International Journal of Economics and Finance; Vol. 16, No. 2; 2024
ISSN 1916-971X E-ISSN 1916-9728
Published by Canadian Center of Science and Education

Forecasting of the Waste Generation in Jordan: Alternative Econometric Approaches

Omar Jraid Alhanaqtah[1]

[1] College of Business, Tafila Technical University, Tafila, Jordan

Correspondence: Omar Jraid Alhanaqtah, College of Business, Tafila Technical University, 66110 Tafila, Jordan. Tel: 962-772-110-181. E-mail: omarhanaqtah@yahoo.com

Received: October 23, 2023 Accepted: December 17, 2023 Online Published: December 20, 2023

doi:10.5539/ijef.v16n2p35 URL: https://doi.org/10.5539/ijef.v16n2p35

Abstract

The main purpose of the article is to predict the household waste generation in Jordan in the short-run using alternative methods and explain factors highly likely impacting its generation. The results of comparative analysis made by three methods – regression technique, time series modelling and the annual growth rate method – are provided. The results of time series approach take a compromised position between the other ones. It is concluded that time series modelling with the help of ARIMA(0,1,0) with drift is more reliable for the short-run forecasting of the waste generation in Jordan while the regression is more suitable for explaining the effect of input variables on an outcome.

Keywords: ARIMA models, consumption, cross-correlation, GDP, log-linear regression, population, waste

1. Introduction

Million tons of the household waste are produced in Jordan every year. Along with other factors this situation has been dramatically worsened by the alerting refugee problem. Jordan has become a safety island in the economically and politically unstable Middle East. The objective of the research is to predict the household waste generation in Jordan in the short-run using alternative economic and statistical methods and explain factors highly likely impacting its generation. It has been hypothesized that 'gdp', "consumption' and 'population' are factors that impact the household waste generation.

The main findings of the research are as follows. The results of time series approach take the middle position between other methods. It is concluded that time series modelling with the help of ARIMA(0,1,0) with drift is more reliable for the short-run forecasting of the waste generation in Jordan while the regression is more suitable for explaining the effect of input variables on an outcome.

The research outcomes are useful for policy makers to realize the scale of the household waste problem and to optimize capital expenditures into the waste management system of Jordan. For the future research, it is supposed to make forecasting of the household waste generation with the help of the Regression with ARIMA Errors combining two powerful instruments – ARIMA and linear regression. To apply this technique, high-quality, disaggregated and regularly collected data is required.

Qualitative data as statistical facts are very important for any research. We need actual data and forecast to realize the scale of the household waste problem and to optimize capital expenditures into the waste management system of Jordan. The research outcomes are expected to contribute to the solution of the household waste management problem in Jordan by providing policy makers data background to put forward appropriate plans.

2. Literature Review

The literature analysis revealed many factors that might influence the household waste generation. There are many remarkable, interesting and rather sophisticated studies attempting to find out the most influencing, the most informative variables. As for any research the access to qualitative data is of a great importance. The output of analysis is as good as the input good. Usually the access to data is the most significant obstacle on the way of a researcher. Notably, it is also important to be critical of the methodology of different statistical techniques used for data processing and predictions. Particularly, econometric tests have its assumptions and conditions for application.

Among the large body of literature about the factors impacting on the household waste generation we focus on commenting the several ones, being the most informative for the current research. Consumption expenditures, population numbers, disposable personal income, the size of the household, gross domestic product are among typical factors mentioned by the researches (Abdoli, Falahnehzad, & Behboudian, 2011; Afroz, Hanaki, & Tudin, 2010; Cheng et al., 2020; Chung, 2010; Hage et al., 2018; Herianto, Maryono, & Budihardjo, 2019; Hockett, Lober, & Pilgrim, 1995; Liu and Wu, 2010; Liu et al., 2019; Lu et al., 2017; Samson et al., 2017; Zhao, Diunugala, & Mombeuli, 2021). For example, Liu and Wu (2010) focus on urban development as an important factor influencing municipal solid waste generation.

As far as rural areas there are groups of factors mentioned in the work of Han et al. (2018): economic and law (administrative levels, industrial development of a village, energy structure, fuel used, per capita income and consumption expenditure), social (population numbers, age, training and education), natural (geographic location, climatic characteristics), cultural (traditions, living habits, personal attitude).

Unique consumption patterns impacting solid waste generation are considered in the work of Liu et al. (2019). The authors considered such interesting factor as the housing rent. It is a large part of a household expenditures that influences the actual purchasing power parity. They established the negative relationship between the financial expenditure on housing and the waste generation, and the positive relationship between the last and the household expenditures on food. This study affirms the importance of the consumption expenditure factor for the waste generation. The study of Alhanaqtah (2020) also confirms the significance of the consumption factor for the household waste generation. Usually this factor is highly correlated with the population numbers and economic growth of a country. Aldayyat et al. (2019) also indicates the population factor as significant for the waste generation and emphasizes its significance in relationship with the refugee problem in Jordan.

Institutional, technical, socio-economic, cultural factors as having strong relationship with the waste generation are considered in the works of Abir, Datta, and Saha (2023); Adeleke et al. (2021).

Atieno, Oinde, and Bosire (2017), Chhay et al. (2018), Khan, Kumar, and Samadder (2016), Prades. Gallardo, and Ibanez (2015), Trang et al. (2017) employ socio-economic data (particularly, demographic) to model solid waste generation.

3. Research Methodology

The comparative analysis made by three methods – log-linear regression technique, time series modelling (ARIMA) and the annual growth rate method. The data set consists of annual time series for Jordan in the period 2000-2022: household waste, total population, GDP at purchaser's prices, household final consumption expenditures.

3.1 Research Plan

First, data is described in statistical terms providing numerical characteristics of variables as well as compute cross-correlation between series 'waste–gdp', 'waste–population', and 'waste–consumption'. Data trends have been visualized and predictions of the dynamics of the waste generation have been made with the help of the regression technique. Second, a set of linear and non-linear models with different combinations of independent variables 'population', 'consumption' and 'gdp' have been analyzed. Then models with the highest R-squared-adjusted have been selected, and models demonstrating high multi collinearity between variables have been excluded. By this models suitable for further econometric tests have been selected. Then alternative models based on values of the probability-value (p-value), R-squared-adjusted, information criteria Akaike (AIC) and Schwartz (BIC) have been analyzed. The following econometrics tests have been conducted: Ramsey test to check for omitted regressors, Goldfeld-Quandt test to check for the presence of heteroscedasticity, Durbin-Watson and Breusch-Godfrey tests to check for the autocorrelation. Based on econometric analysis the best fit model to explain the influencing factors on the waste generation in Jordan have been chosen. Third, time series modelling with the help of ARIMA models for the short-run forecasting of the waste generation have been conducted. Here the behavior of the Auto-correlation function (ACF) has been analyzed, and the distribution of residuals (to exclude autocorrelation) has been considered by performing the Leung-Box test. This allows to opt for the best fit model for forecasting. Forth, forecasting of the waste generation is made with the help of the annual growth rate method. The results of comparative analysis made by three methods (regression technique, time series modelling and the annual growth rate method) are provided. Finally, conclusion is made. The research is accompanied by R-scripts for computation.

3.2 Data Set

The data set is an annual time series for Jordan in the period 2000–2022. The following variables are used:

waste – total household waste generation, million tones;

population – total population, million people;

gdp - GDP at purchaser's prices, million, current US$;

consumption - household final consumption expenditure (formerly private consumption in the World Bank definitions), million, current US$.

The data for variables are determined in the Table 1 (Note 1).

Table 1. Data set

year	population	gdp	consumption	waste
2000	5.06	8460.79	6820.59	1.39
2001	5.16	8975.81	7274.36	2.21
2002	5.28	9582.51	7322.00	2.23
2003	5.4	10195.63	7844.29	2.27
2004	5.53	11411.71	9307.33	2.31
2005	5.68	12589.00	11055.43	2.36
2006	6.08	15056.98	12801.41	2.31
2007	6.47	17110.44	14826.94	2.21
2008	6.63	22658.73	16855.8	2.11
2009	6.78	24537.88	16818.31	1.92
2010	6.93	27133.80	17892.96	2.07
2011	7.11	29524.15	21784.51	2.02
2012	7.21	31634.56	24440.85	2.24
2013	7.69	34454.44	28939.44	2.57
2014	8.66	36847.64	30083.10	2.79
2015	9.49	38587.02	30719.72	3.37
2016	9.96	39892.55	31622.54	3.39
2017	10.22	41608.44	32867.61	3.41
2018	10.46	43370.86	33223.94	3.47
2019	10.7	44503.01	33240.85	3.44
2020	10.93	43579.92	NA	3.56
2021	11.15	45116.32	NA	NA
2022	11.29	47451.50	NA	NA

Note. Author's development based on data for 'population', "gdp', and 'consumption' from (World Development Indicators, 2023), data for 'waste' from (Ritchie & Mathieu, 2023).

Additionally, the daily per capita waste generation, using data from the Table 1, has been computed. The average daily per capita waste generation in Jordan in 2000-2020 was 0.947 kg. The estimated value correlates with estimates made for the middle-income countries (Your guide to waste management 2016).

4. Modelling Factors Impacting Waste Generation

4.1 Descriptive Statistics

Based on the analysis of the literature and available data, it has been hypothesized that 'gdp', "consumption' and 'population' are factors that impact the household waste generation. Descriptive statistics is in the Table 2.

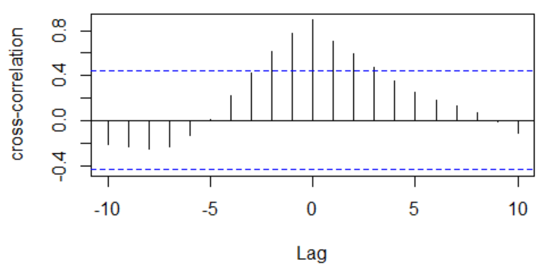
Table 2. Numerical characteristics of variables

variables	mean	median	std.dev.	min	max	skew	kurtosis
population	7.82	7.11	2.21	5.06	11.29	0.30	–1.57
gdp	28012.33	29524.15	13877.37	8460.79	47451.50	–0.12	–1.64
consumption	19787.10	17374.38	10028.50	6820.59	33240.85	–0.09	–1.69
waste	2.55	2.31	0.63	1.39	3.56	0.35	–1.17

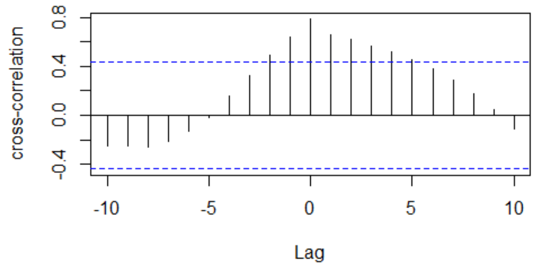
Note. Author's computation in R/R-Studio (R-script is in Appendix, Section 1).

The distribution of the variable 'population' is close to symmetric: mean is close to the median; the skew is rather low. But the high value of kurtosis indicates that it is not a normal distribution: the probability mass is mostly concentrated in the 'tails' of the distribution. The same is true for the variables "gdp' and 'consumption': the skew is almost zero that witnesses about symmetric distribution, but the high and negative value of kurtosis indicates that there is not a feebly varying 'core' of values in the distribution, it is not normal.

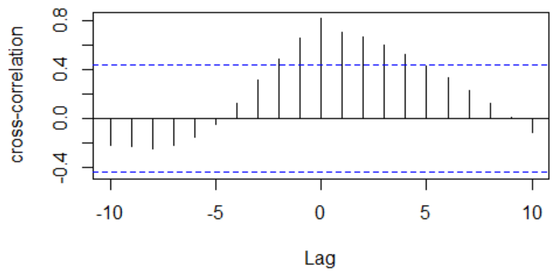
The cross-correlation between series 'waste – population', 'waste – gdp', and 'waste consumption' has been computed (Figure 1).



(a) Cross-correlation between 'waste' and 'population'



(b) Cross-correlation between 'waste' and 'gdp'



(c) Cross-correlation between 'waste' and 'consumption'

Figure 1. Cross-correlation function

Note. Computed by the author in R/R-Studio (see Appendix, Section 2).

Values of the cross-correlation for all pairs of series are close to 0.8: all variables demonstrate high correlation with 'waste'. Analysis of cross-correlation functions (CCF) for every pair of series shows that there is no shift in time of one series relative to the other (time lag is zero); series tend to move in one direction.

4.2 Data Trends and Prediction of the Waste Generation Using Regression Technique

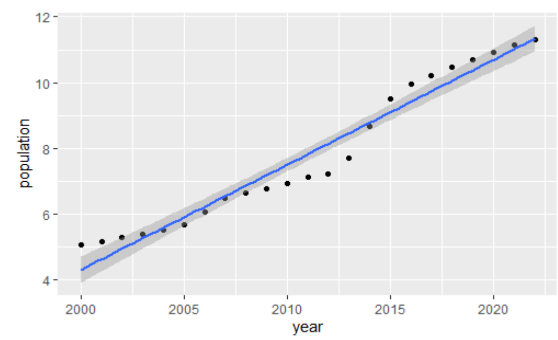
The Figure 2 represents the dynamics of the variables.

The Figure 2(a-c) shows that the 'population', 'consumption' and 'gdp' follow not perfect linear trends. Perhaps, it is reasonable to use exponential function to model this type of a pattern. The standard approach to model exponential function is to use logarithms. That is why it has been selected more than 20 linear, logarithmic and semi-log models to simulate the factors influencing the waste generation (see Appendix, Section 4). Looking at the Figure 1(d) we may notice that the trend for the waste generation is non-linear, more likely exhibiting the random work with drift. First, a log-linear function with one independent variable 'time' to explain the average annual growth rate has been applied. The regression formula (1) for the waste generation is as follows:

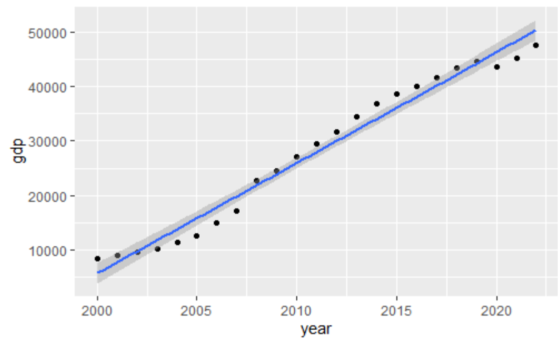
$$\log(waste) = -64.96 + 0.0327 \cdot year \tag{1}$$

Interpretation: every year the waste generation in Jordan increases on average by 3.3 %.

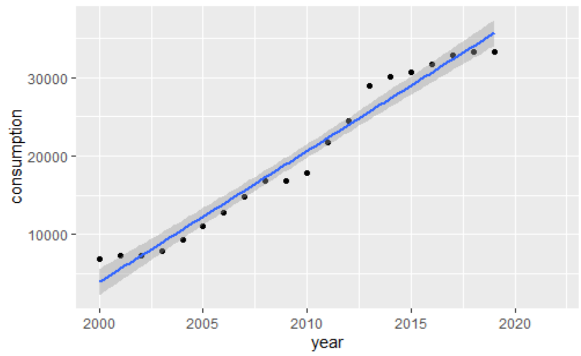
Consequently, using this model for prediction of the waste generation gives us the following results: in 2025 – 3.832 million tons, 2030 – 4.258 million tones, 2035 – 4.683 million tones (author's computation in R-Studio, R-script is in Appendix, Section 5). However, it is expected that these numbers are underestimated because R-squared-adjusted for the model explains its variance only 65.6 %.



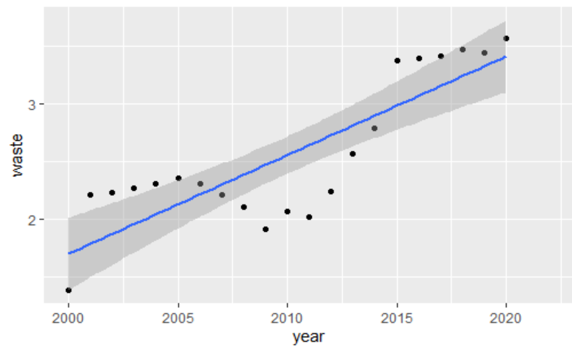
(a) population



(b) GDP



(c) household final consumption expenditures



(d) waste generation

Figure 2. Data trends

Note. Computed by the author in R/R-Studio (see Appendix, Section 3).

4.3 Choice of the Best Fit Model

Second, having analyzed linear and non-linear models with different combinations of independent variables ‘population’, ‘consumption’ and ‘gdp’, two models with the highest R-squared-adjusted have been selected. Models demonstrating high multi collinearity between variables (using the Variance Inflation Factor, VIF) have been excluded from the consideration. For example, ‘gdp’ and ‘consumption’, put together in models, exhibited high correlation. Every variable, put alone in the model, didn’t explain its variance, exhibiting low values of R-squared-adjusted. Thus, the following suitable models for further econometric tests have been selected:

Model 1: $\log(waste) = \log(gdp) + \log(population)$

Model 2: $waste = \log(gdp) + \log(population)$

VIFs for both models are less than 10 so the variables ‘gdp’ and ‘population’ don’t exhibit linear correlation.

4.4 Econometric Tests

At the initial stage there is a comparison of alternative models based on values of the probability-value (p-value) for the coefficients of the model; R-squared-adjusted and residuals sum of squares (RSS) or deviance, that are both measures of model fit; information criterion AIC and BIC that are both ‘penalty’ criteria for high RSS and large number of regressors. Numerical characteristics are in the Table 3.

Table 3. Summary statistics for two alternative models

Criterion	Model 1	Model 2
p-value (Intercept)	0.0792 .	0.005065 **
p-value (log(gdp))	0.0132 *	0.000432 ***
p-value (log(population))	9.93e-05 ***	9.07e-07 ***
R^2 -adjusted	0.7655	0.8676
RSS	0.2291975	0.8152125
AIC	-24.62053	0.7567662
BIC	-20.6376	4.739695

Note. Author’s computation in R/R-Studio (R-script is in Appendix, Section 6).

Comparing p-values for beta-coefficients, the Model 2 looks better: all coefficients are significant at the level of significance less than 1 %. The coefficient of determination (R-squared-adjusted) is higher in the Model 2. It means that the value of the variance of this model is smaller, which is better for the model. However, RSS is higher for the Model 2. Conspicuously, AIC in the Model 2 is suspiciously close to the deviance. AIC and BIC for the Model 1 are negative and lower than for the Model 2. It means that the Model 1 is better in the framework of the information criterion. Its relatively large negative values indicate less loss of information than positive values of AIC in the Model 2. The comparative analysis of numerical characteristics of two models shows that it is difficult to opt for the best model at this stage. Thus, additional econometric tests are required.

Primarily, it is necessarily to check for omitted variables in the models for which we don't have observations. For this purpose, the Ramsey test (RESET test) with the null-hypothesis 'no omitted regressors' has been applied. Test statistics is in the Table 4.

Table 4. Econometric tests

Econometric test	Model 1	Model 2
	Ramsey test	
test-statistic	0.768	0.435
p-value	0.481	0.656
	Goldfeld-Quandt test	
test-statistic	0.052	0.113
p-value	0.998	0.984
	Breusch-Godfrey	
test-statistics	0.246	0.293
p-value	0.884	0.863
	Durbin-Watson test	
test-statistics	1.178	0.016
p-value	1.259	0.026

Note. Author's computation in R-Studio (R-script is in Appendix, Section 7).

The p-values for both models don't allow to reject the null-hypothesis of the Ramsey test. Thus, the models don't have lost informative variables.

Since two models (as the ones without multi collinearity problem) have already been selected, there is the need to check for the presence of the heteroscedasticity, i.e. for every observation the variance of residuals is not constant. For this purpose, the Goldfeld-Quandt test (for small samples), the null-hypothesis of which implies the absence of heteroscedasticity, has been applied. Since p-values are very high in both models, the null-hypothesis is not rejected, and there is no heteroscedasticity in both models.

Finally, there is the necessity to check for the autocorrelation of residuals. For this purpose, the Breusch-Godfrey (LM) test for serial correlation, suitable for any order of the autocorrelation, has been applied. The null-hypothesis of the test is 'no autocorrelation'. The statistics from the Table 5 shows that p-values for both models are higher than 5% so null-hypothesis is not rejected: there is no autocorrelation in the models. The Durbin-Watson test for the autocorrelation of the first order has additionally been applied. The null-hypothesis of the test is 'there is no autocorrelation of the 1^{st} order'. Since p-values for both models are less than 5 %, then the null-hypothesis is rejected: there is the autocorrelation of the 1^{st} order in both models.

Econometrics tests showed very similar results for both models. This is not surprising, since the models are very similar to each other. Nevertheless, it is reasonable to opt for the Model 2 as a better model since it has more significant beta-coefficients and a higher value of R-squared-adjusted.

5. Time Series Modelling of the Waste Generation

Based on the data trends values for the waste generation for 2025, 2030 and 2035 years have been predicted. For this purpose, time series technique – ARIMA models – has been applied. A key feature of ARIMA is that these models don't consider exogenous variables, in their basic forms, time is used as a predictor variable. ARIMA models as a subset of linear regression models are more suitable for short-term forecasting while regression is more suitable for explaining the effect of input variables on an outcome.

In this paragraph the dynamics of an annual waste generation in Jordan using the latest available data in 2000–2019 has been analyzed. This is a univariate time series of the length 20. The Figure 3 represents its dynamics.

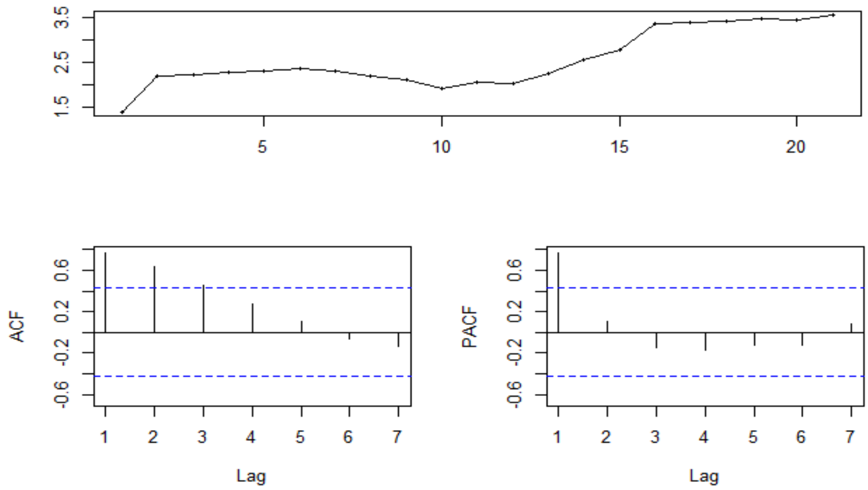


Figure 3. Time series for the waste generation

Note. Computed by the author in R/R-Studio (see Appendix, Section 8).

The Figure 3 shows that our series takes compromised position between stationary and non-stationary processes. In some way it looks like a short random walk. The auto-covariance function (ACF) is decreasing but no that quick as it takes place in an ideal stationary process. So there are two options: either to simulate this process as stationary, or simulate this process as non-stationary. The augmented Dickey-Fuller test (unit-root test) has been conducted. The null-hypothesis of the test is “series are non-stationary’. The p-value of the test is 0.4593 that doesn’t allow to reject the null-hypothesis. Thus, the series for the waste generation is a non-stationary process.

In order not to simulate at random it is possible to conduct auto simulation in R. The command auto.arima() iterates over the set of possible models to opt for the best result based on AIC criteria: ARIMA(2,1,2) with drift, ARIMA(0,1,0) with drift, ARIMA(1,1,0) with drift, ARIMA(0,1,1) with drift, ARIMA(0,1,0), ARIMA(1,1,1) with drift. The auto.arima() with default parameters suggests ARIMA(0,1,0) with drift (the AIC is minimum).

It is necessarily to look at the residuals of the model because residuals inform whether the model captures all the information from the data provided (Figure 4).

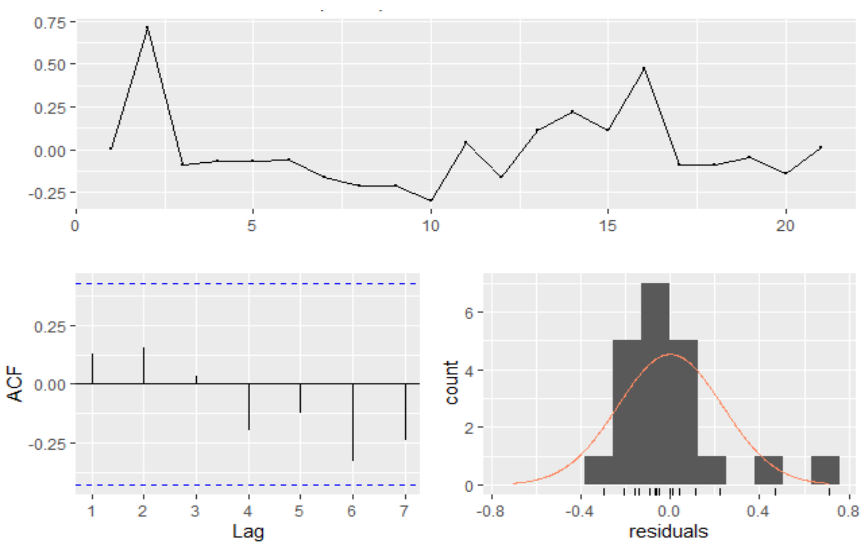


Figure 4. Residuals from ARIMA(0,1,0)

Note. Computed by the author in R/R-Studio (see Appendix, Section 8).

In the Figure 4 it looks like that the model ARIMA(0,1,0) with drift left some information in its residuals. The first graph informs us that residuals are not the ideal ‘white noise’ but the second graph shows that all lags go beyond the threshold established by the Auto-correlation function (ACF). The last graph confirms that the distribution of residuals tends to be normal, although it is a bit right-skewed. Since residuals follow nearly normal distribution, the point forecasts and forecasts for prediction intervals will be rather accurate. To make sure that there is no autocorrelation of residuals the Leung-Box test has been performed. The p-value of the test is 0.5385 so the null-hypothesis is not rejected, and the data don't exhibit serial correlation. This allows to use the resulting model for forecasting (Figure 5).

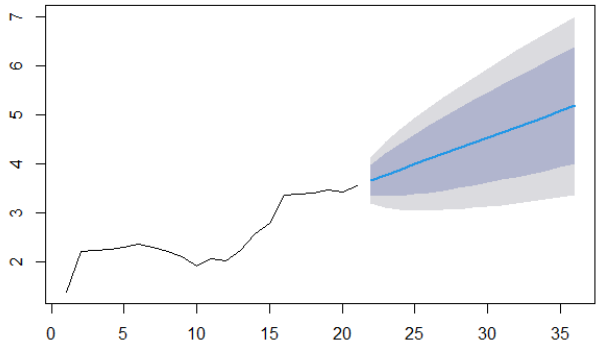


Figure 5. Waste generation forecasting from ARIMA(0,1,0) with drift

Note. Computed by the author in R/R-Studio (see Appendix, Section 8).

The Figure 5 shows that the waste generation in Jordan will steadily increase in the nearest years. Predicted values are in the Table 5.

Table 5. Waste generation forecasting in Jordan

Year	Point forecast, million tons	Confidence interval	
		Lower 95 %	Higher 95 %
2025	4.103	3.054	5.152
2030	4.645	3.161	6.129
2035	5.188	3.370	7.004

Note. Computed by the author in R/R-Studio (see Appendix, Section 8).

Noticeably, forecast made with the help of ARIMA(0,1,0) correlates with predicted numbers for the waste generation from the regression formula (1) above.

6. Annual Growth Rate Method for the Waste Generation Forecasting

Another method used for forecasting in a large number country reports and business statistics is an annual growth rate method. This method uses the annual output of domestic waste in the previous year (Y_{n-1}) as the forecast base, and calculates the annual output of the waste in the forecast year (Y_n) according to the annual average growth rate (r). The calculation formula is as follows:

$$Y_n = Y_{n-1} \cdot (1 + r)$$

Based on data from the Table 1 the average growth rate of the waste generation in Jordan has been computed. Its value is 5.11 % for the period 2000–2020. Results of forecasting of the waste generation made by the annual growth rate method are in the Table 6.

Table 6. Annual growth rate of the waste generation, million tons

year	waste	year	waste	year	waste	year	waste
2000	1.39	2009	2.18	2018	3.41	2027	5.34
2001	1.46	2010	2.29	2019	3.58	2028	5.61
2002	1.54	2011	2.40	2020	3.77	2029	5.90
2003	1.61	2012	2.53	2021	3.96	2030	6.20
2004	1.70	2013	2.66	2022	4.16	2031	6.52
2005	1.78	2014	2.79	2023	4.37	2032	6.85
2006	1.87	2015	2.94	2024	4.60	2033	7.20
2007	1.97	2016	3.09	2025	4.83	2034	7.57
2008	2.07	2017	3.24	2026	5.08	2035	7.95

Note. Computed by the author in MS Excel.

It is assumed that computations by the annual growth rate method don't capture all the information hidden in the data, in comparison with the time series modelling with the help of ARIMA models. In accordance with the considering method the growth rate is higher because it is based simply on the moving average (MA) calculation. As a result, the forecast values are overestimated. The comparison of computations made by three different methods (regression technique, time series modelling, annual growth rate method) are represented in the Table 7.

Table 7. Waste generation forecast made by alternative methods, million tons

Year	Regression	Time series	Annual growth rate
2025	3.832	4.102	4.83
2030	4.258	4.645	6.20
2035	4.683	5.188	7.95

It is observed from the Table 7 that computations made by the time series modelling take a compromised position between the other methods. It is considered that time series modelling with the help of ARIMA models is more reliable for the short-run forecasting while regression is more suitable for explaining the effect of input variables on an outcome.

7. Research Findings

Qualitative data as statistical facts are very important for any research. The output of forecasting is as good as the input good. Statistical inference is largely influenced by the quality of data collection. Time series have dependences. Thus, any omitted value might affect the output dramatically. During data processing it is important to be critical of the methodology behind. This remark is important to keep in mind while conducting data analysis and making predictions.

Based on the analysis of the literature and available data, it has been hypothesized that 'gdp', "consumption' and 'population' are factors that impact the household waste generation.

Analysis of the cross-correlation function (CCF) for every pair of series shows that all variables demonstrate high correlation with 'waste'. There is no shift in time of one series relative to the other; series tend to move in one direction.

The trend for the waste generation in Jordan is non-linear, more likely exhibiting the random work with drift. First, a log-linear regression to explain the average annual growth rate has been applied. The model estimation showed that every year the waste generation in Jordan increases on average by 3.3 %.

Having analyzed linear and non-linear models with independent variables 'population', 'consumption' and 'gdp' in different combinations, two models with the highest R-squared-adjusted and lowest VIFs. Have been selected These models are supposed to explain the impact of different factors on the waste generation. Econometrics tests performed very similar results for both selected models. It is reasonable to opt for the model $waste = \log(gdp) + \log(population)$ as a better model since it has more significant beta-coefficients and a higher value of R-squared-adjusted.

The time series technique – ARIMA models – for the waste generation forecasting has been used. A key feature of ARIMA is that these models don't consider exogenous variables, in their basic forms, time is used as a predictor variable. ARIMA models as a subset of linear regression models are more suitable for short-term forecasting while regression is more suitable for explaining the effect of input variables on an outcome.

The analysis showed that the series took the compromised position between stationary and non-stationary processes. The auto simulation in R suggested ARIMA(0,1,0) with drift based on the minimum value of AIC criterion. The analysis of residuals of this model showed that they follow nearly normal distribution, so the point forecasts and forecasts for prediction intervals are rather accurate. To exclude autocorrelation of residuals performed the Leung-Box test has been performed. On balance, the econometric analysis confirmed that ARIMA(0,1,0) with drift may be used for forecasting of the waste generation.

The annual growth rate method for forecasting of the waste generation has been applied. It is concluded that predicted values of this method are overestimated because it is based just on the moving average (MA) calculation.

Finally, compared computations made by three methods – regression technique, time series modelling, annual growth rate method – have been compared. It is observed that the time series modelling took a compromised position between the other methods. It is considered that time series modelling with the help of ARIMA(0,1,0) with drift is more reliable for the short-run forecasting of the waste generation in Jordan while the regression is more suitable for explaining the effect of input variables on an outcome.

8. Conclusion

To sum up, the time series modelling with the help of ARIMA(0,1,0) with drift is more reliable for the short-run forecasting of the waste generation in Jordan while the regression is more suitable for explaining the effect of input variables on an outcome.

The research outcomes are useful for policy makers to realize the scale of the household waste problem and to optimize capital expenditures into the waste management system of Jordan.

The research contributes to the solution of the waste management problem in terms of forecasting, necessary to put forward appropriate plans.

Regarding future research, it would be useful to make forecasting of the household waste generation with the help of such a technique as 'Regression with ARIMA errors' combining two powerful instruments – ARIMA and linear regression. To apply this technique, high-quality and regularly collected data is required.

References

Abdoli, M., Falahnehzad, M., & Behboudian, S. (2011). Multivariate econometric approach for solid waste generation modelling: A case study of Mashhad, Iran. *Environmental Engineering Science, 28*(9). https://doi.org/10.1089/ees.2010.0234

Abir, T., Datta, M., & Saha, S. (2023). Assessing the factors influencing effective municipal solid waste management system in Barishal metropolitan areas. *Journal of Geoscience and Environmental Protection, 11*(1), 49-66. https://doi.org/10.4236/gep.2023.111004

Adeleke, O., Akinlabi, S., Jen, T., & Dunmade, I. (2021). An overview of factors affecting the rate of generation and physical composition of municipal solid waste. *IOP Conference Series Materials Science and Engineering, 1107*(1). https://doi.org/10.1088/1757-899X/1107/1/012096

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. https://doi.org/10.1007/s10661-010-1753-4

Aldayyat, E., Saidan, M., Abu Saleh, M., & Hamdan, S. (2019). Solid waste management in Jordan: impacts and analysis. *Journal of Chemical Technology and Metallurgy, 54*(2), 454-462.

Alhanaqtah, V. (2020). Econometric modelling of solid waste generation in Jordan. *Journal of Economics and Sustainable Development, 11*(16), 204-220. https://doi.org/10.7176/JESD/11-16-22

Atieno, J., Oindo, B., & Bosire, E. (2017). Modeling household solid waste generation in urban estates using socioeconomic and demographic data, Kisumu City, Kenya. *Scholars Journal of Arts, Humanities and Social Sciences, 5*(5), 468-477.

Cheng, J., Shi, F., Yi, J., & Fu, H. (2020). Analysis of the factors that affect the production of municipal solid waste in China. *Journal of Cleaner Production, 259*(4), 120808. https://doi.org/10.1016/j.jclepro.2020.120808

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. https://doi.org/10.1007/s10163-018-0743-4

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. https://doi.org/10.1016/j.resconrec.2009.11.012

Hage, O., Sandberg, K., Söderholm, P., & Berglund, C. (2018). The regional heterogeneity of household recycling: A spatial-econometric analysis of Swedish plastic packing waste. *Letters in Spatial and Resource Sciences, 11*(2). https://doi.org/10.1007/s12076-017-0200-3

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. https://doi.org/10.1016/j.wasman.2017.11.039

Herianto, H., Maryono, M., & Budihardjo, M. (2019). Factors affecting waste generation: household study in Palangka Raya City, Central Kalimantan. *The 4th International Conference on Energy, Environment,* *Epidemiology and Information System (ICENIS 2019), 125*, 07007. https://doi.org/10.1051/e3sconf/201912507007_

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. https://doi.org/10.1006/jema.1995.0069

Jordan Green Building Council. (2016). Your guide to waste management in Jordan. *Waste sorting informative booklet.* Retrieved from https://mena.fes.de/fileadmin/user_upload/pdf-files/publications/Your_Guide_to_Waste_Management_in_Jordan.pdf

Khan, D., Kumar, A., & Samadder, S. (2016). Impact of socioeconomic status on municipal solid waste generation rate. *Waste Management, 49*, 15-25. https://doi.org/10.1016/j.wasman.2016.01.019

Liu, C., & Wu, X. (2011). 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. https://doi.org/10.1089/cpb.2008.0321

Liu, J., Li, Q., Gu, W., & Wang, C. (2019). The impact of consumption patterns on the generation of municipal solid waste in China: Evidences from provincial data. *International Journal of Environmental Research and Public Health, 16*(10), 1717. https://doi.org/10.3390/ijerph16101717

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. https://doi.org/10.1016/j.wasman.2017.04.014

Prades, M., Gallardo, A., & Ibàñez, M. (2015). Factors determining waste generation in Spanish towns and cities. *Environmental Monitoring and Assessment, 187*, 4098. https://doi.org/10.1007/s10661-014-4098-6

Ritchie, H., & Mathieu, E. (2023). Waste management. *Our World in Data.* Retrieved from https://ourworldindata.org/waste-management

Samson, M., Akinlabi, E., Muzenda, E., & Aboyade, A. (2017). A review on factors affecting municipal solid waste generation. *2nd International Engineering Conference*. Federal University of Technology, Minna, Niger State (October 17-19, 2017). Retrieved from https://www.academia.edu/35943085/A_Review_on_Factors_affecting_Municipal_Solid_Waste_Generation

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. https://doi.org/10.1016/j.egypro.2016.12.144

World Development Indicators. (2023). *The world Bank*. https://doi.org/10.1016/j.egypro.2016.12.144

Zhao, Y., Diunugala, H., & Mombeuli, C. (2021). Factors affecting household solid waste generation and management in Sri Lanka: An empirical study. *Environmental Monitoring and Assessment, 193*, 838. https://doi.org/10.1007/s10661-021-09633-7

Note

Note 1. 'NA' – data for this year are not available.

Appendix

R-scripts for computation

```
# Section 1. Descriptive statistics
describe(d)
hist(population)
hist(gdp)
hist(consumption)
hist(waste)
```

```
# Section 2. Cross correlation
# create new vectors without 'NA' (2000-2019)
year2<-c(2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018,2019)
population2<-c(5.06,5.16,5.28,5.4,5.53,5.68,6.08,6.47,6.63,6.78,6.93,7.11,7.21,7.69,8.66,9.49,9.96,10.22,10.46,10.7)
gdp2<-c(8460.79,8975.81,9582.51,10195.63,11411.71,12589,15056.98,17110.44,22658.73,24537.88,27133.8,29524.15,31634.56,34454.44,36847.64,38587.02,39892.55,41608.44,43370.86,44503.01)
consumption2<-c(6820.59,7274.36,7322,7844.29,9307.33,11055.43,12801.41,14826.94,16855.8,16818.31,17892.96,21784.51,24440.85,28939.44,30083.1,30719.72,31622.54,32867.61,33223.94,33240.85)
waste2<-c(1.39,2.21,2.23,2.27,2.31,2.36,2.31,2.21,2.11,1.92,2.07,2.02,2.24,2.57,2.79,3.37,3.39,3.41,3.47,3.44)
d2<-data.frame(year2,population2,gdp2,consumption2,waste2)
w<-d2$waste2
c<-d2$consumption2
g<-d2$gdp2
p<-d2$population2
cor <- ccf(w,c, ylab = "cross-correlation")
cor <- ccf(w,g, ylab = "cross-correlation")
cor <- ccf(w,p, ylab = "cross-correlation")
# Section 3. Visualization of data trends
qplot(data=d,year,population)+stat_smooth(method="lm")
qplot(data=d,year,gdp)+stat_smooth(method="lm")
qplot(data=d,year,consumption)+stat_smooth(method="lm")
qplot(data=d,year,waste)+stat_smooth(method="lm")
# Section 4. Model simulation
# Log-models
model1 <- lm(log(waste2) ~ log(population2)+log(gdp2)+log(consumption2), data = d2)
summary(model1)
vif(model1)
model2 <- lm(log(waste2) ~ log(gdp2)+log(population2), data = d2)
summary(model2)
vif(model2)
model3 <- lm(log(waste2) ~ log(consumption2)+log(population2), data = d2)
summary(model3)
vif(model3)
model4 <- lm(log(waste2) ~ log(gdp2)+log(consumption2), data = d2)
summary(model4)
vif(model4)
model5 <- lm(log(waste2) ~ log(consumption2), data = d2)
summary(model5)
model6 <- lm(log(waste2) ~ log(population2), data = d2)
summary(model6)
model7 <- lm(log(waste2) ~ log(gdp2), data = d2)
```

```
summary(model7)
# Linear models
model1 <- lm(waste2 ~ population2+gdp2+consumption2, data = d2)
summary(model1)
vif(model1)
model2 <- lm(waste2 ~ gdp2+population2, data = d2)
summary(model2)
vif(model2)
model3 <- lm(waste2 ~ consumption2+population2, data = d2)
summary(model3)
vif(model3)
model4 <- lm(waste2 ~ gdp2+consumption2, data = d2)
summary(model4)
vif(model4)
model5 <- lm(waste2 ~ consumption2, data = d2)
summary(model5)
model6 <- lm(waste2 ~ population2, data = d2)
summary(model6)
model7 <- lm(waste2 ~ gdp2, data = d2)
summary(model7)
# Log-linear models
model1 <- lm(log(waste2) ~ population2+lgdp2+consumption2, data = d2)
summary(model1)
vif(model1)
model2 <- lm(log(waste2) ~ gdp2+population2, data = d2)
summary(model2)
vif(model2)
model3 <- lm(log(waste2) ~ consumption2+population2, data = d2)
summary(model3)
vif(model3)
model4 <- lm(log(waste2) ~ gdp2+consumption2, data = d2)
summary(model4)
vif(model4)
model5 <- lm(log(waste2) ~ consumption2, data = d2)
summary(model5)
model6 <- lm(log(waste2) ~ population2, data = d2)
summary(model6)
model7 <- lm(log(waste2) ~ gdp2, data = d2)
summary(model7)
# Linear-to-log models
model1 <- lm(waste2 ~ log(population2)+log(gdp2)+log(consumption2), data = d2)
summary(model1)
```

```
vif(model1)
model2 <- lm(waste2 ~ log(gdp2)+log(population2), data = d2)
summary(model2)
vif(model2)
model3 <- lm(waste2 ~ log(consumption2)+log(population2), data = d2)
summary(model3)
vif(model3)
model4 <- lm(waste2 ~ log(gdp2)+log(consumption2), data = d2)
summary(model4)
vif(model4)
model5 <- lm(waste2 ~ log(consumption2), data = d2)
summary(model5)
model6 <- lm(waste2 ~ log(population2), data = d2)
summary(model6)
model7 <- lm(waste2 ~ log(gdp2), data = d2)
summary(model7)
# Other combinations
model1 <- lm(log(waste2) ~ population2+log(gdp2), data = d2)
summary(model1)
vif(model1)
model2 <- lm(log(waste2) ~ log(consumption2)+population2, data = d2)
summary(model2)
vif(model2)
# Section 5. Forecasting using the regression formula (1)
model_waste <- lm(log(waste) ~ year, data = d)
summary(model_waste)
nd_waste<-data.frame(year=c(2025,2030,2035))
nd_waste
predict(model_waste,nd_pop)
# Section 6. Summary statistics for two alternative models
# Select better models
model1 <- lm(log(waste2) ~ log(gdp2)+log(population2), data = d2)
summary(model1)
deviance(model1)
AIC(model1)
BIC(model1)
model2 <- lm(waste2 ~ log(gdp2)+log(population2), data = d2)
summary(model2)
deviance(model2)
AIC(model2)
BIC(model2)
# Section 7. Econometric tests
```

```
# Multicollinearity
vif(model1)
vif(model2)
# Ommitted regressors
resettest(model1)
resettest(model2)
# Heteroscedasticity
install.packages("broom")
library("broom")
# Model 1
h<-augment(model1,d2)
glimpse(h)
resid<-h$.resid
describe(resid)
hist(resid)
qplot(data=d2,x=population2,abs(resid))
qplot(data=d2,x=gdp2,abs(resid))
qplot(data=d2,x=consumption2,abs(resid))
qplot(data=d2,x=waste2,abs(resid))
# Model 2
h<-augment(model2,d2)
glimpse(h)
resid<-h$.resid
describe(resid)
hist(resid)
qplot(data=d2,x=population2,abs(resid))
qplot(data=d2,x=gdp2,abs(resid))
qplot(data=d2,x=consumption2,abs(resid))
qplot(data=d2,x=waste2,abs(resid))
# GQ test
gqtest(model1,data=d2,fraction=0.2)
gqtest(model2,data=d2,fraction=0.2)
# Autocorrelation
dwt(model1)
dwt(model2)
bgtest(model1,order=2)
bgtest(model2,order=2)
# Section 8. Time series simulation and forecasting
# create new vector for the waste without NA
Y<-c(1.39,2.21,2.23,2.27,2.31,2.36,2.31,2.21,2.11,1.92,2.07,2.02,2.24,2.57,2.79,3.37,3.39,3.41,3.47,3.44,3.56)
tsdisplay(Y)
adf.test(Y)
```

```
mod_a <- auto.arima(Y,trace=TRUE,ic="aic")
summary(mod_a)
checkresiduals(mod_a)
tsdisplay(residuals(mod_a))
prediction_a <- forecast(mod_a, h=15)
prediction_a
plot(prediction_a)
```