Forecasting Manufacturing sector electricity consumption in South Africa:

A Probabilistic Causality approach using bayesian neural networks

Lufuno Marwala
School of electrical and electronic engineering
University of Johannesburg
Johannesburg, South Africa
marwalal@gmail.com

Bhekisipho Twala
School of electrical and electronic engineering
University of Johannesburg
Johannesburg, South Africa
Btwala@uj.ac.za

Abstract— **The South African manufacturing industry is a significant consumer of electricity. Energy consumption in this sector increases with increased production levels and decreases with decreased production levels. It can be asserted, therefore that there is a probabilistic causal relationship between manufacturing production level and electricity demand. The aim of this work is to develop a mathematical model for this causal relationship to assist in forecasting the future energy demand in the manufacturing sector. Neural networks are used to build the causal model. It is assumed that in the causal relationship between these two variables the cause occurs before the effect. In line with this assumption, lagged values of the production index are used to build a neural networks model which is used for assessing its effect on the demand. The results show the causality neural network models created can be used to predict electricity demand with accuracy**.

Keywords: Causality; Manufacturing production index; Electrictity demand; neural networks, markov chain monte carlo

I. INTRODUCTION

Electricity is a commodity that drives all sectors of the South African economy and as such there are multiple causes of increase/decrease of electricity consumption. These sectors include manufacturing, household, agriculture and mining. Each of these causes is independent and therefore, each induces its own particular effect. Manufacturing in South Africa is a significant part of the economy estimated to be contributing 15.3% of the GDP second only to finance, real estate and business. The south African economy or the gross domestic product (GDP) grew at an average of 5% between 2003 and 2007 [1]. And during this period the percentage share of electricity consumed by the manufacturing sector increased significantly. Early in 2008, South African electricity supply experienced a shortage of supply and thus began a load shedding programme. It is no wonder that the study of the relationship between GDP and electricity demand has been of interest to scholars.

The end-use of electricity in South Africa is currently divided between domestic (17.2%), agriculture (2.6%), mining (15%), industrial (37.7%), commercial (12.6%), transport (2.6%) and general (12.3%) [2]. This breakdown clearly illustrates that the manufacturing industry is the largest consumer of electricity in South Africa. As already mentioned fluctuation of electricity demand has multiple causes (divided into sectors) and therefore, the effect of the causal information from the manufacturing production level is limited to electricity consumption of the manufacturing sector. Understanding the nature of the relationship between these two variables is critical for policy-makers during capacity planning and for electricity pricing. Many studies have been conducted that seek to understand the link between economic variables and electricity demand. Inglesi-Lotz studied the electricity price sensitivity of the industrial/manufacturing sector electricity consumption in South Africa "unpublished" [3]. Using the Kalman filter methodology for analysis and modelling, the study found that the price sensitivity has been changing over time. Odhiambo studied the causal relationship between electricity consumption and economic growth in South Africa [4]. The study determined that there is distinct bidirectional causality between electricity consumption and economic growth in South Africa.

More broadly, Kraft and Kraft published a seminal paper in 1978 arguing that there is evidence in favour of causality running from GNP to energy consumption in the United States, using data for the period 1947-1974 [5]. However, further studies in this area of study have yielded mixed results. Akarca and Long, Erol and Yu, Yu and Choi , and Yu and Hwang evaluated different data sets and found that there exists no causal relationships between GNP and energy consumption [6][7][8][9]. Asafu-Adjaye argues that one of the reasons for the disparate and often conflicting empirical findings on the relationship between energy consumption and economic growth lies in the variety of approaches and testing procedures employed in the analyses [10]. Early studies focused on using linear methods for modelling with the assumption of a stationary time series and later studies started introducing more complex methods that could incorporate the non-stationary time series data set. One such method is the causality procedures introduced by Granger [11].

The use of artificial intelligence techniques such as neural networks falls within the logic of introducing complex methods that are able to deal with the non-stationary data sets. Neural networks can be viewed as a general framework for representing non-linear mappings between multi-dimensional spaces in which the form of the mapping is governed by a number of adjustable parameters [12]. For this reason neural networks have the ability capture and model the dynamics of non-linear data and its complexities. This paper is based on the work that seeks to model a causal relationship between the manufacturing production level proxied by the manufacturing production index and electricity demand in this sector.

II. CAUSALITY AND FORECASTING

A. Causality

David Hume is one the famous philosophers who studied causality and he wrote that "We have no other notion of cause and effect, but that of certain objects, which have been always conjoined together, and which in all past instances have been found inseparable. We cannot penetrate into the reason of the conjunction. We only observe the thing itself, and always find that from the constant conjunction the objects acquire a union in the imagination [13]." And since Hume, the challenge is to use mathematical tools to model causal relationships.

Causality is used to define a relationship between two events such that one event called the cause is believed to have caused another event called the effect. In a causal relationship the cause is necessary for the occurrence of effect. If the likelihood of the cause is non-zero, and then the likelihood of occurrence the effect given that the cause has occurred is bigger than the likelihood of the effect occurring alone [14]. Selltiz, Wrightsman, and Cook have outlined three conditions for the existence of causality [15]:

1. There must be a concomitant co-variation between the cause and the effect.

2. There should be a temporal asymmetry or time ordering between the two observed sequences (the cause should happen before the effect).

3. The covariance between the cause and the effect should not disappear when the effects of any confounding variables are removed.

Covariation implies that there should be an association between the cause and the effect statistically called correlation. However, correlation does not imply causation.

Before quantum mechanics causality had largely dealt with deterministic variables or events. However, there was a realization that certain variables cannot be described using deterministic mathematical representation and therefore, probability was used for representing these types of events. This led to a formulation of an area of study called probabilistic causality. Causation should be differentiated from correlation. It is common to obtain between quantities varying with the time (such as time series) quite high correlations to which we cannot attach any physical significance whatever; although under the ordinary test the correlation would be held to be significant [16].

Wiener wrote that "For two simultaneously measured signals, if we can predict the first signal better by using the past information from the second one than by using the information without it, then we call the second signal causal to the first one [17]." This idea was further developed by Granger. Clive Granger describes two components of causality as follows [11]:

- The cause occurs before the effect; and
- The cause contains information about the effect that is unique, and is in no other variable.

From these two statements Granger concludes that causal variable can help forecast the effect variable. This is expressed in a linear regression model:

$$Y_t = a_0 + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \sum_{k=1}^{L} b_{2k} X_{t-k} + \varepsilon_t \qquad (1)$$

where ε_t is an uncorrelated random variable with zero mean and variance, L is the specified number of time lags, and $L = L + 1, \dots, N$. The null hypothesis that X_t does not Granger cause Y_t is supported when $b_{2k} = 0$ for $k = 1, \dots, N$ which reduces (1) to:

$$Y_t = a_0 + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \varepsilon_t \qquad (2)$$

It is within this context that this paper attempts to create a causal model between manufacturing production index and electricity consumption. There are many ways in which the production levels in the manufacturing sector can be increased or decreased. For example, constructing a factory that will consume an x amount of electricity a month will cause the overall electricity demand to increase by x amount. Similarly, shutting down manufacturing plant consumes x amount of electricity a month will cause the overall electricity demand to decrease by x amount. The consumption levels of electricity are also caused to fluctuate by the fluctuation of production levels of individual manufacturing plants.

Another scenario is one in which factories increase their production volume by using unused capacity which does require increased electricity usage. This case illustrates that it is not of necessity that in all cases where there is an increase in production volume there will be an increase in electricity usage. This is what makes the concept of probabilistic causation a relevant approach to this problem. The basic idea of this tradition is the cause increases the probability of the effect. This idea can be expressed formally using conditional probability:

$$P(E|M) > P(E) \qquad (3)$$

Which tells us that probability that *E* occurred given the occurrence of *M* is greater than the probability of *E* occurring without the occurrence of *M*. *M* may raise the probability of *E* even though instances of *M* are not invariably followed by instances of *E*.

The type of data considered in this study is time series data. The two time series under consideration are co-integrated as noted by Granger that when the idea of co-integration was developed, the mathematics made it clear immediately that if a pair of series was co-integrated then at least one of them must cause the other. The thesis for this paper is that fluctuation in manufacturing production index causes fluctuation in electricity demand:

$$\delta Y \rightarrow \delta X \qquad (4)$$

where δ represents change, *Y* is the production index and *X* is the electricity demand. Standard statistical analysis, typified by regression and other estimation techniques, is used to infer parameters of a distribution from samples drawn of that population so long as experimental conditions remain the same. Causal analysis deals with a different problem; its aim is to infer aspects of the data generation process which makes it possible to deduce not only the likelihood of events under static conditions, but also the dynamics of events under changing conditions [18].

Statistically, getting the right model to represent the facts of association forms the logic of the modelling exercise. These models vary in their degree of specificity (parametric, semi-parametric, nonparametric), in the degree of accuracy or approximation claimed for them, and in whether they deal with discrete or continuous variables [19]. To establish a causal relationship the time series pair under consideration in this study a forecasting model has to be created.

B. Forecasting

We say *Y* causes *X* if we are better able to predict *X* using all available information than if information apart from Y had been used according to Granger. Furthermore, he noted that if *Y* causes *X*, we can define an integer for lag *n* to be the least lagged value of *Y* such that any lagged value of *Y* beyond *n* will be of no use in improving the prediction of *X*. Therefore, the forecasting framework is defined as:

$$X_t = f(Y_{t-1}, \dots Y_{t-n}) \qquad (5)$$

where X_t is the demand in month *t* and is dependent variable and *f(.)* is a function with lagged monthly manufacturing production index as the independent variables of the model. . The modelling technique used in this work is the nonparametric technique called neural networks. Ancona et al have used radial basis approach to nonlinear Granger causality of time series [20]. Neural networks is regarded as black box, in that it is not easy to analyze the manner in which the neural network model is structured after training.

A more detailed explanation of neural networks is provided in the next section.

III. NEURAL NETWORKS

McCulloch and Pitts developed the idea of a very simple artificial neuron, a computational neuron, which can have multiple inputs and a single branching output and a threshold for firing, and with the inputs that are either inhibitory or excitatory [21]. Through this work they were able to prove that in principle a neural network made of these logical neurons could compute anything that could be computed [22]. It meant that the brain could be treated as a computer and therefore, the neuron could be treated as a sort of basic switching element in the computer. Brain studies have revealed that this is an over-simplification, however this formulation remains critical for the creation of artificial intelligence models.

The interconnection of neurons as defined by McCulloch and Pitts is called neural networks [22]. Neural networks can be viewed as a general framework for representing non-linear mappings between multidimensional spaces in which the form of mapping is governed by a number of adjustable parameters. These parameters are called weights within a neural network structure and they are adjusted using Bayesian learning algorithm. In mapping the input and the output the learning algorithm helps to find probable parameters given the input data.

The simplest form of neural networks is a two layer feed forward network called multilayer perceptron. This is expressed in form given in (4).

$$f(x_t, w, \omega) = g_2\left(\sum_{j=0}^{N} w_j g_1\left(\sum_{i=0}^{k} \omega_{ji} x_i\right)\right) \qquad (6)$$

Graphically it can be illustrated as in Fig 2. Neural networks uses statistical pattern recognition which is best described through Bayes' theorem as in (7) .
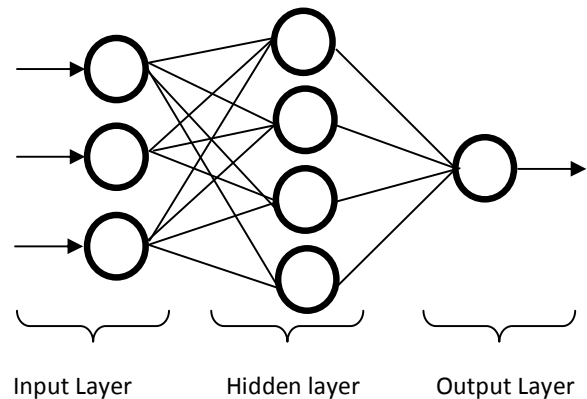

Figure 1: A graphical illustration of an MLP neural network

$$P(w|X) = \frac{P(w)P(X|w)}{P(X)} \qquad (7)$$

where $P(w)$ distribution of the weight vector before the data is presented and it is called a prior probability. $P(w|X)$ is a posterior probability and it corresponds to the probability distribution of the weight vector after observing the data. $P(X)$ is a normalising factor to ensure that:

$$\sum_n P(w|X) = 1 \qquad (8)$$

The posterior predictive distribution of output O_t for the new input x_t given the training data $D = \{(x_t, O_t), \dots, (x_{t+n}, O_{t+n})\}$, is obtained by integrating the predictions of the model with respect to the posterior distribution of the model,

$$P(O_t|x_t, D) = \int P(x_t|O_t, w)\, P(w|D)dw \qquad (9)$$

Markov Chain Monte Carlo (MCMC) was introduced Neal in the implementation of Bayesian learning for MLPs to solve the integral [23]. In MCMC the complex integrals in the marginalization are approximated via drawing samples from the joint probability distribution of all the model parameters. This is approximated using a sample of values drawn from the posterior distribution of parameters which is written as:

$$O_t = \frac{1}{N}\sum_t^N f(x_t, w_t) \qquad (10)$$

where N is the number of samples and $f(.)$ is a neural network function. The Metropolis algorithm is the commonly used method in MCMC approach, to generate samples from the posterior distribution of the weights. It is generally difficult to sample from a complex posterior distribution, and therefore the Metropolis method uses a simpler distribution to generate candidate weight vectors. An adaptive acceptance-rejection condition was used for the weight vector to converge to a posterior distribution over many iterations. The details of the implementation of the Metropolis method are not the focus of this paper and they can be found in Thyer, Kuczera, and Wang [24].

IV. EXPERIMENT

A. Data analysis

The reduction of the manufacturing production index can be caused by closure of manufacturing plants or by the reduction of the output which means shorter operating time for factories. In South Africa, manufacturing production index measures the total output of industrial/manufacturing sector of the economy. Its fluctuation reflects the performance of the manufacturing sector on month to month basis.

Electricity consumption grew by 106% between 1985 and 2007 in the manufacturing sector. From April 2008 the

manufacturing production index started to decline consistent with the economic recession as shown in Fig 2 and accordingly the electricity consumption declined.

The data used in this experiment was sampled on a monthly basis from January 1985 to December 2011 in South Africa. The demand for electricity has been on the increase over the years in South Africa as illustrated in Fig 3. There are trends in the data; these are seasonal trends and monthly trends. Seasonally the demand for electricity is the highest in winter (May through august) and lowest in summer (November through March). This type of seasonality illustrates the relationship between the electricity demand and the weather conditions in different seasons. In the month of February the demand reaches its lowest level every year and it peaks in the month of July. Seasonal changes can affect industrial activity. For example, in the refining industry, different seasonal slates of petroleum products as well as different seasonal processes may affect electricity needs.

The time series pair was tested for stationarity using the Dickey-fuller unit root test. The test involves testing for a unit root in this definition:

$$\Delta y_t = \delta y_{t-1} + \varepsilon_t \qquad (11)$$

where ε_t is a random error. If δ is a unit root which means it is equals to one, then the series of y_t is non-stationary. The test does not reject the hypothesis that both the time series have a unit root. This means that this pair of time series is non-stationary.
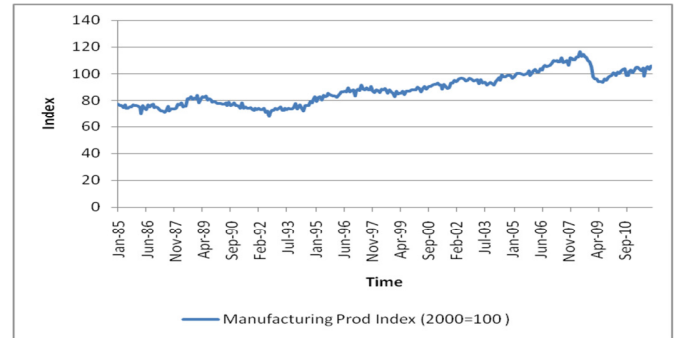

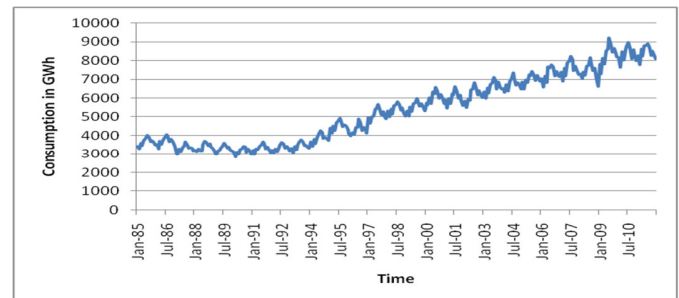**Figure 2: Monthly manufacturing production index**

Figure 3: Monthly electricity consumption in the manufacturing industry

An assessment of the statistical association between the time series was conducted using Pearson's correlation coefficient method. The Pearson's correlation coefficient can be defined as follows:

$$\rho = \frac{cov(x,y)}{\sqrt{var(x)var(y)}} = \frac{E[(x-\mu_x)(y-\mu_y)]}{[E(x-\mu_x)^2(y-\mu_y)^2]^{\frac{1}{2}}} \qquad (12)$$

where $cov(..,..)$ is the covariance, defined by the numerator of the right hand term, $var(...)$ is the variance; defined by the denominator; $E(...)$ is the mathematical expectation operator and μ_j is the mean of j ($j = x$ or y). The test shows that the two time series are highly correlated with $\rho = 0.925$. This means that there is a strong association between the two variables but it does not imply that there is a causal relationship between them. And also due to the non-stationary nature of the two data series, the correlation of the samples considered in this study cannot be considered as conclusions about the correlation in the populations.

B. Data preprocessing

All the data in the input series is normalized to fall between zero and one, which allows the different algorithm and the activation functions used to comprehend the data more intelligently and make valid deductions from the input series. The data retains its inherent characteristics.

The manufacturing production index data was partitioned into windows of sizes [n] from the time series data. The window is then evaluated from month [t - (n - 1)] to month [t]. This is making a hypothesis that the demand or load on any given month is an unknown function of the manufacturing production levels from the recent [n] months. In other words, the function takes in the window size of manufacturing production index values up to month n and outputs the predicted demand for month $n + 1$. As noted by Kaastra and Boyd, and by Kolarik and Rudorfer, that a popular method to use a is sliding window approach [26, 27]. To obtain n examples, we have to slide the window n steps, extracting an example at each step. Window sizes of 5,6,7,8,9,10,11, and 12 months were used for the extraction of the features that were used to train and test the models.

One hundred and fifty training examples were used for training and one hundred instances were used as out-of-sample data for testing the models. Various neural networks models were trained with the sigmoid activation function on the hidden layers neurons and the linear activation function on the output layer.

C. Experimental results

The Mean Absolute Percentage Error (MAPE) is an appropriate performance measure because of its closer relationship to decision making, is reliable and protects against outliers. Hippert, Pedreira, and Souza observed that, although the MAPE has become an industry standard, it is

advisable to also consider other error measurement methods [28]. Various accuracy measurement methods such as Mean Squared Relative error (MSRE) and Root Mean Squared Error (RMSE) were considered for this experiment to compare the differences in their accuracy measurement. Table 1 presents the forecasting errors as measured by the various error measurement methods. Five inputs (lag of 5) model has the most accurate results.

D. Discussion

The graph in Fig 4 shows that neural networks models trained by manufacturing production index lagged values give accurate forecasts of the electricity consumption in the manufacturing sector. The error of the forecasts increases as the lags increases beyond five. This is also illustrated in Table 1 below. This shows that the recent fluctuations of the production index are more relevant to the forecasting of the electricity consumption. Therefore, the lag of five is the least lagged value of production *index* such that any lagged value beyond five is of no use in improving the prediction of the consumption. The essence of this exercise was deal with the problems of how to reason in the presence of uncertainty and how to learn from experience.

Table 1: A table of prediction errors of the various error measurement techniques

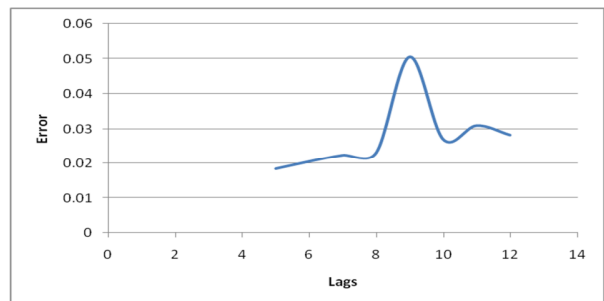| No of Inputs	MAPE	MSRE	RMSE
5	12.38%	0.0248	0.0185
6	12.8%	0.0273	0.0206
7	13.3%	0.0295	0.0225
8	13.6%	0.0305	0.0232
9	26.5%	0.0813	0.0560
10	15.0%	0.0366	0.0271
11	17.7%	0.0429	0.0310
12	16.8%	0.0390	0.0283

Figure 4: Root mean squared error

The historical data of the manufacturing production index was used to determine a causal link to the electricity

consumption in the manufacturing industry in South Africa. The distributions of the two variables co-vary which suggests that there is an association and a strong correlation between them. The results (such as a MAPE of 12.38%) show that there is still work to be done to optimize the model for improved prediction accuracy. In addition, further studies might require an investigation of other confounding variables that could be responsible for the less accurate results. The models constructed in this exercise can assist policy makers to make more informed decisions about the rate of industrialization and the capacity of electricity supply in South Africa. Equally, adequate investments to make in the power supply industry can be determined through these models so that a preferred rate of industrialization can be pursued.

V. CONCLUSION

In this paper neural network models were used to model the probabilistic causal relationship between the manufacturing production index and the consumption of electricity in the manufacturing sector. The results show that a neural network model with a lag a five months of the production index values gives more accurate forecasts that lags beyond five. However, further work is required to optimize the model and also to investigate if there are other confounding variables that could be affecting the accuracy of the results.

REFERENCES

[1] K. Ogujiuba, and N. Stiegler, "Policy Variables and Economic Growth in South Africa: Understanding the Nexus," Mediterranean Journal of Social sciences, ISSN 2039-9340, Vol. 3(11), 2012.

[2] Eskom Annual report 2011, www.eskom.co.za.

[3] R. Inglesi-Lotz, "The Sensitivity of the South African Industrial Sector's Electricity Consumption to Electricity Price Fluctuations," University of Pretoria Working Paper: 2012-25, 2012.

[4] N.M. Odhiambo, "Electricity consumption and economic growth in South Africa: A trivariate causality test," Energy Economics 31, 635–640, 2009.

[5] J. Kraft and A. Kraft, "On the relationship between energy and GNP," Journal of Energy Development 3, 401–403, 1978.

[6] A.T. Akarca and T.V. Long, "On the relationship between energy and GNP: a re-examination.", J. Energy Dev. 5, 326-331, 1980.

[7] U. Erol and E.S.H. Yu, "Time series analysis of the causal relationships between US energy and employment," Resources Energy 9, 75-89,1987.

[8] E.S.H. Yu and J.Y. Choi, "The causal relationship between energy and GNP: an international comparison," J. Energy Dev. 10, 249-272, 1985.

[9] E.S.H. Yu and B.K. Hwang, "The relationship between energy and GNP: further results," Energy Econ. 6, 186-190, 1984.

[10] J. Asafu-Adjaye, "The relationship between energy consumption, energy prices and economic growth: time series evidence from Asian developing countries," Energy Economics 22, 615–625, 2000.

[11] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," Econometrica, Vol. 37, No. 3., pp. 424-438. 1969.

[12] C. Bishop, "Neural Networks: A Pattern Recognition Perspective," Non-linearity and Complexity Research Group, Aston University, Birmingham, UK, 1996

[13] D. Hume (1740), A Treatise of Human Nature, Oxford University Press, Oxford, 1967.

[14] P. Suppes, "A Probabilistic Theory of Causality", Amsterdam: North-Holland, 1970.

[15] C. Selltiz, L.S. Wrightsman, and S.W. Cook, Research Methods in Social Relations, Holt, Rinehart and Winston, New York, 1959.

[16] G.U. Yule, 'Why Do We Sometimes Get Nonsense Correlations Between Time Series? A study of Sampling and the Nature of Time series (with discussion), Journal of the Royal Statistical Society, 89, pp. 1-64, 1926.

[17] N. Wiener, The theory of prediction. In E.F. Beckenbach, editor, Modern Mathematics for Engineers, McGraw-Hill, New York, 1956.

[18] P. Judea, "Causal Inference in Statistics: A Gentle Introduction," TECHNICAL REPORT, Computing Science and Statistics, Proceedings of Interface '01, Volume 33, 2001.

[19] K.D. Hoover, Nonstationary Time Series,Cointegration, and the Principle of common cause , Brit. J. Phil. Sci, 54: 527-551, 2003

[20] N. Ancona, D. Marinazzo and S Stramaglia, Radial basis function approach to nonlinear Granger causility of time series, Physical Review E 70, 056221, 2004

[21] D.C. Denett, "The Normal Well-Tempered Mind," A conversation with Daniel C. Denett, www.edge.org, 2013

[22] W.S. McCulloch and W.H. Pitts, "A Logical calculus of the ideas imminent in nervous activity," Bull. Math. Biophy. 5: 115-133, 1943.

[23] R.M. Neal, Bayesian Learning for Neural Networks, Lecture Notes in Statistics No 118, New York: Springer-Verlag, 1996.

[24] M. Thyer, G. Kuczera, and Q.J. Wang, "Quantifying parameter uncertainty in stochastic models using the Box-cox transformation," Journal of Hydrology 265:246-257, 2002.

[25] Energy Information Administration (EIA), "Homes show greatest seasonal variation in electricity use" March, 4, 2013. http://www.eia.gov/todayinenergy/detail.cfm?id=10211, last accessed on 15 May 2013.

[26] I. Kaastra and M. Boyd,. "Designing a neural network for forecasting financial and economic time series," Neurocomputing, vol. 10, pp. 215–236, 1996.

[27] T. Kolarik and G. Rudorfer, "Time series forecasting using neural networks, department of applied computer science," Vienna University of Economics and Business Administration, no. 1090, pp. 2–6, 1997.

[28] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short term load forecasting: A review and evaluation," IEEE Transactions on Power Systems, 16, 44– 55, 2001.