An adaptation of a prototype recursive model for evaluating and predicting micro-macro economic data in Botswana

Kassim S. Mwitondi[19]

Abstract

Naturally arising patterns in data are known to present potential sources of useful information at both micro and macro-economic levels. We carry out unsupervised and supervised modelling of Botswana's macro-economic data attributes obtained from disparate sources. Both techniques, commonly used to detect inherent patterns in data, have adjustable parameters which inevitably vary across applications. Thus, we propose a sequential unsupervised-supervised modelling approach in which Exploratory Data Analysis (EDA) is used to detect basic structures in data which are then passed on an algorithm based on the Expectation-Maximisation (EM) mechanics. The EM convergent values are then used to guide data labelling before applying the neural networks model. We demonstrate how future economic structures may be detected, monitored and managed by iteratively focusing on conditional checks of a generic algorithm. For the purposes of modelling robustness, we propose setting up an integrated data repository and source that would provide data-based guidelines to policy makers in addressing the country's economic issues while providing economic researchers access to data and/or information resources. Outstanding issues are identified and discussed and potential future directions are clearly highlighted.

Key words: *Data mining, data over-fitting, data recycling, EM algorithm, neural networks, knowledge extraction from data, supervised learning, unsupervised learning.*

1.0   Introduction

With a population of just under 2 million people over an area of approximately 581,730 square kilometres, Botswana has one of Africa's lowest population densities. Over the last few decades, the country has sustained an economic growth unprecedented on the African continent mainly due to the discovery of large diamond reserves in the wake of its independence in 1966. Its economic prosperity and political harmony have prompted many to refer to it as a true African success story. The country's economic achievements surprised many economists especially as they were brought about by conventional economic policies which were virtually politically engineered (Acemoglu *et al.*, 2002). Yet its rapid economic growth in recent years has been described as paradoxical as it has failed to appropriately address key issues such as unemployment and poverty (Hillbom, 2008). This phenomenon has been widely described as Botswana's economic duality.

The rationale for this study is two-fold. Firstly, the study is motivated by issues around measuring economic growth Snowdon and Stonehouse (2006). Like with any empirical data-dependent measurement investigators are primarily concerned about the accuracy and reliability of the results. Secondly, the study is motivated by Botswana's economic duality - a micro-macroeconomic issue. The over dependence on diamonds, is widely perceived as a sign of failing economic diversification policies, the strong correlation between the private sector and the government has been blamed for the rising economic and political elites (Good, 1999) while employment and poverty levels are still growing at an unacceptable rate. It is therefore reasonable to investigate the essence of naturally arising structures in the economy as a way of explaining the phenomenon as well as streamlining economic policies.

The paper's overall aim is to apply a combination of data mining methods in investigating the essence of Botswana's economic duality. The specific objectives of the study are as follows:

---

[19] Sheffield Hallam University; Computing and Communications Research Centre; E-mail: k.mwitondi@shu.ac.uk

1.   To critically evaluate the relevant theories and practices in existing literature. To carry out an empirical investigation into the micro and macro-economic attributes and assess them based on economic and data analysis theories and practices.

2.   To propose an accurate and reliable prototype data-driven model for guiding the setting of economic policies and predictions of their outcomes.

3.   To generate theoretical and empirical evidence (on the basis of objectives 1, 2 and 3) to guide recommendations for enhancing the country's economic policy directives.

Based on the four objectives, the paper seeks to answer the following research question: ***How can Botswana optimally harness potentially useful information inherent in its economic data?*** Answering the research question above requires a critical consideration of the relevant economic theory, an assessment of Botswana's past and current economic practices and addressing the relevant fundamental issues in data-driven decision making.

    This paper uses Botswana's micro and macro-economic data attributes as model inputs for transforming sample spaces into describable partitions and highlights avenues for using the emerging partitions as inputs in decision-making such as helping in devising new economic policies for the country. The paper is organised as follows. This introductory section provides the study background, study aims and objectives and defines the research question. It is followed by the background of Botswana's economy in Section 2.0 and an outline of the study methodology in Section 3.0. Data analyses and discussions of results are presented in Section 4.0 and concluding remarks, recommendations and potential future directions are in Section 5.0
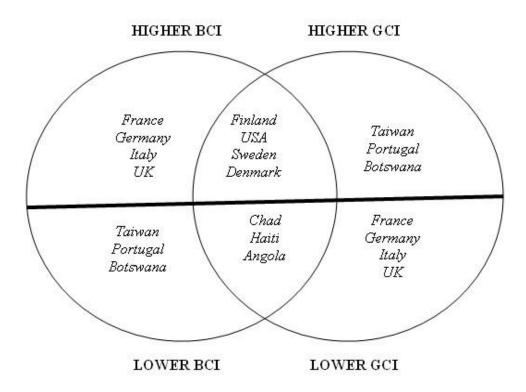
2.0   Background of Botswana☐s economy

In a study conducted over two decades ago, Curry (1987) concluded that poverty in Botswana was a direct consequence of the mining sector's great success which created an economic elite and political elite. According to the author, the elite had more favourable access to the revenues from mining prosperity than the rest of the population. Over a decade later, Good (1999) related the rise of wealth and power to the creation of poverty and weakness among the traditionally cattle-rearing population which is systematically marginalised by the economic and political elite. According to the author, despite the state possessing the financial resources and developmental capacities to alleviate poverty, the poor were marginalised by the state's prioritising controlling institutions. Hope (1996) investigated the country's growth, unemployment and poverty levels and established that only over a 7-year period - from 1984 to 1991 - the proportion of employment in the agricultural sector contracted by about 55% - from 33% of the labour total force in 1984 to 15% in 1991. This systematic decline in conventional sector within such a short time can only be interpreted in a binary scenario – either a sign of great prosperity or economic danger. Acemoglu *et al.,* (2002) attributed Botswana's high per capita growth rate to the presence of strong institutions of private property – particularly their relatively inclusive pre-colonial nature, constrained political influence and the political elite's post-colonial interest in sustaining and strengthening them. They argued that the country's large diamond reserves with its small-sized population motivated the political system to work towards economic greatness – which is probably what the two post-independence presidents Khama and Masire have achieved.

    Apparently, the four studies, carried out over a period of 25 years, present Botswana as a country divided along economic lines. Empirical evidence from many studies suggests that the diamond boom, over the last four decades, has had a dramatic impact on the country's socio-economic structure. Hillbom (2008) carried out a structural analysis of the country and concluded that despite the country being repeatedly depicted as a success story, what the country had actually achieved was a "pre-modern growth" - a concept used to describe an imbalance of resources and income amidst plenty - typical of countries benefiting from the possession of natural resources. Central to the analysis of economic growth is the prediction of economic policy for which, according to the well-known Lucas critique (Lucas, 1976), predicting the

effects of an economic policy experiment should be conditioned on key parameters such as preferences, technology and resource constraints.

In Kuznets (1973, pp 247), the Nobel Economics Laureate describes economic growth as a country's long-term capacity to supply "…diverse economic goods…based on advancing technology and the institutional and ideological adjustments…". The foregoing parameters play a pivotal role in individual behaviour and therefore form the basis upon which Botswana's recent growth ought to be evaluated. The critique of predicting the effects of a change in economic policy on the basis of aggregated historical data (Lucas, 1976) has led to the development of dynamic econometric models for modelling policy dynamics by researchers such as Evans and Ramey (2006). It is not the purpose of this paper to review the mechanics and drawbacks of econometric models since the Lucas critique, but suffice it to say that our inability to formalise economic model parameters inevitably implies that they will typically vary with the policy - from a machine learning perspective, this is known as concept drift[1]. Being able to track the margins of concept drift is potentially helpful in the prediction of individuals' behaviour which, given appropriate quantification, can be used to assess the impact of policy dynamics on micro and macro-economics.

Tracking, measuring and adapting national economic parameters are central to understanding the economy's structural dynamics (Mwitondi, 2009). The main challenge here is that the two key parameters of any predictive model - accuracy and reliability – are typically inversely related (Mwitondi and Ezepue, 2008). The inverse behaviour affects both predictive and diagnostic model parameters. For instance, Snowdon and Stonehouse (2006) interviewed Michael Porter (Porter, 2004) on the use of Business Competitiveness Index (BCI) and the Growth Competitive Index (GCI) in measuring business competitiveness. The former is based on the predictors of the GDP per capita and as such it may be used as a measure of the country's standard of living. Clearly, there will always be a wide range of predictors – ranging from science, technology and innovation levels to strategic, geographical, political and environmental. Apparently, some of the predictors will be strongly correlated and others may completely be missed out for many different reasons. The other index (GCI) is based on the economic dynamics and it provides an indication of whether or not an economy meets global competitiveness conditions. The key attributes here include investment, science, technology and innovation levels, the political climate and the general infrastructure. Porter (2004) believes that both indices are worth looking at despite being strongly correlated as research has shown that a country may be prosperous and lack growth and vice versa. He bases his argument on the global ranking report of 2003–2004 in which the two indices disagree on ranking a number of countries including Botswana. For clarity, we have created a Venn-like diagram of this disagreement as shown in Figure 1 in which the solid horizontal line demarcates a high from a low business competitive index by each of the two indices. The ranking disagreement applies to the seven countries (including Botswana) outside the intersection region. In theory, the intersection region representing a consensus between the indices would imply pure ranking while we may be concerned about the remaining sections of the diagram. Despite "higher" and "lower" not corresponding to distinctive categories the fact that one index "outperforms" the other implies that there are potential risks of generating and utilising unreliable economic statistics. Porter (2004) was aware of this and noted particularly that since a two-way table would contain countries in every cell, then merging the two indices would lead to loss of potentially useful information.
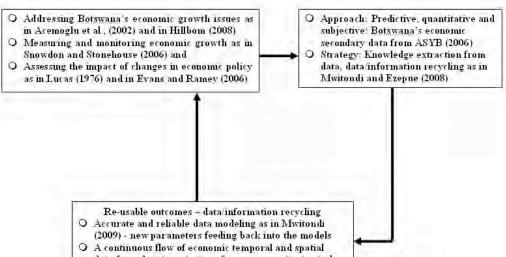
Figure 1: A Venn presentation of the global 2003-2004 BCI and GCI ranking of nations indicating a disagreement between the two indices in ranking the countries outside the intersection region

We believe that separating the two indices potentially leads to loss of information. Reliable results will be attained with less fragmentation and more coherence in handling data and/or information. In other words, both BCI and GCI should produce the same ranking. Our proposed study methodology seeks to attain robust results based on a near-continuous monitoring of the dynamics of economic data and related parameters as detailed below.

3.0   Study Methodology

The long-term socio-economic research across Botswana is a problem not amenable to laboratory experiments and so to coherently bridge up its findings we propose the theory mapping model (TMM) in Figure 2. Each of the rectangles is a target/source of data showing what is to be passed on (or received). Its main idea derives from the principle of systematicity which prefers connected knowledge to independent facts (Gentner, 1983 and Faye, 2006) and which we find quite appealing to our goal of achieving accurate and reliable results in data modelling. Analogous to the coupling and cohesion discussions in Section 2.0 each of the rectangles is cohesive as the change in one enhances a change in the others with the overall change having a more valuable contribution to the economic research system.

Figure 2: A theory mapping model for a systematic economic research in Botswana

Although the contents of the rectangles in Figure vary in time, they are preserved for future model training, validation and comparison purposes. The study seeks to extract knowledge from Botswana's historical data, carry out predictive modelling of the data and propose a continuous and integrated temporal and spatial data modelling strategy. It adopts a data mining approach to data exploration, detection of systematic patterns and rules and applying them to new data. Effectively, it amounts to sophisticated data analysis combining Exploratory Data Analysis (EDA) and classical and Bayesian statistics. While EDA seeks to infer what model to use given the data patterns and the absence of any prior information, the classical approach makes distributional assumptions about the data - such as assuming that the data come from a population with a particular mean and standard deviation. The model in Figure is Bayesian in nature as each rectangle is updated as new information arrives. Its implementation relies on parameters from the following posterior probability density function

$$P(\Omega|C_1, C_2, \dots, C_K) = \frac{P(C_k|\Omega)P(\Omega|C_1, C_2, \dots, C_{K-1})}{\int P(C_k|\Omega)P(\Omega|C_1, C_2, \dots, C_{K-1})d\Omega}$$

Equation 1

where Ω represents economic data and $C_{k=1,2,\dots,K}$ are the economic structures or groupings. Starting with prior probability of observing the fundamental economic parameters $P(\Omega)$ the recursive expression in Equation 1 is performed **K** times to yield the posterior probability. Apparently, one of its basic assumptions is that the K economic structures are independent.

3.1   Data Sources and selection of attributes

Table 1 provides a summary and description of the secondary data attributes obtained from sources such as the African Statistical Year Book (ASYB, 2006), the World Development Indicators (WDI), the Global Development Finance (GDF) and the United Nations database[1].

Table 1: Selected variables (Source: AFYB (2006), WDI, GDF and UN Data)

Var ID	Details	Metric	Variable Type
AGPM	Aggregated precious metal exports	(USD)	Numeric
AGMEO	Aggreg. meat and edible offal exports	(USD)	Numeric

AGDPB	Aggreg. dairy, poultry and bee exports	(USD)	Numeric
DI	Industrial Diamond	('000 carats)	Numeric
DG	Gem Diamonds	('000 carats)	Numeric
NI	Nickel	('000 tonnes^2)	Numeric
CU	Copper	('000 tonnes^2)	Numeric
LIV	Livestock (Cattle and Goats)	('000 count)	Numeric
MMP	Meat and Milk	('000 tonnes)	Numeric
CL	Class Labels[1]	Derived	Binary

Natural log transformation of most of data attributes exhibited at least near-normal behaviour as exemplified by the plot in Figure 1 which shows the aggregated value in US dollars of the export of precious metals and allied products (AGPM) over the period 2000 to 2008.
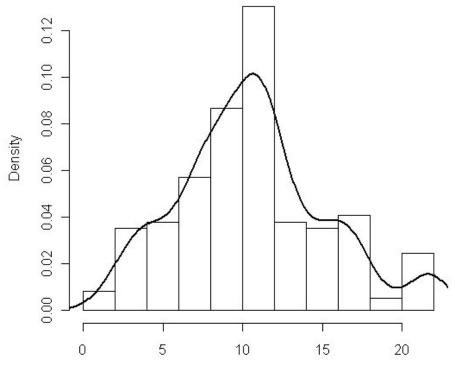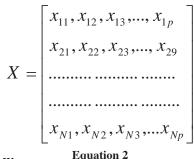

Figure 1: Histogram of the natural log transforms of aggregated precious metal exports (2000-08)

The data attributes in Table 1 can be expressed in an *N x p* matrix in Equation 2 which we can use for both unsupervised and supervised modelling by setting the attribute directions.

$$X = \begin{bmatrix} x_{11}, x_{12}, x_{13}, ..., x_{1p} \\ x_{21}, x_{22}, x_{23}, ..., x_{29} \\ .... \\ .... \\ x_{N1}, x_{N2}, x_{N3}, ... x_{Np} \end{bmatrix}$$

Equation 2

3.2   Methods of data modelling

We carry out a two-fold analysis - unsupervised economic data modelling to examine the validity of the duality claim and supervised modelling for allocating new cases to pre-defined

groups. The two complement each other in that the former helps to describe groups into which new cases can be allocated (Mwitondi, 2003). The general mechanics of the respective methods used in the study are briefly described below under the conditions of normality.

Let the economic structure consist of **K** initial groups each with a prior probability of occurrence π_k where $\sum_{k=1}^{K} \pi_k = 1$. Then if we define the matrix of data attributes by **X**, we can then adopt Webb (2005) and define the economy's finite mixture model distribution as

$$P(X) = \sum_{k=1}^{K} \pi_k p(X; \omega_k) \leftrightarrow P(X; \Sigma_k; \mu_k) = \frac{1}{\left(\sqrt{2\pi}\right)^p \sqrt{|\Sigma_k|}} e^{-\left\{-\frac{1}{2}(x-\pi_k)^T \Sigma_k^{-1}(x-\mu_k)\right\}}.$$

Equation 3

where $p(X, \omega_k)$ is a p-dimensional probability function described by the parameters in ω_k. For continuous data, the parameters are the mean and covariance matrices of the normal mixtures (McLachlan and Peele, 2000) which produce the implied *p*-dimensional probability function on the right. For a random binary *X,* the equivalent of Equation 3 is Equation 4.

$$P(X; \omega_i) = \prod_{j=1}^{p} \omega_{ij}^{Xj} \left(1 - \omega_{ij}\right)^{1-Xj}$$

Equation 4

Effectively, Equation 3 describes the probability distribution of X given the class to which the parameters apply. We can therefore use it to describe a related supervised scenario as follows. Let us denote by C_k a class associated with the parameter vector ω_k. We can then write the conditional probability of observing this class given data as follows.

$$P(C_k|X) = \frac{P(C_k)P(X|C_k)}{P(X)}$$

Equation 5

Taking logarithms to both sides of Equation 3 we can generate the allocation rule as follows

$$\log[P(C_k|X)] = \log[P(X|C_k)] + \log P(C_k) - \log P(X)$$
$$= -\frac{1}{2}(X-\mu_k)^T \Sigma_k^{-1}(X-\mu_k) - \frac{1}{2}\log|\Sigma_k| = \frac{p}{2}\log 2\pi + \log P(C_k) - \log P(X)$$

Equation 6

Since $P(X)$ does not depend on C_k the classification rule will be to allocate X to C_k when $R_i > R_j \; \forall_{i \neq j}$ where $R_i = \log P(C_k) - \frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(X-\mu_k)^T \Sigma_k^{-1}(X-\mu_k)$ (Webb, 2005).

3.3   Data-based parameter and density estimation

Identifying and keeping track of the key parameters characterising patterns of any given phenomenon is central to modelling it. Let $i = 1, 2, \ldots, N$ observations correspond to $C_i = \{1, 2, \ldots, K\}$ initial classes and an unobservable indicator variable of class labels

$$z_i = \{0, 1\}^K = (z_{i1}, z_{i2}, \ldots, z_{iK}) \text{ such that } z_{ik} = \begin{cases} 1 \text{ if } C_i = k \\ 0 \text{ if } C_i \neq k \end{cases}$$

Equation 7

Then, the probability of an observation belonging to a class given data is given as follows

$$P(C_i = k|X_i) = P(z_{ik} = 1|X_i) = \frac{\pi_k f_k(X_i)}{\sum_{k=1}^{K} \pi_k f_k(X_i)} = \hat{\pi}_k$$

Equation 8

If the parameters and densities were observable, we would compute them as follows

$$\hat{\pi}_k = \frac{\sum_{i=1}^{N} z_{ik}}{N}; \quad \hat{\mu}_k = \frac{\sum_{i=1}^{N} z_{ik} X_i}{\sum_{i=1}^{N} z_{ik}}; \quad \hat{\sigma}_k^2 = \frac{\sum_{i=1}^{N} z_{ik}(X_i - \hat{\mu}_k)^2}{\sum_{i=1}^{N} z_{ik}}$$

Equation 9

In practice those key parameters are not observable and will have to be estimated from data. We use an adapted version of the EM algorithm (Dempster, et al., 1977) to estimate the parameters. The algorithm involves two steps - at the E-step it estimates the proportion as

$$\hat{\pi}_{ik}^{(m+1)} = E\left[z_{ik} = 1 | X_i, \quad \hat{\mu}_k^{(m)}, \hat{\sigma}_k^{(m)}, \quad \hat{\pi}_k^{(m)}\right] = \frac{f_k\left[X_i, \hat{\mu}_k^{(m)}, \hat{\sigma}_k^{2(m)}, \hat{\pi}_k^{(m)}\right]}{\sum_{k=1}^{K} f_k\left[X_i, \hat{\mu}_k^{(m)}, \hat{\sigma}_k^{2(m)}, \hat{\pi}_k^{(m)}\right]}$$
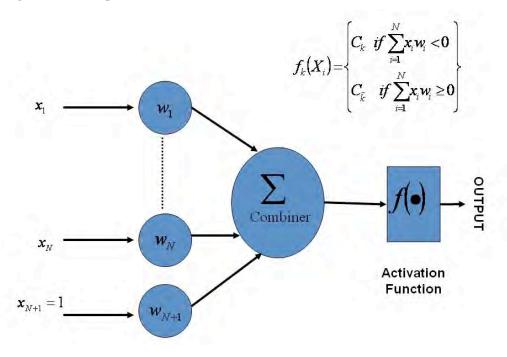
Equation 10

while at the M-step it maximises the two parameters as follows

$$\hat{\mu}_k^{(m+1)} = \frac{\sum_{i=1}^{N} \hat{\pi}_{ik}^{(m+1)} X_i}{\sum_{i=1}^{N} \hat{\pi}_{ik}^{(m+1)}} \; ; \quad \hat{\sigma}_k^{2(m+1)} = \frac{\sum_{i=1}^{N} \hat{\pi}_{ik}^{(m+1)} \left(X_i - \hat{\mu}_k^{(m+1)}\right)^2}{\sum_{i=1}^{N} \hat{\pi}_{ik}^{(m+1)}}$$

Equation 11

In both unsupervised and supervised cases, varying the value of **K** yields different outcomes which can guide policy makers provided the models used are reliable. Finally, the foregoing mechanics are amenable to implementation in a wide range of predictive models including logistic regression, decision trees, support vector machines and other models. One of our implementation is based on the recursive neural networks model the simplest form of which is illustrated in Figure showing how the data and weights are combined to yield output.

Figure 4: The simplest form of a neutral networks model

$$f_k(X_i) = \left\{ \begin{array}{l} C_k \; if \; \sum_{i=1}^{N} x_i w_i < 0 \\ \\ C_{\bar{k}} \; if \; \sum_{i=1}^{N} x_i w_i \geq 0 \end{array} \right\}$$

In its simplest form the neural networks model in Figure is similar to multiple regression modelling. However, the power of neural networks derives from their ability to recursively propagate the prediction error back into the modelling process for prediction enhancement. By adding in one or more layers of nodes between the weights layer and the combiner a multilayer neural networks model is formed and by allowing it to run recursively a back propagation model (Ripley, 1996) is formed. Its mechanics can be summarised as follows.

Let a neural networks architecture consist of n input neurons, a hidden layer with h input neurons and an output with p neurons. Then the weights $w_{ik} = (1, 2, \dots, h, -1, h)$ link the input layer nodes to the hidden layer nodes while the weights $\lambda_{kj} = (1, 2, \dots, p-1, p)$ link the hidden layer nodes to the output layer nodes. Then if we denote the output from the hidden layer nodes by $\beta_k = f(X, w_k)$ and recognising that because the output layer receives inputs which are outputs from the hidden layer weighted by λ_{kj} its output can be defined as

$$Y_j = F(h, \lambda_j) = F\left(\sum_{k=1}^{h} \beta_k \lambda_{kj}\right) = F\left(\sum_{k=1}^{h} \lambda_{kj} f\left(\sum_{i=1}^{n} X_i w_{ik}\right)\right)$$

<div align="center">Equation 12</div>

By working forth and back adjusting the case weights, the neural networks model in Equation 12 seeks to minimise the prediction error for which the initial weights are crucial. Hence, model complexity may be addressed through tracking and monitoring the weights, architecture and the outcomes from across applications of similar nature.
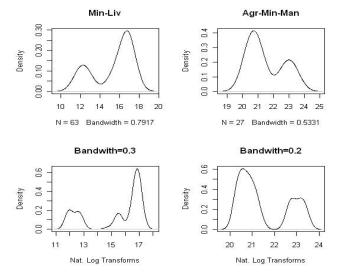
4.0   Analyses and results

The analyses are mainly two-fold – those focusing on detecting naturally arising structures in data and those focusing on predictive modelling of the data defined in Equation 2. The former is initially Exploratory Data Analysis (EDA) from which gained insights are used in carrying out parameter estimations and the later is based on Equation 5 through Equation 12.

4.1   Exploratory Data Analysis

Let us start by looking at two one-dimensional data vectors one for mining and animal husbandry activities in Botswana covering the period 1997 through 2005 and the other for the total agricultural, mining and manufacturing output. Mining output for diamond (both industrial and gem), nickel, copper and cobalt averaged 16,750,466 metric tonnes while the data for animal husbandry – aggregated over the number of animals and meat and milk products - averaged 2,615,666. Their respective natural log transform values were 16.00015602 and 13.67245721. The total sectoral output values for agriculture, mining and manufacturing averaged BWP 783,444,444; 10,281,333,333 and 1,345,666,667 respectively with corresponding natural log transforms 20.4624087, 23.01156508 and 20.9938726.

Figure 5: Initial insights into the underlying structural construct of Botswana's economy
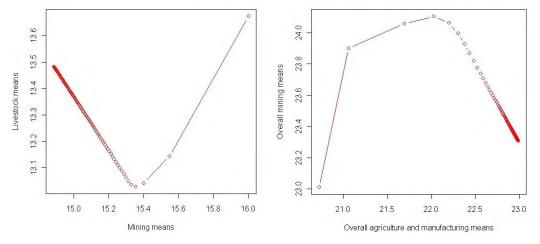
The kernel densities for the two one dimensional vectors are presented in figure 5 based on their natural log transform values. The panels on the left and right hand side columns correspond to the former and latter vectors respectively and are shown at different bandwidths. In the top two panels the kernels are scaled to allow for the bandwidth used to correspond with the standard deviation of the smoothing kernel (Venables and Ripley, 2002) while the bottom two exhibit the same data at much lower bandwidths. While both plots exhibit clearly separable modes centred on the group means above, varying the bandwidth allows the subtle inter-sectoral modes to be detected. The mode centres are the group means which we need to focus on in line with the parameter estimations using the EM algorithm.

4.2   Estimating economic parameters using the EM algorithm

The plots in Figure 5 show that the central tendency of the economy can be characterised by several parameters depending on the bandwidth. In simple terms this means that the variable nature of data sources may lead to different data-based inferences - more plainly, economic policies. We use the EM algorithm to estimate the means in assuming constant variance. Both McLachlan and Krishnan (1996) and McLachlan and Peele (2000) provide comprehensive details of the convergence features of the algorithm. The application of the EM algorithm in this section will only highlight its observed general convergence features and their relevance to Botswana's economic policy decisions. Note that in this case both the mean and common variance estimates are maximized to reveal discernible economic structures.

Figure 6: The EM algorithm converging means for the mining and non-mining sectors

The left hand side panel in Figure 6 are based on a small size univariate data set containing the output of diamond (industrial and gem) and livestock and meat-related products over the period 1997 through 2005. The right hand side panel is based on the overall mining output of diamond, copper, nickel and cobalt as well as the overall agricultural and manufacturing output. The parameter estimates correspond to the estimated densities in Figure 5. The corresponding values the algorithm generated after 100 iterations are presented in Table 2 in which the converging behavior of the algorithm in this case can clearly be seen.

Table 2: Estimated parameters after 100 iterations

Iteration	Mining means	Livestock means	Pooled variance
0	16.07762	13.67245	4.058130
1	16.00016	13.67246	4.058130
2	15.54824	13.14468	4.339844
3	15.54824	13.14468	4.339844
5	15.39953	13.04230	4.370131
6	15.35176	13.02852	4.408635
....	....	....	....
....	....	....	....
96	14.89460	13.47405	5.253469
97	14.89270	13.47597	5.256177
98	14.89082	13.47786	5.258848
99	14.88895	13.47973	5.261484
100	14.88711	13.48158	5.264085

Iteration	Overall agriculture and manufacturing	Overall Mining	Pooled variance
0	20.72814	23.01157	1.311045
1	20.72814	23.01157	1.311046
2	21.05753	23.89996	2.435497
3	21.69210	24.05884	2.682178
5	22.02519	24.10365	2.935643
6	22.19855	24.06329	3.139776
	....	....	
	....	....	
96	22.98237	23.31158	3.981740
97	22.98329	23.31066	3.982042
98	22.98420	23.30975	3.982339
99	22.98509	23.30886	3.982628
100	22.98683	23.30711	3.983189

Note that the final estimated parameter points do not necessarily have to be local maxima as it is possible for the algorithm to converge to local minima and so the algorithm can be used to search for minimum parameters for attributes such as unemployment or inflation.

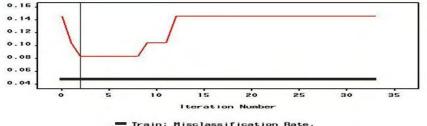### 4.3   Predictive modelling using neural networks

Without loss of generality let us assume that the data in Equation 2 define the key structures of Botswana's economy such as mining, manufacturing, agriculture, animal husbandry and tourism and that the vector X_{i1} represents the mining sector. Typically, an unsupervised modelling will use $q \leq p$ vectors as model inputs while supervised modelling will use $p - r \geq 1$ inputs and at least one output vector. For supervised modelling we need labelled data and so we propose the following algorithm for labelling the structure matrix X_{ij}.
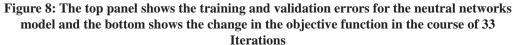
```
Access Data Xij where i=1, 2, 3, …,N-1,N and j=1=1, 2, 3, …, p-1, p

CL=CL( )# Initialise the vector of class labels

VM=VM( )# Initialise the vector of attribute means

        While j <= Vector Length (Xi) Do

                Append the vector of attribute means VM

                        For i = 1: Vector Length (Xi) Do

                                If Xi1 >=VM[1] AND/OR/XOR X[j+1] <=VM[j+1] Class1

                                        Else Class2

                                Append the vector of class labels CL with class output

                                End If

                        End For

        End While

Generate an updated Xij with CL such that i=1, 2, 3, …,N-1,N and j=1=1, 2, 3, …, p, p+1
```

The conditional test $If\ X_{i1} \ge VM[1]\, AND/OR/XOR\, X_{j+1} \le VM[j+1]\, Class1$ is typically conditional on economic policies and may be implemented in different combinations. We implement the above algorithm using the neural networks model as graphically illustrated in Figure 4 and defined in Equation 12. For simplicity, we carry out conditional checks on the mining and non-mining related variables AGPM, AGMEO and AGDPB in Table 1. Labelling the data derives from Equation 7 and it is implemented 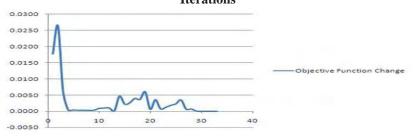by the above algorithm - that is, to label a data item "M" for mining if the corresponding mining entry is greater than the variable median and if the corresponding entry in one of the non-mining attributes AGMEO and AGDPB is lower than its median or label it "D" for diversified otherwise.

Figure 7: An iterative algorithm for labelling the economic data matrix

Figure 8: The top panel shows the training and validation errors for the neutral networks model and the bottom shows the change in the objective function in the course of 33 Iterations

The plots in Figure 8 show the output obtained by predicting the country's economic structure based on the given data. The final training, validation and test errors were 4.8%, 8.3% and 14.2% respectively. Without dwelling on the architectural setup of the network, we only highlight that the plot in Figure 9 was obtained after adding a hidden layer to the network with three neurons. Apparently, the model badly over-fits the data after only 10 iterations while the training error continues to slide downwards. Data over-fitting arises when the model adapts itself to the data so well that it considers even small random variations as meaningful information in modelling the data. It is an issue of great concern because it impinges on model reliability as the same model applied to new sets of data is most likely to yield completely different outcomes. As predictive models in any economy are data-dependent, such issues will arise as long as there is no unified approach to modelling.
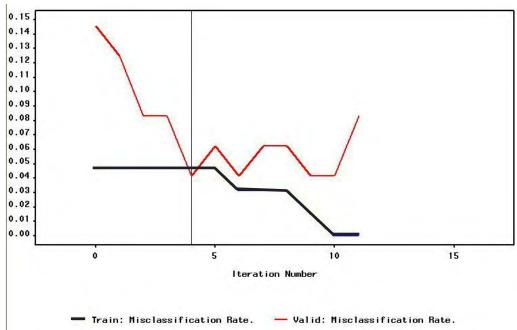
Figure 9: An added hidden layer minimises the trianing error but causes the model to over-fit the data

Similar to the scenario in Figure 1, data analysts use various methods of model selection but there is sufficient empirical evidence to show that within the same research domain different methods may be applied - a practice which potentially encourages ambiguity.

5.0   Concluding remarks, discussions and potential future directions

The paper's motivation was two-fold – theoretical, focusing on the overall issue of modelling economic data and an empirical assessment of Botswana's economic duality. The motivation provided a natural appeal for data-driven investigation into the country's economic structures. This section focuses on the two motivating aspects as well as on potential new directions.

5.1   Key implications of the theoretical and empirical findings

Apparently, the overall task of modelling economic data is strongly attributed to model accuracy, reliability and complexity. The Venn diagram in figure 1 was used to highlight potential inconsistencies in using economic predictors in different settings. To answer the stated research question, we needed to view the Venn diagram from the perspective of the two familiar concepts

in information theory – coupling and cohesion. The former describes the between module interdependency while the latter describes the within module strength with a highly cohesive module possessing desirable features such as robustness and reliability (Press and Constantine, 1979). Thus, BCI and GCI would be coupled if the change in one is conditional on the change in the other and cohesive if the change in one enhances a change in the other with the overall change having a more valuable contribution to the system. The disagreement between the two indices in ranking the countries outside the intersection region implied a potential risk in drawing conclusions from data. The analogy in data mining is that of data clustering in which overlapping (heterogeneous) clusters will usually be hard to describe and while well-separated (homogeneous) clusters may be easy to describe, their usefulness will depend on their interrelationships. Given the strong correlation between the indices it would be reasonable to pool together their known attributes and set off working towards detecting and explaining naturally arising patterns from the data. Thus, our proposed study methodology in Section 3.0 sought to attain robust results based on a near-continuous monitoring of the dynamics of economic data and related parameters.

The EDA results in Figure 5 provided some good insights into Botswana's underlying economic structures. Further, guided by the EDA results the EM algorithm was used to detect simple structures such as group means and variability. Estimating group parameters in this way provides a way for identifying inherent structures in economic data. If the specific sectors, timings and the estimated parameters are properly documented they may be used as inputs in subsequent years. For instance, old parameters may be passed on as starting points for the EM algorithm in which case the new convergence points may be used to assess the structural dynamics of the economy. The neural networks model was used to predict outcomes of attribute relationships effectively involving knowledge representation through learning rules from Botswana's economic data. Learning a model of the whole economy (environment) and maintaining it sufficiently consistent across different applications in the real world (reliability) is difficult. Forms of knowledge representation from the inputs to the internal network parameters are highly diverse making the development of a reliable solution using the model a real design challenge. Both models have adjustable parameters which inevitably vary across applications and yet crucial economic policies typically depend on outcomes influenced by them. It is therefore imperative that the data generating the parameters as well as the models associated with them are carefully monitored in time.

Across applications the key modelling challenge is how to strike a balance between accuracy and reliability which effectively amounts to addressing model complexity issues for which there are no clear-cut rules governing the choice of implementation methods. However, as noted above, most of these issues can only be addressed through long-term recursive modelling, tracking, monitoring and recycling of economic data, parameters and concept drifts as in Figure 2. Apparently, Botswana economic data analysts will have access to different tools and methods for modelling different datasets at different times. The conditional test in Figure 7 may take in the foregoing parameters as inputs and because the conditional test is itself typically conditional on economic policies, model accuracy and reliability results should be used as tools for persuading policy makers to adapt economic policies to existing conditions. Hence, the need to make use of data archives and monitor model parameters across applications as we seek to build robust models.

The study sought to develop an information sharing platform for economic researchers to use as a basis for drawing guidelines for implementing the country's macro and micro-economic policies. According to Hillbom (2008) Botswana's macroeconomic management policies seem to have focused mainly on mining and government investment. Diversification of the economy through the six hubs – education, innovation, agriculture, diamond, medical and transport (Faye et al., 2009) will be crucial as it is expected that in the years to come, government expenditure – particularly investments into power generation schemes - will increasingly be financed via

commercial loans. As diversification gathers momentum, economic analysts should keep track of the changing structures in data and use the emerging parameters to build appropriate performance indices iteratively as recommended in Figure 2.

5.2   Addressing outstanding issues and potential future directions

Figure 3 pools together operational data, technical tools, techniques, human skills to form an economic research life-cycle amenable to both observable and unobservable flow of data. It describes the link between unsupervised and supervised modelling carried out in this study as an input-output interface for realising versatile, accurate and reliable models for measuring the country's socio-economic aspects. However, despite the emerging patterns from data analytics forming fundamental bases for decision making processes in almost all disciplines, they are typically defined on a finite scope with respect to time, concept, data and location. So, like many others, the study was constrained by both the data and the models used. For instance, both unsupervised and supervised modelling techniques are associated with a number of issues - in both cases the sets of parameters in   will typically vary across samples. Different datasets and different model parameters are almost guaranteed to yield different outcomes. For instance, two neural networks models with different settings applied to the same data set are likely to yield different results reiterating the issue of model reliability (Mwitondi and Ezepue, 2006; 2008). Such issues which arise because of the typical attribution of the overall empirical error randomness due to the allocation region and randomness due to the assessment of the rule by random training and validation data can only be addressed through data archiving (Mwitondi, 2003). Data archiving and recycling provides a diagnostic environment in which researchers can look back and forth as they build and evaluate enhanced predictive models (Mwitondi and Ezepue, 2008).

Finally, the study is concluded on the basis of its fourth objective as stated in Section 1.0. In particular, it is expected that the theoretical and empirical evidence generated above will provide some useful insights into enhancing Botswana's economic policy directives. More generally, it is expected that our study will contribute to the existing knowledge base which in turn will be employed in interpreting the primary and secondary data and provide a critical evaluation of the theory upon which the analyses in this study were made. More specifically, to develop a unified approach, as proposed in this study, more work needs to be done on a wide range of models on much more comprehensive datasets than used in this study.

References

Acemoglu, D, Johnson, S. and Robinson, J. (2002). An African Success Story: Botswana; CEPR Discussion Paper No. 3219; Available at SSRN: http://ssrn.com/abstract=304100

ASYB (2006). African Statistical Yearbook; United Nations Economic Commission for Africa (UNECA); United Nations Publications.

Curry, R. L. (1987). Poverty and Mass Unemployment in Mineral-rich Botswana; American Journal of Economics and Sociology, Vol. 46, No. 1., pages

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm; Journal of the Royal Statistical Society, Vol. 39, pages 1-38.

Evans, G. W. and Ramey, G. (2006). Adaptive expectations, underparameterization and the Lucas critique; Journal of Monetary Economics, Vol. 53, Issue 2, pages 249-264, Elsevier.

Faye, J. (2006). Principles and Quantum Revolutions; Journal of Metascience, Vol. 15, No. 3, ISSN 0815-0796, Springer Netherlands.

Faye, I., Olanrewaju, S., Kanga, M. Negatu, G. and Beileh, A. (2009). Economic Diversification Support Loan Country: Botswana Appraisal Report; African Development Bank.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy; Journal of Cognitive Science, 7, 1X5-170, Bolt Beranek and Newman Inc.

Good, K. (1999). The state and extreme poverty in Botswana: the San and destitutes; The Journal of Modern African Studies, Vol. 37, pages 185-205; Cambridge University Press.

Hillbom, E. (2008). Diamonds or development? A structural assessment of Botswana's forty years of success; The Journal of Modern African Studies Vol. 46, pages 191-214; Cambridge University Press.

Hope, K. R. (1996). Growth, unemployment and poverty in Botswana; Journal of Contemporary African Studies, 1469-9397, Vol. 14, Issue 1, pages 53 – 67.

Kuznets, S. (1973). Modern Economic Growth: Findings and Reflections; The American Economic Review, Vol. 63, No. 3, pp 247-258, American Economic Association.

Lucas, R. E. (1976). Econometric policy evaluation: A critique; Carnegie-Rochester Conference Series on Public Policy; Vol. 1, pages 19-46, Elsevier.

McLachlan, G. Krishnan, T. (1996). The EM Algorithm and Extensions; John Wiley.

McLachlan, G. and Peel, D. (2000). Finite Mixture Models; John Wiley.

Mwitondi, K. S. and Ezepue, P. O. (2008). How to appropriately manage mathematical model parameters for accuracy and reliability: A case of monitoring levels of particulate emissions in ecological systems; International Conference on Mathematical Modelling of Some Global Challenging Problems in the 21st Century; Proceedings of NMC-COMSATS Conference on Mathematics Modeling of Global Challenging Problems - 26th-30th, Nov. 2008; pp 24-36, ISBN 978-8141-11-0.

Mwitondi, K. S. and Ezepue, P. O. (2006). Searching for Clusters in Data - Implications for Teaching Business Intelligence; In the Proceedings of the 5th Annual Hawaii International Conference on Statistics, Mathematics and Related Fields; January 16-18, 2006.

Mwitondi, K. (2003). Robust Methods in Data Mining; PhD Thesis, School of Mathematics, University of Leeds; Leeds University Press.

Porter, M.E. (2004). Building the Microeconomic Foundations of Prosperity: Findings from the Business Competitiveness Index; In Porter, M. E., Schwab, K., Sala-i-Martin, X. and Lopez-Claros, A. (2004-2004); The Global Competitiveness Report; Palgrave Macmillan, World Economic Forum: New York, pages 19–50.

Press, Y. and Constantine, L. (1979). Structured Design: Fundamentals of a Discipline of Computer Program and Systems Design; Prentice Hall, PTR.

Snowdon, B. and Stonehouse, G. (2006). Competitiveness in a globalised world: Michael Porter on the microeconomic foundations of the competitiveness of nations, regions, and firms; Journal of International Business Studies, Vol. 37, pages 163–175; Academy of International Business.

Venables, W. N. and Ripley, B. D. (2002). Modern Applied Statistics with S-Plus, Springer.

Webb, A. (2005). Statistical Pattern Recognition; Second Edition, Wiley, ISBN 0470845147.