

# The Effects of Power Usage on Five Power Modes in South Africa Using Multivariate Techniques

<sup>1</sup>Busanga Jerome Kanyama and <sup>2</sup>Solly Matshonisa Seeletse

<sup>1</sup>Department of Statistics, University of South Africa, Florida Campus, South Africa

<sup>2</sup>Department of Statistics and Operations Research,

Sefako Makgatho Health Sciences University, Ga-Rankuwa, South Africa

## Article history

Received: 20-10-2017

Revised: 26-12-2017

Accepted: 09-01-2018

Corresponding Author:

Solly Matshonisa Seeletse

Department of Statistics and

Operations Research, Sefako

Makgatho Health Sciences

University, Ga-Rankuwa,

South Africa

Email: solly.seeletse@smu.ac.za

**Abstract:** South Africa contends with the lack of access to secure affordable energy. This affects the country's ability to provide reliable power to consumers. This paper investigates the source of energy 'power mode' such as electricity, gas, paraffin, solar and other (firewood and cow dung) and the effects of power usage in cooking, heating and lighting at household consumption level using three multivariate techniques, namely; Multivariate Analysis of Variance (MANOVA), discriminant analysis and factor analysis. These methods are used to determine the mostly used source of energy and usage based on the nine Provinces in South Africa. According to all the three techniques, electricity and paraffin were the most used type of energy source. However, electricity usage was ahead of paraffin. Rationalization and use of the power of the optimum of these powers are required. MANOVA was the preferred method in terms of ease of use and interpretation of the results.

**Keywords:** Discriminant Analysis, Factor Analysis, Power Usage

## Introduction

From 1994 of democratization, South Africa has been subjected to many challenges. Due to past apartheid policies, many areas endured a lack of access to basic services such as electricity. About two-thirds of South African population did not have access to electricity before 1994 (Ziramba, 2008).

The South African government considers electricity provision as vital for developing the country. Therefore, South Africa's electricity consumption has increased sharply since the early 1990s (Inglesi-Lotz and Bignon, 2011). Eskom, a company that generates about 95% of South Africa's electricity and about 45% of electricity used in Africa, encouraged the reduction of energy usage at household level (Netshiava, 2014).

Several studies (Altman *et al.*, 2008; De Lange, 2008; Magadla and Holloway, 2011) have been done in the electricity sectors of South Africa since democratic rule in 1994. They all confirm that electricity demand is seasonal, is more costly for higher income and relatively more affluent communities and most electricity income is generated from industrial consumers, among others.

Eskom has potential to increase power production because it generates, transmits and distributes electricity

to industrial, mining, commercial, agricultural and residential customers (Prinsloo, 2009). Eskom requested a budget to build new stations around 2014, which was denied. In December 2007, this rejection decision was found to be an error. It later adversely affected the South African economy. Mining companies estimate that plentiful ounces of both gold and platinum production were lost annually. Preventing these losses would greatly improve the economy of South Africa. The sources of generating electricity power are coal, nuclear and solar energy, gas and paraffin (Amusa *et al.*, 2009).

Eskom management stirs consumers to conserve power during peak periods in order to reduce load shedding (Bignon *et al.*, 2005). Indeed, energy household consumption problems are, by definition, multivariate.

The energy consumption has attracted interest in the energy literature over the past two decades (Amusa *et al.*, 2009) in which electricity was found to be a vital energy source.

Multivariate techniques are applicable in exploring the effects of factors with levels. In this study, these factors are concurrently used instead of their individual effects. The objective was to compare three methods in analysing two-way layout studies: Discriminant analysis, factor analysis and MANOVA.

To make fitting choices among these methods, researchers should know the statistical models underlying them (Bianco *et al.*, 2009). Each statistical method presented the results using power consumption in the household.

This paper investigates the source of energy ‘power mode’ such as electricity, gas, paraffin, solar and other (firewood and cow dung) and the effects of energy usage ‘power usage’ in cooking, heating and lighting at household consumption level in South Africa using these three multivariate techniques (discriminant analysis, factor analysis and MANOVA).

Previous statistics-based studies on electricity were not based on the above methods, but on other methods. They mainly used artificial neural networks and Kalman filter (Khosravi *et al.*, 2013; Yamin *et al.*, 2004; Zhang and Luh, 2005); fuzzy regression and logic (Gładysz and Kuchta, 2011); Markov regime switching models (Janczura, 2014; Janczura *et al.*, 2013; Janczura and Weron, 2009; 2010; 2012; 2014); probability (Gneiting *et al.*, 2007); structural methods (Lanne *et al.*, 2010; Sinton and Levine, 1994); times series and forecasting (Khosravi *et al.*, 2013; Weron, 2014; Weron and Misiorek, 2008; Zhao *et al.*, 2008); among others. The methods were studied individually, for all consumers and no comparisons were made. In this study, the methods used (discriminant analysis, factor analysis and MANOVA) are studied for households, at individual levels and also compared. Hence, this study is an innovation.

## Model Description

### Data

The experimental design was motivated through a survey of household power consumption in South Africa during 2007. The household expenditure dataset was used, which was obtained from the unit record file of Statistics South Africa (Stat SA). This was presented by two factors. The first factor was the ‘power usage’ with three levels. The three levels of power usage were known to be a source of power consumption that was apparent in several measures of the expenses in the household. Household energy consumption is considered as the energy consumed in homes to meet the needs of the residents themselves.

The energy consumption of households is often called the residential energy consumed in household dwellings. It is thought that the power consumption might be different in either the rural areas compared to urban areas or from one Province to another. Another factor was the ‘province’ with nine levels, namely; Gauteng, Eastern Cape, Free State, KwaZulu-Natal, Limpopo, Mpumalanga, North West, Northern Cape and Western Cape.

Using MANOVA, the factor energy source ‘power mode’ with five levels was used as the dependent variables and the two categorical factors ‘province’ and ‘power usage’ were considered as independent variables. When using the discriminant analysis, the factor energy source ‘power mode’ was used as independent variable and the factor ‘province’ and ‘power usage’ was the dependent variables. The factor analysis used the factor energy source ‘power mode’ as the dependent variable. The statistical packages used to perform the three techniques were SPSS (IBM SPSS Statistics for Windows version 20.0) and SAS version 19.3 of 2012. These appear in Jackson (2015) and USI (2015)

### Statistical Methods

Multivariate techniques are the most common applications in social science to identify and test the effects from the analysis. Use of multivariate techniques is unusual in identifying the effects of two factors such as ‘power usage’ with three levels and ‘province’ with nine levels based on the type of ‘power mode’. This study needed both to choose a number of group comparison of the ‘power mode’ and to study the nature of group differences between the two factors ‘power usage’ and ‘province’.

Most statistical techniques require the variables to be measured quantitatively and in some cases to also be normally distributed (Frank, 2009; Jaynes, 2003; Leon-Garcia, 2008). In order to offset limitations that could be due to these confines, this study applied discriminant analysis, factor analysis and MANOVA.

### Discriminant Analysis

According to Lawrence *et al.* (2006), Descriptive Discriminant Analysis (DDA) is often used to support a significant MANOVA to determine the structure of the linear combination of the dependent variables. DDA focuses on revealing major differences among the factors to answer questions:

- Can a province be a useful factor to classify the energy consumption in the household?
- Can the power usage be used as a factor to classify the energy consumption in terms of the utilization in the household?
- What are the chances of making mistakes when using these factors?

In the real data used, mistakes occur whenever a source of energy of type ‘power mode’ is classified into the wrong category expense in the household. Thus, an error will occur when, for example, household expenses for cooking is predicted to be caused by lighting or heating. Alternatively, a consumption of the energy in a particular ‘province’ is allocated to another ‘province’. It

is also noted that these two kinds of errors are probably not equally serious.

Discriminant analysis is a multivariate technique useful to control the power consumption and classify expenses of the households into the appropriate effects either 'Province' or 'power usage'. The independent variable that does not contribute a significant amount of prediction could be considered for deletion from the model (Abdi, 2007; Almeida, 2002; Ennett *et al.*, 2001).

### Factor Analysis

Factor analysis examined the pattern of correlations between the energy sources of the 'energy source' variables. Variables that are highly correlated, either positively or negatively, are likely to be influenced by the same factors, while those that are considered as relatively uncorrelated are likely to be influenced by different factors (O'Brien and Marakas, 2007; Vanlier *et al.*, 2012).

Factor analysis provides information regarding:

- The number of different factors needed to explain the pattern of the relationship among variables
- The nature of these factors
- The level of quality of the way the hypothesised factors explain the 'power mode'
- The amount of unique variance that each type of energy source variable includes?

### MANOVA

The study assessed the differences across the combinations of the factor 'power mode' considered as dependent variable, as this constructed linear relationship between the five dependent variables using Multivariate Analysis of Variance (MANOVA). In addition, the MANOVA technique assessed the level of 'province' and 'power usage'. In particular, MANOVA was used to identify which of 'province' and 'power usage' differentiated the most set of energy source 'power mode' by using a correlation matrix.

The investigation of the relationships between the five 'power mode' (electricity, paraffin, gas, solar and cow dung) at each level of factor 'province' and factor 'power usage' provided statistical guidance to reduce the dimension. The type of energy source that produced the most 'power usage' or 'province' separation was identified.

The idea behind 'factor' is that there are two variables which affect the dependent variable called 'power mode'. The study was conducted on 'province' and 'power usage' and 27 independent observations were been detected at each of the (9×3) combinations of levels. The two-way layout was with one observation per cell for a variety of interaction effects. The experiment procedure considered the multivariate two-way model in which, in turn, the interaction of the factors was examined.

The two-way fixed effects model for a vector response consisted of five (5) components and the  $k^{\text{th}}$  observation at level 'I' of factor 'province' and 'j' of factor 'power usage' was denoted by  $X_{ijk}$ ,  $i = 1, 2, \dots, 9$  and  $j = 1, 2, 3$  and  $k = 1, 2, \dots, 27$ . The two-way fixed-effects for a vector response consisting of five (5) complements, using adapted equations from Bökeoğlu and Büyükoztürk (2008), is given as:

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad (1)$$

where,  $\sum_{i=1}^9 \alpha_i = \sum_{j=1}^3 \beta_j = \sum_{i=1}^9 \gamma_{ij} = \sum_{j=1}^3 \gamma_{ij} = 0$  and the vectors are all of order  $5 \times 1$  and  $\varepsilon_{ijk}$  is assumed to be a  $(N_5(0, \Sigma))$  random vector.  $\mu$  represents an overall level,  $\alpha_i$  represents the fixed effect of factor 'province',  $\beta_j$  represents the fixed effect of factor 'power usage' and  $\gamma_{ij}$  is the interaction between factor 'province' and factor 'power usage'. The interaction term represented the joint effect of two or more treatments. Interaction terms were created for each combination of treatment variables.

The expected response at the  $i^{\text{th}}$  level of factor 'province' and the  $j^{\text{th}}$  level of factor 'power usage', from McLachlan (2004), is therefore:

$$E(X_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij} \quad (2)$$

Considering that the two factors were both between the groups design, the appropriate model takes the form as defined in Equation 1. Addressing the questions posed earlier, a mathematical exploration of the model takes place below.

According to Venables and Ripley (2002), measuring variation between groups:

$$\begin{aligned} X_{ijk} = & \bar{X} + (\bar{X}_i - \bar{X}) + (\bar{X}_j - \bar{X}) \\ & + (\bar{X}_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}) \\ & + (X_{ijk} - \bar{X}_{ij}) \end{aligned} \quad (3)$$

where,  $\bar{X}$  = the overall average of the observation vectors,  $\bar{X}_i$  = the average of the observation vectors at  $i^{\text{th}}$  level of factor 'province',  $\bar{X}_j$  = the average of the observation vectors at  $j^{\text{th}}$  level of factor 'power usage' and  $\bar{X}_{ij}$  = the average of the observation vectors at the  $i^{\text{th}}$  level of factor 'province' and at  $j^{\text{th}}$  of factor 'power usage'.

The effects of two factors, 'province' and 'power usage' were examined simultaneously on 5-dependent variables 'power mode'.

The model consisted of two types of components: Main effects that described the impact of an individual variable value on the results and Interaction effects that consider combinations of variable values (Ash, 2011; Montgomery *et al.*, 2006).

This study considered two factors case, ‘province’ at nine levels as the nine provinces of South Africa and ‘power usage’ at the three levels in developing the model. Possible combinations of various levels of these two factors were assessed for their joint effect on power consumption.

Following the tradition, when a number of group comparison strategies need to be chosen, first MANOVA was conducted and then ANOVA would follow at 5% level of significance if necessary. MANOVA on the multiple dependent variables ‘power mode’ was statistically significant (p-value <0.05). Then ANOVA became necessary.

Statistical procedures of each model for analysing data were used when the experimental design included combinations of two-factors. The hypotheses of interest for a two-factor experiment concerned the main effects and the combined effect. The strengths and limitations of each approach were identified by using an application and comparison of method employing real data.

## Results

A statistically significant multivariate effect showed that the independent variable ‘province’ and ‘power usage’ were associated with differences between the vectors or sets of means. Thus, the study presumed that factor effects existed. If effects existed, the next step in this process was to discover which specific dependent variables were affected.

### Main Effects of Province

The main effects of Province accounts for about 76.6% of the total variance as indicated in last column of the Table 1. As expected for Province, the study found that gas, solar and other (firewood or cow dung) did not differ in terms of Province (p-value >0.05) as shown in Table 2, whereas the consumption of electricity and

paraffin differed in terms of Province (p-value <0.05) where Gauteng was the highest.

### Main Effects of Power Usage

The main effects of power usage accounts for about 75.8% of the total variance (Table 1). Second, the next question was of course to determine the importance of type of energy source ‘power mode’ to the overall effects.

These estimated marginal means were displayed in Table 2.

Table 2 indicates that the energy source ‘power mode’ differed significantly in terms of the type of utilization such as cooking, lighting or heating (p-value <0.05). It was noted that electricity scored the highest of the power usage group with a mean of 949850.444 as this indicated the strength of association between the types of energy source that contributed the most to the significant overall effects.

Table 3 demonstrates that when ‘power usage’ was the dependent variable, electricity and paraffin did not differ in terms of the way that ‘power usage’ (cooking, lighting and heating) was used while other types of energy source differed (p-value <0.05). Two discriminant functions were obtained. However, the first function accounts for 71.2% of the variation of ‘power usage’.

**Table 1:** Main effects, multivariate test

Effect		F	p-value	Partial eta squared
Intercept	Wilks' Lambda	588.96	0.000	0.996
Province	Wilks' Lambda	5.915	0.000	0.766
Usage	Wilks' Lambda	7.533	0.000	0.758

**Table 2:** Grand mean

Dependent variable	Mean
Electricity	949850.4
Gas	14809.6
Paraffin	153307.4
Solar	1584.3
Other)	587388.7

**Table 3:** Test for corrected model and main effects

Source	Dependent variable	F	P-value	Partial Eta square
Corrected Model	Electricity	2827.744	0.000	0.994
	Gas	61.988	0.000	0.795
	Paraffin	157.135	0.000	0.908
	Solar	13.970	0.002	0.466
	Other	33.237	0.000	0.675
Province	Electricity	170.676	0.000	0.988
	Gas	2.514	0.055	0.557
	Paraffin	11.699	0.000	0.854
	Solar	1.264	0.327	0.387
	Other	0.980	0.485	0.324
Usage	Electricity	23.652	0.000	0.747
	Gas	15.526	0.000	0.660
	Paraffin	11.064	0.001	0.580
	Solar	4.570	0.027	0.364
	Other	5.343	0.017	0.400

A simultaneously discriminant analysis of ‘province’ and ‘power usage’ was constructed to determine whether the five independent of type of energy source ‘power mode’ could predict the consumption of energy in the household. Table 4 indicates that electricity and paraffin differed significantly in terms of the Province (p-value <0.05) while no difference was observed when using other type of energy source (p-value >0.05).

The next question determined the discriminate function that explains the dependent variable ‘province’. Tables 4 and 5 showed that 96.4% of variation of the effects ‘province’ was explained by the discriminant function 1 through 5 (p-value <0.05). Two discriminant functions altogether account for about 99%.

Table 6 exposes the superiority of Function 1 from the eigenvalue column. Out of the six Functions offered, the contribution of Function 6 is over 96% on the original data. On the canonical correlation, there is still a high contribution of over 93% from this Function alone.

About the importance of the type of energy source in predicting the effects of ‘province’ electricity (1.680) was found to be the best in prediction Function 1. The other sources’ importance measures, which appear in Table 7, are:

- Function 2: Electricity (1.218)
- Function 3: Solar (1.134)

- Function 4: Gas (0.748)
- Function 5: Other (1.021).

Table 8 shows that when ‘power usage’ was the dependent variable, electricity and paraffin did not differ in terms of the way that ‘power usage’ was used while other types of energy source differed (p-value <0.05). Two discriminant functions were obtained but the first function account for 71.2% of the variation of ‘power usage’ (Table 9).

Alternatively, Lawrence *et al.* (2006) suggested that the smaller Wilks’s lambda signals a higher importance of the independent variable to discriminant function.

**Table 4:** Test of equality of group means

Dependent variable	Wilks’ lambda	p-value
Electricity	0.044	0.000
Gas	0.701	0.494
Paraffin	0.289	0.001
Solar	0.713	0.534
Other	0.773	0.718

**Table 5:** Wilks’ lambda

Test of function	Wilks’ lambda	p-value
1 through 5	0.002	0.000
2 through 5	0.167	0.202
3 through 5	0.473	0.713
4 through 5	0.880	0.992
5	0.982	0.987

**Table 6:** Eigenvalues

Fun	Eigen-value	% of variation	Cum %	Can cor
1	74.607	96.4	96.4	0.993
2	1.822	2.4	98.7	0.804
3	0.863	1.1	99.8	0.681
4	0.116	0.1	100.0	0.322
5	0.018	0.0	100.0	0.134

\*Fun = Function, Eig = Eigenvalue, Cum = Cumulative, var = Variation, Can cor = Canonical correlation

**Table 7:** Standardized canonical discriminant function coefficients

	Function				
	1	2	3	4	5
Electricity	1.680	-0.252	-0.098	-0.057	-0.081
Gas	-0.374	-0.421	0.967	0.746	-0.374
Paraffin	0.668	1.218	-0.117	-0.009	0.035
Solar	-0.818	0.556	1.134	-0.376	0.009
Other	0.702	-0.041	0.041	-0.078	1.021

**Table 8:** Test of equality of group means

Dependent variable	Wilks’ lambda	P-value
Electricity	0.967	0.668
Gas	0.538	0.001
Paraffin	0.832	0.110
Solar	0.741	0.027
Other	0.691	0.012

**Table 9:** Wilks' lambda

Test of function	Wilks' lambda	p-value	% of var	Cum %
1 through 2	0.185	0.000	71.2	71.2
2	0.553	0.011	28.8	100.0

**Table 10:** Total variance: Principal component analysis

C	Initial eigenvalues			Rotation sums of squared loadings		
	Total	% of var	Cum %	Total	% of var	Cum %
1	2.515	35.923	35.923	2.157	30.819	30.819
2	1.397	19.956	55.879	1.696	24.224	55.044
3	1.248	17.829	73.707	1.306	18.664	73.707
4	0.947	13.534	87.241			
5	0.487	6.956	94.197			
6	0.226	3.225	97.422			
7	0.180	2.578	100.000			

Key: C = Component

Table 10 essentially employs the Kaiser-Guttman criterion of eigenvalues greater than 1.0. The table shows that a three-factor solution accounted about 74% of the total variance. The selection of the number of factors is vital. Any factor with an eigenvalue less than 1 is not as important. In practice, a robust solution should account for at least 50% of the variance (Fidell and Tabachnick, 2007). The present three-factor model was deemed the best solution because of its conceptual clarity and ease of interpretability.

## Discussion

The study investigated whether or not the 'power usage' and 'province' had significant effects on one another. If these two factors do not interact, then their individual effects could be investigated separately. The three multivariate techniques indicated earlier were used and the results obtained were then compared to identify the best technique.

A two-way between subjects MANOVA was conducted on the energy source factor 'power mode'. This factor was significantly affected by the main effects of 'power usage' and 'province'. Differences and similarities were found between the three techniques. The data involved more than one variable and all the techniques focused on terms such as correlation, linear combinations, factors and functions.

The three techniques had analysed a complex array of variables, providing greater assurance to get conclusions with less error and more validity. These methods were linear combination of variables indicating whether the independent or dependent variables formed a linear combination of variables to interpret the data.

The primary concern of multivariate techniques was first to predict outcomes based on prior information, such as being able to accurately predict group membership of a given number of variables.

Secondly, the techniques were used to answer the question:

- Which variables are the most important in the prediction of some outcome?

The results from this investigation suggested that MANOVA was the best technique because of:

- Its adaptability
- Its ease of use and result interpretation
- Its overall methodology

When looking at the power consumption, however all the three techniques indicated that electricity and paraffin were the most used types (that is, source of energy) with electricity being the highest among the two. In terms of province, high consumption of type of source of energy was consistently high in Gauteng proportion in cooking compared to heating and lighting.

Electricity showed to be playing a significant role in power source energy among the nine Provinces of South Africa. This is especially factual for those with more industries such as Gauteng Province with a large number of households. In terms of power usage, it was indicated that most of power energy was used for cooking which takes a lot of household's expenses.

The necessity of using the source of energy appeared to be supreme among all the three techniques. The initiatives campaign of using wisely the energy source (electricity) has to be supported by the government up to the lower level in the country. South African government considers electricity provision as very important for the growth and the development of the country (DME, 2003). For each technique there was variation in the way that energy source was frequently consumed.

## Conclusion

A correct model can be proposed to each of the three techniques. Details of the influences of each energy source type could enlighten more regarding the effects of power usage. The study recommends that another study should be undertaken after eliminating the type of energy source such as gas, solar and other (firewood and cow dung) from the analysis.

## Acknowledgement

The researchers acknowledge the support from the departments from their respective universities.

## Funding Information

The National Research Foundation (NRF) funded the postgraduate study from which this paper emerged.

## Author's Contributions

**BJ Kanyama:** Prepared the study samples, conducted the literature study, performed the calculations and wrote the manuscript. Basically, this entails having conducted the main study and initiated the first draft of the article.

**SM Seeletse:** Was the project leader, made conceptual contributions on the research, updated the literature, formatted the paper for AJAS and also submitted the paper.

## Ethics

There are no ethical worries or concerns regarding this paper.

## References

- Abdi, H., 2007. Discriminant Correspondence Analysis. In: Encyclopedia of Measurement and Statistics, Salkind, N.J. (Ed.), Thousand Oaks (CA): Sage., ISBN-10: 1412916119, pp: 270-275.
- Almeida, J., 2002. Predictive non-linear modeling of complex data by artificial neural networks. *Curr. Opin. Biotechnol.*, 13: 72-76.  
DOI: 10.1016/S0958-1669(02)00288-4
- Altman, M., R. Davies, A. Mather, D. Fleming and D Harris, 2008. The impact of electricity price increases and rationing on the South African economy. Final Report to the National Electricity Response Team (NERT), HSRC, RSA.
- Amusa, H., K. Amusa and R. Mabugu, 2009. Aggregate demand for electricity in South Africa: An analysis using the bounds testing approach to cointegration. *Energy Policy*, 37: 4167-4175.  
DOI: 10.1016/j.enpol.2009.05.016
- Ash, T., 2011. Landing Page Optimization: The Definitive Guide to Testing and Tuning for Conversions. 1st Edn., John Wiley and Sons, New York, ISBN-10: 1118058933, pp: 384.
- Bianco, V., O. Manca and S. Nardini, 2009. Electricity consumption forecasting in Italy using linear regression models. *Energy*, 34: 1413-1421.  
DOI: 10.1016/j.energy.2009.06.034
- Blignaut, J.N., R.M. Mabugu and M.R. Chitiga-Mabugu, 2005. Constructing a greenhouse gas emissions inventory using energy balances: The case of South Africa: 1998. *J. Energy Southern Africa*, 16: 105-116.  
DOI: 10.17159/2413-3051/2005/v16i3a3111
- Bökeoğlu, Ç.Ö. and Ş. Büyüköztürk, 2008. Discriminant function analysis: Concept and application. *Eğitim Araştırmaları Dergisi*, 33: 73-92.
- De Lange, E., 2008. The impact of increased electricity prices on consumer demand. MSc Dissertation, University of Pretoria, South Africa.
- DME, 2003. Electricity basic services support tariff (free basic electricity) policy. Department of Minerals and Energy, Government Printers, Pretoria, South Africa.
- Ennett, C.M., M. Frize and C.R. Walker, 2001. Influence of missing values on artificial neural network performance. *Medinfo*, 10: 449-453. PMID: 11604780
- Frank, S.A., 2009. The common patterns of nature. *J. Evolut. Biol.*, 22: 1563-1585.  
DOI: 10.1111/j.1420-9101.2009.01775.x
- Fidell, B.G. and L.S. Tabachnick, 2007. Using Multivariate Statistics. 5th Edn., Pearson Boston, ISBN-10: 0205465250, pp: 980.
- Gładysz, B. and D. Kuchta, 2011. A method of variable selection for fuzzy regression: The possibility approach. *Operat. Res. Decis.*, 2: 5-15.
- Gneiting, T., F. Balabdaoui and A.E. Raftery, 2007. Probabilistic forecasts, calibration and sharpness. *J. Royal Stat. Society*, 69: 243-268.  
DOI: 10.1111/j.1467-9868.2007.00587.x
- Inglese-Lotz, R. and J.N. Blignaut, 2011. Estimating the price elasticity of demand for electricity by sector in South Africa. *South African J. Economic Manage. Sci.*, 14: 449-465. DOI: 10.4102/sajems.v14i4.134
- Jackson, J., 2015. SAS enlarges its palette for big data analysis. IT World, London.
- Janczura, J., 2014. Pricing electricity derivatives within a Markov regime-switching model: A risk premium approach. *Math. Meth. Operat. Res.*, 79: 1-30.  
DOI: 10.1007/s00186-013-0451-8
- Janczura, J., S. Trueck, R. Weron and R. Wolff, 2013. Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling. *Energy Econom.*, 38: 96-110.  
DOI: 10.1016/j.eneco.2013.03.013

- Janczura, J. and P. Weron, 2009. Regime-switching models for electricity spot prices: Introducing heteroskedastic base regime dynamics and shifted spike distributions. Proceedings of the 6th International Conference on the European Energy Market, May 27-29, IEEE Xplore Press, Leuven, Belgium, pp: 1-6.  
DOI: 10.1109/EEM.2009.5207175
- Janczura, J. and P. Weron, 2010. An empirical comparison of alternate regime-switching models for electricity spot prices. *Energy Econom.*, 32: 1059-1073. DOI: 10.1016/j.eneco.2010.05.008
- Janczura, J. and P. Weron, 2012. Efficient estimation of Markov regime-switching models: An application to electricity spot prices. *Adv. Stat. Anal.*, 96: 385-407. DOI: 10.1007/s10182-011-0181-2
- Janczura, J. and P. Weron, 2014. Inference for Markov-Regime Switching Models of Electricity Spot Prices. In: *Quantitative Energy Finance*, Benth, F.E., P. Laurence and V. Kholodnyi (Eds.), Springer, London, pp: 137-155.
- Jaynes, E., 2003. *Probability Theory: The Logic of Science*. 1st Edn., Cambridge University Press, London, ISBN-10: 0521592712, pp: 727.
- Khosravi, A., S. Nahavandi and D. Creighton, 2013. A neural network-GARCH-based method for construction of prediction intervals. *Electric Power Syst. Res.*, 96: 185-193.  
DOI: 10.1016/j.epsr.2012.11.007
- Lanne, M., H. Lütkepohl and K. Maciejowska, 2010. Structural vector autoregressions with Markov switching. *J. Economic Dynam. Control*, 34: 121-131. DOI: 10.1016/j.jedc.2009.08.002
- Lawrence, S.M., G. Glenn and A.J. Guarino, 2006. *Applied Multivariate Research: Design and Interpretation*. 1st Edn., Sage Publications, Thousand Oakes, ISBN-10: 1412904129, pp: 722.
- Leon-Garcia, A., 2008. *Probability, Statistics and Random Processes for Electrical Engineering*. 1st Edn., Pearson/Prentice Hall, London, ISBN-10: 0131471228, pp: 818.
- Magadla, T. and J. Holloway, 2011. The effect of load shedding on South African electricity consumption. A case study in ARIMA and Structural Time Series Modeling. CSIR Technical Report No CSIR/BE/LQM/IR/2011/0042/A, Pretoria.
- Netshiava, M.D., 2014. Statistical modelling of electricity demand and consumption in South Africa. MSc Dissertation, University of Limpopo (Medunsa Campus), South Africa.
- McLachlan, G., 2004. *Discriminant Analysis and Statistical Pattern Recognition*. 1st Edn., John Wiley and Sons, Hoboken, ISBN-10: 0471691151, pp: 526.
- Montgomery, D.C., E.A. Peck and G.G. Vining, 2006. *Introduction to Linear Regression*. 4th Edn., John Wiley and Sons Inc., New York, ISBN-10: 0471754951, pp: 640.
- O'Brien, J.A. and G.M. Marakas, 2007. *Introduction to Information Systems*. 12th Edn., McGraw-Hill/Irwin, New York, ISBN-10: 0072890428, pp: 452.
- Prinsloo, J.D., 2009. Long term electricity demand forecasting in Eskom (1987 to mid-2008, available as part of the proceedings of the First Annual Forecasting Conference 'Forecasting in an ever changing environment', held in Midrand on 17 & 18 November 2009.
- Sinton, J.E. and M.D. Levine, 1994. Changing energy intensity in Chinese industry: The relatively importance of structural shift and intensity change. *Energy Policy*, 22: 239-258.  
DOI: 10.1016/0301-4215(94)90162-7
- USI, 2015. IBM SPSS data collection divestiture. Unicom Systems, Inc., IBM, New York.
- Vanlier, J., C. Tiemann, P. Hilbers and N. van Riel, 2012. An integrated strategy for prediction uncertainty analysis. *Bioinformatics*, 28: 1130-1135. DOI: 10.1093/bioinformatics/bts088
- Venables, W.N. and B.D. Ripley, 2002. *Modern Applied Statistics with S*. 4th Edn., Springer Verlag, London, ISBN-10: 0387954570, pp: 495.
- Weron, F., 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.*, 30: 1030-1081.  
DOI: 10.1016/j.ijforecast.2014.08.008
- Weron, R. and A. Misiorek, 2008. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *Int. J. Forecast.*, 24: 744-763. DOI: 10.1016/j.ijforecast.2008.08.004
- Yamin, H.Y., S.M. Shahidepour and Z. Li, 2004. Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets. *Int. J. Electrical Power Energy Syst.*, 26: 571-581. DOI: 10.1016/j.ijepes.2004.04.005
- Zhang, L. and P.B. Luh, 2005. Neural network-based market clearing price prediction and confidence interval estimation with an improved extended Kalman filter method. *IEEE Trans. Power Syst.*, 20: 59-66. DOI: 10.1109/TPWRS.2004.840416
- Zhao, J.H., Z.Y. Dong, Z., Xu and K.P. Wong, 2008. A statistical approach for interval forecasting of the electricity price. *IEEE Trans. Power Syst.*, 23: 267-276. DOI: 10.1109/TPWRS.2008.919309
- Ziramba, E., 2008. The demand for residential electricity in South Africa. *Energy Policy*, 36: 3460-3466. DOI: 10.1016/j.enpol.2008.05.026