INVEST: Ontology Driven Bayesian Networks for Investment Decision Making on the JSE

Rachel Drake[1] and Deshen Moodley[1]

[1] University of Cape Town, Cape Town, South Africa
racheldrk@gmail.com
Deshen.moodley@uct.ac.za

Abstract. This research proposes an architecture and prototype implementation of a knowledge-based system for automating share evaluation and investment decision making on the Johannesburg Stock Exchange (JSE). The knowledge acquired from an analysis of the investment domain for a value investing approach is represented in an ontology. A Bayesian network, developed using the ontology, is used to capture the complex causal relations between different factors that influence the quality and value of individual shares. The system was found to adequately represent the decision-making process of investment professionals and provided superior returns to selected benchmark JSE indices from 2012 to 2018.

Keywords: Ontology · Bayesian Network · Portfolio Management · Share Evaluation · Value Investing · Fundamental Analysis

1 Introduction

Stock market fluctuations are usually the result of complex phenomena, which are often erratic and difficult to predict. However, despite this complexity many well-known factors including fundamental financial indicators of a company, market analysis, and macroeconomic variables have shown to have a high degree of influence on market movements and can provide a certain level of forecast capability in the stock market [1]. Investment professionals typically engage in complex analysis and modelling of these factors to understand and reduce the level of uncertainty in their forecasting. The process of share evaluation involves the analysis and identification of individual shares with optimal risk-return characteristics to hold in a share portfolio [2]. There are different investment approaches and perspectives for investing which makes acquiring and representing expert knowledge for share evaluation challenging. Current decision models often do not adequately reflect the real investment decision making process used by the broader investment community or may not be well-grounded in established investment theory.

Semantic technologies, such as ontologies have been widely used for acquiring and representing domain knowledge in a graphical form [6]. However, standard ontology languages like OWL, do not have explicit and intrinsic support for representing uncertainty, causal relations and the processes involved in decision making [6][13][17].

Bayesian decision networks are simple models for decision making using expected utility theory [14]. The combination of ontologies and Bayesian Networks have been shown to be effective for knowledge acquisition, knowledge representation and automated decision making in diverse domains [13][14][15].

This research proposes an architecture and prototype implementation of a knowledge-based system for automating share evaluation and investment decision making on the Johannesburg Stock Exchange (JSE). The primary contribution of this research is the analysis and formal representation of expert knowledge and process used for share evaluation from a value investing perspective. The knowledge acquired from an analysis of the literature and iterative engagement with domain experts is represented in an ontology. A Bayesian network, developed using the ontology[1], is used to capture the complex causal relations between the key factors that influence the quality and value of individual shares.

2 Background and Related Work

This section provides a background to share evaluation following the value investing approach and describes related work on decision support systems that incorporate ontologies and Bayesian networks.

2.1 Share evaluation using value investing

Share evaluation is used to identify individual shares with certain risk-return characteristics for inclusion in a share portfolio [2]. There are several approaches for share evaluation. The technical approach focuses on analysing the historical price movement of a share. Fundamental approaches, like value investing, evaluate various factors pertaining to shares beyond price such as profitability, quality of management and growth [3]. Value investing is premised on the idea that undervalued shares will deliver an investment return greater than the market return. Value investing is one of the dominant approaches used by investment professionals.

Value investors prefer to (1) find quality companies and (2) buy them at "reasonable prices" [1] which align to two primary categories of factors used for share evaluation, namely "value" and "quality" [1]. Quality companies are those with a high present value of future residual income or cash flows. A high-quality company is more likely to deliver a higher excess risk-adjusted return as opposed to a low-quality company [1]. Useful measures for achieving this are future profitability and growth [1] which influence future Return on Equity, good cash flows and higher pay-outs while maintaining profitability and growth rates and safety through lower risk and stable earnings. Value (or "cheapness") of a share is confirmed when the intrinsic value of a share is below the current share price by a reasonable margin of safety. The intrinsic value of a share is the total estimated value of equity divided by the number of shares [4]. This research

[1] The ontology and the Bayesian network models are publicly available and can be accessed at: https://github.com/RachelThomson/INVEST-System.

employs the price-to-earnings (PE) multiple, as it is one of the most widely used valuation models in practice [5]. The PE multiple indicates the amount an investor can expect to invest in a company in order to receive one rand of that company's earnings.

2.2 Automated Analysis and Decision Support utilizing Ontologies

Ontologies are a modelling tool used to encapsulate and convert domain knowledge into a formal, unambiguous and machine understandable form [6]. Ontologies can serve as a 'common point of reference' for which there is 'one entry per meaningful concept', and an accompanying definition representing the consensus view of domain experts [7].

Financial decision making can be found across a multitude of categories: stock return prediction, portfolio management and optimization, bankruptcy prediction, foreign exchange rate prediction, detection of fraud, trading models and analysis, and loan risk analysis and payment prediction [8]. In most categories there are several examples of ontology-based approaches that have been used to automate the analysis and decision-making process. Kanellopoulos et al. [9] propose an ontology for predicting whether companies have published fraudulent financial statements through the logical evaluation of twelve financial ratios and the use of a decision tree model. Hu et al. [10] modelled rare risk events to evaluate their effect on banking systems through the Banking Event-driven Scenario-oriented Stress Testing (or simply, BESST), which is a non-probability-based approach for modelling and analysing exceptional but plausible stress testing scenarios without historical data. Chowdhuri et al. [11] propose an Ontology-based Framework for XBRL-mapping and Decision-making (OFXD) which attempts to resolve interoperability between different XBRL filings for seven financial items and describes how this can then be used for meaningful automated analysis. The Fundamental Analysis System for Trading (FAST) proposed by Colomo-Palacios et al. [12] employs semantic technologies for share evaluation. This system, which is most closely related to this work, is described in more detail below.

2.3 Fundamental Analysis System for Trading

FAST utilizes several ontologies and a reasoning tool to reach an investment recommendation for a given share. Financial data is enriched using a financial ontology and stored in a database repository, then accessed using the financial data reader. A financial calculator applies a set of rules to create the financial ratios which are stored in the financial reasoning ontology. Further rules are applied to make a long-term investment recommendation for a company based on these ratios. Two recommendations follow from the financial reasoning ontology: (1) whether a company is a good company to invest in and (2) whether one should buy, sell or maintain shares in the company.

The rules linked to these recommendations are as follows:

• *Medium Term Prediction Rules* facilitate the decision as to whether the company is a "good company to invest in" on a medium-term basis. The rules compare the price-to-book (PTB), price-to-earnings (PER) and price-to-cash-flow (PCF) ratio of a specific share relative to the average of the sector and if below the average, then it is a "good company to invest in".

- The *Long-Term Prediction Rule* fulfils the objective of comparing the calculated value called "actual share calculated value" (ASCV) as defined by the FAST ontology against the current price of a share. Depending on this comparison (one is greater than or equal to the other), one of the three investment options will be returned (sell, buy or maintain). This rule determines when ASCV differs sufficiently, that is by a margin of more than 10%.

FAST only includes the necessary financial ratios like PER, PCFR, PTB, Share Value and ASCV in the inference process and excludes intermediate financial values. We also follow this approach, which is simpler and reduces the number of computations. FAST employs a rule hierarchy to reach an investment decision. Each rule serves to eliminate shares which do not meet the criteria for that rule prior to moving onto the next rule, alleviating the burden of weighing up several criteria at once. The rules are represented as Semantic Web Rule Language (SWRL) rules [21]. SWRL requires crisp logical rules for inference and has limited support for reasoning with multiple variables at once. However, this means that FAST does not take into account that the decision process is not only a multi-criteria problem but also that all forecasts and decisions incorporate a degree of uncertainty.

To our knowledge FAST is the most advanced ontology driven system for share evaluation. However, it has two major limitations. The share evaluation factors are not explicitly linked to a clear strategy and the related underlying investment theory. The use of SWRL for decision making limits the ability of the systems to reason with uncertainty and reasoning with multiple variables at once. This research uses Bayesian networks over SWRL to reflect the uncertainty inherent in the investment decision process.

2.4 Bayesian Networks in Finance

Bayesian networks (also known as Belief Networks) are a compact, flexible and interpretable representation of a joint probability distribution through directed acyclic graphs [6][14]. They are used to capture belief relations, that is informal or uncertain knowledge, between a set of variables which are relevant to some problem. Bayesian networks are adaptable; they are able to be started off small with limited knowledge about a domain [6][14]. One does not require complete information about the world to perform inference in Bayesian Networks. As one acquires more information, the probabilities in the network automatically adjust to cater for the new information. Bayesian networks circumvent several limitations that exist with respect to how information can be processed through ontologies. While ontologies are exceptional at representing organizational structures of large complex domains, their application remains bounded by their inability to deal with uncertainty [6]. Ontology driven Bayesian networks have been proposed to provide more holistic knowledge acquisition, representation and reasoning models for decision making [6]. They have been shown to be effective in use in different applications across diverse domains, including earth observation [13], biodiversity [15][17] and medicine [14][19]. However, to our knowledge there are no systems that integrate ontologies and Bayesian networks for share evaluation and investment decision making.

3 Model Design /Approach

3.1 The INVEST System

The main goal of the INVEST system is to capture and reason with explicit knowledge to support share evaluation. The architecture draws from and is inspired by both FAST [12] and the SWAP system architecture [13][18]. The SWAP architecture proposes a three-layered system architecture and a conceptual knowledge representation and reasoning framework that integrates ontologies and Bayesian networks for developing sensor-based applications on the Sensor Web. The SWAP architecture consists of a sensor layer, which uses ontologies to deal with semantic interoperability and data fusion from heterogenous data sources, a knowledge layer for situation analysis, and an application layer which incorporates Bayesian networks for decision support. The architecture of the INVEST system reflecting the key components of the system is shown in Figure 1. Similar to the layers in the SWAP architecture, the architecture comprises of three layers, i.e. a data layer, an analysis layer and a decision layer.

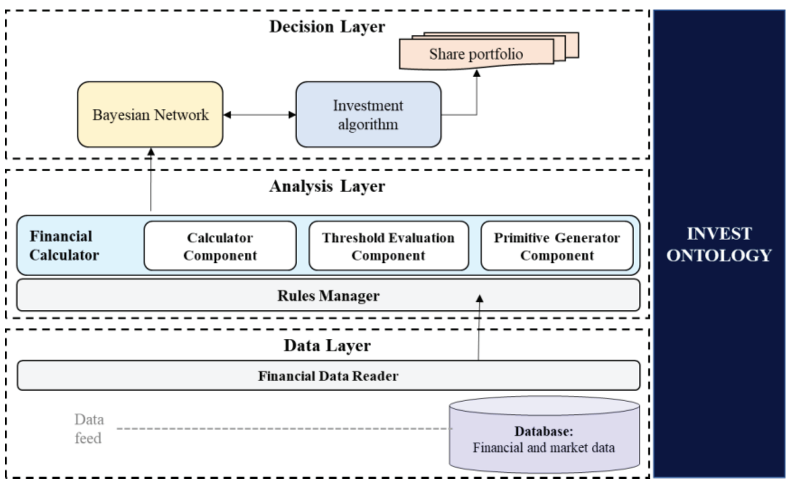

Figure 1. The INVEST system

Data Layer: The Database stores raw fundamental data and is updated through a link to a live data feed. The Financial Data Reader (FDR) provides access to the database.

Analysis Layer: The Financial Calculator (FC) requests data from the Database via the FDR and performs calculations utilizing the data through a conventional programming language like Python in the Calculator Component to produce calculated figures and ratios. Within the FC, an intermediate rule set is housed in the Threshold Evaluation Component (TEC) to evaluate the calculated figures and ratios against a threshold or against another figure or ratio to produce discrete states which represent instances of the **Factor** class in the Invest Ontology (IO). Further to this, the TEC includes straightforward rules to evaluate certain discrete states which determine whether a share is acceptable to be included in the investment universe. The conversion of factors and related instances into useful primitives which are then mapped to the Bayesian network variables is performed by the Primitive Generator Component (PGC). The INVEST ontology (IO) captures concepts across the entire INVEST system.

Decision Layer: In the decision layer, the Bayesian network (BN) is executed for each share that has been included in the acceptable investment universe to determine whether it holds sufficient value and quality to be held in the INVEST portfolio. The investment algorithm iterates through all shares in the investment universe to identify those that meet the minimum investment criteria set by the investment professional. These shares are then included in the INVEST portfolio.

3.2 The INVEST Ontology

Approach. The IO provides a clear structure for information that may be useful to an investment professional and articulates concepts and properties which are required as evidence for the BN. The Unified Process for Ontology Building (UPON) was adopted for the design approach with adaptations as first proposed by Ogundele et al. [14][19] to integrate the development of BNs. Iterative feedback from experts and extensive research on value investing ensured the ontology design was realistic. An OWL ontology was developed using Protégé-OWL (version 5.5.0) [16]. The main classes of the ontology were designed such that they are common to any share evaluation process; and are not confined to the value investing approach. In representing the `Factor` class a design decision was taken to represent factors as the main class in the model and not as properties of a share. This aligns with approaches taken by existing categorizations (see [12]) and with the perspective taken by the investment community. The `Factor` class provides a bridge between financial ratios emanating from the financial calculator and the states of variables in the BNs. This mapping provides seamless translation between financial and market data and evidence nodes in the BNs.

Conceptual Model. This research introduces a new conceptual model which defines a clear approach for share evaluation factor categorization. While FAST (detailed in section 2.5) attempts to articulate this necessary categorization, it fails to adequately identify clear objectives of the decision model, for example: *is the share reasonably priced?* To our knowledge, there is no concrete, unambiguous and comprehensive computer based conceptual model that effectively categorizes factors that are predictive of future share performance.

Two key concepts and ensuing categorizations of the conceptual model were defined and serve to inform the `Factor` class hierarchy in the IO: (1) The `Evaluation Objective` categorization tries to ascertain what question a certain factor will answer. The evaluation objective is guided by the investment approach (e.g., *is the price reasonable* serves to determine whether there is *"Value"* and would be an evaluation objective under the Value Investing approach) and (2) the `Factor Type` categorization groups factors based on the main subject or theme that defines their similarities. For instance, all factors which are return metrics may be grouped together. The conceptual model elevates the `Evaluation Objective` categorization which provides a framework whereby the decision-making process is analysed first to determine and select the factors, under the `Factor Type` categorization, that are most appropriate for the given strategy and objective. The hierarchical approach is inspired by the factor hierarchy used by Ogundele et al. in their ontology of factors affecting tuberculosis patients' treatment adherence behaviour [14][19].

Ontology. The INVEST ontology (IO) formalizes the conceptual model described above. The ontology consists of eight main classes: Factor; Formula; FundamentalData; ModelData; InvestmentAsset; FormalOrganisation and Classifier. These determine the structure of the database for the INVEST system. Figure 7 and Figure 8 in the Appendix provides an overview of the key classes, properties and relations of the ontology.

We focus on the Factor class which is any ratio, figure or qualitative variable that is believed to be predictive of future share performance. The two types of categorizations employed in the conceptual model were transposed to the Factor class hierarchy as depicted in Table 1. The four Evaluation Objective categories, namely *Value; Quality; Elimination* and *Preference*, correspond to abstract classes contained in the second level of Factor class hierarchy (the "Main class"). The seven Factor Type categories, namely *Present Discounted Value*, *Valuation Multiple, Relative Ratios, Profitability Ratios, Growth Ratios, Financial Risk Ratios* and *Systematic Risk Ratios,* correspond to the lower levels of the Factor class hierarchy. The bottom class of the hierarchy represents concrete and measurable factors.

It is important to highlight that the list of factor categories under the main class and sub-classes is not exhaustive but details factor categories specific to value investing as guided by research studies and the guidance of experts. The intention of the model is to provide a framework for extension and to illustrate the application using a specific evaluation approach.

Table 1: Sub-classes of the Factor class

Main Class	**Middle Class**	**Bottom Class**	**Instances**
Evaluation Objective	**Factor Type**	**Measures**	**Values**
Value Factor	Present Discounted Value	IntrinsicValue $_{\text{Discounted Cash Flow}}$ IntrinsicValue $_{\text{Dividend Discount Model}}$	above, equalTo, below
	Valuation Multiples	HistoricalPE_CurrentvsHistory ForwardPE_CurrentvsHistory	cheap, fairValue, expensive
	Ratios	PE Relative$_{\text{Share:Market}}$ PE Relative$_{\text{Share:Sector}}$	cheap, fairValue, expensive
Quality Factor	Profitability Growth Financial Risk	ROEvsCOE CAGRvsInflation Relative Debt to Equity	above, equalTo, below
Preference Factor	Systematic Risk	SpecifiedBeta	above, equalTo, below
Elimination Factor	N/A N/A	Negative Earnings Negative Shareholders Equity	True, False

Delving further, Table 1 reflects the types of instances which belong to the concrete classes of the Factor class. For example, the PE relative$_{\text{Share:Market}}$ has three discrete instances which indicate whether a share is "cheap", "fairValue" or "expensive". These discrete instances have been created through the evaluation of the current value of the PE relative$_{\text{Share:Market}}$ against a threshold; for this particular factor it would be the historical PE relative$_{\text{Share:Market}}$. Each concrete factor class will have a different set of instances depending on the threshold against which it is evaluated.

3.3 Decision making with Bayesian Networks

The share evaluation process involves four evaluation objectives as outlined in the main class of the IO in Table 1. These are assessed in two stages as shown in Figure 2 below. The first stage serves as a filter to eliminate shares that do not meet a minimum set of investment criteria articulated by investors (EliminationFactor and PreferenceFactor**)**. These are represented as simple rules within the Financial Calculator component. This yields a list of acceptable shares for further evaluation. Two examples to illustrate this stage are: (1) Negative Earnings is True under the EliminationFactors would serve to filter out shares where this is the case, and (2) where the SpecifiedBeta under PreferenceFactors (note the investor would set their preference on risk using beta e.g., 0.5) is below the beta of the share being evaluated; this would result in the share being eliminated given that the risk is too high for that particular share.

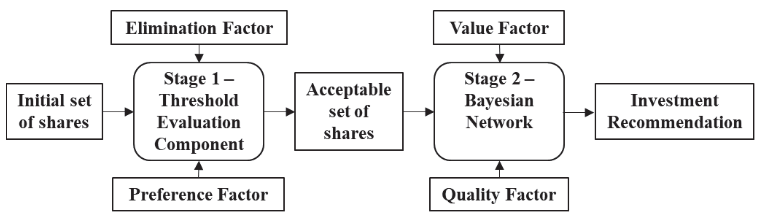

Figure 2. Process diagram of Share Evaluation mapped to the modelled system

The second stage represents the complex analysis undertaken by value investors (see section 2.1) to assess the value and quality of a share. Three BNs are used in this stage. Each BN is employed to address a specific evaluation objective as articulated in the corresponding Factor class (ValueFactor and QualityFactor) where the knowledge is uncertain. As we introduce each BN below, one will note that the categories in the bottom/concrete class of the Factor class correspond to the variables used in the BN.

The **Investment Recommendation Bayesian Network (IR_BN)** as depicted in Figure 3 reflects a one-step decision process on the final investment recommendation for a specific share. The IR_BN is modelled to reflect how investors simultaneously reason with the two decision outcomes from **VE_BN** (Value: cheap, fairValue or expensive) and **QE_BN** (Quality: high, medium, low) which are aligned to the variables and states of the variables for **IR_BN**. The **IR_BN** reflects the final trade-off between Value and Quality. For example, an investor may be willing to pay for a share that is trading at fair value provided it is a high-quality stock but be unprepared to pay for a share trading at fair value should it be a low-quality stock.

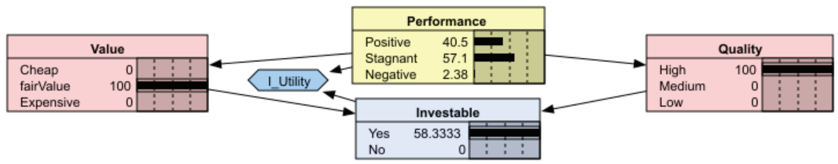

Figure 3. IR_BN Bayesian Network

The **Value Evaluation Bayesian Network (**VE_BN) is used to evaluate the value of the share relative to price. The VE_BN reflects a two-step decision process with two decision nodes modelled to reflect how investors reason with the respective variables which allow one to evaluate a share's value:

(1) **Decision Node 1: Expensive?** Two variables in the Bayesian network; PE_Relative$_{\text{Share:Market}}$ and PE_Relative$_{\text{Share:Sector}}$; are first evaluated to determine whether the share is expensive or not relative. If the decision is "No"; one continues to the second decision. If the decision is "Yes", no further evaluation of the share is required with respect to value; the share is expensive.

(2) Decision Node 2: Value Relative to Price? A third variable, the current ForwardPE_CurrentvsHistory, is evaluated to reach a conclusion on the value of the share relative to its current price.

Each of the variables has a set of states. For example, ForwardPE_CurrentvsHistory node has three discrete states, namely "cheap", "fairValue" and "expensive".

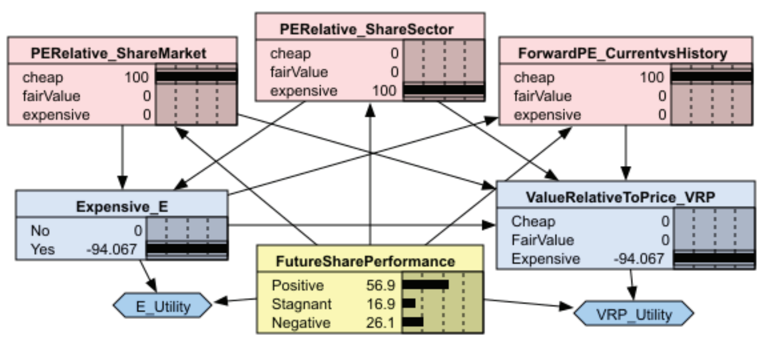

Figure 4. VE_BN Decision Network

The **Quality Evaluation Bayesian Network (QE_BN)** reflects a one-step decision process with one decision node modelled to reflect how investors reason with the respective variables which allow one to evaluate a share's quality. Three variables in the Bayesian network; **Growth_CAGRvsInflation**, **ROEvsCOE** and **Risk_RelDE**; are evaluated using the **decision node** to determine whether a share is high, medium or low quality.

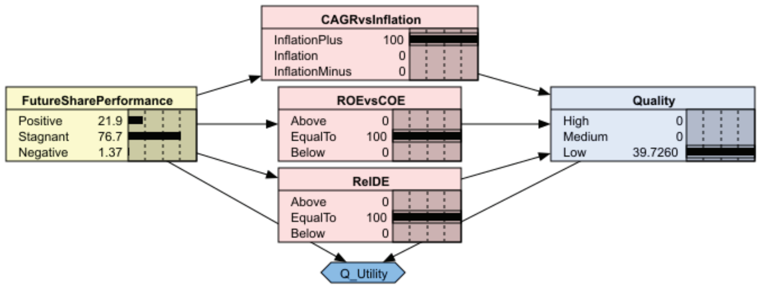

Figure 5. QE_BN Bayesian Network

4 Model Evaluation

4.1 The Importance of the INVEST Ontology

The IO serves the entire INVEST system. The development and structure of the BN draws from the IO, notably the Factor class. The ontology provides clear definitions, categorization and a hierarchy for the factors which may be employed within the Bayesian network to ensure the appropriate selection of factors and related instances for the decision model to address the desired evaluation objective/s. The approach described below draws from the mapping approach used for ontology driven Bayesian networks in the SWAP architecture [13] and Ogundele et al's approach for decision making in the health domain [14][19].

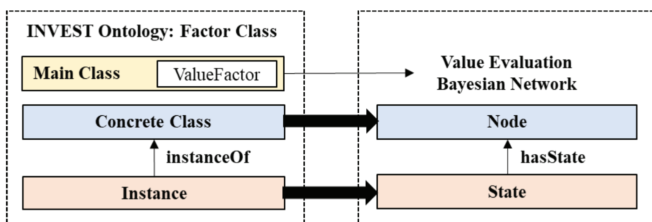

Figure 6. Mapping of Factor Class of the INVEST Ontology to the Bayesian Network

Figure 6 above, reflects the mapping between the Factor class of the IO and the BN and draws us to the second usage of the IO; namely to facilitate the flow of financial and market data between the components of the INVEST System and seamless execution of the Investment algorithm. Concrete sub-classes of the Factor class (e.g., $\text{PE\_Relative}_{\text{Share:Market}}$, $\text{PE\_Relative}_{\text{Share:Sector}}$, ForwardPriceToEarnings) are mapped to a node of the BN which is composed of variables and states. The instances of the Factor class (cheap, fairValue, expensive) are correspondingly mapped to the states of the BN.

Third, the IO ensures that any decision made by the INVEST System and more specifically, the BN, is presented in a form and with a terminology that aligns to the perspective of the investment community.

4.2 Expert Evaluation of the Bayesian Network

The IO and BNs were designed iteratively with continuous engagement with domain experts. A total of 21 value evaluation scenarios, 27 quality evaluation scenarios and 9 investable scenarios were crafted for each of the BNs, with an emphasis on boundary conditions, to reflect different combinations of factors.

The nodes in the three BNs described in section 3.3. were parameterised with appropriate CPT values which reflected reasonable weightings and trade-offs between the different factors. The scenarios were then presented to the experts. For each scenario, the expert's decision and the BN decision was recorded. Any deviations between the system's decision and the consensus decision from the experts were investigated. The domain experts either re-examined their recommendations or the BN models were adjusted appropriately. Several rounds of interaction and feedback between the experts, the modeler and the system occurred until a satisfactory convergence was reached between the expert's decisions and the model's decisions. A total of 24 differences between the experts and the BNs were recorded. The BNs were adjusted 18 times to align them closer to the expert's perspective. Certain adjustments were not made to the model given that the experts' recommendations were split for these scenarios; often due to the presence of boundary conditions.

Table 2: Model and Expert Knowledge Convergence

	Scenarios	Differences	Adjusted	Not Adjusted
Value Evaluation	21	13	10	3
Quality Evaluation	27	11	8	3
Investable Evaluation	9	0	0	0

The different scenarios, the original and refined model together with the final decisions from the four experts are shown in Tables 4, 5 and 6 in the Appendix. Table 2 above summarises the extent to which the model was adjusted to converge with the consensus decision for all the scenarios.

4.3 Back-testing Evaluation of the System

Following the evaluation with experts, the refined model was validated through back-testing the decision-making performance of the system on real-world data. Two Johannesburg Stock Exchange (JSE) indices, General Industrials (JGIND) and Consumer Services (JCSEV), that comprised of a large number of shares were selected.

The Dataset. The data set for shares in the two selected indices was collected for the validation period: 2012 to 2018. There were 19 shares in JGIND and 25 shares in JCSEV with a total of 44. The first stage of share evaluation employs rules to filter shares based on any preference or elimination factors specified by the investor (see section 3.3). Since different investors may have different risk preferences, no SpecifiedBeta (PreferenceFactor) was set to exclude high risk shares. However, the rules for EliminationFactors were set to exclude shares with negative earnings and negative shareholders' equity. Of the initial 44 shares, 8 were excluded since they were not listed on the JSE for the full evaluation period. The final data set consisted of a total of 36 shares and are listed in tables 7, 8 and 9 in the Appendix.

Investment Algorithm. The INVEST portfolio (IP) comprises of shares that were found to be investable by the IR_BN (see Section 3.3., Figure 3). The system does not provide a weighting for each stock so the portfolio comprises of a single share of each share that was determined to be investable. Dividends were not accrued during the evaluation period. For calculation of relative ratios, i.e. debt-to-equity and PE relative to sector, the shares were aggregated into either Consumer Services (JCSEV) and General Industrials (JCIND), and each component share was evaluated relative to the average performance across the shares in the sector.

The investment period is one year, in line with a medium to long term investment horizon from the 1st of January to the 31st of December. On the 1st of January each year the portfolio all shares are re-evaluated. Each of the 36 shares is evaluated on the most recent financial data. For example, the financial data for a company in 2012 is used in the evaluation of the company on the 1st of January 2013. For some metrics, such as historical growth of earnings per share, three years of the preceding historical financial data is required for evaluation. In this example, 2012, 2011 and 2010 financial data would be required. Any share that is investable and not already in the portfolio are bought while any shares that are no longer investable are sold out of the portfolio on this day. The evaluation period for this experiment covers the company financial year-ends from 2012 to 2017.

Portfolio Performance Measures. An active portfolio manager is expected to derive above-average returns for a given risk class. More specifically the portfolio should provide a return that exceeds the return of a passive benchmark; referred to as the active return. The performance of the IP is compared with the respective sector performance either JCSEV or JCIND, which serves as the benchmark.

Three pure return metrics (see formulae in Figure 9 of the Appendix) are used to measure portfolio performance: *annual return, average annual return* and *compound return*. Minimising risk within the portfolio is crucial since it affects the volatility of returns. Two risk adjusted performance metrics, the *Treynor ratio* and the *Sharpe ratio,* are also provided (see formulae in Figure 10 of the Appendix). The Treynor ratio relates excess return over the risk-free rate to the additional risk taken; however, systematic risk is used instead of total risk. Similarly, the Sharpe ratio measures the performance of an investment compared to a risk-free asset, after adjusting for its risk and is defined as the difference between the returns of the investment and the risk-free return, divided by the standard deviation of the investment which represents the portfolio's volatility. The higher the Treynor ratio and the higher the Sharpe ratio, the better the performance of the portfolio.

Results. The performance of the INVEST portfolio IP is measured for the validation period by using the return and risk adjusted return measures detailed above. Performance of the portfolios are compared with the selected benchmark indices.

Table 3: Performance Measures

Period	IP.JGIND AR*	Benchmark JGIND AR*	Active Return	IP.JCSEV AR*	Benchmark JCSEV AR*	Active Return
2012 – 2013	**30.10%**	26.20%	3.90%	**23.30%**	18.04%	5.26%
2013 - 2014	**16.92%**	8.30%	8.62%	**12.73%**	9.28%	3.45%
2014 – 2015	**- 0.38%**	- 9.32%	8.94%	**45.42%**	25.09%	20.33%
2015 – 2016	- 1.96%	**0.04%**	-2.00%	**4.86%**	2.34%	2.52%
2016 – 2017	**20.92%**	14.55%	6.38%	**- 0.26%**	- 0.33%	0.06%
2017 – 2018	**13.89%**	- 1.17%	15.06%	- 2.01%	**1.19%**	-3.20%
Measure	**IP.JGIND**	**Benchmark JGIND**	**Delta**	**IP.JCSEV**	**Benchmark JCSEV**	**Delta**
Compound Return	**13.25%**	6.43%	6.82%	**12.90%**	8.87%	4.03%
Average Annual Return	**6.25%**	-0.57%	6.82%	**14.01%**	9.27%	4.74%
Standard Deviation	12.44%	12.70%	-0.26%	17.99%	10.32%	7.67%
Treynor Ratio	**0.10**	-0.01	0.11	**0.13**	0.03	0.10
Sharpe Ratio	**0.50**	-0.04	0.54	**0.39**	0.22	0.17

* Annual Return

Table 3 reflects that both the IP.JGIND and IP.JSCEV portfolios outperformed the benchmark index in terms of annual return for five one-year periods from 2012 to 2018 respectively and yielded higher compound returns and average annual returns relative to their respective benchmark index. With respect to risk-adjusted measures, Table 3 reflects the Treynor ratio for both IP.JGIND and IP.JCSEV was higher than the respective benchmark index and reflects a more favourable risk/return outcome. The results for the Sharpe ratio reflects the same outperformance. That being said, the absolute Sharpe ratio for both investment portfolios was below 1 which is sub-optimal; a ratio above 1 is acceptable.

Analysis of Evaluation and Results. The two indices selected served as proxies for larger share groupings like the All Share Index (ALSI) on the JSE. The choice to evaluate the BNs' performance against sector/industry indices is reflective of contextual fundamental analysis carried out by investors where shares are evaluated within their sector/industry groupings. While the results are promising a larger investment universe (e.g. ALSI on the JSE) may be required to conclude that the results are statistically significant and may provide more decisive results. The performance of the INVEST decision model is superior to the selected benchmark indices for the period 2012 to 2018 based on return and risk-adjusted return measures. Some well-performing shares within the dataset were not selected in certain years; this suggests that the model could be improved with the addition of further factors or model refinement. The risk-adjusted return measures support this view given that on an absolute basis the performance of the model is below the desired levels. This suggest one of two things: (1) the model performance could be refined or (2) the industry or sector of the indices evaluated have inherently poor risk-adjusted return characteristics. To avoid selection of shares with poor risk-adjusted returns, this could be alleviated by setting a risk threshold using SpecifiedBeta in the filter employed in the first stage of share evaluation. The investment algorithm specified a 12-month holding period for shared deemed investable in line with the full-year financial results cycle of companies. Future evaluation could be extended to different holding periods.

5 Discussion & Conclusion

This research proposed an architecture and prototype implementation of a knowledge-based system for automating share evaluation and investment decision making to support investment professionals. The combination of ontologies and Bayesian Networks was shown to be highly effective for acquiring and representing the necessary knowledge and process of share evaluation and selection. The ontology defines the key factors that influence share performance and explicitly links these factors to different evaluation objectives. This allows an investor to select the correct factors for their decision model based on their beliefs and investment strategy and to make explicit the target objectives and metrics to develop their customized Bayesian decision networks to reflect this. The construction of appropriate Bayesian networks in this regard was illustrated and evaluated using a standard value investing approach. The Bayesian network was able to deal with the uncertainty and complex causal relationships that is inherent in predicting the quality and value of shares using a value investing approach. Our experience aligns with the findings in the study undertaken by Coetzer et al. [17] that showed that ontology driven Bayesian networks can be a highly effective tool for eliciting and representing expert knowledge in the biodiversity domain.

The system's recommendations were refined and evaluated using an expert panel and through back testing on real world data on the JSE. The initial refinement with experts revealed that automated share evaluation provides an investor with an explicit and transparent decision framework for making more informed and objective decisions. In addition, while experts agreed on the inputs, they sometimes disagreed with the model's recommendations especially for boundary conditions. These disagreements do not necessarily negate the usefulness of the model but highlight differences between investors and the inherent uncertainty and subjectiveness associated with investment decisions. Future studies may aim to tackle these boundary conditions through the refining or adding more evaluation factors or expanding the number of discrete states associated with each factor. The back-testing results are promising; the system's portfolio was superior to benchmark indices for the evaluation period.

Future work on developing share evaluation models can draw on this work as a framework for share evaluation. Even though the decision-making model focused on the value investing approach, it can be easily adapted and extended to alternative investment approaches for share evaluation.

References

1. Lee, C.M., 2014. Value investing: Bridging theory and practice. China Accounting and Finance Review, 16(2), pp.1-29.

2. Marks, H., 2018. Mastering the market cycle: getting the odds on your side. Houghton Mifflin Harcourt.
3. Williams, J.B., 1938. The theory of investment value (No. HG4521 W48).
4. Penman, S., 2016. Valuation: The state of the art. Schmalenbach Business Review, 17(1), pp.3-23.
5. Pinto, J.E., Robinson, T.R. and Stowe, J.D., 2019. Equity valuation: A survey of professional practice. Review of Financial Economics, 37(2), pp.219-233.
6. Yang, Y. and Calmet, J., 2005, November. Ontobayes: An ontology-driven uncertainty model. In International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06) (Vol. 1, pp. 457-463). IEEE.
7. Bennett, M., 2011, May. Semantics standardization for financial industry integration. In 2011 International Conference on Collaboration Technologies and Systems (CTS) (pp. 439-445). IEEE.
8. Zhang, D. and Zhou, L., 2004. Discovering golden nuggets: data mining in financial application. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(4), pp.513-522.
9. Kanellopoulos, D., Kotsiantis, S. and Tampakas, V., 2007, September. Towards an ontology-based system for intelligent prediction of firms with fraudulent financial statements. In 2007 IEEE Conference on Emerging Technologies and Factory Automation (EFTA 2007) (pp. 1300-1307). IEEE.
10. Hu, D., Yan, J., Zhao, J.L. and Hua, Z., 2014. Ontology-based scenario modeling and analysis for bank stress testing. Decision Support Systems, 63, pp.81-94.
11. Chowdhuri, R., Yoon, V.Y., Redmond, R.T. and Etudo, U.O., 2014. Ontology based integration of XBRL filings for financial decision making. Decision Support Systems, 68, pp.64-76.
12. Rodríguez-González, A., Colomo-Palacios, R., Guldris-Iglesias, F., Gómez-Berbís, J.M. and García-Crespo, A., 2012. FAST: Fundamental analysis support for financial statements. Using semantics for trading recommendations. Information Systems Frontiers, 14(5), pp.999-1017.
13. Moodley, D., Simonis, I. and Tapamo, J.R., 2012. An architecture for managing knowledge and system dynamism in the worldwide sensor web. International Journal on Semantic Web and Information Systems (IJSWIS), 8(1), pp.64-88.
14. Ogundele, O.A., Moodley, D., Pillay, A.W. and Seebregts, C.J., 2016. An ontology for factors affecting tuberculosis treatment adherence behavior in sub-Saharan Africa. Patient preference and adherence, 10, p.669.
15. Coetzer, W., Moodley, D. and Gerber, A., 2017. A knowledge-based system for generating interaction networks from ecological data. Data & Knowledge Engineering, 112, pp.55-78.
16. Horrocks, I., Patel-Schneider, P.F. and Van Harmelen, F., 2003. From SHIQ and RDF to OWL: The making of a web ontology language. Journal of web semantics, 1(1), pp.7-26.
17. Coetzer, W., Moodley, D., & Gerber, A. (2016). Eliciting and Representing High-Level Knowledge Requirements to Discover Ecological Knowledge in Flower-Visiting Data. PloS one, 11(11)
18. Moodley D and Simonis I (2006), A new architecture for the sensor web: the swap-framework. In Semantic Sensor Networks Workshop, A workshop of the 5th International Semantic Web Conference ISWC 2006, November 5-9, Athens, Georgia, USA.
19. Ogundele, O.A., Moodley, D., Seebregts, C.J. and Pillay, A.W., 2015, September. An ontology for tuberculosis treatment adherence behaviour. In Proceedings of the 2015 Annual

Research Conference on South African Institute of Computer Scientists and Information Technologists (pp. 1-10).
20. Moodley D and Simonis I (2006), A new architecture for the sensor web: the swap-framework. In Semantic Sensor Networks Workshop, A workshop of the 5th International Semantic Web Conference ISWC 2006, November 5-9, Athens, Georgia, USA.
21. Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B. and Dean, M., 2004. SWRL: A semantic web rule language combining OWL and RuleML. W3C Member submission, 21(79), pp.1-31.

A. Appendix

A.1. The INVEST Ontology

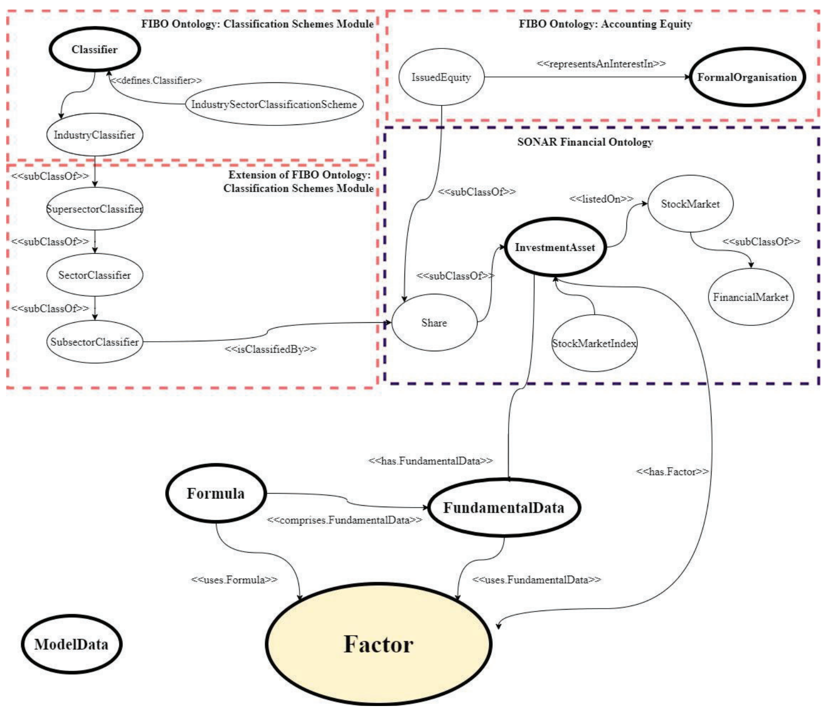

Figure 7. Overview of the key concepts, properties and relationships of the ontology

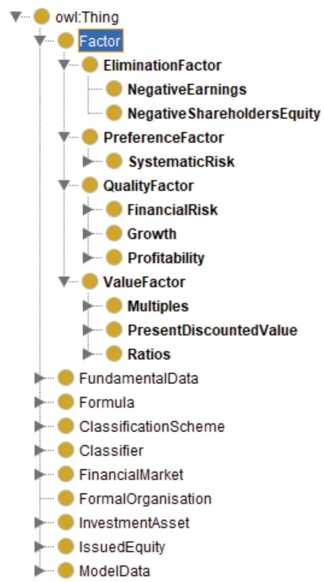

Figure 8. Classes of the INVEST ontology as represented in Protégé ontology editor

A.2. Expert Evaluation of the Bayesian Network

Table 4: Scenario Analysis for Investable Evaluation

Scenario	Outcome_Value	Outcome_Quality	Original Model	DE1	DE2	DE3	DE4	Refined Model
1	Cheap	High	Yes	Yes	Yes	Yes	Yes	Yes
2	Expensive	High	No	No	No	No	No	No
3	fairValue	High	Yes	Yes	Yes	Yes	Yes	Yes
4	Cheap	Medium	Yes	Yes	Yes	Yes	Yes	Yes
5	Expensive	Medium	No	No	No	No	No	No
6	fairValue	Medium	No	No	No	No	No	No
7	Cheap	Low	No	No	No	No	No	No
8	Expensive	Low	No	No	No	No	No	No
9	fairValue	Low	No	No	No	No	No	No

In Table 5 and 6 below, the differences (see cells in pink) are highlighted and where appropriate, adjustments were made to reduce the differences between the model and experts' decisions (see cells in orange where an adjustment was made and cells in green where no adjustment was made).

Table 5: Scenario Analysis for Value Evaluation

			Expensive?							Value relative to price?					
Scenario	**ShareMarket _PERelative**	**ShareSector _PERelative**	**Original Model**	**DE1**	**DE2**	**DE3**	**DE4**	**Refined Model**	**Share_PE _CurrentVsHistory**	**Original Model**	**DE1**	**DE2**	**DE3**	**DE4**	**Refined Model**
1a	Cheap	Cheap	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Cheap	**Cheap**	Cheap	Cheap	Cheap	Cheap	**Cheap**
1b	Cheap	Cheap	**Yes**	Yes	Yes	Yes	Yes	**Yes**	fairValue	**Cheap**	Cheap	fairValue	Cheap	Cheap	**Cheap**
1c	Cheap	Cheap	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Expensive	**Expensive**	fairValue	Expensive	fairValue	fairValue	**fairValue**
2a	Cheap	fairValue	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Cheap	**Cheap**	Cheap	Cheap	Cheap	Cheap	**Cheap**
2b	Cheap	fairValue	**Yes**	Yes	Yes	Yes	Yes	**Yes**	fairValue	**fairValue**	fairValue	fairValue	fairValue	fairValue	**fairValue**
2c	Cheap	fairValue	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Expensive	**Expensive**	fairValue	Expensive	fairValue	fairValue	**fairValue**
3a	Cheap	Expensive	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Cheap	**Cheap**	Cheap	Cheap	Cheap	Cheap	**Cheap**
3b	Cheap	Expensive	**Yes**	Yes	Yes	Yes	Yes	**Yes**	fairValue	**fairValue**	fairValue	fairValue	fairValue	fairValue	**fairValue**
3c	Cheap	Expensive	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Expensive	**Expensive**	Expensive	Expensive	Expensive	Expensive	**Expensive**
4a	fairValue	Cheap	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Cheap	**Cheap**	Cheap	Cheap	Cheap	Cheap	**Cheap**
4b	fairValue	Cheap	**Yes**	Yes	Yes	Yes	Yes	**Yes**	fairValue	**Cheap**	Cheap	fairValue	Cheap	Cheap	**Cheap**
4c	fairValue	Cheap	**Yes**	Yes	Yes	Yes	Yes	**Yes**	Expensive	**Expensive**	fairValue	Expensive	fairValue	fairValue	**fairValue**
5a	fairValue	fairValue	**No**	Yes	Yes	Yes	Yes	**Yes**	Cheap	**Cheap**	Cheap	Cheap	Cheap	Cheap	**Cheap**
5b	fairValue	fairValue	**No**	Yes	Yes	Yes	Yes	**Yes**	fairValue	**fairValue**	fairValue	fairValue	fairValue	fairValue	**fairValue**
5c	fairValue	fairValue	**No**	Yes	Yes	Yes	Yes	**Yes**	Expensive	**Expensive**	Expensive	Expensive	Expensive	Expensive	**Expensive**
6	fairValue	Expensive	**Yes**	Yes	No	No	No	**No**	**-**	**Expensive**	-	-	-	-	**Expensive**
7a	Expensive	Cheap	**No**	Yes	Yes	Yes	Yes	**Yes**	Cheap	**Cheap**	Cheap	Cheap	Cheap	Cheap	**Cheap**
7b	Expensive	Cheap	**No**	Yes	Yes	Yes	Yes	**Yes**	fairValue	**fairValue**	fairValue	fairValue	fairValue	fairValue	**fairValue**
7c	Expensive	Cheap	**No**	Yes	Yes	Yes	Yes	**Yes**	Expensive	**Expensive**	Expensive	Expensive	Expensive	Expensive	**Expensive**
8	Expensive	fairValue	**No**	Yes	No	No	Yes	**No**	**-**	**Expensive**	-	-	-	-	**Expensive**
9	Expensive	Expensive	**No**	No	No	No	No	**No**	**-**	**Expensive**	-	-	-	-	**Expensive**

Table 6: Scenario Analysis for Quality Evaluation

				What is the quality of the stock?					
Scenario	**ROEvsCOE**	**Risk_RelDE**	**Growth_CAGRvsInflation**	**Original Model**	**DE1**	**DE2**	**DE3**	**DE4**	**Refined Model**
1	Below	Below	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
2	Above	Above	InflationPlus	**High**	Medium	Medium	Medium	Medium	**Medium**
3	EqualTo	EqualTo	Inflation	**Medium**	Medium	Medium	Medium	Medium	**Medium**
4	Below	EqualTo	Inflation	**Low**	Low	Low	Low	Low	**Low**
5	Below	Above	InflationPlus	**Low**	Low	Medium	Low	Medium	**Low**
6	EqualTo	Above	InflationPlus	**Medium**	Medium	Medium	Medium	Medium	**Medium**
7	EqualTo	Below	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
8	Below	Below	Inflation	**Low**	Low	Low	Low	Low	**Low**
9	Above	Above	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
10	Above	Below	InflationPlus	**High**	High	High	High	High	**High**
11	EqualTo	Above	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
12	EqualTo	Below	Inflation	**Medium**	Medium	Medium	Medium	Medium	**Medium**
13	Below	EqualTo	InflationPlus	**Low**	High	Medium	Medium	High	**Medium**
14	EqualTo	EqualTo	InflationPlus	**Medium**	High	Medium	High	Medium	**Medium**
15	Below	EqualTo	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
16	EqualTo	Above	Inflation	**Medium**	Low	Medium	Low	Low	**Low**
17	Below	Below	InflationPlus	**Low**	Medium	Low	Medium	Medium	**Medium**
18	EqualTo	EqualTo	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
19	Above	EqualTo	Inflation	**Medium**	Medium	High	Medium	High	**Medium**
20	Above	Above	Inflation	**High**	Medium	Medium	Medium	Medium	**Medium**
21	Above	EqualTo	InflationMinus	**Low**	Medium	Medium	Medium	Medium	**Medium**
22	Below	Above	InflationMinus	**Low**	Low	Low	Low	Low	**Low**
23	Above	EqualTo	InflationPlus	**High**	High	High	High	High	**High**
24	Above	Below	InflationMinus	**Low**	Medium	Low	Low	Medium	**Medium**
25	Below	Above	Inflation	**Low**	Low	Low	Low	Low	**Low**
26	Above	Below	Inflation	**Medium**	High	High	Medium	High	**High**
27	EqualTo	Below	InflationPlus	**Medium**	High	Medium	High	High	**High**

A.3. Back-testing Evaluation of the Bayesian Network

Table 7: Shares within the General Industrials Index

No	Stock Code	Company
1	ADH	Advtech Ltd
2	CLH	City Lodge Hotels Ltd
3	CLS	Clicks Group Ltd
4	COH	Curro Holdings Ltd
5	CSB	Cashbuild Ltd
6	FBR	Famous Brands Ltd
7	ITE	Italtile Ltd
8	LEW	Lewis Group Ltd
9	MRP	Mr Price Group Ltd
10	MSM	Massmart Holdings Ltd
11	PIK	Pick n Pay Stores Ltd
12	SHP	Shoprite Holdings Ltd
13	SPP	SPAR Group Ltd
14	SUI	Sun International Ltd
15	SUR	Spur Corp Ltd
16	TFG	The Foschini Group Ltd
17	TRU	Truworths International Ltd
18	TSG	Tsogo Sun Gaming Ltd
19	WHL	Woolworths Holdings Ltd
20	MCG	MultiChoice Group
21	DCP	Dis-Chem Pharmacies Ltd
22	TGO	Tsogo Sun Hotels Ltd
23	PPH	Pepkor Holdings Ltd
24	MTH	Motus Holdings Ltd
25	SDO	Stadio Holdings Ltd

Table 8: Shares within the Consumer Services Index

No	Stock Code	Company
1	AFT	Afrimat Ltd
2	BAW	Barloworld Ltd
3	BVT	Bidvest Group Ltd
4	GND	Grindrod Ltd
5	HDC	Hudaco Industries Ltd

No	Stock Code	Company
6	IPL	Imperial Logistics Ltd
7	IVT	Invicta Holdings Ltd
8	KAP	KAP Industrial Holdings Ltd
9	MPT	Mpact Ltd
10	MUR	Murray & Roberts Holdings Ltd
11	NPK	Nampak Ltd
12	PPC	PPC Ltd
13	RBX	Raubex Group Ltd
14	RLO	Reunert Ltd
15	SPG	Super Group Ltd
16	TRE	Trencor Ltd
17	WBO	Wilson Bayly Holmes-Ovcon Ltd
18	CTK	Cartrack Holdings Ltd
19	TXT	Textainer Group Holdings Ltd

Table 9: Sector Indices selected and number of companies in each

No	Sector Class	Number of Companies	Included in Benchmark Index
1	General Industrials	19	17
2	Consumer Services	25	19
	Total	**44**	**36**

A.4. Formula

$$\text{Annual Return} = \frac{\text{Portfolio Value}_{\text{Year}+1}}{\text{Portfolio Value}_{\text{Year}}} - 1$$

$$\text{Compound Return} = \left(\frac{\text{Portfolio Value}_{\text{Year}+\text{N}}}{\text{Portfolio Value}_{\text{Year}}}\right)^{1/\text{N}} - 1$$

$$\text{Average Annual Return} = \frac{\left(\frac{\text{Portfolio Value}_{\text{Year}+\text{N}}}{\text{Portfolio Value}_{\text{Year}}}\right)}{\text{N}}$$

Figure 9: Return Metrics

$$Sharpe\ Ratio = \frac{R_p - R_f}{\sigma_p}$$

where:

R_p = return of portfolio

R_f = risk-free rate

σ_p = standard deviation of the portfolio's excess return

$$\text{Treynor Ratio} = \frac{r_p - r_f}{\beta_p}$$

where:

r_p = Portfolio return

r_f = Risk-free rate

β_p = Beta of the portfolio

Figure 10: Risk-adjusted Return Metrics