2nd IIMA International Conference on

Advanced Data Analysis, Business Analytics and Intelligence

January 8 - 9, 2011, Ahmedabad, India

ABSTRACT BOOKLET

Table of Contents

KN1- Next Generation Statistics

Kanti V. Mardia

Senior Research Professor

University of Leeds, UK

I, with Wally Gilks, in 2005 wrote a paper in Significance (a Royal Statistical Society Journal), entitled 'Meeting the statistical needs of 21st-century Science'. We considered the future role of statistics in scientific exploration and prediction. One thing has become clear, that large amounts of data are coming through (such as from high throughput experiments, from climate monitoring). Few new answers are emerging. The monograph by Efron 'Large Scale Inference', Cambridge University Press (2010), gives an empirical Bayes approach to handle large-scale data, such as in micro-arrays. Another new area is Uncertainty Quantification (UQ) which is needed where computer models are utilized for complex real-world processes. Corresponding issues range from inaccuracies in the models to uncertainty in the parameters or intrinsic stochastic features, such as in climate modelling. We will describe these new developments.

The key question is 'What is the status of statisticians in the 21st century?' The presentation by Terry Speed to LASR 2010 described the status of the statisticians in science very clearly. Here is a summary from my sketchy notes. In all the key areas, statisticians perhaps are not the front runners. Initially, in very top journals like Nature, Science, PNAS, one sees new methods that work in a challenging problem. Then comes a phase where applied statisticians do a little bit better than the first phase, producing perhaps more lasting solutions. Then it goes into the hands of statisticians outside the scientific field who aspire to develop new theory and methods. In this talk we will muse on the following conclusion by Mardia and Gilks (2005) who identified three themes for statisticians to be front runner!

"First, statistics should be viewed in the broadest possible way for scientific explanation or prediction of any phenomenon. Second, the future of statistics lies in a holistic approach to interdisciplinary research. Third, a change of attitude is required by statisticians – a paradigm shift – for the subject to go forward."

KN3- BI for IT: Transforming Enterprise IT using Analytics

Harrick Vin,
Vice President and Chief Scientist
Tata Research Development and Design Centre (TRDDC),
Tata Consultancy Services (TCS), India

Over the past decade, we have witnessed the emergence of a plethora of technologies for transforming or optimizing enterprise IT. Yet, most enterprises are struggling to design and operate IT environments that can achieve the elusive goals of efficiency, agility and stability. In this presentation, we will explore the fundamental reasons behind this fact.

- First, I will argue that the complexity and scale of modern enterprise IT environments render the task of obtaining a holistic understanding of the as-is state of an enterprise quite difficult. In most cases, IT environment designers and operators only have a silo-based understanding (along dimensions such as application, infrastructure, operations, and technology stack, among others). This makes it very difficult to predict the impact of a change, leading to resistance for introducing change and hence lack of agility.
- Second, each transformation technology as well as its implementation has costs, benefits, risks and side effects; further, their applicability is governed by several technology and business constraints. This makes the process of deriving and executing a custom transformation strategy for an enterprise quite difficult. Today, most enterprises rely upon manual processes, intuition and experience to derive a transformation plan. Such an approach not only takes too long, but often also results in many implementation surprises leading to many failed transformation programs.

To overcome these challenges, I will describe an analytics-led approach for transforming enterprise IT. This approach (1) relies upon facts (data collected from an operational environment) rather than intuition and experience, and (2) systematizes and automates the process of deriving a custom transformation strategy for each enterprise.

P170- Can family-owned groups attract a loyal shareholder base? Evidence from India

Nikhil Arora, Jyoti Gupta, Alain Chevalier

ESCP Europe, 79 avenue de la République, 75543 Paris Cedex 11, France

This paper examines the rationale behind shareholder loyalty in the case of family owned groups. We have analyzed the case of "Reliance Group", the largest private sector company in India, which is headed by the Ambani Family. We have compared its performance with its competitors in the Indian market. Our results show that in spite of having a weak position with respect to voting rights and managerial influence, the minority shareholders prefer to invest in Reliance due to its consistently positive abnormal returns, low leverage, and other factors that consolidate the family brand.

P082- Determinants of IPO Syndicate Size: An Empirical Investigation

Seshadev Sahoo, Prabina Rajib

Vinod Gupta School of Management (VGSOM),

Indian Institute of Technology, Kharagpur.

We examine syndicates for 154 IPOs issued during the period 2002-2007. We find that the syndicate size is significantly influenced by prestigious investment banks, initial return, leverage, offer size, and ex-ante uncertainty. Variables i.e. market adjusted initial return, and leverage ratio shows an inverse association with the magnitude of the syndicate, while the prestige of the lead bank, offer size, and ex-ante uncertainty are positively influencing the syndicate size.

Among the taken variables we document offer size, followed by prestige of top rated bank, and initial return are relatively more superior in estimating the syndicate size. Further, the differences of the syndicate size across underpricing of IPOs are found statistically significant, explaining syndicate structure helps in reducing underpricing and hence improve the pricing accuracy. Over the sample period syndicate structure is not uniform across operational history of IPO firm, indicating young firms design a syndicate of more number of investment banks than matured IPO firms. Our findings have useful implication for issuing firms, investment banks and other participating agencies in general and investors in particular.

P050- Investigating presence of Nonlinearity in Indian Commodity Markets

Kousik Guhathakurta, Army Institute of Management, Kolkata

Basabi Bhattacharya, A. Roy Chowdhury - Jadavpur University, Kolkata

While there are many evidences of nonlinearity in developed markets, there has not been many works in this direction in Indian financial markets. In this study we wish to bridge this gap by testing for nonlinearity in the Indian commodity market. We consider the index movements in Indian financial markets. The indices under consideration are MCX-COMDEX , MCX-ENERGY, MCX-METAL, MCX-AGRI indices based on trading data from the Multi Commodity exchange of India from June 2005 to August 2010. We take the time series representing the daily log return of the daily close value of the indices as our input value for the tests. We first use the test method developed by Brock, Dechert, and Scheinkman (BDS) and test for nonlinearity in each of the time series. Additionally we also perform the Keenan's test for nonlinearity. Another popular non-linear test is the Hinich bispectrum test, which involves estimating the bispectrum of the observed time series. We also use this test to find out whether it detects nonlinearity in these tine series. To reinforce our findings we also conduct the White's neural Network tests on the same data set. Another linearity test for time series was introduced based on concepts from the theory of neural networks. Teräsvirta et al. developed its power fully. We use this Terasvirta

Neural Network test as a final reinforcement of our findings. Our findings lead us to the conclusion that all the time series developed from data in the Indian Commodity Market exhibit significant nonlinearity. On one hand, the results highlight the fact that researchers cannot take the linear assumption as granted, especially dealing with Commodity market time series data. On the other hand, it points to the need to test for nonlinearity as a preliminary diagnostic tool to determine the nature of the data generating process before any further empirical analysis.

P021- Forecasting Inflation with Artificial Neural Networks

Rudra P. Pradhan, Vinod Gupta School of Management,

Indian Institute of Technology, Kharagpur

The paper presents an application of Artificial Neural Network (ANN) to forecast inflation in India during the period 1994-2009. The study presents four different ANN models on the basis of inflation (WPI), economic growth (IIP) and money supply (MS). The first model is a univariate model based on past WPI only. The other three are multivariate models based on WPI and IIP, WPI and MS and WPI, IIP and MS. In each case, the forecasting performance is measured by mean squared errors and mean absolute deviations. The paper finally concludes that multivariate models showed better forecasting performance over the univariate model. In particular, the multivariate ANN model using WPI, IIP and MS resulted in better performance than the rest of other models to forecast inflation in India.

P005- Modelling of Surface Air Temperature and Pricing of Weather Derivatives

Anandadeep Mandal, KIIT School of Management

This study attempts to formulate a pricing model for the weather derivatives, whose payoffs depend on surface air temperature. Daily temperature data for the last thirty years is closely analyzed for four cities in U.K. to model a temperature process which captures the daily temperature fluctuations including the seasonal patterns and the year-on-year up-ward trend behaviour of the temperature. This work further evaluates an arbitrage-free option pricing using a Gaussian Ornstein-Uhlenbeck model. Keeping in mind that temperature, the underlying variable of the weather derivative, is non-tradable we consider a risk premium estimator to find the price of a weather derivatives contract. Finally, the study provides results based on these models as well as based on Monte Carlo Simulations.

P160- Modeling the effects of extreme price points on consumer choice.

Sanjoy Ghose, University of Wisconsin-Milwaukee

Oded Lowengart, Ben-Gurion University of the Negev

The aim of this paper is to consider the proposition that extreme price values act as internal reference prices, and significantly affect consumer choice and quantity purchased. Scanner panel data provides values for actual prices paid by consumers as well as information about the minimum and maximum prices paid by them in the past. We argue that these minimum and maximum prices not only capture range but also reflect high degrees of incongruity of information. High incongruity should lead to higher recall (Hastie, 1980), and we thus reason that these extreme points will be utilized as reference prices by consumers in making brand choice decisions. We formulate the extreme value points in a multinomial mixed logit model,

and test its viability with choice data for a consumer nondurable. We compare the performance of our model with the performances of benchmark models in the literature and find that our model's performance is superior. Our empirical analysis leads to findings that have managerial implications. To account for the impact of dual internal reference price points, managers might need to consider modifying their traditional pricing strategies; these considerations could be applicable for both EDLP (every day low price) and Hi-Lo type stores. In addition to these, we discuss other managerial implications such as those related to the pricing of new products.

P023- Students' Perception Of Academic And Institutional Service Quality In University Of Jammu

Jaya Bhasin, Vinod Kumar

University of Jammu

Higher Education Institutions leverage knowledge to spur innovation, improve customer service and help in achieving excellence. Regarding this essential and strategic role of quality in Higher Education Institutions, the interventions in academic institutions service quality becomes important. The administrative executives and managers, along with the faculty of the higher education institutes should pay attention in developing their educational institutes in the light of various dimensions of students' quality perception. They should comply with all the necessities, standards and requirements of quality education needed by their students. Accreditation process and the law is not a solution for the problem instead the involvement of the stakeholders in every step is essential. This paper broadly focuses on the efforts and strategies of Higher Educational Institutions for transforming them into progressive educational institutions of higher learning driven by innovative strategies and standards of quality and excellence as seen by the students. Analysis of the collected data has been done by using perceptual mapping for the responses and applying Chi-square test to determine the dependence of attributes. The determining factors have been studied in detail to investigate the perception of quality amongst the students of University of Jammu. It is seen that any Institution which has to progress needs to analyze the end users' perceptions for making necessary improvements. Specifically, it looks into the blueprint of transformation of the University

through Quality Management System. This study also outlines administrative solutions to the problems and analyzes the dynamics of change by proposing a strategic intervention for achieving quality standards in Higher Educational Institutions.

P038- INTMKORIENT: Internal Market Orientation Scale Development, Refinement, Validation and Measurement

Gurjeet Kaur, R.D.Sharma, Nitasha Sheli

University of Jammu

Recent years have witnessed tremendous interest in the concept of internal market orientation (IMO), which is firmly believed to have a strong positive impact on the success of the external market. Despite the growing attention of marketing scholars towards the relevance of IMO as an effective mechanism to counter internal customers' defection, not many research efforts have been directed towards its instrument development, refinement and validation. Apropos, an endeavor has been made in this regard, particularly in context to a fast developing country. The validated Internal Market Orientation Construct (INTMKORIENT) is best represented by intelligence generation, intelligence dissemination and responsiveness. Further, intelligence Generation of internal market is best predicted by "segmenting internal market" and "identifying exchange value". Similarly, Intelligence Dissemination is significantly determined by dissemination of internal market intelligence between "management and employees", "employee and employee" and "management about wants and needs of employees". Finally, Responsiveness is significantly predicted through "job design", "management consideration" and "training".

P070- Consumer Perceived Value and Consumer Loyalty in Healthcare Sector

Hardeep Chahal, Neetu Kumari

University of Jammu

Objectives: The study is undertaken with two primary objectives. First, to develop and validate CPV and CL scales for health care services using three stage criteria (exploratory factor analysis, inter – item analysis and confirmatory factor analysis). Second to examine the impact of CPV on CL for public healthcare services, private health care services and overall health care services.

Methodology: The data is collected from 515 hospitalized patients of two tertiary hospitals namely Government Medical College and Hospital (public) and ASCOMS (private) sector operating in Jammu, North India, during July 2009 to October 2009.

Results: The results suggest that CPV is a function of acquisition value (AV), transaction value (TV) (Grewal, Manroe and Krishan (1988), efficiency value (EV), esthetic value (ESV), social interaction value (SI) and self gratification value (SG). Similarly CL is composite of preference loyalty (i.e. using provider again for services (UPAS), using provider again for different services (UPAD)), recommending provider to others (RP), price indifference loyalty (PL) and dissatisfaction response (DR). Ruyter and Bloemer (1999) also stated that RP, PL and DR are the important dimension of loyalty. Hence hypotheses 1 and 2 are accepted. The study also confirms that delivering superior consumer value enables service provider to associate consumers for long –term through their favourable behavioural intentions or loyalty (Parasuraman and Grewal, 2000). Hence hypotheses 3 is also accepted.

P087- Predictive Analytics for Churn modeling in Telecom – A case study

Jay B. Simha, Sathya Karthik R

Abiba Systems

Most of the literature identifies churn as the loss of a client to a competitor. This is a problem facing every company with subscription based products or services. Churn not only results in a direct financial loss due to loss of customer, but it also calls for higher sales and marketing costs, since it is much more expensive to attract new customers than it is to retain existing ones. Hence, it is widely accepted in telecom industry that it is important to measure, analyse and manage customer churn. The recent explosive growth in prepaid subscribers has created an additional complexity in churn/retention management. Prepaid mobile phone customers have no exit barrier since they are not bound by a contract. Hence, prepaid customers can churn at their convenience and without notification. . In addition the presence of lifetime validity packages and wide prevalence of multi SIM customers makes the churn definition and prediction more complex and difficult for prepaid subscription.

P032- Customer repeat purchase modeling: A Bayesian hierarchical framework

Jayanta Kumar Pal, Subhasish Misra, Abhisek Saha

Hewlett Packard Company, Bangalore.

Two of the major queries in database marketing are of predicting the churn of a customer and the frequency of his/her repeat purchases. Data mining techniques have often been used to approach these, at times with limited success. In this paper we develop a methodology, using a Bayesian analysis framework, specifically to answer these questions. Using this solution we predict the likelihood of customers to make a transaction within a time span (in the next six months/one year etc) in the future. This likelihood/propensity to buy can in turn can be used to rank customers. This will lead to more efficient targeting for marketers. The proposed model when tested on customers of two internal HP databases has yielded very satisfactory results in terms of model validation statistics like concordance, lift, rank-ordering etc.

P063- Application of Propensity Scoring for Establishing Causality in the Area of Direct Marketing

Harsha Kodandaram, Chetan Mahajan Fidelity Investments

D. Karthik, Indian Institute of Management Ahmedabad

Propensity score methods have taken care of casualty more seriously than some traditional data mining methods. There have been numerous applications of propensity scores in medical and health care fields to establish causality. The propensity score, defined as the conditional probability of assignment to a treatment given a vector of covariates. The estimated propensity score, along with matching algorithm can be used to reduce the bias in two groups.

In this paper, we describe the application of propensity score in the area of direct marketing. In direct marketing pre/post and test/control design is the most accurate and comprehensive measure of lift and estimating causal effects. We describe a propensity matching technique to identify customers that can serve as a control group in absence of random control group because of practical business reasons. We illustrate pre/post and test/control design with the matched control group analysis in a case study where a financial service provider company is interested in evaluating the impact of a seminar event.

P081- Using analytics to pitch the right product to customers at inbound channels

Yudhishtar Tiwari, HDFC Bank

Outbound calling is the most commonly used channel for Service organizations in order to engage customers and cross-sell their products. However it's also one of the most intrusive ways of customer engagement. Most of the instances the customer is not in the right frame of mind to receive the communication that service provider wants to send to him / her. With more and more customers registering for DNC and NDNC, the accessibility of such customers is also reducing. In such a challenging environment, it is most appropriate to engage with the customer when he / she's proactively interfacing with the service provider thru the inbound channels (e.g. Customer Service). For HDFC Bank, Customer Power is one such channel which provides an access to a customer while he / she's interfacing across touch points like ATM, Internet Banking and Phone Banking. In order to map customer needs with HDFC Bank offerings, we have used multinomial regression to arrive at the product having highest propensity out of around 12 products to be sold to the customer. The propensity is arrived at by considering various parameters like customer profile, customer behavioral characteristics, customer's extent of engagement with the bank etc. Two levels of correction to the final sequencing of score at customer level were done. The Bank started using this solution for inbound campaigns from Aug'09 and is being used on a regular basis. Over 6 CP (Customer Power) campaigns launched using the scorecard. These campaigns have resulted into an incremental asset booking of over

40%. The multiproduct scorecard helped Bank reduce the outbound calls and replace the same with inbound interactions. Significant numbers of the outbound calls were replaced with inbound interactions using the scorecard. This has also resulted into an improved product holding at customer level.

P124- Building Impactful Segments Using Defined Objective Functions

Guha Athreya, Vamsi Krishna
AbsolutData Research & Analytics Pvt. Ltd.

Segmentation is a widely used marketing approach for solving a variety of business problems. In the case where the focus is to drive growth, an impactful segmentation must clearly identify what are the key consumer and customer behaviours that we need to leverage or change in order to drive towards the growth goals.

Objective Driven Segmentation (ODS) ensures that people within a given market segment are likely to respond similarly to defined marketing strategies. It provides a heat map of opportunities that informs decisions of “Where to Play” vis-à-vis each growth objective.

The ODS approach is one of defining objective functions that represent growth strategies and these objective functions then guide the selection and use of segmenting variables. ODS uses a multitude of techniques including but not limited to Univariate, Bivariate Analyses as well as Multivariate Techniques including Correspondence Analysis, CHAID etc.

This paper provides a detailed understanding of the context & principles of ODS as well as the analytical approach to applying ODS. We look at two case studies - a multi brand beverage manufacturer & a hospitality giant; to understand ODS and the value delivered by it.

P125- Marketing Mix Modeling Application & Implementation in Data Deficient Countries

Imran Saeed, Tarun Mahajan

AbsolutData Research & Analytics Pvt. Ltd.

The global markets are shifting, with manufacturers reaching close to saturated demand levels within the developed countries. The new economy countries likes of BRIC are gaining huge focus. Both manufacturers and marketers are running to have a bigger share of the pie, though with relatively small investment pockets. The business objective of "Growth with constrained investments" demands the businesses to perform on high volume/high ROI matrix. Doing deep dive analysis to identify opportunities might have been relatively simpler in the developed countries, but it is not the same across these new economy countries/data deficient environment. The biggest challenge faced by them is the availability of the data in the required form, followed by understanding of the concept- something which is fairly new to these countries.

P033- Statistical Analysis of Marketing Data

Winfried Stute, University of Giessen

In the marketing research literature one may find famous models which aim to explain the purchase behavior of consumers. In the so-called Ehrenberg model it is assumed that the purchase times follow a Poisson process. To cope with heterogeneity among the households one basic assumption is that the observed processes are mixtures of a Gamma and Poisson process. So far no serious goodness-of-fit test of this model has been achieved. In the talk we present a new method to test for the Poissonian assumption. This is based on the Principal Component decomposition of the Poisson process and the finite sample distribution of its

components. The new test is applied to real micro market data obtained from the US market. As an alternative we also briefly discuss a new purchase time model featuring so-called self-exciting phenomena. This model is applied and fitted to German market data.

P162- Density Forecast Evaluation for Dependent Data

Aurobindo Ghosh, Singapore Management University

Anil K Bera, University of Illinois

In this paper, we propose a formal test for density forecast evaluation in presence of dependent data. Apart from accepting or rejecting the tested model, this approach provides possible sources (such as the location, scale and shape of the distribution) of rejection, thereby helping in deciding possible modifications of the assumed model. We also proposed how to augment the smooth test to investigate explicit forms of dependence in the data in the same test. Our applications to S&P 500 returns indicate capturing time-varying volatility and non-gaussianity significantly improve the performance of the model.

P161- Conditional beta pricing models: A nonparametric approach

Eva Ferreira, Susan Orbe - University of the Basque Country

Javier Gil-Bazo, University Carlos III de Madrid

The implementation of beta pricing models has traditionally relied on the assumption of constant betas and constant MPR. This assumption contradicts the mounting empirical evidence that risk premia vary through time (e.g., Keim and Stambaugh, 1986, Fama and French, 1989, Ferson, 1989, Ferson and Harvey, 1991). As an alternative, some researchers have proposed conditional beta pricing models in which the linear relation holds period by period and both time-varying factor sensitivities and MPR are allowed to vary through time. A drawback of conditional models is that estimation requires additional assumptions about the dynamics of risk exposures and/or MPR.

The method we develop in this paper can be seen as an extension of the popular Fama-MacBeth two-pass method (Fama and MacBeth, 1973), originally developed in the context of unconditional models. In the first stage of the Fama-MacBeth method, asset betas are computed for every asset and period using a time series regression of several periods of previous data, typically spanning between 3 and 5 years. In the second stage, a cross sectional regression of returns on betas is run at every period, which gives a time series of estimated

slope coefficients. The constant slope estimator is finally obtained as the sample mean of the corresponding series of estimated slope coefficients. Similarly, we propose to estimate conditional covariances nonparametrically for each asset and period using previous information. However, unlike the Fama-MacBeth procedure, conditional covariances are assumed to be smooth (but possibly nonlinear) functions of the state variables. In the second stage, time-varying MPR are estimated at each point in time from the cross-section of returns and estimated covariances (the regressors), but instead of running a single cross-sectional regression, the method uses the entire sample. More specifically, in the second pass we use a Seemingly Unrelated Regression Equations (SURE) model, introduced by Zellner (1962), with each equation in the system corresponding to one asset. Time-varying slope coefficients (MPR) are treated as free parameters that vary smoothly through time and are estimated nonparametrically subject to the constraint of equality of slopes across assets, allowing for heteroscedastic and cross-sectionally correlated errors. The method, therefore, enables us to estimate time-varying MPR in conditional models under no specific parametric structure.

Although the Fama-MacBeth procedure was derived to estimate and test unconditional asset pricing models, it also yields a time series of conditional factor sensitivities and MPR. Our method exhibits a number of important advantages with respect to Fama-MacBeth. First, in our method the weight of observations used in the estimation process is driven by the data, that is, it is determined optimally for each data set rather than established ex-ante by the researcher. Second, although both methodologies allow for time variation in betas (covariances) ours is more efficient when betas (covariances) are believed to be functions of a set of variables capturing the state of the system. Third, we derive the asymptotic distribution of the time-varying MPR, rather than that of the constant MPR, which enables us to conduct inference on MPR at each point in time and not only for the constant MPR. Fourth, under the assumption that MPR vary smoothly through time, there is a substantial efficiency gain in our estimators of MPR relative to the time series of slope coefficients since in order to estimate MPR at each point in time we use the entire sample rather than a single cross section of asset returns and covariances. Finally, we assume locally stationary variables as defined in Dalhaus (1997), which permit time-varying mean and, therefore, enable us to drop the usual strong hypothesis of stationarity. Our work is closely related to that of Stanton (1997), Jones (2006), Wang (2002, 2003), and Lewellen and Nagel (2006).

To evaluate the performance of the method in practice, we first carry out a Monte-Carlo simulation and then apply the method to data on stock returns. We base both analyses on the Fama and French (1993) three-factor model. More specifically, for the purpose of the simulation study we consider different specifications for the dynamics of beta, all of which assume that beta is a function of observable state variables. Results indicate that the nonparametric estimator clearly outperforms the traditional rolling estimator under all specifications. When we apply the method to the 25 Fama-French portfolios sorted on size and book-to-market for the 1963-2005 period, we find that nonparametrically estimated MPR exhibit substantial time variation, which supports the use of flexible estimation methods. Further, the nonparametric

method proposed in this paper is clearly superior to different parametric alternatives in terms of its ability to forecast the cross-section of future returns. A purely empirical model, however, appears to dominate even our flexible version of the Fama-French model.

P139- A New Test Of Exponentiality Against IFR Alternatives

M. Z. ANIS, Indian Statistical Institute

MURARI MITRA, Bengal Engineering & Science University

In this note we consider the problem of testing exponentiality against IFR alternatives. A measure of deviation from exponentiality is developed and a test statistic constructed on the basis of this measure. It is shown that the test statistic is an L□statistic. The exact distribution as well as the asymptotic distribution of the test statistic is obtained and the test is shown to be consistent.

P044- Modified Satterthwaite Bootstrap tests for inflation parameter in Zero Inflated Negative Binomial Distribution

Vasudeva Guddattu, Manipal University

K .Aruna Rao, Mangalore University

In this paper the authors propose seven tests for testing inflation parameter in a zero inflated Negative Binomial Distribution. Here performance of modified satterthwaite bootstrap test of likelihood ratio , Wald and Score test with its perturbed versions were investigated. The results indicate that the perturbed version of Wald test with restricted MLE of inflation parameter ,covariates for mean parameter maintain type I error rates and has reasonably good power compared to other test.

P048- Statistical Quality Control with Directional Data

Arnab Kumar Laha, Indian Institute of Management Ahmedabad

Abhishek Gupta, Indian Institute of Technology, Kharagpur

In this paper the properties of circular control chart for mean direction of an angular quality characteristic following von-Mises (a.k.a. circular normal) distribution is studied in detail. Another new log-likelihood based control chart for the mean direction is introduced and its properties are compared with that of the circular control chart. It is found that for small to moderate values of the concentration parameter the circular control chart performs better than the log-likelihood based control chart while for large values of the concentration parameter the log-likelihood based control chart performs better in terms of in-control average run length. However the conclusion is reversed when out-of-control average run length is considered. The robustness of these charts when the underlying distribution is wrapped Cauchy, wrapped normal, and cardioid is examined. A control chart for the concentration parameter is also developed and its properties are studied.

P149- Shape Restricted Regression for Econometric Models

Sujit K Ghosh, Jiangdian Wang

NC State University

In many real business applications economic theory often provides us with qualitative information about the required shape of the regression functions. For instance, cost functions are often required satisfy non-negativity, monotonicity and concavity constraint over a set of prices. Although various parametric shape restricted regression models (e.g., generalized

Leontief function, translog versions of Cobb-Douglass forms) are routinely used in the current literature, often the estimated curves fail to satisfy the desired constraint due the lack of suitable estimating methods. The estimation of such shape restricted regression functions becomes challenging when there are multiple predictors and the problem remains largely unresolved and computationally intensive. This article considers a suitable class of multivariate Bernstein polynomials and proposes a sieved estimator obtained from a nested sequence of shape-restricted multivariate Bernstein polynomials. Three key features of the proposed method are: (i) the regression function estimate is shown to be the solution of a quadratic programming problem; making it computationally attractive (ii) the nonparametric estimator is shown to be universally consistent under some mild regularity conditions and (iii) the estimation methodology is flexible in the sense that it can be easily adapted to accommodate many popular multivariate shape restrictions. Numerical results derived from simulated data sets and real data analysis are used to illustrate the superior performance of the proposed estimator compared an existing estimator in terms of various goodness of fit metrics.

P150- A Robust Bayesian Approach to the Analysis of Volatilities and Value at Risk in Financial Time Series

Pulak Ghosh, Indian Institute of Management, Bangalore

M. Concepcion Ausn, Pedro Galeano, Universidad Carlos III de Madrid, Spain

Financial time series analysis deals with the understanding of data collected on financial markets. Investors and financial managers need to understand the behavior of asset prices to have good expectations about future prices and the risks they will be exposed to. For that, the usual approach is to give insights into the probability distributions of future values, which also allow for the derivation of measures of investment risk. Several parametric distribution models have been entertained for describing, estimating and predicting the dynamics of financial time series. Alternatively, this article considers a Bayesian semi-parametric approach. In particular, the usual parametric distributional assumptions of the GARCH-type models are relaxed by entertaining the class of location-scale mixtures of Gaussian distributions with a Dirichlet process prior on the mixing distribution, leading to a Dirichlet process mixture model. Although in different settings than considered here, Dirichlet process mixture models have an extensive literature in Bayesian analysis and provide a broad and flexible class of distributions. The proposed specification allows for a greater flexibility in capturing both the skewness and

kurtosis frequently observed in financial returns. Also, the Bayesian methodology offers a natural way to introduce parameter uncertainty in the estimation of in-sample volatilities and to obtain predictive distributions of future returns and volatilities. Furthermore, it is also possible to obtain predictive distributions for the Value at Risk (VaR), which has become the most widely used measure of market risk for practitioners. The developed methodology offers a convenient specification of the return distribution, which is crucial to give accurate estimations of the VaR, and provides with a measure of precision for VaR estimates via predictive intervals. The Bombay Stock Exchange Index (BSE-30) and the Hang Seng Index (HSI) are analyzed using the proposed methodology. The results are compared with those obtained assuming the usual Gaussian assumption. The deviance information criterion (DIC) is considered for model selection.

P065- Analysis of High Volatility Financial Data through Circular Statistics

Ashis SenGupta, Indian Statistical Institute, Kolkata

Financial data often involve high volatility. Models based on probability distributions for such data may not admit of any analytical (closed form) representation. Further, the usual moments may not even exist for such distributions e.g. Stable, Linnik, etc. Thus statistical inference for such models, though of prime importance, faces uncommon hurdles. In this talk, we will exhibit how circular statistics can be exploited to overcome such problems and even provide optimal procedures. Real-life examples will be given to illustrate the methods proposed.

P165- Bivariate Cardioid Distributions and Tests of Isotrophy and Independence

S.H.Ong, University of Malaya, Malaysia

Ashis SenGupta, Indian Statistical Institute, Kolkata

In this paper mixture models are proposed for the construction of bivariate cardioid distributions where the marginal distributions are cardioid distributions. There does not seem to be any bivariate cardioid distribution reported in the literature. An advantage of these mixture models is that they are simple in form. Tests of isotropy and independence for the bivariate cardioid distributions will be examined by constructing locally most powerful test and the score test.

P061- A Multi-Dimensional Visualization Approach towards Analyzing and Correlating FORTUNE 100 Stocks

Amit Prakash Sawant, NetApp, Inc.

We present a visual tool to perform interactive multi-dimensional visualizations of current and historical stock data. We visualize FORTUNE 100 stocks and represent the correlation between them using Multi-Dimensional Scaling (MDS) technique. We use simple 2D geometric objects called “glyphs” that vary their spatial position, color, and texture properties to encode a company’s daily and weekly stock prices over the past decade. The result is a display that can be used by viewers to rapidly and accurately analyze, explore, compare, and discover within the stock data. We represent the stock data in ways that facilitate generation of both high-level overviews and low-level details for subsets of the data. This visualization tool will aid stock investors and fund managers with their portfolio theory, investment strategies, and financial analysis.

P084- Imputing the income band using neural network technique for credit card customers

Abhijit Watve, Rajesh Sabapathy, Chiranjibi Dipti Ranjan Panda

ICICI Bank Ltd.

When income information is not available, a traditional work-around to determine the customer income band is to use behavioral (i.e. transactional) information once he or she starts using the credit card. This approach needs data collected over a long period of time to ensure that the acquired information is sufficient, consistent and valid. This approach might yield an income band that is relatively more accurate to the one obtained using application-time information. However, time required to collect the required information is too long (due to activation, dormancy and NUNP status of cardholders), but the initial years of the relationship between customer and card-provider are significant from the point of view of loyalty and retention. This is because most of the communication from the customer, in the form of queries, happens during this period. If the appropriate income band is not available with the card-provider during this period, it might adversely affect the relationship.

Through neural networks, an attempt will be made to establish a multi-layered relationship for segmenting the customer base on the basis of income bands. There is a possibility that the income might get exhibited at various levels of analysis. Post that the segments need to be targeted based on the business requirement. There will be multiple factors that affect the process of customer level relationship and neural network has the capability to study and establish a logic based relationship in a given set of customers.

P015- Price Casual Relations of Oil Markets in Selected Countries of OPEC & Non OPEC

I.Krishnamurthy

University of Petroleum & Energy Studies (UPES)

Crude oil price dynamics determines the order of the world oil market in which OPEC and Non OPEC countries plays major role as exporters and importers. The co integration and information flow of crude oil price between these groups of countries is helpful in understanding the oil price dynamics in terms of oil rich and oil dependent countries. In this empirical study we explored whether or not the selected OPEC and non OPEC countries are co integrated, if so what kind of causal information implies between each other and country's dependence or independence in terms of price giver or price taker. Econometrics tools and causal data analysis software is used in the analysis of oil price time series data.

P159- Nonparametric Control Chart using Discrete Wavelet Transformation

Arnab Kumar Laha, Indian Institute of Management Ahmedabad

Biswaroop Mukherjee, Independent Researcher, Kolkata

Partha Banerjee, Indian Statistical Institute

In this paper we develop a distribution free (nonparametric) control chart using Discrete Wavelet Transformation (DWT) which can efficiently detect shift in the measure of dispersion of the quality characteristic. The performance of this new chart is compared with that of Median Absolute Deviation about Median (MAD) based control chart and Range based Control Chart. Further, the robustness of performance of this chart when the data are autocorrelated is also studied.

P115- Effective Spend Data Management using Machine Learning and Expert Knowledge

Prasanna Venkatesh, Bintu Vasudevan and Rajesh Balakrishnan

SETLabs Infosys Technologies Limited

Spend Data Management(SDM) involves the analysis and classification of a company's spend data, enabling it to understand its purchasing process and to identify saving opportunities, which assist the organization in making informed decisions on its transactional operations with other organizations. We propose a method for spend data management, which involves the use of the structured (vendor, cost center codes, etc.) and unstructured (like item description) elements of the spend data to categorize into a standard taxonomy. The process is improved by effective utilization of the domain expert's knowledge. Using Machine learning techniques the system is train to categorize the sped data along with the heuristics, extracts patterns and rules from the data to map the spend data to standard taxonomy. The initial experiment on the spend data show a promising result with 89% accuracy. We demonstrate these procedures in an experiment, and outline the lessons learnt from the current work.

P008- A Simplified Algorithm for Multi-objective Portfolio Optimisation

Madhu Vij, Saurabh Agarwal

University of Delhi

The paper attempts to first empirically apply Lee and Lerro and Kumar, Philippatos and Ezzell Goal Programming (GP) Model formulations which incorporate existing theories. Secondly, paper contributes theoretically by improving existing GP Modelling framework. Alternate GP formulations are created for classically recommended Max. Min. exact goal achievement and for recommended flexible Q3-Q1 Min. Un-desirable deviation Model. GP Models enabled us to achieve specific targets of multiple goals

related to Capital Gain, Beta, Covariance among securities, Dividend Yield, Unsystematic Risk, Budget Constraint, Industry Diversification and maximum investment in a particular equity. Graphical comparison with Markowitz's efficient frontier shows performance of GP portfolios in risk-return space. Performance evaluation of the Investment portfolios has been done using Sharpe Ratio (S_p), Treynor Ratio (T_p) and excess return to unsystematic risk ratio (VA_p). The use of quartiles for defining the aspiration level in our empirical analysis on BSE -30 was found to be superior to the arbitrary method of setting unrealistically high level of goals and targeting exact achievement.

P012- Impact of Index Futures on the Index Spot Market: An Empirical Study at National Stock Exchange

Y. P. Singh, Megha Agarwal
University of Delhi

The study is an empirical work conducted to assess the impact of trading of index futures on the returns and volatility of the index by examining the nature and strength of relationship that exists between Nifty Index and Nifty futures. The lagged futures returns have forecasting power in explaining current spot index returns as the lag one coefficient is 0.1103. The subsequent lead/lag coefficients are diminishing and the results suggest that the cross correlation coefficients at longer leads/lags are not significant. The cross correlation coefficients indicate that the current spot returns are correlated to the current future returns and one-lead/lag futures returns. Futures thus lead the Index by one lead/lag in Nifty market. Granger Causality shows that returns on Nifty Futures cause returns on Nifty Index while the reverse is not true. There exists both the ARCH Effect (due to recent news) and GARCH effect (due to old news). GARCH effects are stronger in spot index markets for Nifty. ARCH effects are stronger in Index futures market. Futures absorb recent information whereas index markets absorb old information. Thus, empirically it may be concluded that there is a bi-directional flow of information from futures market to the index market and vice versa.

P049- Modeling the Indian Stock Market using Jump Diffusion Processes

Arnab K. Laha, Indian Institute of Management Ahmedabad

Ruchir Gupta, Indian Institute of Technology Bombay

In this paper an attempt is made to model the Bombay Stock Exchange's Sensitivity Index (Sensex). It is seen that the Geometric Brownian Motion model having continuous sample paths is not a good fit for the observed data. The inclusion of jumps through use of Jump Diffusion processes lead to better models. Two Jump Diffusion based models one having jump sizes normally distributed and another with jump sizes double exponentially distributed are considered and their parameters are estimated. Both of these models are found to fit the given data adequately. The parameter estimates of these models can be interpreted easily and the findings are consistent with the stylized facts known for stock markets in advanced economies.

P054- Collective Intelligence Paradigm For Parameter Estimation And Global Optimization of a Hyperchaotic Finance System

Sumona Mukhopadhyay, Army Institute of Management

Santo Banerjee, Politecnico di Torino

Collective Intelligence refers to the synergistic cooperative strategy followed by the multi agent systems inspired from nature-inspired computing. These are primarily employed to optimize combinatorial problems having multiple objectives to be accomplished in order to yield a cost effective and efficient feasible solution. The rapid development in the area of non linear science have lead to the emergence of the study on the complex properties of chaotic system. This is now extended to economy and finance systems whose investigation reveal that these are

complicated nonlinear systems, concerned with real life entities and containing several complex decision making factors which are quite impossible to predict. A detailed study reveals that these systems under certain conditions demonstrate aperiodic motion and uncertainty with non differentiable functions having several discontinuity which depend on initial conditions. Therefore, the control of nonlinear systems and estimation of parameters is a vital task in nonlinear science. A calibration of the parameters with an effective model would subsequently enable a near precise estimation of the unknown parameters of a similar system from its time series. Parameter estimation entails the fulfillment of multiple objectives where it is necessary to not only estimate the unknown parameters but also to optimize them keeping in view the various constraints and hence is primarily a multi objective optimization (MOO) task. However, traditional optimization methods are inadequate to handle the non linearity and complexity embodied by real life entities and systems. This makes it imperative to explore optimization techniques which are self-learning, adaptable and faster in computation. Multi-Agent systems possess the potential to yield effective and optimal result in choosing the best design. In this scenario, we propose a hybrid system based on the underlying principle of particle swarm optimization (PSO). The modified PSO divides into sub-swarms and communicate indirectly through the process of stigmergy. Each swarm is delegated with the task of optimization a parameter through a commensalism relationship to handle the task of global optimization. The proposed system encapsulates the merits of PSO as well as the ergodicity and uncertainty rendered by a chaotic system for achieving improved convergence resulting in an exploratory and exploitation oriented search for multi-dimension estimation and optimization of the unknown parameters of a finance system exhibiting highly aperiodic fluctuations. The outcome of the simulation results is vital from the perspective of control theory and prediction.

P004- Likeability of successful film and differentiating factors for unsuccessful films – A study of select Telugu films using multivariate analysis

A.Ramesh, V.Jayashree, Arvind Gandhi

Vignana Jyothi Institute of Management

Telugu film industry, popularly known as 'Tollywood' contributes to the 45% of south Indian film industry revenue. (Ficci & E&Y Report 2009). The failure rate is precariously high, out of 123 movies released in 2009, only ten of them are successful. The paper attempts to define new construct of 'likeability' of successful film using multi-item scale. Ten popular and nine unsuccessful films in 2009 are selected for the analysis. Likeability is measured using one movie "Maghadhera' which is the most expensive movie made in the history of telugu film industry.(Rs 35 crores) . Underlying key factors for the likeability of the popular or successful movie are derived using factor analysis. The derived factors are used for differentiating the successful and unsuccessful movies using discriminant analysis. The paper attempts to build the model for discrimination and prediction of successful and unsuccessful films. Six factors for the likeability of the successful film derived using exploratory factor analysis are Entertainment value, Direction and Genre, Unique family entertainer, Technical aspects, Hero & Coherence and Screenplay. The variables which discriminate between unsuccessful and successful film are (in the order of importance) Hero and coherence, technical aspects, unique family entertainer, screenplay, direction, genre, and entertainment value.

P101- Voice of the Customer in Emerging Fresh Food Retail in India

Sanhita Athalye, Indian Institute of Management Bangalore

Sangeeta Sahney, VGSOM, Indian Institute of Technology, Kharagpur

Fresh foods retail is a relatively new and budding business area in India. A large difference between the perspective of Indian customers and their western counterparts exists in the case of fresh foods' purchase in a variety of ways. This has made the incorporation of the voice of customer in the structure and design of this industry imperative. This study has been conducted on the customers of such fresh food retail stores in India with an aim to recognize the priorities for quality improvement in the stores. It uses total quality management tools such as SERVQUAL and Quality Function Deployment to analyze the customer needs and wants and their relations to the design parameters of the fresh food retail stores. The study progresses first though a pilot study to identify the most important items in the customer requirements and design characteristics. The SERVQUAL methodology is then applied to identify the gap between the expectations of the customers and their actual experiences. The Quality Function Deployment technique was used to generate the House of Quality and compare the significance of the various design parameters of the fresh foods retail stores in India.

P058- Customer Repeat Purchase Analysis in B2B Context: A Bayesian Framework

Anindya Sankar Dey, Jayanta Kumar Paul, Subhasis Mishra, Paul Abraham

Hewlett Packard, Global Business Services: Decision Support and Analytic Services

Various analytic techniques have been developed to mine historical transaction data of existing customers for predicting future purchase patterns. In the B2B scenario, we consider the

problem of estimating the future transactions in a given period for accounts. We developed a Bayesian hierarchical framework for modeling the customer transactions in a particular business area(called HPSS), to help the sales and marketing groups to target most likely customer accounts.

P107- Should I Build a Segmented Model? A Practitioner's Perspective

Krishna Mehta, Varun Aggarwal
EXL Service

Modelers are often faced with the million dollar question: when should one abandon the convenience of an aggregate level model, to go with a more complex segmented modeling approach? A reasonable response to this query would be that a segment level strategy would be more appropriate when the business needs or data dynamics call for it. While this high level guideline is very useful, there are a lot of blanks that need to be filled in to make this concept operational. In this paper we attempt to fill in some of those blanks and develop guidelines about when segmentation is appropriate.

Our approach looks for evidence of need for segmentation throughout the model development process and provides the reader with a process flow to make this important decision. In making this decision we consider (a) business needs (b) data coverage (c) stability issues due to over dependence on some variables (d) consistency of relationship among predictors and predicted across sub samples and (e) residual analysis of the aggregate model to identify segments of par performance.

All long the paper, we present the reader with some situations where we implemented our approach and share the associated improvement in model performance.

P002- How Product Importance Perception, Need for Cognition and Enduring Involvement Influence Investors' Knowledge of Investment Products?

Sanjay Kr. Mishra, Shri Mata Vaishno Devi University

Manoj Kumar, Indian Institute of Management Lucknow

The lack of proper knowledge about investment products can have a large negative influence on the financial well being of investors. Inspite of that, there is a dearth of studies conducted, specifically to investigate the factors influencing the knowledge of investors about investment products (Vople, Kotel, and Chen, 2002). This paper propose a model and investigates the direct and indirect (mediated through investor's enduring involvement with product) influence of investor's product importance perception (PIP) and need for cognition (NFC) on investor's knowledge of investment products. The hypothesized relationship is empirically validated in the context of mutual fund schemes (MFs). Survey conducted on the sample of 268 MF investors suggests that there is an insignificant direct influence of investor's PIP and NFC on investor's knowledge of investment product (KIP). However, when mediated through enduring involvement with product (EIP), the influence was found to be significant. The results suggest that EIP perfectly mediates the influence of PIP and NFC on investor's KIP. The relevance of the results to the policy makers is also discussed.

P120- Value of Web Analytics for Organizations to Drive Web-Based Strategies

Anindita Paul, Sanda Erdelez

University of Missouri

This paper provides insights to the use of web analytics metrics to support non-profit organizational decisions. The study is conducted in an academic library setting to understand how analytics can be useful for evaluation of their services. Data was collected using a focus group interview with six members of the library's web committee and reviewing the web

analytics data for two semesters before and after the interview. Follow up interviews were also done where necessary. Three important themes are discussed - decisions about services, decisions about web design, affect of management on analytics use. This study can help guide decision-making strategies in organizations that are governmental, commercial, non-profit etc. that would like to utilize the internet to achieve the organizational goals. Managers will be able to consider web analytics as one of the readily available tools that provide support for decision-making to achieve organizational objectives.

P121-Evolving Best Practices for Training Text Data in a Social Media Monitoring Framework

Sujatha R Upadhyaya

Infosys Technologies Ltd

Text mining tools are inevitable in the context of social media monitoring. Text mining employs a combination of techniques for data analysis of which the combination of statistical algorithms with Natural Language Processing (NLP) techniques has been one of the most popular strategies employed. However, the accuracy of analytics on top of text data depends on how well the algorithms are trained to learn from the data and how well the test data responds to the training. Unlike in structured data analytics, accuracy of modeling text analytics depends on the features extracted from text and the subsequent training methodology. This paper discusses the a framework for text mining in the Social Media Monitoring context and also elaborates on the experiments that are conducted to come up with best practices of training methodologies.

P143- Market Basket Analysis using Association Rules and Clustering

Suresh Veluchamy, Target Corporation

Gopal Govindasamy, University of Madras

Market Basket Analysis is used to understand customer purchasing behavior in large store chains. Large stores chains have millions of items and hence the number of item combinations is also very large (i.e. 8 E+12). Metrics like support, confidence and lift are used to evaluate their relative importance. Threshold or cut-off values of support metric can be used to reduce the number of item combinations. Still, marketing and store managers have to go over a large number of item combinations to derive actionable information.

Cluster Analysis is a data mining tool that uses correlation among different items in stores to form items clustering. Association rule is the most commonly used Market Basket Analysis technique. Large store chains using this combined Association Rules and clustering techniques for encouraging cross-shopping, store planning and promotions will lead to increased potential for growth.

In this paper, we describe a method for Market Basket Analysis for large store chains by combing Association Rules and Clustering techniques. We illustrate this method using a simulated example.

P127- Harnessing Social Network with Link Data Mining for Predictive Analytics: an Extended Approach

Tarun Kumar, O.P.Vyas

Indian Institute of Information Technology Allahabad

The social networking websites are connecting people with their common interests, hobbies and professional carrier/business objectives. Social network analysis is posing many severe

challenges in terms of the appropriate methodologies, algorithm and effective implementations. The problem gets more complicated because of data capturing, data analytics and extracting useful pattern from the data is not easy. Many attempts for social network analysis are not so successful because they were lacking in terms of appropriate approach, direction, objective and suitable methodology.

Our research work is targeted to explore and analyze the various techniques for social network analysis and propose a suitable model for predictive analytics to harness social networks. In our proposed study of social network analysis it is believed that collaborations will hold the key for survival in competitive world, we have considered the academic social network for research collaboration. There are various categories of communities being formed as visible in the web. Our proposed method uses the stepwise approach and begins with case study of bibliographic database DBLP having various links in form of academic collaboration between authors of various categories of research papers. Our proposed methodology framework forms the foundation for predicting future collaborations and considers significant aspect of social network being diverse and having different features. These various features can be well represented in our proposed feature measure matrix. Finally, we have described our initial results of generated co-authorship network in the form of basic network metrics which serve as foundation of social network analysis.

P099- Extraction of Suggestions from Opinionated Text – When Opinion is not just a Sentiment

Amar Viswanathan, Prasanna Venkatesh, Bintu G.Vasudevan, Umadas Ravindran, Swarnalatha Ramalingam, and Rajesh Balakrishnan

Infosys Technologies Ltd.

In this Internet Age, the customer is the biggest contributor of product related data and information. The abundance of customer reviews on all kinds of products or services, on various platforms like blogs, articles and discussion forums, provide manufacturers or sellers with a

good understanding of how their products are being received in the market. Many sentiment analysis techniques and applications have been implemented to mine customer reviews and present these to the user, but they primarily focus on extracting individual features and their associated sentiment polarity. Sentiments on the product and product features are of interest to consumers and businesses alike. Businesses are also interested in any actionable feedback available from users on the internet, that is, suggestions for improvements or changes made by the customers of their products. In this paper, we examine the state-of-the-art Opinion Mining techniques, and establish a case for extracting suggestions from actionable user feedback.

P171- Two sample testing and data-dependent allocation for angular responses.

Atanu Biswas, Somak Dutta - Indian Statistical Institute, Kolkata

Arnab K Laha – Indian Institute of Management Ahmedabad

Partho Bakshi – Disha Eye Hospital, Barrackpore, West Bengal

Circular data can occur in many biomedical studies, e.g. some measurements in ophthalmologic studies, degrees of rotation of hand or waist, etc. With reference to a real data set on astigmatism induced in two types of cataract surgeries we carry out some two-sample testing problems including the Behren-Fisher type of test in the circular set up. Response-adaptive designs are used in phase III clinical trials to allocate a larger proportion of patients to the better treatment. There is no available work on response-adaptive designs for circular data. In this present work we provide some response-adaptive designs where the responses are of circular nature, first an ad-hoc allocation design, and then an optimal design. Detailed simulation study and the analysis of the data set including resigning the data using the proposed data-dependent allocation designs are carried out.

P078- Odds ratio for 2 × 2 contingency tables in the presence of surrogate responses

Buddhananda Banerjee, Atanu Biswas

Indian Statistical Institute, Kolkata

The use of surrogate outcome to improve the inference in medicine or biology is an area of research with growing interest. In this present paper we study the asymptotic distribution of log-odds ratio for a 2 × 2 table in the presence of surrogate end-points. We also study the Mantel-Haenszel estimator and also the profile maximum likelihood estimator in the presence of surrogate data when k (> 1) groups are available. As a by-product we prove that the the difference in variation of the two estimators are of order $O((km)^{-3})$ or less where m is the minimum number of true end-points available for any treatment in any group.

P105- Count Distributions for Autoregressive Conditional Duration Models with Application to Financial Data

S.H. Ong , University of Malaya

Atanu Biswas, Indian Statistical Institute

S. Peiris, The University of Sydney

In this presentation the count distribution arising from the autoregressive conditional duration (ACD) model of Engle and Russell (1998) is derived when the duration in the ACD model has a generalized Weibull distribution (Mudholkar et al, 1996). The computation of its probabilities is examined. Applications of the derived count distribution to analysis of price changes in simulated and real stock transaction data and the number of polio cases are then considered. Using a simulation study we show that the method of moments provides better estimates of the parameters of interest.

P141- An Optimal Covariate Adjusted Response Adaptive procedure for Normal Treatment Responses

Rahul Bhattacharya, West Bengal State University

Uttam Bandopadhyay, University of Calcutta

An optimal covariate adjusted adaptive allocation procedure is developed for two treatment continuous response clinical trial. The proposed procedure maximizes the benefited number of subjects and maintains a specified level of power for testing equality of treatment effects. Assuming normality of the response variables several exact and asymptotic properties of the proposed design are studied and compared with the covariate ignored counterpart.

P168- Development of the Utility Function of an Airline Travel: A Logarithmic Goal Programming Approach

Goutam Dutta, Priyanko Ghosh

Indian Institute of Management Ahmedabad

An airline passenger has a number of choices before his/her travel decision. Competing airlines are also interested regarding the preference set of the traveler. Passenger choice modeling is an essential component of any Revenue Management System. In this paper we attempt to develop a linear utility model for an airline travel using a logarithmic goal programming method. The utility score of the competing airlines can help the passengers to choose the best feasible option among several alternatives. The rank also can provide a system of mechanism that compares the competing airlines in a common framework. We also derive the choice probability of the airlines by a multinomial logit choice model. This facilitates the airlines to make a relative comparison among them and to estimate the market share.

P085- A Rough Set Approach to develop an Efficient l-diversity Algorithm based on Clustering

B. K. Tripathy, VIT University

G. K. Panda, MITS

K. Kumaran, VIT University

Most of the organizations publish micro data for a variety of purposes including demographic and public health research. To protect the anonymity of the entities, data holders often remove or encrypt explicit identifiers. But, released information often contains quasi identifiers, which leak valuable information. Samarati and Sweeney introduced the concept of k-anonymity to handle this problem and several algorithms have been introduced by different authors in recent times. Lin et al put forth a new clustering-based method known as OKA for k-anonymization. But, k-anonymity can create groups that leak information due to homogeneity attack. This problem is tackled by the notion of l- diversity introduced by Machanavajjhala et al. Recently, the OKA algorithm is improved by Tripathy et al by making some modifications in the adjustment stage and introducing distinct l-diversity into it. But, in most of the modern databases impreciseness has become a common characteristic, which is not handled by any of the above algorithms. The primary purpose of this paper is to use MMeR, an algorithm introduced by Tripathy et al, in developing a suitable anonymisation algorithm which is applicable to any database having precise or imprecise heterogeneous data and satisfies both k-anonymity as well as l-diversity properties.

P036- Efficient algorithm to predict rare events in a temporal domain

KoteswaraRao Kolli, Alekhya Viswanadhuni, Shilpa Kadam

Prithvi Information Solutions Limited

In general, efficiency and accuracy of algorithms are the major concerns while working with time series data. Here our focus is to identify potential patterns that precede rare events and be able to predict occurrences of rare events within the network efficiently and accurately. A literature review shows that defining temporal window size is considered as user defined. This might be a setback as there is a possibility of losing some information which might actually be of value in detecting hidden information. Here, we propose new algorithms to overcome the problem stated by defining Dynamic window concept. Existing algorithms in the literature use Apriori algorithm for predicting the frequent patterns in the temporal data, which is computationally time consuming. To increase the efficiency of these algorithms we employ Prefix tree based search method. We applied these proposed techniques on event log data of IT infrastructure as well as used synthetic data sets to explore additional patterns, which were not found by the existing algorithms. However, to evaluate if the additional patterns obtained might be of value to the network operators or in general created noise, we have confirmed the same from domain expert and observed that some of the patterns were extremely useful. Our statistics shows that the proposed algorithms have out-performed the existing techniques in both efficiency and accuracy aspects.

P163- Financial Evaluation of Pipeline Projects.

Samir K. Barua, T. Madhavan

Indian Institute of Management, Ahmedabad.

The supply of petroleum products to users for a vast country like India is a challenging exercise. The market forces would determine the pattern of production and distribution of petroleum products. In such a scenario, in all probability, flow of petroleum products would be such as to minimize the total production and distribution costs for the country. The search for financially

viable pipeline projects and the evaluation of proposed pipeline projects should therefore be based on the assumption that such optimal flow pattern would prevail in the sector. The optimal flows can be determined by using a comprehensive optimization model for the entire country.

Two possible approaches for building such an optimization model are: a) network-flow model, b) transshipment model. The major advantage of the transshipment model is that unlike the network flow model, it provides complete information on the route and the mode combinations used by every source-destination pair that is connected in the optimal solution. The attributes of the model used are as follows:

- it minimizes the production and distribution costs for petroleum products for the entire country for a specified year
- it takes into account the demand and supply constraints for petroleum products
- it takes into account the port capacities for import of products
- it is a multi-modal model with the modes considered being combinations of road, rail, pipeline and sea (for coastal movement)
- it assumes that the last segment of transportation to the final consumer (demand point) is always by road
- it uses the districts as the demand points or final destinations
- it uses the existing refineries and the proposed refineries as the sources
- it uses the existing ports as being available for imports
- it takes into account the port capacities
- it assumes that imports would be resorted to only if there is a shortage of products in the country in aggregate
- it uses existing and planned storage as transshipment points
- it assumes that adequate road and rail transportation capacity would be available if the capacity requirements are planned
- it takes into account the pipeline capacities

The model has the flexibility to accommodate changes in the number of sources, ports, destinations and transshipment points. The model also uses the idea of a “super source” or the source of last resort that is to be used in case it is infeasible to meet the demand. The super source supplies at a prohibitively high cost and therefore is used only when all other sources operate at full capacity. The incorporation of this option ensures that an optimal solution is always reached from running the optimization routine.

P001- A Rough Set Approach for Mining Multi-Level Association Rules in Large Databases

Virendra Kumar Shrivastava, Praveen Kumar

Asia Pacific Institute of Information Technology

Jigyasa Bisaria, Indian Institute of Planning and Management

Nishant Nambison, Nambison Softronix

K. R.Pardasani , MANIT

In recent years, the problem of discovering association rules has received considerable research attention and several algorithms for mining frequent itemsets have been developed. Mining association rules at multiple concept levels may lead to the discovery of more refined knowledge in terms of discovery of generalized and specialized rules. This paper presents a rough set perspective for discovery and analysis of multi- level association rules. The novel method RMA discovers multi-level association patterns by heuristic based partitioning of search space. A rough set based framework for analysis of emerging patterns is presented using indiscernibility mapping concept with knowledge base of known patterns. An application of the above method is illustrated for diagnostic decision support and pattern discovery system in clinical databases. Key results include discovery of obvious and novel patterns of disease associations.

P086- Loss Forecasting Techniques: A comparative analysis of Global banking practices

Arti Aggarwal, Chiranjibi Dipti Ranjan Panda

ICICI Bank Ltd.

Most forecasting methodologies are based upon current data (previous 12-24 months). This data is obviously founded on the prevailing macro-economic conditions. As loss rates are

sensitive to changes in economic fortune , this is a factor that the majority of the survey banks reported taking into account. Most firms reported that an adjustment was made on the basis of macro-economic expectations, primarily to ensure that economic capital reflects realistic risk forecasts rather than to stress-test models against a recession scenario. For most, the adjustment process is in essence judgmental but based upon consideration of core data variables such as projected interest rates and property price index. A number of firms reported the development and application of macro-economic models.

P095- A Study on Determination of Turning Points of Newsprint Prices

Rimi Banerjee, Arpit Kumar Laha

Statconsulta, Kolkata

The quarterly newsprint price data was analysed with a view to predict turning points. Price forecasting was made using both time series and dynamic regression models. From these regression models it is found that a strong relation exists between all the study variables and newsprint price. Exponential smoothing and ARIMA methods were also applied to the newsprint price data with a view to find turning points. Analysis suggests that it may be also fruitful to estimate two-regime economic models, one for upswings and one for downswings. This two-regime model could then be combined with a turning-point model that indicates the probability of being in one regime versus the other, period by period.

P108- Interest Rate Sensitivity Analysis for Retail Mortgage Loans

Prateek Prashant, Sandipan Ray

ICICI Bank

Majority of retail mortgage account's interest rate is linked to a floating reference rate (FRR). Once FRR is increased, the tenure of the loan is altered to keep the equated monthly installment

(EMI) amount same. However accounts for which only tenure increase will not suffice, the EMI amount is readjusted post an interest rate change. The objective of this project is to measure the marginal impact of EMI increase on retail residential mortgage portfolio delinquency. All relevant factors that can potentially impact delinquency rate need to be considered and controlled for while measuring the marginal impact of interest rate change. This analysis has been done with respect to one of the largest Indian bank's retail residential mortgage accounts that underwent an EMI reset due to interest rate increase in second quarter of 2008. Results show that EMI reset indeed has an impact on delinquency but in comparison to other variables, its magnitude wise impact is relatively moderate.

P089- Estimation of Operational Value at Risk using Advanced Measurement Approach

Chaitanya Rajulapati, Sandipan Ray, Alok Kumar

ICICI Bank

Basel Accord II allows bank to estimate the capital required for operational risk under the Advanced Measurement Approach (AMA). The present study tries to provide a framework for estimating Operational Value at Risk (OpVaR) for a large bank in India. We have explored the Loss Distribution Approach in this context. The committee suggests the banks to divide the loss events into 7 categories in each of 8 business lines. Due to limitations in the quality of data, we have clubbed all loss events other Internal and External frauds into "Other Loss" category. Similarly, all the business lines are clubbed into one. LDA is then applied to these 1x3 losses separately. Loss frequency distribution is then modelled using the best fit (negative binomial) distribution. While modelling the loss severity, Extreme Value Theory (EVT) is widely used. We used the EVT to distinguish between body and tail distributions. We then observe that outliers in the loss data lead to a high increase in the capital charge for Operational Risk. Hence we deal with outliers separately. Also, since only losses above a cutoff point are used for modelling, we propose to model severity by using a mixture of four distributions – namely – Cutoff, Body, Tail and Outlier distributions. Extreme Value Theory (EVT) is used to find the boundaries between

these distributions. Finally, loss frequency and loss severity distributions are convoluted using monte carlo simulations and then 99.9th percentile is computed to find the OpVaR.

P158- A Bayesian analysis to provide better estimates for surveys with sensitive issues

Sayan Deb Kundu

The Nielsen Company, Mumbai, India

We know that people often tend to play safe with sensitive questions when he has to rate it for himself. However he has a higher propensity to answer that question truthfully when it is projected to other people. Thus this paper draws help from the projective techniques in market research where the respondent is asked about his perception of others regarding the sensitive question. The respondent is also asked some of the indicator questions about himself and some about his opinion of others. Thus we intuitively derive a latent sensitivity score for each respondent. Further using the techniques of game theory, we apply BTS (Bayesian Truth Serum) methodology to get approximate reliability scores for each respondent which is also based on relative entropy. Thus we device an individual response level weighting mechanism where each response is weighted according to its reliability score. Hence this method attempts to apply lesser weightage to those respondents thereby improving the overall quality of the estimate. As a further extension to this model, the triangular NRR (Non Randomized Response) model is combined with this model where we have tried to model it as a multi-group scheme of response models.

P048-Case Study on Time Series Analysis of Hits and Amount withdrawn

D.K.Ghosh

Saurashtra University

Most business problems contain variables whose value are random over time. Time, either as an independent variable or as an added dimension to the existing variables may help the decision making in business. In decision making problem time may be one of the most important variables. This usually occurs in problems where we wish to estimate the expected value to predict the new value at a future point of time after having observed the pattern of the values during past and present time periods. Considering this in this investigation, we have collected data from ATM machine of a reputed bank in respect of date and time of visit of the customer and amount withdrawn by him for a period of one year.

Here we have used various method of time series analysis to estimate the average amount of money withdrawn per day. We have shown that how much money should be kept in the ATM So that customer should not return without collecting the money. We have shown that different amount should be kept at different occasion.

P079- Analysis of spontaneous adverse drug reaction (ADR) reports using supplementary Information

Palash Ghosh, Anup Dewanji

Indian Statistical Institute

Assessment of safety of newly-marketed drugs is an important public health issue. Once the drug is in the market, clinicians and/or health professionals are responsible for recognizing and reporting suspected side effects known as adverse drug reaction (ADR). Such reports are collected in a so-called spontaneous reporting (SR) system. The primary purpose of spontaneous

ADR reporting is to provide early warnings or suspicions, which have not been recognized prior to marketing of a drug because of limitations of clinical trials. We shall discuss the existing work to analyze SR database and their drawbacks and also suggest methodologies to tackle these drawbacks by defining a source population and considering the problem of under-reporting, with the help of supplementary data. Unbiased estimate of population odds-ratio has been obtained and the corresponding asymptotic results are derived.

P083- On the robustness of tests of mean direction for circular normal distribution: A breakdown approach

Arnab K Laha, Indian Institute of Management Ahmedabad

Mahesh K.C, Som-Lalit Institute of Management Studies, Ahmedabad

A statistical procedure is said to be robust if its performance is not sensitive to small deviations from the postulated theoretical model. Several tests for the mean direction of the circular normal distribution have been developed in the literature but their robustness aspects has not explored. In this paper we study the robustness of likelihood ratio test statistic, the directional mean as a test statistic and the γ - circular trimmed mean certain as a test statistic for the mean direction of circular normal distribution using the concept of breakdown functions such as level and power breakdown functions.

P047- SB-robust estimator for directional mean of three parameter symmetric Kato –Jones distribution on circle

Arnab Kumar Laha, Pravida Raja A.C

Indian Institute of Management Ahmedabad

Mahesh K.C

Som-Lalit Institute of Business Management, Ahmedabad

The most often used distribution for modeling directional data has been the von-Mises (a.k.a. circular normal) distribution. Recently Kato and Jones (K-J) proposed a four parameter family of distribution which includes the von-Mises distribution along with wrapped cauchy distribution and uniform distribution. We look at a symmetric unimodal subset of the K-J family of distributions (F) and study the SB-robustness of the estimate of the mean direction at this family. We use the concept of equivalent measures of dispersion in an essential way for proving SB-robustness at F for different measures of dispersion. We also study the SB-robustness of the above said estimator for mixtures of K-J distributions.

P173- Technical Efficiency of Indian Cement Industry

R. N. Mishra, Patna university

A. K. Pandey, Amity university

Data Envelopment Analysis is an efficient technique for deciding the relative efficiency of a DMU (Decision Making Unit) with other unit s engaged in making the same outputs from the same inputs. The DEA model used a mathematical programming technique to estimate the efficiency frontier. Some important studies in the area are due to Charnes, Cooper and Rhodes (1978) and Banker, Charnes and Cooper (1984).

P174- Biases in Peers' and Supervisors' Ratings

Manjari Singh

Indian Institute of Management Ahmedabad

Anita Sarkar

XLRI Jamshedpur School of Business and Human Resources, Jamshedpur

We have explored the possible leniency/ strictness biases in the ratings in a non-managerial multi-rater context. The study deals with the differences in ratings given by the peers and the supervisors. The assessment has been done from two perspectives: one is the evaluation perspective and the other is the support perspective. In the evaluation perspective we have looked at the work outcomes of the individual in terms of job involvement and innovative behaviour. The support perspective deals with the social support provided to the individual by the peers and supervisors. In the first perspective, the focus is on the individual's work as seen from the raters' viewpoint where as in the second perspective, the raters' assessment of the support provided to the individual gets influenced by their role as providers of that support. Peers and supervisors show interesting differences while rating the same parameters. Also, both peers and supervisors look from the two perspectives differently and this gets reflected in their ratings.

P075- Employee Attrition Risk Assessment using Logistic Regression Analysis

Rupesh Khare, Dimple Kaloya, Chandan Kumar Choudhary & Gauri Gupta

Hewitt Associates

Employee Attrition Risk Assessment is receiving significant attention and opening a scope of focused research initiatives. An analytical approach to this assessment aids in prediction of

attrition risk and subsequent action planning. Among the various statistical predictive techniques available, Logistic Regression and Discriminant Analysis come the closest to give a solution. Logistic Regression in this case would give more robust results as it does not assume conditions of multivariate normality and homoscedasticity. In the case presented, Logistic Regression has been employed to predict employee attrition risk based on demographic information and a retention plan has been charted out to target the risk categories derived.

P104- System Dynamics Approach to Manpower Modelling: A Case Study

Santanu Roy, Institute of Management Technology

Sunita S. Upadhyaya, Schneider Electric

Vikrant Bhushan, Wise Cells Learning Solutions

India is emerging as one of the biggest markets for offshore services. Business process outsourcing (BPO) is the delegation of one or more information technology (IT) intensive business processes to an external provider that, in turn, owns, administrates and manages the selected processes based upon defined and measurable performance metrics. Offshore outsourcing is an umbrella term covering a range of IT and business services delivered to companies in developed countries by personnel based in developing countries. Though Indian outsourcing industry is growing, the attrition rate is also rising in this sector. So is the backlash against outsourcing. In order to survive and grow in this scenario, Indian firms must ensure that their services are not only cost-effective but also qualitatively superior. The present study probes into these issues. The study aims to explore the structure of manpower sector of business process outsourcing industry in India through the methodology of system dynamics. A system dynamics model has been developed, validated and simulated over time to understand the trends that characterize this industrial segment. The implications of the results of the study are discussed.

P142- Labour Welfare Measures in the Corporate Sector: A Multivariate Analysis

Anil Kumar
Guru Jambheshwar University of Science & Technology

In this paper we use factor analysis to reveal that various issues relating to labour welfare can be clubbed into nine factors. These are: loans and compensation facilities, education, housing, subsidized food, better working environment, stability of work force and provision of cooperative societies. In order to maintain better industrial relations and stability in the organizations, these types of welfare facilities can go a long way in improving efficiency in the organizations.

P072- A Novel Approach to Identify Potential Business Leads by Behavioral Analysis

Senthil Kumar K, Anil Kumar Gogada, Anutosh Maitra

Infosys Technologies Limited

Many general business practices involve different principles, methodologies and systems to generate new potential business clientele. The work described in this paper proposes a novel approach about how multiple small chunks of information available across the electronic media with the sales executives and with the marketing managers could be processed to provide intelligence in advance about the potential business leads in the coming days to a business organization. In the proposed methodology to identify potential business leads, a rule driven approach has been adopted to analyze the behavior exhibited by an enterprise over a company's website. It's a decision support model and the inferences are usually correlated with information available from other channels like email and telephone conversation before making

a business decision. The aim is to predict the interest of the visiting enterprise well in advance so that marketing teams could approach the potential client with their offerings. The system currently developed has been deployed internally with one of the business units in our organization. The rules have been derived by analyzing past two years of website visits by approximate 2400 organizations globally. A scoring pattern was also determined analytically that grades all potential leads in different categories like strong prospect, good prospect and moderate prospect. In the testing phase, the rule system was able to correctly identify about 90% of the future leads. In most of the cases, the leads were identified about 4 weeks to 4 months prior to the actual date of a request for proposal being raised. Failure analyses have been conducted on the few cases of wrong decision in order to re-tune the rules and also as guidance to data capture and business process modifications.

P067- Product Testing in Financial Services: An Application

Dinabandhu Bag, National Institute of Technology Rourkela

Credit Marketing has come a long way in today's economy of hard-hitting competition and diminishing customer loyalty. With the increasing level of cut-throat competition, decreasing customer loyalty and the increasing commoditization of banking products, it has become essential in today's sluggish economy for banks to proactively understand the changing customer preference. Understanding the changing customer preference can help build a value proposition for the Bank since banks today, are flexible enough to align their products towards the value needs of their customers. Traditional testing by the direct marketers has involved split groups, like an apple to apple, to compare customers' reaction to different offers. Therefore, with changing times, the traditional process of testing has become cumbersome. This study is aimed at demonstrating the benefits of product testing using experimental design to the bank's marketing team. We analyse a factorial model and demonstrates the results of incremental lifts in the market response rates. We conclude that incremental lifts in response rates are much higher against lower interest rates for

home loans and lower late fees for credit cards. It also provides significant insights to the design of Banks' offers.

P080- Channel Productivity Measurement and Expansion through Fusion of External Market data with Internal Bank data

Zohra Ladha, HDFC Bank

Traditionally the Branch expansion exercise in Bank has been happening based on local inputs and past performance of existing branches in the region. Also, the channel level productivity is measured using past performance as benchmarks and using a step-up method. The idea here was to provide analytical assistance to Branch channel for deciding the channel expansion strategy and channel productivity measurement.

The branch expansion recommendations were used by the Branch banking team to arrive at locations for branch expansion, for centers where the population is less than 50,000. Out of over 3006 locations, over 150 locations were recommended for branch expansion and these locations were screened and finalized by the Branch Banking team. The proposed methodology is more scientific and robust as compared to the old methodology on channel expansion. Also the PMI based approach for efficiency measurement is very useful as it takes care of many dynamic parameters which otherwise are impossible to account for while comparing the performances. The appropriate benchmarking is essential for setting up correct business targets at a branch level.

P062- Research Online and Buy Online or Offline (ROBO): Growing Importance of Web and Digital Data in Predictive Models – A Retail Case Study

Ravi Kandikonda, Deborah Balme, David Coppock, Noah Burger

Ogilvy & Mather

Retail purchase predictive models have traditionally focused on the 'in-store' transactions, customer demographic profile and loyalty data. With the growing importance of the digital media and web in every day life, several consumers now do 'pre-shopping' research online on where and which products they intend to buy. Online research includes price comparison, product reviews and availability of the product in store etc. Needless to say, it has a significant impact on the consumer shopping behavior.

In this paper, we have quantified the importance of consumer online behavior data in predicting the likelihood of purchasing a specific product category in the next 30 days. The process flow for this analysis involved three phases:

a) Creating a modeling dataset and an analysis file summarizing data to a household level across 24 months of data. This exercise included about 5 million email addresses. Some of the key variables include past purchase behavior (online and in-store), browse and click history for each product category, campaign response history and demographics.
b) Build and compare 20+ predictive models using different approaches to the model building process - controlled for variable significance, reduced variables through creation of factors and controlled for in-variable variance through transformations
c) Finally, compare the predictive power of the models that included online data versus not. The significance of each online variable was also observed.

The results of this modeling exercise proved that there is a significant lift in the predictive power of the models when the consumer's online research data is included. These findings highlight the level of importance and influence the digital media and web has on the retail industry.

P056- Lead/Opportunity Prediction Analysis

Dwipashray Niyogi, Paul Abraham

Hewlett Packard-GBS-DSAS

In any corporation, marketing divisions through different campaigns generate leads which sales force tries to pursue as opportunities and try to convert them into wins. In a large IT company like HP at any point of time, there exist a huge number of leads and opportunities at different stages of Sales Pipeline. It is not feasible to pursue all of them by Sales. GBS-DSAS, the internal Analytics wing of HP has formulated an analytical technique which can prioritize from the ocean of opportunities, the most valuable ones which have higher chance of converting into wins. The Sales folks can pursue the Opportunities from the prioritized list. The paper illustrates the statistical technique that can help in scoring Leads/Opportunities and give a prioritized list to the Sales Force. This will help the Sales force to employ more resource to the opportunities having higher chance of winning and reducing wastage of resource on bad Leads/Opportunities and thereby improving overall conversion rates of Opportunities.

P123- A Novel Usage of Conjoint Analysis in HR Recruitment

Manica Aggarwal, Raj Ganla, Sandeep Guruvindapalli, Suhale Kapoor

AbsolutData Research and Analytics, Gurgaon, India

Recruitment is a critical HR function and organizations invest substantially in developing robust recruitment practices. Many companies rely on psychometric tests to help them recruit the right people with the right mix of abilities and personal qualities. Psychometric tests, for many years now, claim to provide an accurate assessment of the ability, aptitude and personality of an applicant, enabling employers to assess the candidate's job-fitment as well as his overall organizational fitment.

Psychometric tests are highly standardized tests using scientific tools to measure various aspects of human performance or behavior covering a wide area from motivation to personality. They have an in-built mechanism of detecting false responses by adding similar statements which are structured differently. While screening, one can identify the discrepancies in these responses. Though many companies are using these tools, there is an increasing concern among them regarding accuracy because the candidates have easy access to similar tests which help them to prepare their responses beforehand to best please the recruiter.

Given the limitations of psychometric tools, AbsolutData realized that the selection of candidates entirely based on the results of a psychometric test may not be the right approach. Thus we decided to create a tool that is customized to our needs and is better at predicting actual employee behavior.

We applied the concept of Conjoint Analysis in our recruitment practices. It helps in identifying the combination of multiple attributes of a job opportunity that are most influential on the candidate's choice or decision making in a job-choice situation. Additionally, it is also instrumental in assessing the aptitude, attitude, values and

behavioral traits of the candidate to give an understanding of his/her overall fitment to a particular organization.

P025- An empirical study of Indian call money market

Anurag Joshi, Aastha Singh

Indian Institute of Management, Shillong

The aim of this paper is to provide a better understanding about the inter-relationships between various macroeconomic money market variables in India and their dependence on other external factors. The paper provides an insight into the Indian call money market, identifies factors affecting it and provides ground for further research. The analysis includes interpretation of monetary policy using Taylor's Rule and measuring liquidity though an Index of Market Pressure. Multivariate Analysis is used to explore the Indian Call Money Market using interest rates, liquidity, inflation, FII (Foreign Investment Inflows) as variables. The manner in which various macroeconomic factors tend to affect the behavior of the call money market variables is observed and variables are prioritized in terms of relevance in making policy decisions.

P110-Credit Risk Scoring Analytics- Banks' Lending to MSMEs

Vinay K. Nangia, Rajat Agrawal, Sanjeev Arora

Indian Institute of Technology Roorkee

Banks and other lending institutions that provide financial assistance to MSMEs have developed credit evaluation methodologies to assess the creditworthiness of the enterprises based on different parameters which, usually, include management quality, financial standing, business

risks etc. The banks make use of a variety of quantitative models varying from highly sophisticated ones to simple scoring models. The credit scoring models cover a number of risks that affect the loan repayment capacity of the borrower. Based on literature review, case study of credit scoring models of a large nationalised bank, and expert opinions, some modifications in the models are proposed by including therein some more risks that are not explicitly accounted for in the models studied. These aspects of risk assessment mainly include strategic and some other risks, and judgement of the banker in credit evaluation. The proposed models are expected to give more realistic picture of the risks of business thereby providing better picture of creditworthiness of the prospective borrower. Moreover use of sophisticated modelling, either with the help of softwares or manually, in credit risk assessment is criticised due to their unwarranted use.

P138- Application of Factor Analysis and Logistic Regression Approaches for Analyzing and Predicting the Corporate Dividend Performance: An Empirical Study on Indian Cement Companies.

Anupam De, Gautam Bandyopadhyay, B.N. Chakraborty

National Institute of Technology Durgapur

This study aims at predicting the dividend paying performance of companies with the help of financial ratios. A large number of internal and external factors may influence the dividend paying performance of corporate entities. However, this study is restricted to internal factors which can be evidenced from financial ratios. To get rid of heterogeneous behavior of financial ratios, companies belonging to a single industry i.e. Indian Cement Industry are chosen for the analysis. For this purpose, 42 financial ratios of 38 companies (listed in either Bombay Stock Exchange or National Stock Exchange of India or both) calculated from audited financial statements of respective companies over 9 years (i.e. 1999-2000 to 2007-2008) have been used. At first, Factor Analysis approach is applied to smooth out the effect of multicollinearity on the financial ratios. It resulted in identification of 10 underlying categories (factors) which are

almost independent to each other. Representative financial ratios are identified for each such category (factor) based on the maximum factor scores. Only these financial ratios have been used to develop a model with the help of the Logistic Regression approach. For the purpose of conducting Logistic Regression, dividend paying performance of companies has been classified as 'Good" or "Poor" depending on some appropriate criteria. The developed model has been used to predict the dividend paying performance of the selected companies for the 10th year (i.e. 2008-09). Logistic Regression approach is applied once again to check the predicting capability of the model. It is found that classifications made by the model are correct with perfect accuracy.

P037- Modeling the Symmetric and Asymmetric Volatility for Select Stock Futures in India: Evidence from GARCH Family Models

K. Srinivasan, Shukla Parth, Christ University

Malabika Deo, Pondicherry Central University

This paper examine the modeling and forecasting volatility of stock futures market in India over the period beginning from 1st April 2003 and ending 31st December 2008, for a total of 1440 observations by using Symmetric GARCH and Asymmetric TGARCH, EGARCH and IGARCH model to draw valid conclusion. In sample analysis is carried out for the period from April 1, 2003 to March 31, 2008 and the remaining 184 observations are used to evaluate the out-of-sample forecasting performance of the model. The forecasting performance of two different models was evaluated by considering two forecasting error statistics like Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE). The results of the study indicate that in RMSE statistics, the IGARCH model was performed and it is considered as the best model followed by TGARCH model. Despite its mathematical and statistical simplicity, the IGARCH model provides the most accurate forecast compared to other competing models in the study. Finally, our findings suggest that volatility is a part and parcel of derivative market, which is mainly influenced due to the other key determining factors like inflow of foreign capital into the country like exchange rate, balance of payment, interest rate.

P172- Application of Fractional Brownian Motion to Stock Market Data

Arnab K Laha, Indian Institute of Management Ahmedabad

Abhimanyu Panwar, Indian Institute of Technology Bhubhaneswar

In this paper we study the long range dependence of the daily returns of stocks in the Indian stock market using the Hurst exponent (H). The Hurst exponent is estimated using the rescaled range method. Daily returns data for 44 stocks on National Stock Exchange (NSE) of India are studied for a period of little more than four years between January, 2006 and February, 2010. It is seen that 21 of these stocks showed significant persistent behavior. Since the Hurst exponent is closely associated with Fractional Brownian Motion (FBM) as a next step we attempt to build FBM models of the daily returns of stocks showing significant persistent behavior. We find that for some stocks FBM model fits well. We also attempt to predict some selected stocks for the next 50 trading days using the FBM model by using several different methods and analyze their performance.

P166- How Have Goverenment Policies Driven Rural Credit in India? A Brief Empirical Analysis,1969-2009

Debdatta Pal, Amey Sapre

Indian Institute of Management Ahmedabad

This paper makes a modest attempt to identify structural breaks in outstanding credit of rural branches of Scheduled Commercial Banks (SCB) in India during the period of 1969 to 2009. With the use of endogenous method for finding structural breaks developed by Bai & Perron (1998, 2003), we find three possible structural shifts in growth, i.e. 1981, 1989 and 1999 and thus four

different regimes of growth and performance. These structural changes are further analyzed with respect to branch licensing policy and priority sector lending by the SCBs. Our findings show that break dates closely coincide with the start and withdraw of the 1:4 branch licensing policy. It has also been observed that credit off-take in rural areas is sensitive to network of SCBs in rural areas or stipulation of particular targets for lending to priority sector. Lastly, empirical evidence and growth performance shows that such policies have been instrumental in changing the off-take of rural credit in a significant way. The study also finds corroborative evidence of break dates and growth performance in evaluating the outcomes of the prevailing banking policies.

P136- Designing Intelligent Recommendations for Cross Selling

Kamalika Mazumdar, Satavisha Mukherjee

Genpact India

In a market driven consumer space, typically a lot of communication is seen from the business to the consumer (B2C) in the form of advertisement and offers. But for an ideal B2C set up it is crucial for the business to be knowledgeable about customers to make the communication more effective. Consumers do not explicitly reveal their preferences on the product attributes in a proactive manner; the business only has the past transaction behavior as a resource to understand a consumer. The presence of methods to track the demographic and income level details also makes the business much more resourceful to understand the consumers. If this information or knowledge can be analyzed in an effective manner,it can help businesses offer the right product to the right customers.

This paper offers a solution that enables businesses to design a customized offer for a customer unraveling his/her choices with help of segmentation exercise and cosine distance calculation method.

P057- Enterprise Churn Analysis

Dwipashray Niyogi, Paul Abraham

Hewlett Packard-GBS-DSAS

Churn Analysis estimates the probability of a customer leaving a company; it is the calculation of the rate of attrition in the customer base. It involves identifying those consumers who are most likely to discontinue using your service or product. Churn analysis is extremely helpful in developing a sustainable and robust strategy for customer retention in the company. When company is aware of Churn percentage and identity of these customers in a given time period, makes it easy for the company to analyze causes for churn and developing effective customer retention programs. Churn Analysis is quite prevalent in the B2C domain especially in the Telecommunication and Financial sectors. Due to obvious reasons like life time of enterprise products, smaller base of B2B customers and difficulty of implementing the common techniques practiced, Churn analysis is not a widely used analysis in the B2B space. HP-GBS-DSAS an internal shared services analytics has designed and implemented a new methodology called “Value Movement Analysis” to identify churn effect in HP’s Business Units. In the subsequent section we shall cover in details how the methodology is used to perform Churn Analysis in the B2B space.

P168- A Mathematical Model for Predicting Length of Post-operative Intensive Care Requirement Following Cardiac Surgery in an Indian Hospital

Goutam Dutta, Priyanko Ghosh

Indian Institute of Management Ahmedabad

Ajay Naik,

The Heart Care Clinic, Ahmedabad

Dipa Gosai,

Shri K. K. Shastri Government Commerce College, Ahmedabad

Anil Jain,

SAL Hospital & Medical Institute, Ahmedabad

The Intensive Care Unit (ICU) is a critical resource in a hospital, especially in developing countries like India. The length of ICU stay after a cardiac surgery is an important variable for overall utilization of this resource. If this information is available through a predictive model, the hospital can utilize its ICU occupancy optimally. We conduct a study on ICU patients and collect data for a one year period in a hospital in India. We identify the critical factors for prolonged ICU stay (more than 72 hours) using univariate and multivariate logistic regression and build a predictive index based on model development set. The predictive index is tested on validation set and we show that the mean length of ICU stay increases with an increase in the risk score. We also test the risk score in case of mortality. Efficient utilization of the ICU facility is possible with the help of this predictive index.

P064- A Quantum Inspired Particle Swarm Approach for Multi-Objective Supply Chain Designs

Shruti Gupta, Amity Business School,

Manish Bachlaus, F F Chen,

University of Texas at San Antonio

This paper presents a novel approach that integrates the intangible factors with the tangible ones to model the resource assignment problem in a product driven supply chain. The problem has been mathematically modeled as a multi-objective optimization problem with the objectives of profit, quality, early delivery, and volume flexibility. In this research, product characteristics have been associated with the design requirements of a supply chain. Different types of resources have been considered each differing in its characteristics, thereby providing various alternatives during the design process. The aim is to design integrated supply chains that maximize the weighted sum of the objectives, the weights being decided by the desired product characteristics. The problem has been solved through a proposed Quantum inspired Particle Swarm Optimization (QPSO) metaheuristic. It amalgamates particle swarm optimization with quantum mechanics to enhance the search potential and make it suitable for integer valued optimization. The performance of the proposed solution methodology and its three variants has been authenticated over a set of test instances. The results of the study and the insights derived through it validate the efficiency of the proposed model as well as the solution methodology in the underlying problem.

P122- Analysis of Rating Models of Indian Credit Rating Agencies

Swati Khatkale

Symbiosis International University

The credit ratings are the product of rating process & models, so it is very necessary to understand & judge their model to judge their ratings. India has major five rating agencies which rate numerous products out of them ten products are important. Therefore credit rating models of these 10 products of 5 Indian Credit rating Agencies have been studied. The study showed most of the Credit Rating Agencies use financial statement analysis to rate the debt of an organization. The Credit ratings are based on financial factors like debt equity ratio, interest coverage ratio, net worth, profitability, quality of assets etc. Majority of Credit rating agencies do not go beyond financial statement analysis and management meetings to rate a company though a few rating agencies use some of the sophisticated models of credit rating. In India most of the credit rating agencies give their ratings on subjective criteria at the end rather than any black box model. The problem with subjectivity is rating methodology changes with one person to another. This weakens the due diligence process and thus credit ratings.

P129- Efficiency of Indian Life Insurance Companies: A Non-Parametric Approach

Gokulananda Patel, Monika Mittal

Birla Institute of Management Technology

Indian insurance companies performance were measured based on their net premium received, income from investment as outputs and commissions paid, operating expenses as inputs using BCC model. Projections for inefficient units were calculated through two-phase method. The slack and surplus variables help in determining the reduction of inputs and increasing of outputs to make the inefficient companies to efficient.

P132- Stock Index Prediction using Macroeconomic Time Series: A Neural Network based approach

Rahul Jain
Freelance Researcher

An artificial neural network (ANN) based approach to predict the value of a stock index is presented here. The independent variables consist of time series of macroeconomic variables, namely, inflation rates, GDP, interest rates, index of industrial production, corporate bond yields and business cycle phase. Each of these has been known to possess some relationship with stock prices in general. The dependent variable is the S&P 500 index value, a well-diversified index with a small unsystematic risk component. Additional reasons for choosing this index include its large cap bias, since small stocks have greater firm-specific risk and availability of historical data.
As a first-order approximation, linear regression is done to check for the existence of any linear relationship between the independent variables and the index. The next step is to use a more general, non-linear curve fitting approach using a feedforward ANN with a sigmoidal activation function. To reduce computational overhead, the correlation within the dataset is reduced by using principal component analysis (PCA). Finally, the network is fine-tuned by altering parameters like training set size and number of training cycles. The effects of modifying the training data itself are also analysed. For instance, we find that using lagged values of the dependent variable as an input to the network significantly increases prediction accuracy; also, significantly increasing time series' length has a negative effect on the results.

P134- Exchange Rate Variations and Unanticipated Money Supply in India – A Spectral and Time Domain Study

Biswajit Maitra, Surya Sen College

C. K. Mukhopadhyay, University of North Bengal

This paper examines how far unanticipated money supply changes cause variations in exchange rate in India. A mathematical model on this issue is presented. For the estimation of the model anticipated and unanticipated components of money supply have been identified by applying GARCH (generalised autoregressive conditional heteroskedasticity) based ARIMA (autoregressive integrated moving average) model. Appropriate unrestricted Vector Autoregression (VAR) model estimation testified that variations in exchange rate are Granger caused by those in the unanticipated part of the money supply. Finally, the findings of VAR model have further been confirmed through the spectral analysis.

P151- A SWOT Analysis of the Field of Spatial Data Mining in Precision Agriculture

Kanubhai K. Patel, Ahmedabad University

Kirtan R. Rathod, Shree Swaminarayan Institute of Technology

Sanjay K. Vij, Sardar Vallabhbhai Patel Institute of Technology

The use of spatial data mining in the area of precision agriculture continues to grow, with encouraging results being reported for applications that address crop monitoring, production prediction, fertilizer requirement, field watering, and weather forecasting. This article presents a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis of the field of spatial data mining in the areas of precision agriculture. It reviews the spatial data mining strengths and weaknesses, with illustrations from existing works in precision agriculture. It gives comparative analysis of the various algorithms (like divide-and-conquer, filter-andrefine, ordering, hierarchical structures and parameter estimation, etc.), data structures (like quad trees, k-d trees, R-trees, R*-

trees, etc.), file formats (like tabular geographic information, the ESRI shape-file format, and the KML file format used by Google Earth), spatial statistical and machine-learning techniques (like the support vector machine, spatial cross-validation, and spatial clustering algorithms, etc.), and spatial data mining software in general and specifically with respect to precision agriculture. It also surveys the research work carried out along with tools and solutions available for spatial data mining specifically for precision agriculture. Research centres and organizations active in this field are briefly described along with various tools available. Solutions using convergence of various technologies show a great promise and potential. It is hoped that this structured examination of the factors relevant to the current and future status of spatial data mining in precision agriculture will provide a good overview of the key issues and concerns that are relevant for understanding and advancing this vital application area.

P154- Decision Making Style and Need Pattern as predictors of Team Effectiveness

Neha Verma, Santosh Rangnekar, Mukesh Kumar Barua

Indian Institute of Technology Roorkee

Indian business executives differ in terms of decision making styles and need pattern of their behaviors. The Decision Making Style and Need Pattern also predict Team Effectiveness; however it varies from individual to individual. Work Team has positive synergy through coordinated efforts. It is seen that recently organizations have began to restructure the work processes around teams. It is an empirical study of Decision Making Style and Need Pattern as Predictors of Team Effectiveness. Factors such as Achievement Motive, Rational Decision

Making, Affiliation Motive and Dependent Decision Making have been evaluated to identify their impact on Team Effectiveness. The present study is supported by literature review. The responses of Executives from Public and Private Sector Manufacturing and Service organizations of India are taken and analysed. The objectives of here are - (a) to identify role of Decision Making Style, (b) to identify the role of Need Pattern viz a viz Team Effectiveness. Graphical representation has been used to show the responses and Statistical analysis has been performed to find out the variablity in responses, to diagnose the association between independent and dependent variables. Interestingly the Need Pattern is found to be more significant predictor out of the two independent variables. The study has Practical implications for Indian managers to understand roles of improving decision making style and understanding need pattern in order to improve their team effectiveness. Managerial implications and Scope for further study is incorporated. Finally the conclusion is drawn out of Results and discussion.

P157- Conflict Management and Leadership Style as predictors of Organizational Learning

Aruna B. Bhat,Santosh Rangnekar, Mukesh Barua

Indian Institute of Technology Roorkee

The present study on “Conflict management and leadership style as predictors of organizational learning” is being carried out for senior and middle level executives of Indian Organizations to find out the impact of the two independent variables on a dependent variable. Handling conflicts constructively is one of the greatest challenges in the modern world. Despite the fact that societies have matured, scientific knowledge has increased and a more educated population rise, we still face damaging conflicts. Leadership style refers to the characteristic manner in which an individual leads others. Leader can be people oriented or task oriented. Task oriented leaders tend to be high in need of achievement. Organisational Learning is defined as the intentional use of learning processes at the individual, group and system level to continuously transform the organization in a direction that is increasingly satisfying to its stakeholders. The present study is supported by literature review. The responses of Executives from different Indian organizations are taken and analysed. The objectives of study are - (a) to identify how conflicts in an organization are managed (b) to identify the style of leadership and the effect of these two variables on Organisational Learning. Graphical representation has been used to show the responses and Statistical analysis has been performed to find out the variability in responses, to diagnose the association between independent and dependent variables. The result has shown that Transformational Leadership style is preferred and organizations are learning in terms of Innovation .Finally the conclusion is drawn out of Results and discussion.

P045- A Study on Role Stress among Doctors working in a Government Medical Hospital in Shimla (Himachal Pradesh).

Hirak Dasgupta, Suresh Kumar

Balaji Institute of Modern Management

Objectives: To determine sources of role stress among doctors in Indira Gandhi Medical College & Hospital, a Government Medical Hospital located in Shimla, Himachal Pradesh (India) and to examine the stress levels among Male and Female doctors working in the hospital.

Methodology: Two hundred and fifty three (253) questionnaires were distributed to the doctors and one hundred and fifty (150) duly completed questionnaires were received. Non-probability (Judgment) sampling method was used to select the sampled units within the hospital for study. Statistical treatment included Factor analysis and t-test.

Results: The factors causing role stress among doctors is: (1) Role Overload (2) Self-role distance (3) Role Isolation (4) Inter-role distance (5) Role Stagnation (6) Role expectation conflict (7) Role ambiguity and (8) Role Inadequacy. Roles overload shows 40 percent variance which was found to be a significant factor causing stress among the doctors. t-test indicated that there was no significant difference between the stress levels among male and female doctors except in cases of – Inter-role distance and Role Inadequacy.

Conclusions: The study showed that Role Overload is most significant source or factor causing role stress among the doctors working in the hospital. Male doctors are more stressed than the female doctors in cases of – Inter-role Distance and Role Inadequacy.

P060- An analysis of Housing Market variations through Panel Data approach

Mandeep Dhillon

Everonn IIPS

An analysis of movements in prices of houses is of vital importance to the macro economy as well as to individual. For most Indians, a house is the single largest component of wealth. The housing prices have emerged as a good indicator of output, inflation and financial health for the developing countries like India. So it has become more important to analyze housing price variations in the light of increase in property prices in the resent past (till 2007) and the recent slump in the housing market. Most of the developed countries and some developing countries have housing price indices. These indices are utilized by planners, real estate developers, building material industries, financial institutions as well as the individual home buyers. In India NHB launched an index NHB RESIDEX for tracking prices of residential properties, in July 2007, as the first official housing price index of India. This paper the researcher has studied the housing market variations to analyze the housing dynamics of major Indian cities. This study reviews the prior research on analysis of the housing

market variations. The literature shows that it can be analyzed into four main models namely the fundamental model, the hedonic model, the repeat-sales model and the ripple-effect model.

Based on this, this study analyzed the variations, which are made up of regional information, home market information and time information. The panel data regression method, unit root test and F test are adopted to interpret the housing market variations of the Indian cities. This paper suggests that the Indian home-market information has the elasticity to the housing market variations across cities and time. The time information contributes differently along the observing period, although the similarities are found in certain periods. This paper has tried to answer the questions like what are the determinants of the housing prices. Can we assess a housing bubble? How do we measure the role of liquidity in housing price increases?

2nd IIMA International Conference on Advanced Data Analysis, Business Analytics and Intelligence, January 8-9, 2011

Abstract Booklet

Addendum

Table of Contents

P175- The Antecedents and Consequences of Trust in Online Shopping: The moderating effect of Product Type and mediating role of Trust

Boudhayan Ganguly

Chandragupt Institute of Management Patna

and

Satyabhusan Dash

Indian Institute of Management Lucknow Prabandh Nagar Off Sitapur Road

Research Context: Lack of Trust has been sited as a major reason for the abhorrence of online shopping. Online trust building is a complex phenomenon and the current cross-cultural research on online trust and culture assessment methods are very chaotic. Further, there have been no work till date that comprehensively studies and compares the determinants of trust for customers of different product types. Trust has been proposed to be a mediator between purchase intention and the antecedents of purchase intention. A number of comprehensive models have been developed and tested empirically to get insights into the antecedents and consequences of online trust and how the relationship between trust and these antecedents and also that between trust and its consequences vary for different product categories (Such as *customised vrs standardised*; *search vrs experience* and *tangible vrs intangible*).

Methodology and Results: Student sample from various premier B-schools of India, Canada and USA were used in the study. **Confirmatory Factor Analyses** have been used to test the validity and reliability of the questionnaire items. To test the relationships between trust and its antecedents and also between trust and its consequences, **path analysis** has been used. Moderator regression analysis have been used to test the moderator effects of culture, product type and customer characteristics. Our results indicate that security, privacy, information design, communication with online store, self efficacy and internal norms positively influence trust. Trust was found to reduce perceived risk and increase purchase intention. Visual design was found to be influencing trust among the westerners but not among the Indians. External norm, navigation

design, social presence, vendor repute did not have any impact on trust. **Hierarchical moderated regression** analysis is used the test the moderating customized product and intangible product.

Managerial Implications of Results: This study compared and contrasted the importance attached to the antecedents of trust both at the individual and the country level. So, the results of this study can be used for country centric segmentation of the online consumer market. Online companies are also increasingly dealing with the situation of selling its different product types to a varied range of customers on the Internet.. So, the empirical results of this integrated study can also be used for product centric and country centric segmentation of the online consumer markets.

P013- Beta Estimation Practice And Its Reliability Biasness Towards Aggressive Stocks: An Empirical Evidence From NSE

Neeraj Sanghi, Institute of Management studies

Gaurav Bansal, R.D. Engineering College

In finance literature 'beta' possesses a prominent place as a measurement statistic of systematic risk arising out of economic wide uncertainties. As a matter of fact, Index Model is very common in practice as leading stock exchanges of India make use of this model for beta estimation. Our study aims at this approach of beta estimation for establishing how beta coefficients for aggressive stocks prove to be more reliable than to defensive stocks. Since Index Model is linear and envisage the premise of Simple Linear Regression, we compare reliability of beta coefficients for aggressive and defensive stocks on the basis of R-squared statistic.

P030- Comparative Study of different Sectors of Stock Market Volatility in India: During 2007-2010

Kirti Arekar, Rinku Jain

K.J. Somaiya Institute of Management Studies & Research

The Indian Stock Market started falling from January 2007 to January 2010, the descent accelerating towards the end of 2008, due the global fallout of the U.S. mortgage crisis. After that there is a slowly improvement in the performance of the Indian Stock market relative to the other World markets. In this article, a snapshot of the market performance during the two-year is presented and compared with the major overseas markets. A study considered a market performance of different sectors i.e. Information Technology and Banking with respect to the market. Further we analyzed that which sector is impacted most during the recession period. A number of parameters are used to capture the market performance such as daily return, Volatility of daily return, market capitalization and mutual fund activity. The period from January 2007 to November 2010 showed Indian market's march towards the highest-ever levels of market capitalization and stock indices in 2007, and, thereafter, a precipitous fall in 2008. These include strong economic fundamentals, relatively stable political climate and, hence, large foreign funds inflow. Finally, we interpreted that which sector performing good and bad at this Global recession period and which sector has performed good after the recession or we can say there is no impact of recession for that particular sector.

The study seeks to analyze the following aspects of the Indian stock market for the years 2007 and 2010:

- The extent and pattern of daily returns on Sensex and Nifty and their volatility.
- Comparison of Indian market performance with global markets in terms of price trend, daily returns, and their volatility.

P041- Bayesian Fractal Model for Microarray Data

R.Geetha, S.D.N.B. Vaishnav College for Women

Microarrays provide information to identify the set of genes responsible for the trait under study. DNA microarray measures the expression levels of thousands of genes simultaneously and has emerged as an important tool in biomedical research. The gene expression profiles can discriminate between known cell types or conditions, for e.g. between tumour and normal tissues or between tumours of different types. The widely used techniques to analyze microarray data are clustering (Eisen *et al* 1998; Dudoit *et al* 2000; Tibshirani *et al* 2000) and classification methods (Golub *et al* 1999; Khan *et al* 1999; Hsiao *et al* 2000; Wong *et al* 2003). In this paper, a methodology using fractal dimension to measure the binary disease outcome in patient models based on selective gene expression is proposed. In a microarray data with p genes, which will usually be of high dimension with comparatively fewer samples, the m ($< p$) genes with differential expression under two conditions are first identified by fitting a statistical model $Y_{jk} = a_j + b_j X_k + \varepsilon_{jk}$ (Y_{jk} is the expression level for j_{th} gene in array k, X_k is a binary indicator and ε_{jk} 's are random error with mean zero), under which the j_{th} gene would be differentially expressed whenever $b_j \neq 0$ .The Hausdorff fractal dimension (D), a measure of self similarity for each sample profile with m genes is then estimated. Subsequently the disease outcomes are regressed against fractal dimension using logistic regression as $logit(p) = \beta_0 + \beta_1 D$. Standard microarray datasets are used to demonstrate this method and the results obtained using R code are reported. The analysis is enhanced further by adopting a Bayesian approach with uniform priors for β_0 *and* β_1.

Result: The fractal characteristics of the microarray derived gene-expression profiles may be explored in classification of diseases.

P071- Mutual Funds Selection and Investment: Using AHP and Goal Programming

Shailesh Rastogi, Jaipuria Institute of Management

In this paper 10 diversified equity based mutual funds are considered. The decision criteria used in AHP are capital appreciation (long term), capital appreciation (short term) current income, tax advantage etc. After having selected or ranked the mutual funds, the next step is to build the portfolio of mutual fund investments. Goal programming (GP) model is being used in this study to find out optimum portfolio. In the present study a GP model is being used to determine in what proportion the mutual fund schemes should be invested. The important decision criteria used are standard deviation, the percentage growth in NAV, the Sharpe index and the systematic risk. The selection of the mutual funds has been modeled as an easy process and no technical or advanced knowledge is required to follow the set process but in case of using GP it does require some advance know-how to decide about the optimum portfolio.

There are two important **results** which come out of this study. The first result is a model which can be replicated by the practitioners for selecting the best alternative for mutual fund investors on the basis of investor's own preferences not on the basis of some financial planners best pick. The second result using GP will give optimization of the financial resources using a different technique which uses more than one objective for optimization of the resources along with other constraints. The results which come out of this study have several **implications**. The finding has good practical utility and it can change the way financial planner suggest the mutual fund schemes to the investors.

P076- Cluster Based Web Search Using Support Vector Machine

Rita. S. Shelke

Bharati Vidyapeeth University

Now days, searches for the web pages with a given keyword constitute a notable fraction of queries to Web search engines. Web search is difficult because it is hard for users to construct

queries that are both sufficiently descriptive and sufficiently discriminating to find just the web pages that are relevant to the user's search goal. Queries are often ambiguous: words and phrases are frequently polysemantic and user search goals are often narrower in scope than the queries used to express them. This ambiguity leads to search result sets containing distinct page groups that meet different user search goals. Often users must refine their search by modifying the query to filter out the irrelevant results. Users must understand the result set to refine queries effectively; but this is time consuming, if the result set is unorganized. Web page clustering is one approach for assisting users to both comprehend the result set and to refine the query. Web page clustering identifies semantically meaningful groups of web pages and presents these to the user as clusters. The clusters provide an overview of the contents of the result set and when a cluster is selected the result set is refined to just the relevant pages in that cluster. Using Support vector machines (SVMs), a set of related Supervised learning methods used to divide the total load of queries from the user to the server onto different client machines as per their capacities to keep the system stable. Server maintains the complete database.

P097- Use of advanced data analysis in Indian Politics

Gaurav Bhargava , Neha Bhardwaj, Neelam Saraswat
Lovely Professional University

The objective/focus of this paper is to study the contributions made by Data Analysis in Indian Politics. In this conceptual paper, we are highlighting how various data analytical tools find applications in the political sector and can thus derive benefits for political leaders, voters and government (policy making & policy implementation). This paper aims to identify important factors that voters consider before electing a particular candidate as their chosen representative. This paper also reflects how government can use data analysis a) to make and implement various policies and b) to keep a check on fund allocation & resource utilization by elected representatives.

As there is no datum available on data analysis in Indian politics, we are using exploratory research with primary data analysis. Our research relies upon review of literature, discussion with private political research organizations and actual primary research done by field surveyors in one

assembly. Factor Analysis shall be used to arrive at important factors that influence the decision making of voters and that determine the favorability of a particular candidate in the political space. It shall also help in explaining the variability among various factors in terms of the importance level of each of these factors. Factor Analysis shall thus help in reducing the number of observed variables by merging these into an unobserved variable, and thus shall be helpful in redefining the focus areas for a particular candidate and his campaigning strategy. Regression Analysis shall be used to model the behavior of variables and to define the relationships between various variables using their correlations. The paper shall be helpful in determining the relation between the success of a candidate and the variables that influence this success thereupon. Also, correlation indices determine the factors or the variables that are complimentary to each other and thus act in unison to increase the chances of success for a particular candidate.

Our study concludes caste equation, party network strength, development issues, MLA/MP fund allocation, Image of candidate and interaction of candidate with voters are most important factors that influence decision making of voters. Research shows development work is most important parameter that directly influences decision making of voters. There are observations about volatile mindset and decision making of voters during last days of election campaign. Some observations also indicate changing preference of voters just one night before election. But it can not be volcanic change. Combination of Caste equation and party network strength is must for a political leader to have a strong hold in any booth. Presence of this combination leads to higher chances of voting in a specific booth for a candidate. Age of the Voter is an important factor that influences the decision choice of the voter for a particular candidate. This indicates that the political aspirants also must consider the age of their target population and the strategies can thus be designed accordingly based on the age of the voters.

Accurate Primary research with efficient sampling technique can provide important insights for improvement in Indian political system. Research shows minimum 5% sample population is required for an effective forecasting and representation of political equations in a booth with minimum cost. For political leader or government, voter's profile segmentation in terms of caste, age, occupation and geography is very important to cater their needs and wants.

P102- Analysis of SQL SERVER Log: A Data Mining Approach

Samaya Pillai, Haridas Acharya, Priya Revalkar, Parashar Ranade
Symbiosis Institute of Computer Studies and Research, Pune

Today the business environment without IT is just unthinkable. Databases constitute a major part of any IT system. Generally in the small organizations MySql is largely used. But in large organizations Oracle or SQL server is more widely used. SQL servers provide the much needed back end support for the transaction processing. Performance Analysis and Optimization of IT infrastructure is equally important to any Business to consistently improve its performance. PA & O basically is to see performance trends and avoid bottlenecks. The Paper focuses on analysis of log files with special ref. to Sql Server to improving Database performance and availability. The Analysis has two major aspects. Initially text mining approach is used to determine strengths of association between distinct messages which occur. Text mining is further used to remove the general messages which have no relevance to the performance of the system itself. This reduces the size of the log files to be analyzed considerably. The focus helps in deriving facts related to the performance easily and much faster. Next the trimmed log files are used to generate MDDBs with severity as fact and day, business hrs, month and types of messages as dimensions. Treating the dimensions as factors and severity as the response the MDDBs were subjected to ANOVA to examine the effect of dimensions on the severity. A severity index (0 to 1) has been specially defined for the purpose. The analysis helps to trace the performance of the database server

P113- A Modified Fuzzy Least-Squares Approach for Claim Reserving

Ramakrishna Singh, Exl Service

Broadly two approaches exist in the literature for studying a Fuzzy linear regression (FLR) model. In the former, proposed by Tanaka *et al.* (1982), parameters of FLR model are estimated by minimizing "Total vagueness" of model-data combination, subject to constraints that each data point must lie within estimated value of response variable. This can be visualized as a Linear

programming (LP) problem and solved by using "Simplex procedure". To this end, several software packages, like SAS, LP88, Excel and LINDO are available. However, a criticism of Tanaka's approach is that it is not based on sound statistical principles. Another drawback, as pointed out by Chang and Ayyub (2001), is that as the number of data points increases, the number of constraints in LP increases proportionally, thereby resulting in computational difficulties.

The second approach based on Fuzzy least squares (FLS) method, was pioneered by Diamond (1988), which as its name suggests, is a fuzzy extension of Least squares method based on a new defined distance on the space of fuzzy numbers. D'Urso (2003) initiated a new approach based on Modified fuzzy least-squares approach, to deal with Fuzzy linear regression analysis. A doubly linear adaptive fuzzy regression model was proposed based on two linear models: (i) Core regression model and (ii) Spread (width) regression model. First one explains "centres" of fuzzy observations, while second one is for their "spreads". In this work, this approach is followed. Further, performance evaluation criterion based on "Difference in membership functions" is adopted for computation of error in estimation (Kim and Bishu, 1998). As an illustration in this work, the methodology is applied in insurance for proper claim reserving.

P119- The Effect of Weather on Indian Stock Market – An Empirical Study

Gautam Bandyopadhyay, National Institute of Technology. Durgapur

Arindam Banerjee, Amity University

Parijat Upadhyay, International School of Business and Media.Kolkata

Behavioral finance proposes a relationship between the weather and equity market by way of investors moods. A well-established and diverse literature, primarily in the field of psychology, has investigated the premise that ''…weather variables affect an individual's emotional state or mood, which creates a predisposition to engage in particular behaviours'' Howarth and Hoffman, 1984). Some researchers also examined the significance of moods in economic decision-making (Elster, 1998; Loewenstein, 2000; Romer, 2000; Loewenstein et al., 2001; Hanock, 2002).This

paper intends to examine the impact of environment on Indian stock market. The daily market returns on National Stock Exchange (NSE) price index (NIFTY) are regressed against three daily weather observations e.g. minimum temperature, maximum temperatures, and relative humidity.

The weather data for this study is collected from Indian Meteriological Department (IMD), Government of India, and the stock market data is collected from National Stock Exchange (NSE) . The period of study under consideration is 2005 – 2009. A regression based approach is adopted to assess the impact of weather effects on daily stock market returns. The market data employed are the closing prices of NSE Nifty from 1st January, 2005 to 31st December, 2009. The weather data for the same period used in this study is collected from IMD, Govt. of India. Three weather elements are collected on daily basis e.g. minimum temperature, maximum temperature, and relative humidity.

P039- New Technique for Lookback option valuation using an efficient Monte Carlo Approach based on exit probabilities of the Brownian Bridge

Alpesh Chaddha

Indian Institute of Management Calcutta

Lookback options, also known as Hindsight options are a type of path-dependent option where the payoff is dependent on the maximum or minimum asset price over the life of the option. In this paper, I have used the principles of Brownian Bridge to determine the spot asset prices between discrete time intervals of Monte Carlo simulation approach for Lookback Option valuation. Although, the Monte Carlo method has the advantages of ease of coding and robust numerical method, it has the drawback of slow convergence. To reduce the statistical and time error, I suggest the use of Brownian Bridge principle by taking a uniformly distributed random variable and exit probabilities to estimate if the max/min point is reached between 2 discrete time intervals. The basic Monte Carlo simulation path computes the spot price at 2 discrete time instances t_k and t_{k+1}. Now, for the interval t[k,k+1], I calculate the max and min spot price that can be achieved by the Brownian motion path. It can be seen that when we compare the new 'Alpesh Algorithm' method to Monte Carlo technique for the same number of simulations (30,000), it

was seen that the new technique showed faster convergence and smaller absolute error (compared to analytical method) for calculating continuously monitored floating strike call option.

P046- Incorporating Implied volatility in Pricing Options using implied Binomial Tree

K.V.N.M.Ramesh
JNTU Kakinada

Dr. J.V.R.Murthy
JNTU Kakinada

Dr.J.K.R.Sastry
K.L.University

Pricing of exotic options like Asian, American etc, is done by using Binomial Trees. Ramesh-Subhash Model gave a method to price European Options at a confidence using one-step binomial model by considering only the maximum and minimum values that can be taken by the underling at the maturity. An analyst can expect to obtain only a very rough approximation to an option price by assuming that stock price movements during the life of the option consist of one or two binomial steps. In this work the Ramesh-Subhash Model is extended to an N-step Ramesh-Subhash Model so that the model can be used to price various exotic options. A study of the convergence in European option price with respect to number of steps (N) and variation in price of Asian and American options with respect to confidence factor (k) (proxy for implied volatility) are performed. The maximum and minimum boundaries on the value of k are also determined. The N-step Ramesh-Subhash Model gives the investors the ability to change the value of k so that they can have their own opinions concerning the risk-neutral probability distribution.

P016- Analyzing E-readiness of Countries: a Frontier Analysis Approach

Sourabh Bishnoi, Kapil Garg, A. Bose
Birla Institute of Management Technology, Greater Noida, UP 201306

In this paper we use Data Envelopment Analysis to derive efficiency scores for each of the countries. DEA (CCR model) uses linear programming letting each country, DMU (Decision Making Unit) in DEA parlance, choose a set of six weights that maximizes its summated score with respect to its peer DMUs. This, in a way, circumvents the arbitrariness of assigning a preordained set of weights. It is quite possible that one nation might place greater emphasis on "connectivity and technology infrastructure", while a nation that has a relatively good

connectivity infrastructure might place greater emphasis on “consumer and business adoption”. DEA lets each country choose its set of weights, with suitable normalizing constraints.

Comparing the two sets of scores, one provided by EIU and the other computed by us using the DEA model, we classify the 70 countries in one of two classes. Countries for which the two scores are not too different form one category while the countries with significantly different scores are placed in the second category.

The final part of this paper tries to examine the rationale behind the differences and advocates a case for a revised set of weights and consequently a revised set of scores and rankings which more accurately reflect the state of e-readiness of each of the countries.