imply better policies based on the information about key factors influencing the amount of waste generation in Jordan. We found out that the most informative factor is the level of consumption which is correlated with the number of population and gross national income. Consumption culture itself rather than the population growth stimulates waste production. In order to reduce the level of waste generation, it is necessary to move from an unreasonable culture of unrestrained consumption to a reasonable and responsible attitude to the resources of the country and the environment as a whole. It is also necessary to introduce modern waste management methods, develop technologies for waste recycling and reuse.

Keywords: waste, consumption, population, GNI, employment, econometric tests, regression analysis, time series

DOI: 10.7176/JESD/11-16-22

Publication date: August 31st 2020

1. Introduction

To plan the proper solid waste management (SWM) system, it is essential to know the amount of waste produced in the country. Knowing the amount of waste allows authorities to understand the scale of the problem and put forward appropriate plans. It also provides SWM organizations with the opportunity to optimize their investments to buy new machinery for recycling facilities, equip transition stations, evaluate disposal capacities of landfills and plan other sanitarian measures for the future. The objective of this study is to develop an econometric model that might help decision-makers to imply better policies based on the information about key factors influencing the amount of waste generation in the country. We try to find out the most significant variables explaining the current situation in Jordan and to develop an econometric tool for prediction the amount of waste produced.

1.1 Literature review

Increasing income and economic growth influence the rate of solid waste generation. Hockett *et al.* (1995) discovered that income is an important factor in solid waste generation. Chung (2010a) also found out that the level of income of individual waste generators is a variable that usually has no reliable source of data. To obtain the permission to collect these data is generally complicated. Thus, other approaches should be utilized instead. Joutz (1996) used the ratio of two economic indices computed at the national level so as to approximate the real incomes. Bach *et al.* (2004) and Gay *et al.* (1993) found that the level of employment can be a good factor to predict commercial waste. Samson *et al.* (2017) suggest that GDP and population are the two compelling factors affecting municipal solid waste generation. They conclude that without an in-depth understanding of the variables affection waste generation waste reduction policies may be ineffective and unsuccessful. Afroz *et al.* (2010) found out that the waste generation of the households is significantly affected by the concern about the environment, income, household size and willingness to separate the waste.

Alongside with economic factors, changes in social variables may influence waste generation levels. For example, change in population numbers and household size. In the article of Abdoli & Falahnehzad (2011), similar to the study of Dennison *et al.* (1996b) and Beigl *et al.* (2004, 2008), the average household size and the municipal solid waste have a reverse relationship. Population and household size are highly correlated. Thus, for preventing multicollinearity, Abdoli & Falahnehzad (2011) select population for the model estimation. McBean & Fortin (1993) has also considered population as one of the most important variables affecting total waste generation rate.

In the research of Liu & Wu (2010) twelve items were selected: urban population, the proportion of urban population, annual per capita disposable income of urban households, annual per capita consumption expenditure of urban households, total energy consumption, annual per capital consumption for households, GDP, per capita GDP, the area of urban gardens and green areas, the number of the large cities, the area of urban construction, the

area of paved roads. Energy consumption was another factor considered in their study. The results showed that urban development and the economy are important factors influencing municipal solid waste generation. Influencing waste factors in rural areas are considered in the study of Han *et al.* (2018).

Al-Rifai & Amoudi (2016) analyzed key factors of construction waste in Jordan. They found out that the most significant factors can be categorized into two groups: workforce-related (lack of skilled workers and subcontractors) and management-related (lack of quality management system).

Some authors analyzed the impact of consumption factors on the generation of municipal solid waste (Liu *et al.* (2019); Lu *et al.* (2017); Mahees *et al.* 2011; Hage *et al.* 2018). The results showed that developing regions generate more waste than developed regions. Thus, different regions have to control waste levels by optimizing consumption pattern, as well as to implement best experience of developed regions. Many scholars also use socio-economic and demographic data to model solid waste generation (Trang *et al.* 2017; Bosire *et al.* (2019); Prades *et al.* 2015; Chhay *et al.* 2018; Khan *et al.* 2016).

1.2 Variables

The data set for this study is an annual time series for the period 1991–2018. It consists of the following variables (World Bank data):

- *popurb* – urban population (Note 1);
- *poprur* – rural population (Note 2);
- *pop* – total population computed as aggregation of urban and rural population;
- *gni* – gross national income (current US \$) (Note 3);
- *gnicap* – gross national income per capita (current US \$) (GNI per capita, Atlas method, current US\$);
- *cons* – final consumption expenditure (current US \$) (Note 4);
- *conscap* – household consumption expenditure per capita (constant 2010 US\$);
- *empagr* – employment in agriculture (% of total employment);
- *empind* – employment in industry (% of total employment);
- *empserv* – employment in services (% of total employment);
- *wasteurb* – waste produced in urban areas (tone/year) computed as follows (Note 5):
$$wasteurb = popurb \cdot 0.99 \cdot 365/1000$$
- *wasterur* – waste produced in rural areas (tone/year) computed as follows:
$$wasterur = poprur \cdot 0.87 \cdot 365/1000$$
- *waste* – total waste computed as aggregation of waste in urban and waste in rural areas (tone/year).

1.3 Research plan

First, in this study a panel data set with selected socio-economic indicators for Jordan is created. The importance of the input variable depends on the linear correlation coefficient (r). Thus, second, we compute linear correlation coefficients and create visualizations of the relationship between selected economic variables: scatter plots and bubble charts. Third, regression analysis is conducted. Several econometric models for solid waste generation are estimated with the help of the ordinary least-squares (OLS) method. Then, based on statistical criteria, the best fitted model is chosen, and every input variable is given an interpretation. Tests for multi-collinearity, heteroscedasticity and autocorrelation are fulfilled. Forth, time series analysis is conducted. We detect whether our time series is a stationary process, then conduct auto simulation in R/R-Studio. Projections for waste generation until 2043 are made. Finally, conclusions are made.

2. Descriptive statistics

There are numerical characteristics of selected variables represented in Table 1.

Table 1. Numerical characteristics of variables

variables	mean	std.dev.	min	max	skew	kurtosis
waste	2247.4	694	1316.8	3558.3	0.57	-1.1
wasteurb	1920.1	711.9	1011.7	3273.1	0.62	-1.1
wasterur	327.2	30.7	283.3	3760.5	0.19	-1.26
pop	6344.2	1914.1	3760.5	9956	0.56	-1.1
popurb	5313.8	1970.2	2799.9	9057.9	0.62	-1.1
poprur	1030.4	96.8	892	1209.7	0.19	-1.26
cons	18514795.6	13290849	4236304.9	42333254.4	0.59	-1.29
cons(cap)	2.6	0.4	1.9	3.3	-0.11	-1.01
gni	18287046.8	12895609.4	3981780.3	42091114.2	0.56	-1.32
gni(cap)	2.5	1.1	1.1	4.2	0.24	-1.66

Source: Author's computation in R-Studio (R-script is in Appendix).

The Table 1 shows that minimum waste value corresponds with minimum values of population, GNI and consumption, as well as maximum value of waste corresponds with maximum values of population, GNI and consumption. This is also approved by high values of linear correlation coefficients between waste and population (0.99) (Figure 1), waste and GNI (0.99) (Figure 2), waste and consumption (0.99) (Figure 3). Additionally, there are high linear correlations coefficients between GNI and consumption (0.99), GNI and population (0.99) and consumption and population (0.99).

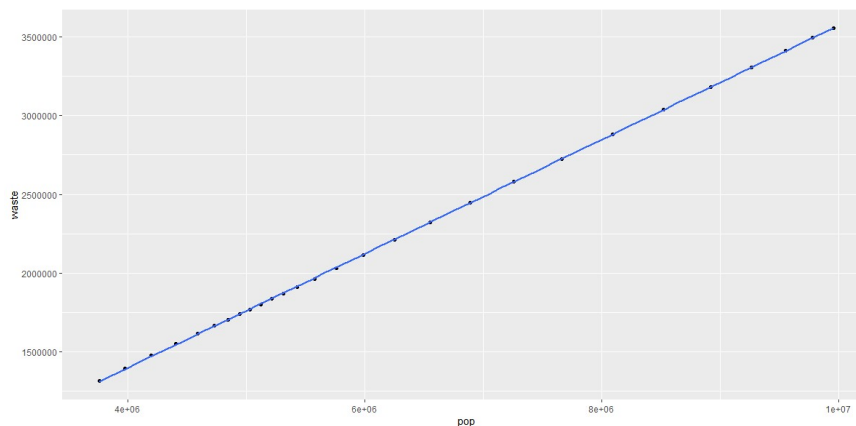


Figure 1. Relationship between waste and population

Source: Author's computation in R-Studio (R-script is in Appendix).

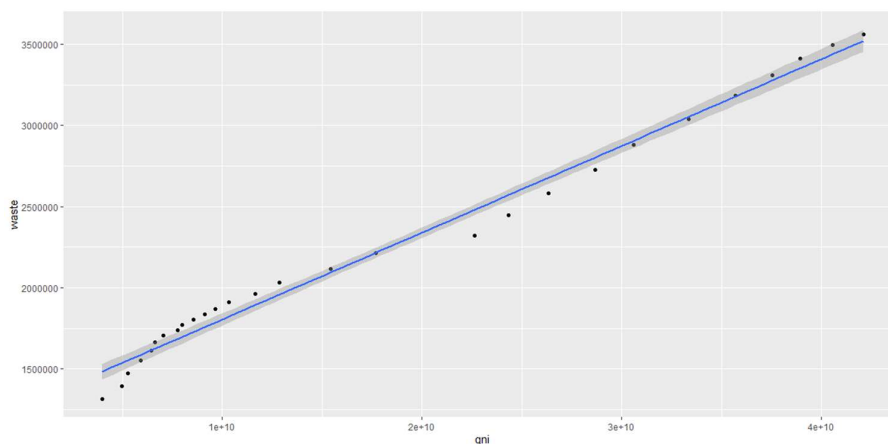


Figure 2. Relationship between waste and GNI

Source: Author's computation in R-Studio (R-script is in Appendix).

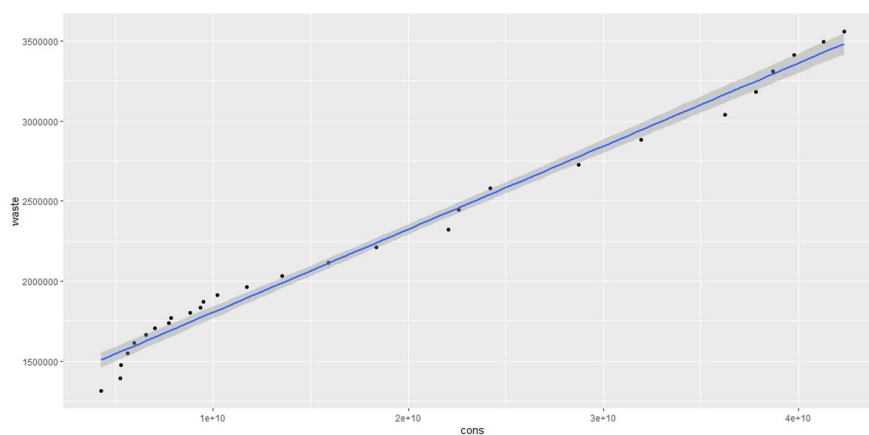


Figure 3. Relationship between waste and consumption

Source: Author's computation in R-Studio (R-script is in Appendix).

The visualizations show that relationships between mentioned variables are strictly deterministic.

There are bubble charts for two numeric variables evaluated by the third numeric variable representing the size of the bubble. Figure 4 shows the relationship between waste in urban and rural areas evaluated by population.

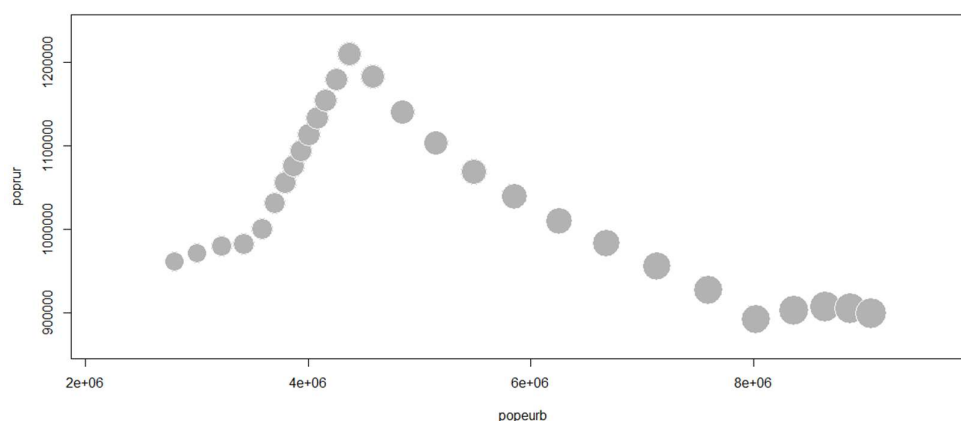


Figure 4. Waste in urban and rural areas evaluated by population (bubble size)

Source: Author's computation in R-Studio (R-script is in Appendix).

The visualization shows that for 14 years, since 1991 to 2004, the relationship between volumes of waste in urban and rural areas was positive: both variables rose. For the next 11 years, since 2004 to 2014, the relationship was negative showing that the amounts of waste produced shifted to urban areas alongside with the population growth. Since 2014 we observe slow increase in volumes of waste both in urban and rural areas.

Figure 5 shows the relationship between waste in urban and rural areas evaluated by consumption.

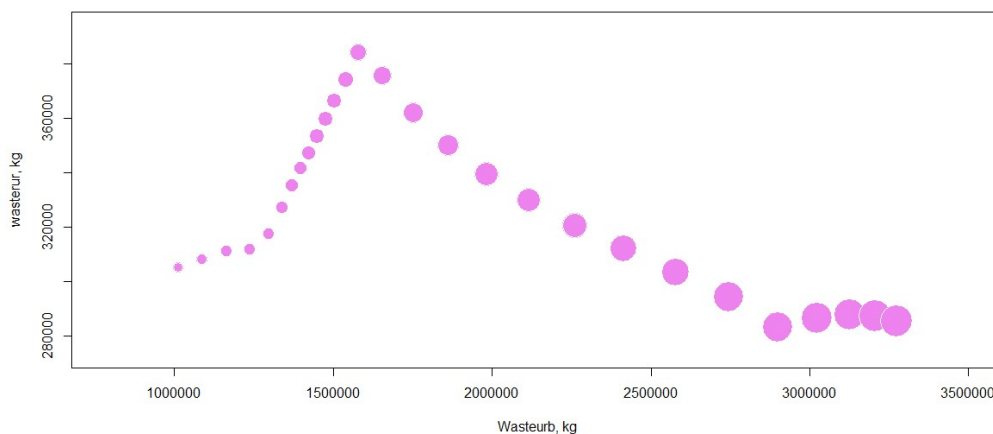


Figure 5. Waste in urban and rural areas evaluated by consumption (bubble size)

Source: Author's computation in R-Studio (R-script is in Appendix).

Visualization in Figure 5 matches visualization in Figure 4 which is explained by high value of a linear correlation coefficient between population and consumption. The visualization shows that for 14 years the relationship between volumes of waste in urban and rural areas was positive: both variables rose. For the next 11 years the relationship was negative showing that the amounts of waste produced shifted to urban areas alongside with the high consumption growth. Since 2014 we observe slow increase in volumes of waste both in urban and rural areas. The further in time the higher values of consumption which we consider the cause of high volumes of waste generation.

3. Model selection

While selecting a better model we base upon the following features of a “good” model: out of two models, reflecting the reality approximately similarly, a simpler model, with a fewer number of variables, is chosen; beta-

estimators must be unique, i.e. have a single meaning; model is admitted to be better if it may explain more variance of a dependent variable (in comparison with other models), and the model with a higher value of R-squared-adjusted is selected; model is qualitative if it is acknowledged by real data, i.e. it has good prognostic qualities.

At this stage we opt for the most qualitative model in order to explain the change in the volumes of waste. Since we have found out that relationships between variables are linear, we will estimate a regression model with the help of a linear function using OLS method. The alternative models are:

Model 1: $waste = cons$

Model 2: $waste = pop + gni + cons$

Model 3: $waste = cons + empagr + empind + empsevr$

Model 4: $waste = cons + empagr + empind + empsevr + poprur + popurb$

Summary statistics for models is shown in the Table 2 (Note 6).

Table 2. Summary statistics for four alternative models

Model 1				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.287e+06	2.654e+04	48.49	<2e-16 ***
cons	5.188e-05	1.172e-06	44.28	<2e-16 ***
Model 2				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.691e+04	9.564e+03	-1.768	0.0897 .
pop	3.527e-01	2.592e-03	136.048	<2e-16 ***
gni	2.764e-07	6.644e-07	0.416	0.6811
cons	1.167e-06	6.464e-07	1.806	0.0835 .
Model 3				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-8.547e+08	1.591e+09	-0.537	0.596
cons	4.187e-05	2.326e-06	17.997	4.76e-15 ***
empagr	8.080e+06	1.589e+07	0.509	0.616
empind	8.575e+06	1.591e+07	0.539	0.595
empsevr	8.584e+06	1.591e+07	0.540	0.595
Model 4				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.338e+01	9.105e+01	-2.570e-01	0.800
cons	-1.109e-12	7.437e-13	-1.492e+00	0.151
empagr	2.140e-01	9.111e-01	2.350e-01	0.817
empind	2.354e-01	9.103e-01	2.590e-01	0.799
empsevr	2.346e-01	9.105e-01	2.580e-01	0.799
poprur	3.175e-01	2.328e-08	1.364e+07	<2e-16 ***
popurb	3.614e-01	4.289e-09	8.425e+07	<2e-16 ***

Source: Author's computation in R-Studio (R-script is in Appendix).

We may compare alternative models and choose the best one based on the following criterions: coefficient of determination (R^2), adjusted coefficient determination (R^2_{adj}), informational criteria AIC (AIC) and Schwartz (BIC), and residuals squared sum (RSS). Summary statistics for four alternative models is represented in a Table 3.

Table 3. Quality criterions for four alternative models

Criterion	Model 1	Model 2	Model 3	Model 4
p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
R-squared	0.9869	1	0.9949	1
R-squared adjusted	0.9864	1	0.994	1
Deviance (RSS)	1.7	202436384	66758349055	0.0001861242
AIC	716.2495	531.685	696.0405	-238.3359
BIC	720.2461	538.3461	704.0337	-227.6782

Source: Author's computation in R-Studio (R-script is in Appendix).

The Model 1 has the smallest value of Residuals Sum of Squares so it is better by the criterion of deviance. Other models have very high values of RSS which is a bad indicator of the quality of a model. All four models have high values of coefficient of determination which is an indication of a good model. "Penalty" criteria AIC and BIC are smallest for the Model 4 and highest for the Model 1. The results of comparative analysis show that at this stage we can't identify the best model yet. To opt for the best model additional investigations are required. At the next stage we conduct econometric tests

4. Econometric tests

First, we test a hypothesis for some *linear restrictions simultaneously*. The null hypothesis (H_0) implies that the whole regression is insignificant. The alternative hypothesis (H_a) implies that at least one of β -coefficients is significant (non-zero) so the whole regression is considered to be significant. The null hypothesis is verified with the help of the Wald test based on the algorithm of the F-test (Table 4).

Table 4. Econometric tests

Econometric test	Model 1	Model 2	Model 3	Model 4
Wald test				
test-statistic	1960.5	513973	NA	NA
p-value	< 2.2e-16 ***	< 2.2e-16 ***	NA	NA
RESET test				
test-statistic	14.631	28.195	17.674	4.6063
p-value	7.006e-05	8.507e-07	3.155e-05	0.02339
Breusch-Pagan test				
test-statistic	1.0806	0.89795	11.529	16.147
p-value	0.2986	0.8259	0.02122	0.01299
Durbin-Watson test				
test-statistics	0.3612651	0.2054333	0.7087056	1.604281
p-value	0	0	0	0.026

Source: Author's computation in R-Studio (R-script is in Appendix).

The Wald test shows that even though some coefficients significant and some are not (Table 2), regression models 3 and 4 are significant. At 5 % level of significance null hypothesis is rejected (p-values are close to zero for all regressions). Thus, at least one of the β -coefficients is not equal to zero, and omission of coefficients is not reasonable (Note 7).

If we have observations, then, in order to verify whether significant variables were omitted or not, we use F-test

(Wald test). However, it can be a situation when we have to include variables into the model, for which we do not have observations at all.

Thus, second, we verify whether we have *omitted regressors* in the model, for which we don't have observations. The null hypothesis (H_0) implies that the model is correct: we included exactly those variables that must have been included. There are not any omitted regressors. The alternative hypothesis (H_a) implies that there are unknown omitted regressors. The null hypothesis is verified with the help of Ramsey (RESET) test (Table 4). The results of computations show that for four alternative models p-values are much higher than the 5 % level of significance. Thus, null hypothesis is not rejected. Regression models don't contain omitted informative variables.

Third, also a possible problem of multi-collinearity (linear relationship between explanatory variables) has to be taken into account. Even though it does not affect the standard set of OLS-assumptions, in particular the Gauss-Markov theorem (beta-coefficients are still unbiased and asymptotically normal, we can test hypothesis and construct confidence intervals), standard errors of beta-estimators become very large. Thus, we may get an over fitted model which is not statistically robust, i.e, it does not predict reliably across numerous samples. In order to check the presence of multi-collinearity variance inflation factors (VIF) are computed for every estimated model (Table 5). When $VIF_j > 10$ there is potentially an indication of multi-collinearity in the model.

Table 5. Detection multi-collinearity

Model 1					
model contains fewer than 2 terms					
Model 2					
pop	gni	cons			
78.8192	234.9662	236.2629			
Model 3					
cons	empagr	empind	empserv		
8.892541e+00	2.258041e+05	8.483443e+06	6.986916e+06		
Model 4					
cons	empagr	empind	empserv	poprur	popurb
297.6425	243233.9370	9094466.3923	7493301.4846	15.4521	217.5441

Source: Author's computation in R-Studio (R-script is in Appendix).

We don't consider multi-collinearity in the Model 1 because it is a simple linear regression. The Model 2 shows high multi-collinearity: values of population, consumption and GNI are highly linearly correlated. The results of calculations for the Model 3, presented in exponential format, also indicate high multi-collinearity. The Model 4 also demonstrates high multi-collinearity. In such cases it is recommended to drop one of the correlated explanatory variables out of the model keeping in mind that we miss some information and beta-estimators for remained correlated variables become biased. In terms of multi-collinearity the Model 1 is the best. Additionally, it has high values of coefficient of determination (Table 3) and satisfactory results in Wald- and RESET tests. At this stage the Model 1 is considered to be the first-best.

Forth, the presence of *heteroscedasticity* (variance of residuals is not constant for every observation) in the model has been verified (Table 4). The null hypothesis (H_0) in the Breusch-Pagan test implies that there is no heteroscedasticity in a model. Since p-values for the Model 1 and the Model 2 are high (more than $\alpha = 10\%$), then null hypothesis is not rejected. These models do not have a problem of heteroscedasticity of residuals. Beta-coefficients are unbiased, consistent and efficient. The Model 3 and the Model 4 have a problem of heteroscedasticity (p-values are less than the 5 % level of significance, usual for economic studies). Beta-coefficients are still linear and unbiased but not efficient, even asymptotically. As a consequence, all statistical inferences based on t- and F-statistics above, as well as confidence intervals, are unreliable. Therefore, they are not applicable for construction of adequate models. At this stage we may exclude the Model 3 and the Model 4 from consideration.

Fifth, the presence of *autocorrelation* (whether residuals are correlated between observations) in the models has

been checked. For this purpose, Durbin-Watson test for the autocorrelation of the first order has been fulfilled (Table 4). The Durbin-Watson test showed that there is autocorrelation of the first order in every alternative model (p -values are less than the level of significance, so null hypothesis on absence of autocorrelation is rejected). This outcome is quite logical because we deal with time series.

All conducted econometric tests show that the regression Model 1 is the most adequate to the data. The volume of waste produced is the most explained by the level of consumption. Consumption culture itself rather than the population growth stimulates waste production. In order to reduce the level of waste generation, it is necessary to move from an unreasonable culture of unrestrained consumption to a reasonable and responsible attitude to the resources of the country and the environment as a whole (Alhanaqtah 2020a, 2020b). It is also necessary to introduce modern waste management methods, develop technologies for waste recycling and reuse.

5. Time series analysis

In this section we will try to predict waste generation for the next 25 years since 2018. First, we test a time series for the variable “waste” for stationarity. Visualization of a time series in the Figure 6.

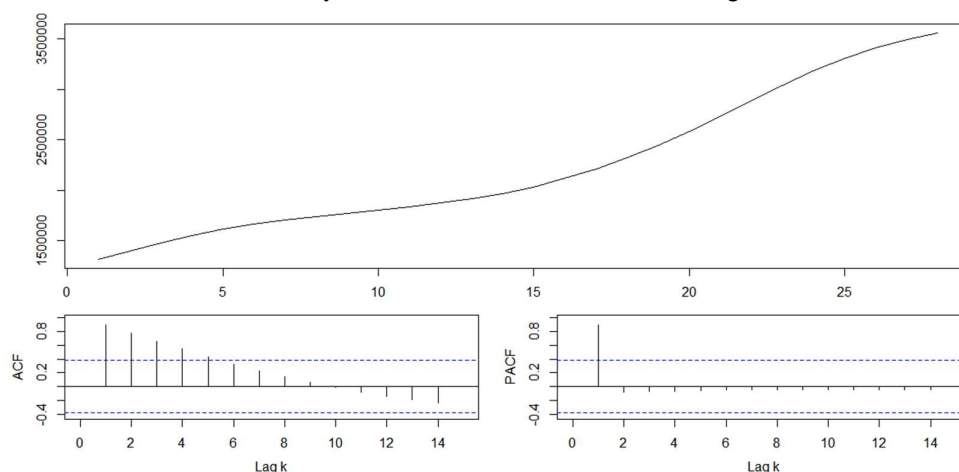


Figure 6. A time series for “waste”

Source: Author’s computation in R-Studio (R-script is in Appendix).

We see that our series takes a compromised position between stationary and non-stationary processes. Auto-covariance function (ACF) is decreasing but not that quick as it takes place in an ideal stationary processes. The first PACF is non-zero, other PACFs are equal to zero (Note 8). In other words, PACF plot has a significant spike only at lag 1, meaning that all the higher-order autocorrelations are effectively explained by the lag-1 autocorrelation. This is the graph of a typical random walk with drift. In order to be certain of that, let’s take differences, i.e. build the same parameters for Δy_t . The outcome of the command `ndiffs()` in R shows that 2 differences is required to make this process stationary.

In the Figure 7 we see that the autocorrelation falls to zero fairly quickly which is typical for a stationary time series.

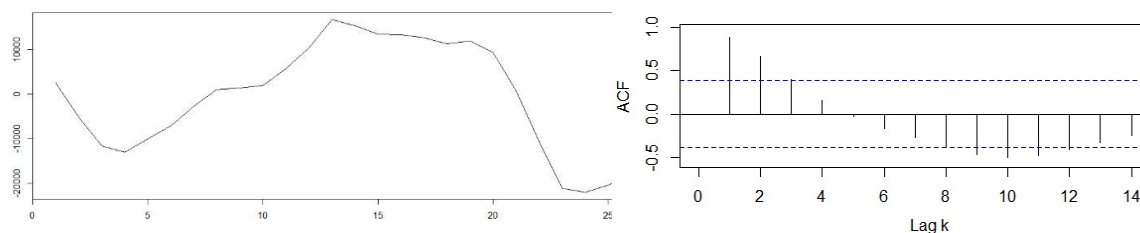


Figure 7. Differenced and stationary

Source: Author's computation in R-Studio (R-script is in Appendix).

It means that our initial series is a non-stationary process with drift.

In order not to simulate at random we may request R-Studio to choose the model with minimum AIC out of the models with a few lags. In other words, we can choose the better model on ACAICE criteria automatically via auto-simulation command `auto.arima()`. The outcome of the command shows that our process is ARIMA(2, 2, 0). The time series for "waste" is non-stationary.

Additionally, in accordance with the Model 1 we compute cross correlation between two time series: waste and consumption (Figure 8).

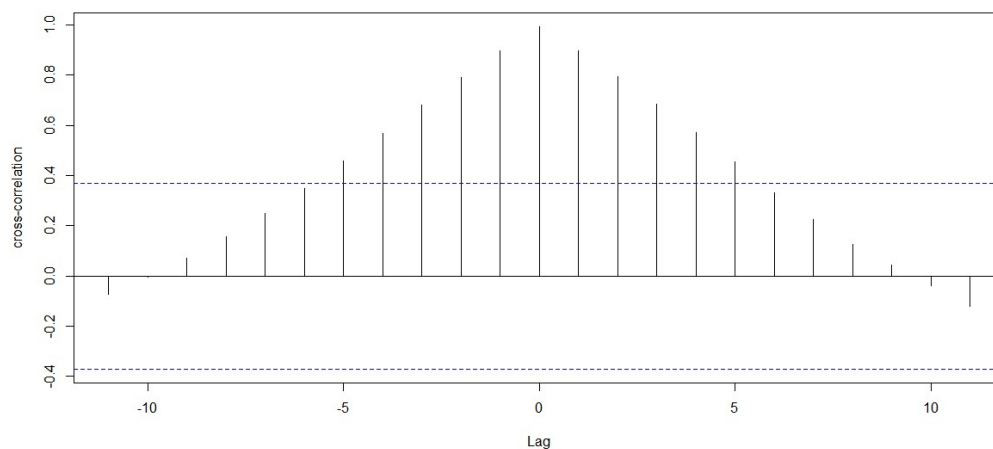


Figure 8. Cross correlation between waste and consumption

Source: Author's computation in R-Studio (R-script is in Appendix).

Usually, a correlation is significant when the absolute value is greater than $\frac{2}{n-|k|}$,

where n is the number of observations and k is the lag. This calculation is a rule of thumb procedure. In the Figure 8 the correlation at lag 0 is approximately 0.94.

Because $0.94 > \frac{2}{\sqrt{28-|0|}}$ the correlation is significant.

We may conclude that levels of waste and consumption change together in one direction with no shift in time.

The forecast by the auto-simulation is in the Figure 9, where grey line is actual numbers, blue line is forecasts, grey interval is 95 % limits.

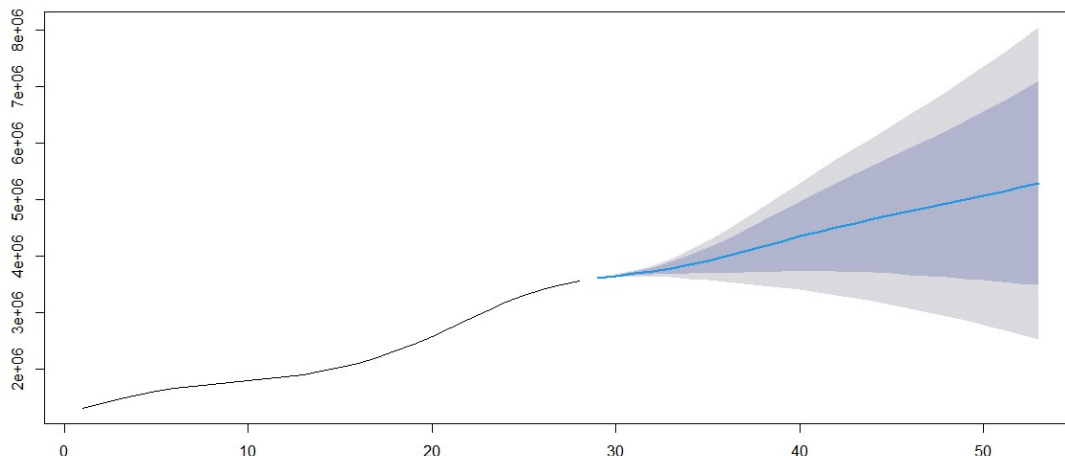


Figure 9. Forecasts from ARIMA(2, 2, 0)

Source: Author’s computation in R-Studio (R-script is in Appendix).

The model predicts that the level of waste in Jordan will increase in the nearest 25 years (Table 6).

Table 6. Point forecast for annual volumes of waste in Jordan until 2043.

Year	Waste, kg	Year	Waste, kg	Year	Waste, kg	Year	Waste, kg	Year	Waste, kg
2019	3608903	2024	3851495	2029	4263880	2034	4656866	2039	4995012
2020	3651852	2025	3922447	2030	4350580	2035	4724700	2040	5066606
2021	3693128	2026	4001663	2031	4433636	2036	4791265	2041	5140251
2022	3737946	2027	4086843	2032	4512246	2037	4857854	2042	5215582
2023	3790070	2028	4175181	2033	4586427	2038	4925533	2043	5292042

Source: Author’s computation in R-Studio (R-script is in Appendix).

The forecasts for the model show a linear upward trend projected into the future. The trend in the long-term forecasts is due to fact that the model includes two non-seasonal differences: this model is basically a random walk with growth fine-tuned by the addition of two autoregressive terms, i.e. two lags of the differenced series.

6. Conclusion

All conducted econometric tests show that the volume of waste produced in Jordan is the most explained by the level of consumption. Consumption culture itself rather than the population growth stimulates waste production. In order to reduce the level of waste generation, it is necessary to move from an unreasonable culture of unrestrained consumption to a reasonable and responsible attitude to the resources of the country and the environment as a whole. It is also necessary to introduce modern waste management methods, develop technologies for waste recycling and reuse.

Projections for waste generation until 2043 are given with the help of ARIMA (2, 2, 0). The forecasts show a linear upward trend projected into the future.

7. Appendix: R-script for data processing

```
# Creation and first view of a data set
```

year<-

c(1991,1992,1993,1994,1995,1996,1997,1998,1999,2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2015,2017,2018)

waste<-

c(1316780.1,1394797.203,1475316.528,1550682.75,1614376.359,1665067.052,1705792.456,1739395.751,1770309.736,1802258.277,1835656.817,1870930.688,1911957.491,1963434.514,2031604.691,2115102.657,2212045.285,2322351.051,2445373.403,2579709.263,2726015.848,2881766.557,3038268.862,3184189.887,3308928.323,3411704.06,3494070.989,3558266.393)

wasteurb<-

c(1011735.554,1086421.54,1164180.085,1238929.308,1296811.435,1337731.431,1370640.66,1397851.038,1422908.131,1448787.656,1475857.107,1504442.422,1537645.428,1579281.259,1656000.2,1753035.323,1861869.245,1982817.065,2115454.196,2258912.314,2413855.942,2578208.762,2744028.302,2900933.699,3022201.314,3123744.639,3206730.473,3273064.577)

wasterur<-

c(305044.5635,308375.663,311136.4427,311753.4423,317564.9249,327335.6208,335151.7965,341544.7131,347401.6053,353470.6209,359799.71,366488.2656,374312.0625,384153.2546,375604.491,362067.3345,350176.0397,339533.9865,329919.2076,320796.9488,312159.9063,303557.7944,294240.5598,283256.1878,286727.0093,287959.4208,287340.5159,285201.8166)

pop<-

c(3760496,3977672,4201555,4410359,4588843,4732854,4848541,4943975,5031762,5122493,5217336,5317506,5434030,5580244,5765635,5991540,6255280,6556478,6893260,7261539,7663131,8090872,8520420,8920049,9266575,9551467,9779173,9956011)

popurb<-

c(2799877,3006563,3221752,3428613,3588796,3702038,3793111,3868413,3937756,4009375,4084287,4163394,4255280,4370503,4582815,4851350,5152537,5487248,5854308,6251314,6680105,7134935,7593824,8028044,8363640,8644651,8874306,9057879)

poprur<-

c(960619,971109,979803,981746,1000047,1030816,1055430,1075562,1094006,1113118,1133049,1154112,1178750,1209741,1182820,1140190,1102743,1069230,1038952,1010225,983026,955937,926596,892005,902935,906816,904867,898132)

cons_cap<-

c(1887.38088,2284.598449,2138.238176,2077.94686,1995.296415,2094.66701,2159.848253,2267.759426,2284.593008,2563.963183,2647.197791,2571.02773,2539.194951,2734.16715,3030.483945,3077.095463,3232.43199,3341.554843,3090.829298,2694.451881,2740.921584,2855.812399,2952.582631,2798.499919,2803.286916,2726.065626,2767.718899,2685.760768)

cons<-

c(4236304891,5231685790,5269014288,5612136826,5932077626,6566431594,6995486601,7727644570,7848236953,8825339492,9331170663,9496332863,10208180536,11737658674,13514104372,15909026798,18352186178,22042313117,22589562764,24209167842,28712834979,31937235543,36226093089,37810719323,38677009873,39787306065,41295761213,42333254443)

gni_cap<-

c(1090,1270,1290,1340,1420,1430,1450,1480,1530,1670,1740,1780,1900,2150,2320,2500,2730,3090,3350,3450,3540,3650,3750,3830,3870,3930,4020,4210)

gni<-

c(3981780291,4963285672,5295884832,5922766566,6449050514,6627371932,7037674894,7774398307,7994693230,8560847532,9137325952,9651423413,10325562059,11642702116,12846211142,15420540480,17686327927,22630606488,24350634697,26336530324,28667418900,30577574048,33336530324,35687023977,37535825106,38946403385,40559802539,42091114245)

empagr<-

c(4.6069998741,4.5079998970,4.440999847,4.3810000420,4.3210000992,4.2610001564,4.2049999237,4.150000954,4.0920000076,4.0359997749,3.9809999466,3.9300000668,4.0349998474,4.1139998436,4.0529999733,3.9879999161,3.9330000877,3.8910000324,3.8340001106,3.7929999828,3.7550001144,3.7090001106,3.6649999619,3.6280000210,3.5929999352,3.5499999523,3.5099999905,3.4670000076)

```
empind<-
c(22.4960003,22.0090008,21.8490009,21.9899998,21.7479992,21.4190006,21.3160000,21.3579998,21.247999
2,21.1509991,21.0629997,21.1219997,23.2439995,25.8120003,25.6630001,25.6509991,25.7889996,25.721000
7,25.5550003,25.3470001,25.2810001,25.1100006,25.0879993,25.0130005,24.9750004,24.8759995,24.823999
4,24.7649994)

empserv<-
c(72.8970032,73.4830017,73.7099991,73.6299973,73.9309998,74.3199997,74.4789963,74.4919968,74.661003
1,74.8119965,74.9560013,74.9469986,72.7210007,70.0729980,70.2839966,70.3619995,70.2770004,70.387001
0,70.6110001,70.8600006,70.9649963,71.1809998,71.2470016,71.3600006,71.4329987,71.5739975,71.666999
8,71.7669983)

data<-
data.frame(year,waste,wasteurb,wasterur,pop,popurb,poprur,cons,cons_cap,gni,gni_cap,empagr,empind,empserv)
head(data)
tail(data)
# Descriptive statistics (Table 1)
install.packages("psych")
library("psych")
describe(data)
# Scatter plot (Figure 1)
install.packages("ggplot2")
library("ggplot2")
qplot(data=data,pop,waste)+stat_smooth(method="lm")
cor(waste,pop)
# Scatter plot (Figure 2)
qplot(data=data,gni,waste)+stat_smooth(method="lm")
cor(waste,gni)
# Scatter plot (Figure 3)
qplot(data=data,cons,waste)+stat_smooth(method="lm")
cor(cons,waste)
# Bubble chart (Figure 4)
# distribution between „wasteurb“ and „wasterur“ evaluated by „pop“ (size of the bubble is determined by „pop“)
radius<-sqrt(data$pop/pi)
symbols(data$popurb,data$poprur,circles=radius,inches=0.2, fg="white", bg="grey70", xlab="popurb",
ylab="poprur",main="Waste in urban and rural areas evaluated by Population")
# Bubble chart (Figure 5)
# distribution between „wasteurb“ and „wasterur“ evaluated by „cons“ (size of the bubble is determined by „cons“)
radius<-sqrt(data$cons/pi)
symbols(data$wasteurb,data$wasterur,circles=radius,inches=0.2, fg="white", bg="violet", xlab="Wasteurb, kg",
ylab="wasterur, kg",main="Waste in urban and rural areas evaluated by Consumption")
cor(cons,pop) # highly correlated variables which explains similar graphs
# Summary statistics (Table 2)
model1<-lm(data=data,waste~cons)
summary(model1)
model2<-lm(data=data,waste~pop+gni+cons)
summary(model2)
```

```
model3<-lm(data=data,waste~cons+empagr+empind+empserv)
summary(model3)
model4<-lm(data=data,waste~cons+empagr+empind+empserv+poprur+popurb)
summary(model4)
# Quality criterions (Table 3)
deviance(model1)
deviance(model2)
deviance(model3)
deviance(model4)
AIC(model1)
AIC(model2)
AIC(model3)
AIC(model4)
BIC(model1)
BIC(model2)
BIC(model3)
BIC(model4)
# Econometric tests (Table 4)
install.packages("lmtest")
library("lmtest")
# Test for some linear restrictions simultaneously
waldtest(model1)
waldtest(model2)
waldtest(model3)
waldtest(model4)
# Test for omitted regressors
resettest(model1)
resettest(model2)
resettest(model3)
resettest(model4)
# Heteroscedasticity test
bptest(model1)
bptest(model2)
bptest(model3)
bptest(model4)
# Test for autocorrelation of the first order
dwt(model1)
dwt(model2)
dwt(model3)
dwt(model4)
# Multicollinearity test (Table 5)
install.packages("car")
```

```
library("car")
vif(model1)
vif(model2)
vif(model3)
vif(model4)
#Time series
#Test for stationarity of a time series (Figure 6, 7)
install.packages("DescTools")
library("DescTools")
install.packages("forecast")
library("forecast")
y<-data$waste
PlotACF(y)
ndiffs(y)
dy<-diff(y, differences= 2)
PlotACF(dy)
# Auto-simulation
mod_a <- auto.arima(y)
summary(mod_a)
# Cross correlation (Figure 8)
z<-data$cons
cor <- ccf(y,z, ylab = "cross-correlation")
# Prediction (Table 6, Figure 9)
mod_a <- auto.arima(y)
summary(mod_a)
prediction_a <- forecast(mod_a, h=25)
prediction_a
plot(prediction_a)
```

References

- Abdoli, M., & Falahnezhad, M. (2011). Multivariate econometric approach for solid waste generation modelling: A case study of Mashhad, Iran. [Online] Available: https://www.researchgate.net/publication/258697406_Multivariate_Econometric_Approach_for_Solid_Waste_Generation_Modeling_A_Case_Study_of_Mashhad_Iran (August 27, 2020)
- Afroz, R., Hanaki, K., & Tudin, R. (2010). Factors affecting waste generation: A study in a waste management program in Dhaka city, Bangladesh. *Environmental Monitoring and Assessment*. 179(1–4), 509–519.
- Alhanaqtah, V. (2020a). Economic and ecological efficiency with respect to waste treatment policies: Literature review. *European Journal of Business and Management*. 12(1), 149–158.
- Alhanaqtah, V. (2020b). Economic instruments to fund solid waste management services: Analysis and recommendations for Jordan. *Journal of Economics and Sustainable Development*. 10(14), 144–153.
- Al-Rifai, J., & Amoudi, O. Understanding the key factors of construction waste in Jordan. *Jordan Journal of Civil Engineering*. 10(2), 2016–244.
- Bach, H., Mild, A., Natter, M., & Weber, A. (2004). Combining socio-demographic and logistic factors to explain the generation and collection of waste paper. *Resource Conservation Recycling*. 41, 65–79.

- Beigl, P., Lebersorger, S., & Salhofer, S. (2008). Modelling municipal solid waste generation: A review. *Waste Management*. 28(1), 200–14.
- Beigl, P., Wassermann, G., Schneider, F., & Salhofer, S. (2004). Forecasting municipal solid waste generation in major European cities. [Online] Available: https://www.researchgate.net/publication/250457873_Forecasting_Municipal_Solid_Waste_Generation_in_Major_European_Cities (August 27, 2020)
- Bosire, E., Oindo, B., & Atieno, J. (2017). Modeling household solid waste generation in urban estates using socioeconomic and demographic data, Kisumu City, Kenya. [Online] Available: <https://repository.maseno.ac.ke/handle/123456789/441> (August 27, 2020)
- Chhay, L., Reyad, M., Suy, R., Islam, M., & Mian, M. (2018). Municipal solid waste generation in China: Influencing factor analysis and multi-model forecasting. *Journal of Material Cycles and Waste Management*. 20(3), 1761–1770.
- Chung, S. (2010). Projecting municipal solid waste: the case of domestic waste in Hong Kong special administration region. *Resources Conservation and Recycling*. 54(11), 759–768.
- Dennison, G., Dodd, V., & Whelan, B. (1996). A socioeconomic based survey of household waste characteristics in the city of Dublin, Ireland. – II Waste quantities. *Resources, Conservation and Recycling*. 17(3), 245–257.
- Gay, A., Beam, T., & Mar, B. (1993). Cost effective solid waste characterization methodology. *Journal of Environmental Engineering*. 119(4), 631–644.
- Hage, O., Sandberg, K., Söderholm, P., & Berglund, C. (2018). The regional heterogeneity of household recycling: A spatial-econometric analysis of Swedish plastic packing waste. [Online] Available: <https://link.springer.com/article/10.1007/s12076-017-0200-3> (August 27, 2020)
- Han, Z., Liu, Y., Zhong, M., Shi, G., Li, Q., Zeng, D., Zhang, Y., Fei, Y., & Xie, Y. (2018). Influencing factors of domestic waste characteristics in rural areas of developing countries. *Waste Management*. 72, 45–54.
- Hockett, D., Lober, D., & Pilgrim, K. (1995). Determinants of per capita municipal solid waste generation in the Southeastern United States. *Journal of Environmental Management*. 45(3), 205–217.
- Joutz, F. (1996). Modeling and forecasting municipal solid waste generation in the US energy supply. *Journal of Forecasting*. 15(6), 477–494.
- Khan, D., Kumar, A., & Samadder, S. (2016). Impact of socioeconomic status on municipal solid waste generation rate. *Waste Management*. 49, 15–25.
- Liu Ch., & Wu X. Factors influencing municipal solid waste generation in China: A multiple statistical analysis study. *Waste Management and research: The Journal of a sustainable circular economy*. 29(4), 371–378.
- Liu J., Li Q., Gu W., Wang Ch. (2019). The impact of consumption patterns on the generation of municipal solid waste in China: Evidences from provincial data. [Online] Available: https://www.researchgate.net/publication/333161238_The_Impact_of_Consumption_Patterns_on_the_Generation_of_Municipal_Solid_Waste_in_China_Evidences_from_Provincial_Data (August 27, 2020)
- Lu, J., Zhang, S., Hai, J., & Lei, M. (2017). Status and perspectives of municipal solid waste incineration in China: A comparison with developed regions. *Waste Management*. 69, 170–186.
- Mahees, M., Sivayoganathan, C., & Basnayake, B. (2011). Consumption, solid waste generation and water pollution in Pinga Oya catchment area. [Online] Available: <https://tar.sljol.info/articles/abstract/10.4038/tar.v22i3.3697/> (August 27, 2020).
- Prades, M., Gallardo, A., & Ibáñez, M. (2015). Factors determining waste generation in Spanish towns and cities. *Environmental Monitoring and Assessment*. 187, 4098.
- Samson, M., Akinlabi, E., Muzenda, E., & Aboyade A. (2017). A review on factors affecting municipal solid waste generation. [Online] Available: https://www.academia.edu/35943085/A_Review_on_Factors_affecting_Municipal_Solid_Waste_Generation (August 27, 2020)
- Trang, P., Dong, H., Toan, D., Hanh, N., & Thu, N. (2017). The effects of socio-economic factors on household solid waste generation and composition: A case study in Thu Dau Mot, Vietnam. *Energy Procedia*, 107, 253–258.
- Your guide to waste management in Jordan: Waste Sorting Informative Booklet (2016). Jordan Green Building Council. Friedrich Ebert Stiftung. [Online] Available: <https://library.fes.de/pdf-files/bueros/amman/12729.pdf>

(August 27, 2020)

World Bank data. [Online] Available: <https://www.mdpi.com/1660-4601/16/10/1717> (August 27, 2020)

Notes

Note 1. Urban population refers to people living in urban areas as defined by national statistical offices. It is calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects.

Note 2. Rural population refers to people living in rural areas as defined by national statistical offices. It is calculated as the difference between total population and urban population. Aggregation of urban and rural population may not add up to total population because of different country coverages.

Note 3. GNI (formerly GNP) (current US \$) is the sum of value added by all resident producers plus any product taxes (less subsidies) not included in the valuation of output plus net receipts of primary income (compensation of employees and property income) from abroad.

Note 4. Final consumption expenditure (formerly total consumption) (current US\$) is the sum of household final consumption expenditure (private consumption) and general government final consumption expenditure (general government consumption).

Note 5. In accordance with (Your Guide 2016) average waste generation rate is 0.99 kg/cap/day in the urban areas and 0.87 kg/cap/day in the rural areas in 2015.

Note 6. Significance codes:

. coefficient is significant when the level of significance $\alpha = 0.1$ (10 %)

* coefficient is significant when the level of significance $\alpha = 0.05$ (5 %)

** coefficient is significant when the level of significance $\alpha = 0.01$ (1 %)

*** coefficient is significant when the level of significance $\alpha < 0.01$, i.e. when a probability to make an error of the first type is less.

Note 7. “NA” means that a system is computationally singular: reciprocal condition number = 8.01736e-28.

Note 8. ACF is merely a bar chart of the coefficients of correlation between a time series and lags of itself.

The PACF plot is a plot of the partial correlation coefficients between the series and lags of itself.

Veronika Alhanaqtah was born in Minsk, Belarus, in 1983. She received her PhD in Economics from the Belarusian State University in 2010. She was hired as the Lecturer (2008), Assistant Professor (2011) and then as an Associate Professor (2013) at the Department of Theoretical and Institutional Economics at the Belarusian State University. She is currently the Assistant Professor at the College of Business at Tafila Technical University (Jordan). Dr. Alhanaqtah has been given a special award of the President of the Republic of Belarus for talented young scientists (2012). She is the winner of the international competition for the best author’s research “*Ecology and long-term investments*” (Russia, 2018). She worked as the consultant in working out the bills concerning the environmental protection and rational use of natural resources in the Republic of Belarus (2009), the researcher in UNDP-Belarus on implementation of mechanisms of Kyoto Protocol (2006–2009). Since 2019 she is a member of the InterGlobe Research Network.