Collection of Abstracts of Master Theses

Sampling Design Enhancements for the Quarterly Survey of Philippine Business and Industry by Gloria A. Cubinar[1]

This study proposes nonparametric bootstrap and model-based estimation procedures to provide estimates for the Quarterly Survey of Philippine Business and Industry (QSPBI). The current design does not exactly produce a true probability sample. Only "large" establishments that are expected to lead their respective sectors are included in the sample. Nonparametric bootstrap confirms the severe bias in the sample in representing the population of establishments.

Of the four regression models considered in model-based estimation, a generalized linear model with Gamma distribution and logarithmic link function consistently exhibits better fit to the data and provides estimates with high precision for all the variables.

Intensive stratification (industry and employment size), probability proportional to size sampling (pps) and the current design are simulated assuming the results of the 2005 QSPBI first quarter as frame. Estimates from the current design and probability proportional to size sampling with average total employment (ATE) as the measure of size, gave estimates with higher precision especially when the sample size is small. However, as the sample size increases, the behavior of estimates from these sampling strategies do not suffer much.

The non-sampled and nonresponding segments of the population are best predicted through Gamma regression or Time Series Cross Sectional regression.

Initial Development of a Residential Location Choice Model For Metro Manila Using Microsimulated Household Data by Michel Angelo I. Rivera[2]

Metro Manila's urban structure has changed dramatically over the past decades. Coupled with its urban evolution are problems such as population boom, unchecked land development, increased housing cost and longer commutes. The government on the other hand is still far behind in addressing these problems. Given these present conditions, there is an immediate need to analyze and understand the underlying causes that shape the urban structure of the metropolis. Since household location preferences play an important role in urban development patterns, the research focused on household residential location choice behavior.

However, the unavailability of housing and income-related variables and spatial identifiers in available census-based surveys prohibit our research. In addition, the conduct of private household surveys is very expensive. These limitations prompted the use of artificial means of data generation. Specifically, the study made use of spatial microsimulation to create synthetic household micro data that were later used in the modeling process.

[1] Master of Science in Statistics of UP School of Statistics, graduated in 2005
[2] Master of Science in Civil Engineering of UP College of Engineering, graduated in 2005

The research in fact was able to show the usefulness and effectiveness of artificial data in urban modeling development. For example, household bid functions for neighborhood characteristics were successfully made. The bid functions show high sensitivity to land use mix, local household income composition, and accessibility. Accessibility was found to be positively valued by low and middle income households, in contrast to high income-earning households who are willing to bid more for lower density neighborhoods and larger living spaces located farther from the urban core of Metro Manila.

A residential location choice model was also developed using the previously estimated bid functions. The choice model was based on consumer surplus maximization in multinomial logit framework. It was able to show that households locate to places where they maximize their consumer surplus. However, model predicting capacity is still weak and further improvement is recommended.

Nevertheless, it is perceived that this study will serve as a first step in understanding the choice behavior of households in making long-term decisions particularly their choice of home.

A System of Management and Analysis of Rice Genomics Data For Candidate Gene Discovery by Locedie A. Mansueto[3]

An information system to manage and analyze rice genomics data was developed to facilitate discovery of gene functions following the candidate gene strategy. All predicted genes in the rice genome were analyzed using tools to assign putative functions, chromosomal position and expression to treatments. Various biological (QTL, biopathways, germplasm, treatment) and molecular (gene expression, protein-protein interaction, transcriptional regulation, metabolic reactions) evidence of gene activity were used to constrain the gene set. A domain, schema and store independent middleware platform was developed to integrate these data. New evidence can easily be added by mapping API functions or relational queries as modules. An interactive and flexible client application computation in the domain's logic. The system was applied to experimental data generated by our collaboration at the International Rice Research Institute and from public domain databases to identify candidate genes and pathways for drought tolerance in rice.

A Probabilistic Model For Predicting Reliability Performance of Distribution Substation Using Minimal Cut Set Method by Allen C. Gonzales[4]

This study developed a method to predict the reliability performance of distribution substations using minimal cut set method in building reliability network models and basic reliability parameters such as failure rates and repair rates of substation component. Reliability models for four (4) typical designs of distribution substations were developed.

---

[3] Master of Science in Computer Science of UP College of Engineering, graduated in 2005
[4] Master of Science in Electrical Engineering of UP College of Engineering, graduated in 2005

The effectiveness and accurateness of the proposed approach was validated on existing distribution substations of Manila Electric Company (MERALCO) where predicted performance of the substations are compared with their historical performances. The proposed methodology was also applied to a future substation development project of MERALCO with a single breaker-single bus scheme having one transformer back. Predicted reliability performance of this project was compared with the ring bus and breaker-and-a-half substation configurations as alternate scheme.

The models developed in this study have made possible the quantitative reliability comparisons between alternative design schemes of future distribution substation projects that are useful in conducting cost-reliability tradeoff studies.

Seasonal Properties of Selected Philippine Economic Time Series by Jade Eric T. Redoblado[5]

The importance of seasonality in econometric modeling of economic time series has been downplayed by analysts of business cycles. This paper seeks to determine the validity of the assumption of independence between the seasonal and cyclical component for Philippine economic time series.

Some issues relating to seasonal unit roots, changing seasonality, nonlinearity and structural change in Philippine economic time series are addressed. Results from test for seasonal integration and changing seasonality are mixed. However, both graphical measures and prior observations by analysts tend to provide support to the result that most Philippine time series do not exhibit stability in their seasonal patterns. Using a modeling framework accounting for both nonlinearity and structural change, it was found that most economic time series exhibiting linked seasonal and cyclical behavior tend to cater to the domestic market whereas those with independent seasonal and cyclical components may have been insulated by their greater exposure to the global economy.

Sparse Spatial Autoregressive Modeling of Poverty in the Philippines by Ginalyn B. Suratos[6]

Phenomena and systems interact with one another across space and time. Poverty, just like other socio-economic and environmental systems, is then analyzed and defined in relation to its location in a space-time continuum. The sociological perspective is aptly tallied in the spatial dimension, while the economic perspective is addressed in the temporal dimension. This paper aims to postulate a spatio-temporal model that reflects the geographic structure of poverty in the Philippines from 1985 to 2000. The study postulated a spatio-temporal model for selected poverty indicators and compared the SAR (Spatial Autoregression), OLS (Ordinary Least Squares), and Mixed models. The household level information in the Family Income and Expenditure Survey (FIES) data for six time points (1985, 1988, 1991, 1994, 1997, 2000) were aggregated at the provincial level.

---

[5] Master of Statistics of UP School of Statistics, graduated in 2005
[6] Master of Statistics of UP School of Statistics, graduated in 2005

The improvement in predictive capability and dimension reduction relative to other modeling techniques should earn Sparse Spatial Autoregression a recommendation in the field of spatio-temporal modeling. Spatio-temporal predictions are more pronounced using this method, despite its computational complexity. The results of the out-of-sample prediction through simple cross-validation further prove that SAR is far better than OLS in terms of predictive capability.

There is indeed spatio-temporal clustering among the provinces in the Philippines in terms of poverty distribution. The temporal autocorrelation also suggests that poverty remains a stark reality in the Philippines.

Heterogeneity in Long-Term Survivors: Frailty Model with Immunes by Christian Russel D. Reyes[7]

In many registry data, one encounters the problem of many variables with missing values, making the variables unusable. Excluding these variables in the analysis usually results in excess variability or overdispersion. To compensate for this overdispersion, a simple approach is to introduce a new variable called frailty variable.

In breast cancer data, one comes across the problem that some respondents who have censored survival times are totally cured of the breast cancer, which are sometimes referred to as "immunes". That is, the possibility that the cancer will return is remote or even nonexistent. To incorporate this in the model, an "non-immune" parameter p is introduced, such that the new cumulative survival distribution becomes $F(t) = pF_0(t)$, where $F_0(t)$ is the baseline cumulative survival distribution and p is the immune parameter with value between 0 and 1. When $p = 1$, the cumulative survival distribution is just the baseline cumulative survival distribution. When $p < 1$, $F(t)$ will become an improper cumulative survival distribution. The parameter 1-p can also be viewed as the proportion of immunes.

In this paper, a frailty model with immunes is introduced. Properties of this model were investigated using simulated data. It was then applied to Philippine Breast Cancer Registry Data and compared to other models. Other models considered are simple parametric model; presence of immunes with covariates; and simple frailty model. Weibull distribution was used as the distribution of the survival times, while distributions with mean 1 and variance θ were considered for the frailty variable.