Nonparametric Model-Based Predictive Estimation in Survey Sampling

April Anne H. Kwong-Aquino
University of the Philippines Visayas

A nonparametric model-based estimator of the population total is proposed. The sample data along with the auxiliary information are used in fitting a generalized additive model that is then used in reconstructing the unknown population. The estimates of the population parameters are computed from the predicted population values (for the unsampled part of the population) and the sample values. A simulation study designed to account for different association patterns between the target variable and the auxiliary variable, population size, and sample size was conducted to evaluate the proposed procedure. The method is robust to data-generating model form, population size, and sampling rate, and is generally superior to design-unbiased estimators.

Keywords: model-based estimation, predictive estimation, nonparametric regression, additive model

1. Introduction

In survey data analysis, one of the primary objectives is the estimation of population characteristics and often, design-unbiased, model-assisted, or model-based techniques are used. Design-unbiased methods of estimation are dependent on the sampling distribution induced by the sample selection process where knowledge of the population structure is very important. Design-based estimation requires availability of population frame and sampling weights (alternatively, selection probabilities). Oftentimes, complete and reliable information about the population needed in frame construction is difficult to secure. One way to address this problem is to use model-based estimation techniques. Model-based estimation procedure does not completely depend on the population frame and does not require knowledge of weights to account for the unsampled segment of the population. Instead, the auxiliary information is used to predict the unsampled values. This method requires the estimation of the relationship between the two variables by fitting a regression equation using the pairs (x_i, y_i), for $i \in S$, (Barrios, 2007). Model-assisted approach integrates the design-unbiased and model-based approaches. Estimation does not rely on model assumptions and inferences are based on the survey design alone, however, models are used to specify the parameters of interest (Lohr, 1999). (Rueda and Sanchez-Borrego, 2009) noted the advantage of model-based estimation when there is a strong linear relationship between the target and auxiliary variables.

Model-based inferences have been evaluated using both parametric and nonparametric models. Parametric models necessitate that model assumptions are met for the inferences to be valid. Under this approach, the accuracy of the specified model is a major concern. Also, conditions like linearity, normality and independence of the error terms must be satisfied. However, deviations from these assumptions are to be expected. These deviations and misspecifications lead to erroneous inferences. Robustness of estimation procedures to model assumptions is, thus, desirable. The use of nonparametric methods can possibly achieve this as they allow the data to dictate the relationship curve of the variables (Eubank, 1998). This is an appropriate alternative when there is little a priori information on the structure of the relationship or even when there is doubt about the validity of the parametric model.

Barrios (2007) proposed a model-based estimation technique in estimating the population total for variables which are linearly related to an auxiliary variable. The estimator is given by $\hat{T} = \sum_{i \in S} y_i + \sum_{j \notin S} \hat{y}_j$ where y_i are the sample values and $\hat{y}_j$ are the predicted values of the unsampled population units. Simulation scenarios that included symmetric and skewed populations were conducted. The normal regression model was used for the symmetric population while poisson regression with log link function was used for the skewed population. The proposed estimation procedure yielded comparable results to design-based estimation for small and large populations.

An estimator of the population mean was proposed by Rueda and Sanchez-Borrego (2009) using local polynomial regression and was compared to several existing methods of estimation using simulated populations. The estimator of the population mean is given by $\bar{y}_{MB} = f\bar{y}_S + (1-f)\frac{1}{N-n}\sum_{j \in \bar{S}} \hat{m}_j$ where $\bar{y}_S = \frac{1}{n}\sum_{i \in S} y_i$ is the sample mean and $\hat{m}_j$ are the predicted values of the unsampled part of the population. The $\hat{m}_j$ are computed using a local polynomial regression model that was fitted to the sample data on the pairs (x_i, y_i). The proposed method exhibited satisfactory performance relative to design-based estimators as the sample size decreases.

The combination of the concept of model-based estimation and nonparametric methods promises several advantages. Sampling and estimation can be done even without a reliable and complete population frame. Prior information on the attributes of the population is also not a primary concern.

2. Estimation of the Population Total

Assume that the response variable Y is related to an auxiliary variable whose census is known. Assume further that the values of Y corresponding to the maximum and minimum values of X are known or can at least be estimated prior to analysis. The variable of interest (Y) and the auxiliary variable (X) are assumed to be related through the function $y = f(x) + \varepsilon$. The following algorithm is proposed to estimate the population total of Y.

1. Given that a census on X is available, obtain sample information on Y and X (S).
2. Estimate $f(x)$ through a generalized additive model to the pairs (x_i, y_i), $i \in S$.
3. Predict the unsampled part of the population using the estimated generalized additive model, i.e., $\hat{y}_j = \hat{f}(x)$ for $j \notin S$.
4. Combine the sample values y_i, $i \in S$ and predicted values $\hat{y}_j$, to recreate the population.
5. Estimate the population total using $\hat{T} = \sum_{i \in S} y_i + \sum_{j \notin S} \hat{y}_j$.

The estimator is called Nonparametric Model-Based Estimator (NMBE).

3. Estimation of the Standard Error of $\hat{T}$

The standard error of the proposed estimator $\hat{T} = \sum_{i \in S} y_i + \sum_{j \notin S} \hat{y}_j$ can be computed using the bootstrap. Two hundred replicates are selected from the recreated population given a sampling rate of 50%. The statistic $\tilde{T}_i = N\overline{Y}_i$ for every $i = 1, 2, \ldots, 200$ is computed along with $\hat{\sigma}(\tilde{T}) = \sqrt{\dfrac{\sum_{i=1}^{n}\left(\tilde{T}_i - \overline{\tilde{T}}\right)^2}{n-1}}$ where $\overline{\tilde{T}}$ is the mean of the $\tilde{T}_i$'s and n is the number of replicates (200). The computed value of $\hat{\sigma}(\tilde{T})$ estimates the standard error of $\hat{T}$. This is compared to the estimate of the standard error of the design-unbiased estimate of the population total, $\hat{\sigma}_{SRS} = \sqrt{\left(\dfrac{N-n}{N-1}\right)}\dfrac{Ns}{\sqrt{n}}$.

4. Simulation Study

The algorithm above is implemented on several simulation settings to evaluate the performance of the method. Each scenario postulates a model $y = f(x) + k * \varepsilon$, where $f(x)$ is either linear or nonlinear in x. Populations of variables exhibiting linear, quadratic, and exponential relationships are simulated. The equations used are:

Linear	$y = 1.35x + k * e$
Quadratic	$y = -0.1x^2 + 10x + k * e$
Exponential	$y = 10\exp(x/25) + k * e$

These relationships represent some possible association patterns between variables. The constant multiplier (k) to the error term is used to induce varying degrees of model misspecification, small k implies no or minimal misspecification, while large k implies severe misspecification. For each model above, population size, sampling rate, and multiplier on the error term were varied. Small and large finite populations are represented by populations of sizes $N = 1000$ and $N = 10000$, respectively, from where random samples are taken given the sampling rates: 1%, 3%, 5%, 10%, and 20%. Error terms are assumed to follow the standard normal distribution with multipliers (k) set to 1, 5, 8.75 and 10. For $k > 1$, the error is bloated, destroying the fit of $f(x)$ to the data. This introduces additional variability in Y that is explained by components other than $f(x)$ and lessens the capacity of X to predict the values of Y. Additional considerations include setting the coefficient of X in the linear equation to 1.35 in order to generate data values with $r \approx 0.95$ for $k = 1$, $r \approx 0.50$ for $k = 5$ and $r \approx 0.30$ for $k = 8.75$ when variance of X is set at 5. These values of the correlation coefficient correspond to strong, average, and weak linear relationships, respectively. The auxiliary variable, X, was simulated to follow the normal distribution with mean equal to 50. To investigate the effect of the variance of X on the efficiency of the estimates, variances are set to 5, 25 and 225, and 400.

The frame problem posed by some population units not being accessible to sampling is also considered. For each simulation scenario presented above, NMBE estimates using samples from the middle 50% of the population are also calculated. Under this case, the unsampled segment of the population includes the lower and upper 25% as well as the units in the middle 50% that were not captured during sampling.

To evaluate the performance of the proposed estimator, the absolute percent difference is computed from $PD_{est} = \left(abs\left(\frac{\hat{T}_{est} - T}{T} \right) \right) * 100\%$, where T is the true population total. The percentage advantage of NMBE estimates over design-unbiased estimates is further computed from $PA = \frac{\overline{P}D_{SRS} - \overline{P}D_{NMBE}}{\overline{P}D_{SRS}} * 100\%$, where $\overline{P}D_{NMBE}$

is the average absolute percent difference between the NMBE estimate and the true population total and $\overline{P}D_{SRS}$ is the average absolute percent difference between the SRS estimate and the true population total.

5. Results and Discussions

The different scenarios in the simulation study are considered in order to vary the amount of variability in the population as measured by the coefficient of variation. This characteristic of the population has an effect on the performance and even on the choice of estimation procedures. High coefficients of variation indicate high heterogeneity of population values. For such cases, estimation can be more complex and costly as a large sample is needed to ensure that the variability pattern is captured adequately by the sample. On the contrary, a small sample would suffice to represent a population with low coefficients of variation.

We summarized in Tables 1-3 some values of population parameters resulting from the restrictions imposed during the simulation for $N = 1000$. Similar values are generated for $N = 10000$.

Table 1 Coefficient of Variation (CV) of $Y = 1.35X + k*e$ and Correlation Coefficient Across Varying Values of $Var(X)$ and k

$Var(X)$	5			25			225			400		
k	1	5	8.75	1	5	8.75	1	5	8.75	1	5	8.75
r	0.95	0.50	0.30	0.99	0.80	0.60	0.99	0.97	0.92	0.99	0.98	0.95
CV of Y	4.63	8.34	13.24	10.01	12.09	15.81	29.95	30.59	32.14	40.01	40.44	41.58

Table 2 Coefficient of Variation (CV) of $Y = -0.1X^2 + 10X + k*e$ Across Varying Values of $Var(X)$ and k

$Var(X)$	5			25			225			400		
k	1	5	10	1	5	10	1	5	10	1	5	10
CV of Y	0.48	2.04	4.06	1.46	2.45	4.28	13.81	13.91	14.36	26.61	26.64	26.86

Table 3 Coefficient of Variation (CV) of $Y = 10\exp(x/25) + k*e$ Across Varying Values of $Var(X)$ and k

$Var(X)$	5			25			225			400		
k	1	5	10	1	5	10	1	5	10	1	5	10
CV of Y	9.00	211.45	16.61	19.92	21.34	24.66	62.62	63.31	64.63	87.33	87.90	88.85

5.1 Effect of model form

We summarized in Table 4 the average absolute percent differences of both the design-unbiased (SRS) and NMBE estimates from the true population total over different model forms. The percentage advantage (*PA*) of NMBE over SRS is also summarized in Table 5. This is evaluated only for NMBE estimates that utilized the entire population.

The average absolute percent differences of NMBE estimates from the true population total are lower than that of the design-unbiased estimates for all model forms and across sampling rates. Moreover, the contrast becomes clearer as the sampling rate decreases. With the linear association, the NMBE estimates are better for sampling rates up to 5% and comparable to the design-unbiased estimate for much higher sampling rates. The same trend can be seen for the quadratic model where the NMBE is better for sampling rate 1% only and exhibited performance similar to the design-unbiased estimator for sampling rates 3% up to 20%. For the exponential function, the proposed estimator is superior for all sampling rates.

The dissimilarity in the percent differences of the estimates is larger for the linear and exponential model forms with NMBE estimates exhibiting advantage over design-unbiased estimates. These patterns can be attributed more to the coefficients of variation of the population than to the model form. For the linear and the exponential case, coefficients of variation range from 5% to 89%. On the other hand, resulting coefficients of variation for the quadratic relationship are from 0.5% to 27% only. Coefficients of variation indicate the level of homogeneity of the population values. SRS is more appropriate for homogeneous populations, that is, populations with low coefficients of variation. Hence, they are expected to perform well in such scenarios. The design, however, may fail to give accurate estimates for heterogeneous populations or populations with high coefficients of variation. By inspection of the values, the NMBE estimates seem to be unaffected by this population characteristic.

The model forms considered in this paper yield different types of population characteristic in relation to symmetry. The linear form outputs symmetric data while the quadratic form and exponential form introduce negative skewness and positive skewness, respectively. The percentage advantage of NMBE over the SRS-unbiased estimator does not change much with respect to model form. This indicates that the model forms, and as a consequence type of skewness, does not necessarily determine the cases for which NMBE estimates are more superior.

As the sample size increases, the accuracy of SRS estimates also increases. The NMBE estimates, however, are relatively robust to sample size changes. In addition, the average absolute percent differences of NMBE estimates are significantly lower than that of design-based estimates for smaller samples, highlighting the advantage

Table 4 **Average Absolute Percent Difference of the Estimates With Respect to Model Form**

	1%			3%			5%			10%			20%		
	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)
Linear	4.09	1.30	0.97	2.30	0.72	0.50	1.75	0.53	0.41	1.24	0.37	0.31	0.73	0.24	0.25
Quadratic	2.02	0.57	0.78	1.12	0.32	0.55	0.94	0.27	0.44	0.60	0.16	0.25	0.33	0.10	0.22
Exponential	8.14	1.67	5.69	4.52	0.97	5.16	3.30	0.69	4.81	2.29	0.48	4.32	1.49	0.30	4.36

Table 5 **Percentage Advantage of NMBE Estimate over SRS-unbiased Estimate With Respect to Model Form**

	1%	3%	5%	10%	20%
Linear	68.192	68.823	69.903	70.129	67.123
Quadratic	71.726	71.505	71.868	72.833	69.970
Exponential	79.496	78.526	78.963	79.214	79.584

of NMBE in small samples as also pointed out in (Rueda and Sanchez-Borrego, 2009).

For the Linear model, it can be observed that the percent differences of the NMBE estimates from the true population total when sampling is only from the middle 50% of the population is lower when compared to the percent differences of the other two estimates (SRS and NMBE) which are based on the entire population. The auxiliary variable *X* was simulated to follow the normal distribution. Thus, the Linear model $y = 1.35x + k * e$ also yielded a symmetric distribution for Y. Due to the symmetry of *Y*, sampling on the middle 50% decreased the percent difference of the estimate. When only the middle 50% of the population is sampled, the likelihood that values in the neighborhood of the mean are selected increases. As a result, the estimates are also more accurate.

The quadratic model ($y = -0.1x^2 + 10x + k * e$) and exponential model ($y = 10\exp(x/25) + k * e$) resulted to a skewed distribution for *Y*. Sampling on the middle 50% yields less accurate estimates as extreme values that caused the skewness in *Y* did not have representation in the sample of paired observations (x_i, y_i) .

5.2 Effect of variance of X

We summarized in Table 6 the average absolute percent differences of both design-unbiased (SRS) and NMBE estimates from the true population total over varying variance of *X*. The percentage advantage of NMBE (considering the entire population) over SRS are presented in Table 7.

From Table 6, there are slight changes in the NMBE estimates across different variances of *X*. However, estimates of SRS are highly variable with respect to these adjustments. There is a direct proportional relationship between the variances of the auxiliary and the target variable, an increase in the variation of *X* adds to the heterogeneity of the population of *Y* and as a consequence, affects the performance of SRS estimates. As it increases further, NMBE estimates are more accurate than SRS estimates even as the sampling rate becomes much larger.

The percentage advantage of NMBE does not change much with respect to sampling rate. However, there is an apparent increase in that measure as more variability is introduced to the auxiliary variable. By increasing the variance in *X* while holding all other simulation parameters constant, proportion of variability in *f(X)* that accounts for the variability in *Y* increases. This is tantamount to increasing the predictive ability of the auxiliary variable thus resulting to an increase in the advantage of the proposed estimator.

For low variations in *X*, the simulated frame problem (only the middle 50% of the population is accessible to sampling) actually yields better NMBE estimates. The variability in *Y* proportionally changes with respect to the variability in *X*. When

Table 6 Average Absolute Percent Difference of the Estimates With Respect to Variance of *X*

		1%			3%			5%			10%			20%	
	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)
5	1.370	0.936	0.510	0.749	0.53	0.272	0.55	0.388	0.208	0.394	0.279	0.141	0.255	0.186	0.092
25	2.222	0.929	0.714	1.198	0.529	0.411	0.903	0.387	0.302	0.633	0.277	0.227	0.397	0.186	0.155
225	6.349	1.141	2.800	3.543	0.654	2.388	2.683	0.487	2.157	1.843	0.329	1.826	1.146	0.224	1.734
400	9.042	1.712	5.886	5.091	0.961	5.220	3.855	0.719	4.882	2.667	0.461	4.309	1.651	0.262	4.451

Table 7 Percentage Advantage of NMBE Estimate over SRS-unbiased Estimate With Respect to Var(*X*)

	1%	3%5%	10%	20%	
5	31.679	29.239	29.455	29.188	27.059
25	58.191	55.843	57.143	56.240	53.149
225	82.029	81.541	81.849	82.149	80.45
400	81.066	81.124	81.349	82.715	84.131

there is little variation in X and when the Y values also do not vary much, sampling from only the middle 50% is sufficient to acquire a representative of the population. Given a small variance of X and Y, this scheme is actually advantageous over sampling from the entire population. There is higher probability of sampling around the mean when we focus sampling on the middle 50% than when the entire population is considered. With the latter case, the extreme cases (which are uncommon when the variation is low) still have a nonzero probability of inclusion. When captured during sampling, these extreme values can pull the sample mean in either direction depending on the direction of the extreme case, if it is significantly higher or lower than the mean, resulting to bias in the estimation. Sampling from the middle 50% removes this possible source of inaccuracy. However, the same sampling restriction will not yield desirable results if the population variability is high as can be seen in Table 6 for variances of X equal to 225 and 400. The percent difference of the estimates from the true population total significantly increased for higher levels of heterogeneity in X and, as a consequence, in Y. Since the spread of the population values is wide, sampling from only the middle 50% will not be enough to get a good representative of the population. For the symmetric case with high variability, density of the values will not vary much as you move in either direction away from the mean. Sampling from the middle 50% disregards this and assigns zero inclusion probability to values which are farther from the mean. This will result to lesser accuracy of the estimates.

5.3 Effect of population size

Table 8 shows the average absolute percent differences of both design-unbiased (SRS) and NMBE estimates from the true population total with respect to population size. The percentage advantage of NMBE estimates (sampling over the entire population) over SRS estimates are shown in Table 9.

Simulated populations that vary with respect to population size alone have the same amount of variation. Hence, the same sample size irrespective of the population size will result to similar inferences. This can be observed for the estimates given the sampling rate of 10% for population size of 1000 and for the estimates given the sampling rate 1% for population size of 10000. The absolute percent differences of SRS estimates are similar (2.089 for $N = 1000$ and 2.263 for $N = 10000$). The same is true for the NMBE estimates (0.509 for $N = 1000$ and 0.630 for $N = 10000$).

Regardless of the population size, NMBE is more advantageous than SRS (Table 9). Moreover, there are no notable differences in the percentage advantage of NMBE over the SRS-unbiased estimator considering changes in the population size. This indicates that population size does not contribute to the efficiency of the NMBE estimate relative to the efficiency of the SRS-unbiased estimate.

Even though sampling is only from the middle 50%, NMBE still performed better relative to SRS-unbiased estimation (based on the entire population) for small

Table 8 Average Absolute Percent Difference of the Estimates With Respect to Population Size

	1%			3%			5%			10%			20%		
	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)
1000	7.13	1.729	2.401	4.054	0.991	1.939	2.967	0.743	1.778	2.089	0.509	1.621	1.275	0.317	1.404
10000	2.362	0.630	2.555	1.236	0.347	2.207	1.028	0.248	1.996	0.68	0.164	1.630	0.45	0.112	1.812

Table 9 Percentage Advantage of NMBE Estimate over SRS-unbiased Estimate With Respect to Population Size

	1%	3%	5%	10%	20%
1000	75.750	75.555	74.958	75.634	75.137
10000	73.328	71.926	75.875	75.882	75.111

samples from the smaller population. For the larger population, the performance of the NMBE estimates and that of the SRS-based estimates are comparable when sampling rate is at least 10%. However, for other sampling rates, the design-unbiased estimates have lower percent differences.

5.4 Effect of model fit

Presented in Table 10 are the average absolute percent differences of both design-unbiased (SRS) and NMBE estimates from the population total with respect to the model fit. The percentage advantage of NMBE estimates (using the entire population) over SRS estimates are shown in Table 11.

There are slight changes in the average absolute percent difference of both SRS and NMBE estimates with respect to the increases in the error multiplier *k*. Increasing *k* decreases the amount of variation in *Y* that is explained by *f(X)*. The NMBE estimates are model dependent, thus, their performance is affected by changes in the model fit. For $k = 1$, NMBE estimates yield better performance for sampling rates of at most 10% while for larger values of *k* (greater than 5), NBME estimates are more accurate for sampling rates of at most 5%.

The percent difference of the NMBE estimates from the actual population total increases as the error multiplier also increases. These estimates are model-based and any changes in the capacity of the auxiliary variable *X* to predict values of *Y* will affect their accuracy. These changes are incorporated during simulation through the error multiplier *k*. Increasing k decreases the prediction capability of *X* and as a consequence, increases the percent difference of the estimate from the population value. It can be observed that the increments are larger when the sampling rate is small suggesting that increasing the sample size reduces the effect of the error multiplier.

There is no clear pattern in the percentage advantage of NMBE estimates with respect to the sampling rates. Model-dependent methods rely on the relationship between variable of interest and auxiliary variable to capture the trends in the data. They usually fail as estimation tools when model does not fit the data well (Kalton, 2002). This stresses the importance of that assumption even when using nonparametric methods. While they might be robust to model assumptions, the minimum requirement of an association between the variables *X* and *Y* still needs to be met. Otherwise, the performance of nonparametric model-based techniques will not be optimal.

For sampling from only the middle 50%, NMBE still performed better across all the given values of the error multiplier for sampling rate of 1%. The completely random nature of SRS sampling translates to better representation of the population when the samples are large. Thus for small sample cases, the performance of NMBE

Table 10 Average Absolute Percent Difference of the Estimates With Respect to the Value of the Error Multiplier, *k*

	1%			3%			5%			10%			20%		
	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)	SRS	NMBE	NMBE (50%)
1	4.46	0.572	1.887	2.516	0.326	1.682	1.895	0.242	1.552	1.311	0.156	1.339	0.810	0.087	1.847
5	4.697	1.142	2.281	2.611	0.646	1.999	1.99	0.480	1.846	1.367	0.324	1.631	0.851	0.206	1.358
>8.75	5.081	1.825	2.877	2.808	1.035	2.293	2.123	0.764	2.041	1.476	0.530	1.714	0.927	0.350	1.428

Table 11 Percentage Advantage of NMBE Estimate over SRS-unbiased Estimate With Respect to the Value of the Error Multiplier, *k*

	1%	3%	5%	10%	20%
1	87.175	87.043	87.230	77.101	89.259
5	75.687	75.259	75.696	76.298	75.793
>8.75	64.082	63.141	64.013	64.092	62.244

relies heavily on the relationship between *X* and *Y*. For all the other scenarios, NMBE estimates using only the middle 50% of the population are comparable to SRS-estimates based on the entire population.

5.5 Standard error of NMBE estimates

For each simulation scenario, the estimated standard error and coefficient of variation of the NMBE estimate is lower than that of the SRS estimate. According to Lohr (1999), the standard errors of estimates under model based approaches are generally lower than those of design based estimates. This can be attributed to the differences in the way the standard errors are computed. The variance of model-based estimates is the average squared deviation of the estimate from its expected value with the average computed over all possible values that the model can generate. While for the design-based estimates, the average is over all possible samples using the specified design.

The estimate of the standard error and, correspondingly, the coefficient of variation (CV) of the design-unbiased estimate is a function of the sample variance and the sampling rate. The decreasing trend in the CV as the sample size increases can be observed in the results. However, the effect of the sample size on the CV of the proposed estimator is negligible. The coefficients of variation of the latter estimator also exhibited robustness to the variation in the auxiliary variable and model fit.

6. Conclusions

The performance of the nonparametric model-based estimates (NMBE) is at the least comparable to SRS estimates. The NMBE estimates are superior to design-unbiased (SRS) estimates when the population is very heterogeneous and when a high proportion of variation in *Y* is accounted for by variation in *f(X)*, i.e., when there is a relationship between Y and X. NMBE also provide an alternative strategy in estimating population total where sampling suffers from severe frame problem.

REFERENCES

BARRIOS, E., 2007, Model Based Predictive Estimation with Coverage Error, Bulletin of the 57th Session of the International Statistical Institute, Portugal.

RUEDA, M. and SANCHEZ-BORREGO I.R., 2009, A Predictive Estimator of Finite Population Mean using Nonparametric Regression, *Comput Stat,* 24:1-14.

KALTON, G., 2002, Models in the Practice of Survey Sampling (Revisited), *Journal of Official Statistics* 18:129-154.

LOHR, S., 1999, *Sampling: Design and Analysis*. Brooks/Cole.

EUBANK, R., 1998, *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, Inc.

Investigating the Efficiency of Stratified Ranked Set Sampling Using Nonparametric Bootstrap Estimation

Kevin Carl P. Santos
University of the Philippines Diliman

Jenniebie C. Salagubang
Brady Philippines Direct Marketing Inc.

This paper aims to compare stratified random sampling and stratified ranked set sampling. A simulation study was conducted to evaluate the performance of the parameter estimates on both sampling techniques. Population sizes, sampling rates, stratum sizes, and correlation of the target variable and concomitant variable were varied, nonparametric bootstrap was then used in estimating the mean and its standard error. The coefficient of variation (CV) and the bias of the bootstrap estimates were compared. Stratified ranked set sampling generally outperforms stratified random sampling in terms of bias most especially for small populations. The two sampling designs were used in estimating the average mango production per *barangay* in the country.

Keywords: ranked set sampling, nonparametric bootstrap estimation, stratification, simple random sampling

1. Introduction

Sampling has been very essential in different areas of discipline. However, it becomes critical when the target population is very large and heterogeneous with respect to the variable of interest. To address the problem of heterogeneity, the population is subdivided into non-overlapping groups called strata. Then, independent samples are obtained in each stratum. These subsamples comprise the sample from the population of interest.

Stratification is a common strategy for large and heterogeneous populations. One reason for this as stated by Cochran (1977) is that gain in precision in the estimates can be obtained from stratification. The idea behind this technique is to divide the population into strata such that each stratum is homogenous within but they should be heterogeneous across each other. Thus, it is expected that precise estimates will be computed in each stratum and combining these weighted estimates across strata will give a better estimate for the parameter of interest.

Another reason in resorting to stratification as Kish (1965) mentioned is that strata may serve as domains of the study. This means that there can be acceptable estimates not only in the whole population but also in each stratum. Lohr (1999) added that stratification may lead to convenience in administering the survey and may result to lower costs.

As Kish (1965) pointed out, strata may be formed to employ different methods and procedures within them. The most common sampling design used in obtaining sample in each stratum is simple random sampling without replacement (SRSWOR). This is called Stratified Random Sampling (SSRS).

Furthermore, the most crucial issue in stratification is the choice of the stratification variable. This variable should stratify the population into strata which are homogenous within but heterogeneous across each other. According to Cochran (1977), the ideal variable to be used in stratification is the variable of interest itself. He also pointed out that using the variable of interest itself as stratification variable would result to non-overlapping strata and the variability within strata would be low. However, the target variable is usually unknown in practice. A solution to this problem is to look for an auxiliary variable, which is readily available and is believed to be correlated with the variable of interest, that will be a good stratification variable.

Another issue in stratification is the allocation of the sample sizes in the strata. There are three common allocation rules, namely: equal, proportional, and optimal allocations. Equal allocation is rarely used because of its impracticality. Proportional allocation takes into account the stratum sizes in allocating sample sizes while optimal allocation ensures that the variability and cost in each stratum would be included in the sample size determination. It was shown by Cochran (1977) that if the reciprocals of the stratum sizes are ignored relative to unity, the variance of the estimator of the mean under optimal allocation is the lowest compared to that under proportional allocation and under SRSWOR.

Another statistical technique for data collection that is of interest in this paper is Ranked Set Sampling. This was first introduced by McIntyre (1952) when it was difficult to take the actual measurements for sample observations. He wanted to gain precision in estimating the average yield from the large plots of arable crops. The measurements he needed were costly and tedious to collect. This led him to this approach in data collection.

McIntyre was not able to provide mathematical proofs of the optimal properties of the estimator of the population mean using RSS. Takahasi and Wakimoto (1968) provided the mathematical foundations of RSS showing that its estimator for the mean is unbiased even with different distributional assumptions under the assumption of perfect ranking. In addition to that, the RSS estimator of the mean is more efficient than that of SRS. It was shown that the relative precision (RP) of RSS compared to SRS estimator of the population mean is:

$$1 \le RP = \frac{Var(\bar{X}_{SRS})}{Var(\bar{X}_{RSS})} \le \frac{k+1}{2},$$

where k is the set size. This means that as the set size k increases, the RP also increases. Hence, increasing the set size will result to a more reliable estimate for RSS compared to that of SRS. Nevertheless, it should be noted that taking a large set size would entail higher costs in obtaining samples. This is the reason why k is not usually high in practice.

The ranking criterion is the most important issue in RSS. Before, ranking was done visually or through eye inspection but because of ranking errors, RSS might be worse than other sampling designs. Similar to sampling with probability proportional to size (PPS), an auxiliary or concomitant variable can be used to rank the sampling units. This auxiliary variable should be highly and directly associated with the characteristic of interest or should be a frugal measure of the variable of interest. RSS was extended to ranking using a concomitant variable by Stokes (1977). He concluded that the reliability of the estimates depends on the degree of the relationship of the two variables.

Chen (2007) enumerated applications of RSS in many areas of discipline such as agriculture, environment and ecology, medicine, and genetics. Samawi and Muttlak (1996) introduced the so-called Extreme RSS. This has been applied in genetics for quantitative trait loci (QTL) mapping to measure obesity and cholesterol level. Chen and Wang (2004) used RSS in studying lung cancer. They investigated how lung cancer is affected by smoking through the use of bio-markers. Another novel application of RSS is in comparing treatments in experiments which include many clinical trials.

Since it was shown that RSS is better than SRS in estimating the population mean in terms of efficiency, this paper aims to determine whether Stratified Ranked Set Sampling (SRSS) will perform better than Stratified Random Sampling (SSRS). SRSS performs ranked set sampling in each stratum independently and combining these RSSs will comprise the sample from the whole population.

Ibrahim et al. (2010) already compared SRSS, RSS and SSRS with Stratified Median Ranked Set Sampling (SMRSS). They have shown that SMRSS estimator is an unbiased estimator of the population mean when the population is symmetric and is more efficient in estimating the population mean compared to the other abovementioned sampling designs. However, their simulation study did not consider different stratum sizes in the population and different degrees of correlation of the auxiliary variable used in ranking. Additionally, their sample sizes are very small (*n* = 7, 12, 14, 15, 18). These sample sizes will still be allocated to 2-3 strata in their study. This study wants to investigate these things since Ibrahim et al. did not take these into account. Another difference of this paper and theirs is the assumption of the distribution of the population in the simulation study. This simulation study is limited to normal populations only while Ibrahim et al. investigated other distributions as well.

Different scenarios in estimating the population mean are considered in the simulations done in this study. Section 2 discusses the details of the different sampling designs. Section 3 elaborates on the different cases taken into account in the simulation. Section 4 shows the results of the simulations while Section 5 includes conclusions and directions for future research.

2. Sampling Designs

This section gives a discussion on the sampling schemes that were used in the paper.

2.1. Simple Random Sampling without Replacement (SRSWOR)

Simple Random Sampling (SRS) is the simplest form of probability sampling design. A simple random sample of size *n* is taken from the population wherein all possible samples of size *n* are given the same chance of selection. SRS can be done with replacement or without replacement.

In Simple Random Sampling without replacement (SRSWOR), each possible combination of *n* different elements out of *N* has the same chance of being selected in the sample. To obtain an SRSWOR, each element in the sampling frame will be assigned with a unique number. Afterwards, *n* distinct numbers will be drawn using a random process. The elements associated with the distinct numbers will comprise the sample. The table below shows how the mean is estimated using SRSWOR.

Table 2.1 SRSWOR Estimator of Population Mean, its Variance and Estimator of the Standard Error

Estimator	Variance	Estimated Standard Error
$\bar{y}_{SRSWOR} = \frac{1}{n}\sum_{i=1}^{n} y_i$	$\frac{S^2}{n}\frac{N-n}{N}$	$\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$ where s^2 is the sample variance

The sample mean $\bar{y}_{SRSWOR}$ is an unbiased and consistent estimator of the population mean. The estimator of the variance of the sample mean is also unbiased. However, their square roots or the estimators of the standard errors are slightly biased for the true standard error of the sample mean. This sampling design is used in obtained independent samples in each stratum in SSRS.

2.2. Ranked Set Sampling

Ranked Set Sampling (RSS) requires drawing a simple random sample (SRS) of size k from a sampling frame. Afterwards, these k elements will be ranked using either visual inspection or a readily available concomitant variable. As a consequence, the sample frame should also include auxiliary information that will be used to rank the elements. It must be noted that these k units are not necessarily obtained physically. For instance, a mapping can be done initially on the concomitant variable before actually measuring the variable of interest.

The first order statistic from the ranked units will be the first unit in the sample. Then, another SRS of size k will be drawn and these will be ranked again. The second order statistic from this set will be the second unit in the sample. This is repeated until the k^{th} order statistic is obtained for the k^{th} batch of SRS. This whole process is called a cycle and k is called the set size. If a cycle would be repeated m times, the total number of units in the sample is $n = mk$.

Table 2.2 shows the RSS estimator of the mean. It is an unbiased estimator of the population mean and its variance is smaller than that of SRS. The estimator of the variance of the sample mean under RSS is biased but the bias is a function of the set size and number of cycles. Hence, as the set size or number of cycles increases, the bias is expected to be negligible as stated by Chen et al. (2004).

Table 2.2 RSS Estimator of Population Mean, its Variance and Estimator of the Standard Error

Estimator	Variance	Estimated Standard Error
$\bar{y}_{RSS} = \frac{1}{mk}\sum_{r=1}^{k}\sum_{i=1}^{m} y_{[r]i}$	$\frac{\sigma^2}{mk} - \frac{1}{m^2 r}\sum_{i=1}^{m}(\mu_{[i:m]} - \mu)^2$	$\sqrt{\frac{1}{mk-1}\sum_{r=1}^{k}\sum_{i=1}^{m}(y_{[r]i} - \bar{y})^2}$

2.3 Stratified Simple Random Sampling

When the population is heterogeneous with respect to the variable of interest, SRS should not be used because its standard error is expected to be large. The idea is that there exists a variable that can divide the population to obtain l homogeneous sub-groupings. This variable is called the stratification variable. In which case, SRS can be used in each homogenous grouping. This is the so-called Stratified Simple Random Sampling (SSRS) or simply Stratified Random Sampling.

In this sampling design, SRSWOR is conducted in obtaining samples in each stratum. Sampling is done independently across strata. The estimator of the mean is simply the weighted mean of the stratum means where the weight is the ratio of the stratum size to the population size. Due to the independence of sampling across strata, the variance of the estimator under SSRS is simply the weighted variances of each stratum under SRSWOR as shown in Table 2.3. It is expected that the variance of $\bar{y}_{SSRS}$ is lower than that of SRS when the population is very heterogeneous and an appropriate stratification variable is chosen.

Table 2.3 SSRS Estimator of Population Mean, its Variance and Estimator of the Standard Error

Estimator	Variance	Estimated Standard Error
$\bar{y}_{SSRS} = \sum_{h=1}^{l} \frac{N_h}{N} \bar{y}_h$	$\sum_{h=1}^{l} W_h^{\ 2} \frac{S_h^2}{n_h}\left(1-\frac{n_h}{N_h}\right)$	$\sqrt{\sum_{h=1}^{l} W_h^{\ 2} \frac{s_h^2}{n_h}\left(1-\frac{n_h}{N_h}\right)}$
	where $W_h = \frac{n_h}{N_h}$	

Another issue in stratification is the sample size determination in each stratum. Some of the common allocation rules are equal, proportional, and optimal allocation. The allocation rule used in this study is limited to proportional allocation. This method allocates the samples proportional to size of the stratum. In addition to that, it was shown in Cochran (1977) that the variance of the estimator of the sample mean under proportional allocation is smaller than that using SRSWOR.

2.4 Stratified Ranked Set Sampling

Since stratification is used to homogenize the population into strata and, as shown by Patil (2002) and Wolfe (2004), the variance of the estimator of the mean under RSS is lower than SRS, the authors wanted to investigate the reliability of the estimates when RSS is used to obtain samples in each stratum instead of using SRSWOR.

The process discussed in Section 2.2 is performed in each stratum and the process is repeated *m* times. The derivations of the estimator of the mean and its variance are similar to that of SSRS. The estimator of the mean is simply the weighted RSS estimator in each stratum where the weight is the ratio of the stratum size to the population size. The variance of $\bar{y}_{SRSS}$ and its estimator are derived similarly by substituting the variance of $\bar{y}_{RSS}$.

Table 2.4 SRSS Estimator of Population Mean, its Variance and Estimator of the Standard Error

Estimator	Variance	Estimated Standard Error
$\bar{y}_{SRSS} = \frac{1}{mk}\sum_{h=1}^{l}\sum_{r=1}^{k}\sum_{i=1}^{m}\frac{N_h}{N} y_{[r]ih}$	$\sum_{h=1}^{l} W_h^2 \frac{\sigma_h^2}{m_h k_h} - \sum_{h=1}^{l}\sum_{i=1}^{m}\frac{W_h^2}{m_h^2 k_h}(\mu_{[i:m]h} - \mu_h)^2$ where $W_h = \frac{n_h}{N_h}$	$\sqrt{\sum_{h=1}^{l}\sum_{r=1}^{k}\sum_{i=1}^{m}\frac{W_h^2}{m_h k_h - 1}(y_{[r]ih} - \bar{y})^2}$

3. Simulation Studies

Different scenarios were considered to represent real-life situations and evaluate the behavior of the parameter estimates based on these scenarios using the proposed sampling design. Two sampling designs, namely, Stratified Simple Random Sampling (SSRS) and Stratified Ranked Set Sampling (SRSS) would be compared in this study. Table 3.1 below shows the different cases considered.

Table 3.1 Cases Considered in the Simulation Study

Cases Considered	Scenarios
Population Sizes (N)	(1) 1,200 (2) 12,000
Stratum Sizes	(1) 33-33-33% (balanced) (2) 20-30-50% (moderately unbalanced) (3) 15-25-60% (unbalanced)
Sampling Rate (n)	(1) 1% of N (2) 3% of N (3) 5% of N
Correlation of Concomitant Variable with the Target Variable	(1) High i.e. X = 2+ 5*Y + e, e ~N(0,1) (2) Moderately High i.e. X = 2 + 5*Y + 5*e, e ~N(0,1)

Only two population sizes were considered. The large populations are represented by N=12,000 while the small populations are represented by N=1,200. Usually, SRS performs well in large populations while RSS has an edge in small populations. This paper would like to determine if this will still hold true when stratification is conducted in the population.

To investigate the performance of the two sampling designs, the stratum sizes are also varied. The number of strata is limited to 3 only. Sampling rates are 1%, 3%, and 5% of the population sizes. This is to determine the behavior of the two sampling designs when the sample size increases.

In addition, the set size k is limited to 3-5 only to avoid propagation of ranking error. The set sizes were varied depending on the sample size being divisible by 3, 4 or 5.

This paper also investigates the effect of the correlation of the concomitant variable and the target variable. Since the auxiliary variable is used in ranking, there may be an effect in the results of RSS estimator. In Stratified Ranked Set Sampling, the correlation between target variable and concomitant variable has two cases: high and moderately high. The concomitant variable should be strongly correlated with the target variable so the case of low correlation is not considered in this study.

The study assumed normally distributed simulated populations and applied proportional allocation scheme for the three strata. Two kinds of data set were used. The first one assumes that the elements within strata were homogeneous while the variances within and across strata were varied for the other data set. Descriptions of the simulated data are given below in Table 3.1.

Table 3.1 Simulated Finite Population Means, Variances, and CVs

Data	Stratum sizes	Mean	Variance	CV (in %)
Data Set 1	All	25	10	12.6
		50	10	6.3
		75	10	4.2
Data Set 2	33-33-33	25	100	40
		50	400	40
		75	900	40
	20-30-50	25	156.25	50
		50	225	30
		75	126.5625	15
	15-25-60	25	225	60
		50	506.25	45
		75	56.25	10

As shown in Table 2.4, the estimated standard error of SRSS is quite complicated because it requires the computation of the averages of the order statistics across all cycles, which is needed to be done across the strata. This is why there are three indices in the formula of the estimated standard error. Moreover, the inclusion probabilities would be very hard to compute since RSS is done across strata. Because of these, the authors used variance estimation method. Linearization using Taylor series expansion would be difficult to perform so Nonparametric Bootstrap Method was used instead. This is a re-sampling method to determine the empirical distribution of the estimator. The goal of this paper is to determine the empirical or sampling distribution of the sample mean. The bootstrap method was performed for both SSRS and SRSS for comparison purposes. Along with this method is the estimation of the variance of the bootstrap estimate of the mean.

The nonparametric bootstrap uses simple random sampling with replacement (SRSWR) in re-sampling. Glivenko-Cantelli Lemma justifies the use of nonparametric bootstrap. Nonparametric bootstrap estimation has not yet been used in RSS based on literature.

The procedure for the estimation procedure is as follows: Obtain a sample in each stratum using SRSWOR for SSRS and RSS for SRSS. Only 3 strata were considered in this study. Then, once the samples are obtained, nonparametric bootstrap will be performed 200 times in each stratum. Then, the weighted mean will be computed in each bootstrap sample. In this case, there will be 200 weighted means where the weight is $W_h = N_h/N$. Afterwards, the arithmetic mean of these 200 weighted means will be calculated. This will be the first bootstrap estimate of the mean.

Subsequently, the Monte Carlo variance of these 200 weighted means will be computed. This variance will be the estimated variance of the 1st bootstrap mean. This entire process will be repeated 100 times. After which, there will be 100 bootstrap estimates of the mean and of the variance of the sample mean. The average of these 100 bootstrap estimates will be the estimated mean and its estimated variance using nonparametric bootstrap.

4. Results and Discussion

This section presents the results of the simulation study. The tables include the Monte Carlo variance, bias of the estimate and the coefficient of variation, expressed as percentage, of the bootstrap estimate. The bias and CV of the bootstrap estimate were used as measures of validity and reliability of the estimates. Tables 4.1-4.4 show the results for the first kind of data set which assumes homogeneity of the elements within each stratum. Tables 4.5-4.8 present the results for data sets wherein the variances among strata were varied.

Table 4.1 shows that for the small populations (N=1200), regardless of the stratum sizes, most of the SRSS estimates of the means are closer to the population means. The highlighted biases are those cases in which the SRSS has smaller bias than that of SSRS. However, for all the nine cases, SSRS estimates are more precise since they have smaller CVs as compared to SRSS estimates.

Table 4.11 Bootstrap Estimates of the Mean, their Variances, Biases and Coefficients of Variation (CV) for Small Population (High Correlation)

Population Size 1,200		SSRS			SRSS		
High Correlation							
Stratum Sizes	Sampling Rate	Var	Bias	CV	Var	Bias	CV
33-33-33%	1%	5.98	1.60	**4.97**	8.15	**0.93**	5.76
	3%	2.67	**0.79**	**3.29**	2.74	1.01	3.34
	5%	1.68	1.05	**2.62**	1.70	**1.01**	2.63
20-30-50%	1%	2.65	**0.28**	**2.82**	5.41	0.66	4.06
	3%	2.35	0.23	**2.67**	2.52	**0.11**	2.75
	5%	1.59	0.23	**2.19**	1.65	**0.00**	2.23
15-25-60%	1%	1.63	0.15	**2.07**	5.92	**0.07**	3.96
	3%	2.62	**0.07**	**2.63**	2.69	0.09	2.67
	5%	1.68	0.20	**2.10**	1.70	**0.08**	2.12

Table 4.2 shows that for small population and having an auxiliary variable with moderately high correlation with the target variable; SSRS yield CVs which are lower than that of SRSS. Thus, in general, SSRS still performs a little better than SRSS in terms of reliability of its estimates. Note that their CVs are not so different from each other. In terms of bias, SRSS has an edge over SSRS because SRSS bootstrap estimates have smaller bias in 7 out of 9 scenarios. Even if the CVs of SRSS are a little larger than that of SSRS, the values of its bootstrap estimates are much closer to the true value of the parameter.

In addition, the CVs for SRSS are smaller than that of SSRS when the sampling rate is 5% in the two unequal stratum sizes cases. When the population is small and the stratum sizes are all equal, the bootstrap estimates of the mean are less biased under SSRS if the sampling rates are 1% and 3%.

Table 4.2 Bootstrap Estimates of the Mean, their Variances, Biases and CVs for Small Population (Moderately High Correlation)

Population Size 1,200		SSRS			SRSS		
High Correlation							
Stratum Sizes	Sampling Rate	Var	Bias	CV	Var	Bias	CV
33-33-33%	1%	6.63	**0.45**	**5.13**	7.68	1.05	5.60
	3%	2.55	**1.14**	**3.23**	2.70	1.37	3.33
	5%	1.65	1.22	**2.60**	1.73	**1.06**	2.66
20-30-50%	1%	3.75	2.92	**3.27**	7.27	**0.07**	4.68
	3%	2.72	3.59	2.97	2.50	**0.01**	**2.75**
	5%	1.81	0.23	2.33	1.67	**0.02**	**2.25**
15-25-60%	1%	4.47	0.38	**3.43**	6.04	**0.19**	3.99
	3%	2.43	1.85	**2.49**	2.77	**0.09**	2.71
	5%	1.94	0.58	2.28	1.67	**0.50**	**2.12**

Table 4.3 shows that almost all of the SRSS estimates have smaller biases, hence closer to population means, than their SSRS estimates counterparts. The only two scenarios in which SSRS estimates are closer to the population mean is the equal stratum sizes case with sampling rates of 3% and 5%. On the other hand, for all the stratum sizes considered, SRSS have lower CVs at the 5% sampling rate.

As shown in Table 4.4, only 3 out of 9 CVs under SRSS are lower than that of SSRS. But, it can be noticed that their CVs are not so different from each other. There is just a minimal difference in their CVs. In terms of the bias of the bootstrap estimates, SRSS has lesser bias compared to SSRS in general. Thus, the SRSS bootstrap estimates are closer to the true value of the population mean.

Moreover, large bias, in general, was observed for the equal stratum sizes case for the three sampling rates for both sampling designs. It can be noticed as well that in the two unbalanced cases, the bias of the estimates under SRSS is very small.

Table 4.3 Bootstrap Estimates of the Sample Mean, their Variances, Biases and CVs for Large Population (High Correlation)

Population Size 1,200		SSRS			SRSS		
High Correlation							
Stratum Sizes	Sampling Rate	Var	Bias	CV	Var	Bias	CV
33-33-33%	1%	0.83	1.10	1.84	0.83	**0.95**	**1.84**
	3%	0.28	**0.92**	**1.06**	0.28	1.04	1.06
	5%	0.17	**0.93**	0.82	0.17	0.98	**0.82**
20-30-50%	1%	0.82	0.12	**1.57**	0.84	**0.11**	1.59
	3%	0.28	0.06	0.93	0.28	**0.01**	**0.92**
	5%	0.17	0.03	0.71	0.17	**0.02**	**0.71**
15-25-60%	1%	0.81	0.09	**1.46**	0.85	**0.03**	1.50
	3%	0.28	0.11	**0.86**	0.28	**0.05**	0.86
	5%	0.17	0.11	0.67	0.17	**0.01**	**0.67**

Table 4.4 Bootstrap Estimates of the Mean, their Variances, Biases and CVs for Large Population (Moderately High Correlation)

Population Size 1,200		SSRS			SRSS		
High Correlation							
Stratum Sizes	Sampling Rate	Var	Bias	CV	Var	Bias	CV
33-33-33%	1%	0.80	1.22	**1.81**	0.83	**1.10**	1.84
	3%	0.28	1.12	1.07	0.28	**0.86**	**1.07**
	5%	0.16	1.09	**0.82**	0.17	**0.90**	0.82
20-30-50%	1%	0.82	0.25	**1.57**	0.85	**0.12**	1.60
	3%	0.28	0.15	**0.91**	0.28	**0.06**	0.92
	5%	0.15	0.16	**0.66**	0.17	**0.07**	0.72
15-25-60%	1%	0.84	0.23	1.49	0.82	**0.06**	**1.48**
	3%	0.28	**0.03**	**0.87**	0.29	0.11	0.8
	5%	0.17	0.05	**0.67**	0.17	**0.01**	**0.66**

The succeeding tables show the results when the variances across the strata differed from each other. When the concomitant variable is strongly correlated with the target variable, the bias of SRSS mean estimator is smaller than that of SSRS in general for small populations. The SSRS estimates for the population mean have smaller CVs compared to that of SRSS. This result is consistent with the previous results. Table 4.6 shows that SRSS is still better in terms of the bias of the bootstrap estimates most especially in balanced and unbalanced cases. The result is quite different in the moderately unbalanced case. In general, SRSS performs better in terms of lower CVs compared to SSRS.

Moreover, it can also be noticed that the bias of the estimates in the unbalanced cases decreases in SRSS while it increases in SSRS as the sample size blows up. The moderately unbalanced case has a different behavior in SRSS but in SSRS, the bias declines as the sample size increases.

Table 4.5 Bootstrap Estimates of the Mean, their Variances, Biases and Coefficients of Variation (CV) for Small Population (High Correlation)

Population Size 1,200		SSRS			SRSS		
High Correlation							
Stratum Sizes	Sampling Rate	Var	Bias	CV	Var	Bias	CV
33-33-33%	1%	23328.78	**0.95**	**302.87**	21434.24	23.15	381.37
	3%	8820.80	17.20	**160.41**	7719.61	**1.92**	172.57
	5%	23294.07	80.21	1543.96	4719.82	**5.98**	**146.27**
20-30-50%	1%	1262.82	7.62	**56.69**	1264.09	**0.03**	61.03
	3%	640.19	**4.98**	**41.38**	719.74	5.10	48.54
	5%	458.50	4.31	38.42	451.66	**1.55**	**37.06**
15-25-60%	1%	3695.28	**0.40**	**94.32**	6859.75	25.17	102.26
	3%	1703.53	4.40	**61.09**	1925.82	**1.15**	67.05
	5%	957.09	62.32	**29.45**	2891.76	**0.31**	83.36

Table 4.6 Bootstrap Estimates of the Mean, their Variances, Biases and CVs for Small Population (Moderately High Correlation)

Population Size 1,200		SSRS			SRSS		
High Correlation							
Stratum Sizes	Sampling Rate	Var	Bias	CV	Var	Bias	CV
33-33-33%	1%	21160.46	53.09	620.71	2635.98	**5.05**	**97.83**
	3%	7985.29	13.68	**157.36**	8962.15	**1.66**	192.71
	5%	5317.99	10.91	**163.85**	5472.06	**9.86**	164.27
20-30-50%	1%	1277.33	11.71	69.50	1408.97	**3.68**	**62.16**
	3%	675.81	**0.09**	44.60	684.87	0.98	**44.50**
	5%	443.10	**2.38**	**37.02**	487.48	5.01	39.91
15-25-60%	1%	4278.22	21.52	128.80	5213.31	**4.11**	**107.17**
	3%	1866.69	1.79	**67.99**	1951.74	**0.20**	68.14
	5%	1180.76	2.00	52.06	1148.62	**1.87**	**51.41**

Table 4.7 shows that the CVs of the bootstrap estimates under SSRS are generally lower than that of SRSS when population is large. This is not surprising since in large and heterogeneous populations, SSRS works well. In terms of the bias, the two sampling designs are almost comparable.

When the concomitant variable is moderately correlated with the target variable, SRSS performs better than SSRS in large populations as listed in Table 4.8. Generally, the CVs of SRSS are lower than that of SSRS. Furthermore, the bias and CVs of both sampling designs decrease as the sample size increases in general. In the moderately unbalanced (20-30-50%) and unbalanced (15-25-60%) cases, the CVs of the two sampling designs are comparable.

Table 4.7 - Bootstrap Estimates of the Sample Mean, their Variances, Biases and CVs for Large Population (High Correlation)

Population Size 1,200		SSRS			SRSS		
High Correlation Stratum Sizes	Sampling Rate	Var	Bias	CV	Var	Bias	CV
33-33-33%	1%	2683.09	**2.76**	107.14	2772.65	5.08	**100.79**
	3%	908.87	2.49	**59.16**	915.26	**0.78**	61.33
	5%	550.78	**0.82**	**47.60**	536.65	2.65	47.86
20-30-50%	1%	233.44	1.80	**25.66**	240.41	**0.36**	26.60
	3%	80.58	**0.95**	**15.20**	80.33	2.17	15.66
	5%	46.64	**0.77**	**11.77**	48.06	0.81	11.95
15-25-60%	1%	78.87	14.03	**15.64**	134.77	**12.80**	20.15
	3%	22.81	15.73	**8.58**	50.91	**14.48**	12.63
	5%	15.57	**11.35**	**6.74**	33.47	14.97	10.30

Table 4.8 Bootstrap Estimates of the Mean, their Variances, Biases and CVs for Large Population (Moderately High Correlation)

Population Size 1,200		SSRS			SRSS		
High Correlation Stratum Sizes	Sampling Rate	Var	Bias	CV	Var	Bias	CV
33-33-33%	1%	2646.58	4.95	108.87	2709.01	**0.55**	**104.12**
	3%	905.14	3.07	62.43	910.25	**0.03**	**60.71**
	5%	547.20	**9.50**	51.99	531.98	26.99	**36.53**
20-30-50%	1%	231.41	**0.42**	**25.90**	224.52	2.39	26.24
	3%	78.96	2.26	15.54	78.80	**1.34**	**15.38**
	5%	46.87	**0.41**	**11.75**	48.11	1.38	12.02
15-25-60%	1%	572.41	**5.33**	**36.08**	592.98	5.49	36.67
	3%	202.43	4.99	23.79	204.41	**1.60**	**22.35**
	5%	123.23	1.29	17.87	123.50	**0.06**	**17.64**

It should be noted that nonparametric bootstrap was used in the estimation procedure. Thus, the unbiasedness of the estimators under SRSS and SSRS for the population mean mentioned in Section 2 does not apply anymore because that is under design-unbiased estimation. This is why the estimates are biased even though repeated sampling was done.

In general, the CV of SRSS bootstrap estimate of the mean is larger than that of SSRS because the sample obtained using RSS is spread out. SRSS estimates are closer to the actual value producing small bias. In SSRS, the bootstrap estimates of the mean have smaller CVs generally because the measurements obtained using SRSWOR are near the true mean.

5. Illustration

The two sampling designs were illustrated using the 2002 Census of Agriculture data set. The target variable is the production of mangoes in the country. The concomitant variable used is the corn production. This is because mangoes and corns are usually planted together. Instead of using the information per farmer, the variables were aggregated in the barangay level. Thus, the elementary units considered in this illustration are the barangays. The parameter that will be estimated in this case is the average production of mangoes in the country.

Furthermore, the three major islands namely Luzon, Visayas, and Mindanao were defined as the strata. Only those records with complete information on the necessary variables such as stratum, region, province, municipality, barangay, farm areas for mango and corn production were considered in this illustration; hence, there are a total of 1,144 records used.

In the first stratum, Luzon, there are a total of 460 barangays while the second stratum, Visayas, has 210 barangays and lastly, Mindanao, has 474 barangays. Based on these, the weights used per stratum are 40-18-42% respectively. The two sampling designs were used to obtain 1%, 3%, 5%, and 10% of the population as the sample sizes. Based on Table 4.9, it is evident that SRSS works very well as compared to SSRS since it has lower biases and CVs for all cases except only when the sample size is at 3%. This result further confirms the results of the simulation study. That is, SRSS works well with small population and the bootstrap estimate under this sampling design produces smaller bias.

Table 4.9 Bootstrap Estimates of the Mean, their Variances, Biases and CVs for the 2002 Census of Agriculture

Population Size 1,200		SSRS				SRSS			
High Correlation									
Stratum Sizes	Sampling Rate	Mean	Var	Bias	CV	Mean	Var	Bias	CV
	1%	9.92	27.98	89.34	53.31	3.05	0.80	**41.73**	**29.37**
40-18-42%	3%	2.95	0.24	43.64	**16.51**	6.02	2.65	**14.88**	27.02
	5%	6.86	7.98	30.83	41.20	3.91	0.70	**25.30**	**21.4**
	10%	4.66	2.07	11.09	30.91	5.18	1.05	**1.11**	**19.82**

6. Conclusion and Direction for Future Research

Based on the simulation scenarios, the bias of the bootstrap estimates using Stratified Ranked Set Sampling is generally smaller compared to that of Stratified Random Sampling most especially in small populations regardless of the sample size obtained and the degree of variability in each stratum. This means that whether the stratum sizes are different (balanced, moderately unbalanced or unbalanced) or the elements in each stratum are very heterogeneous, the bias of the bootstrap estimate of SRSS is smaller than that of SSRS. Hence, the bootstrap estimate of the mean under SRSS is expected to be closer to the true mean. The measure of reliability may suffer because the CVs of SRSS are larger compared to that of SSRS. This result is consistent with the fact that the sample obtained using SRSS is spread out in the population giving enough representation of the population of interest.

If the researcher wants an estimated value much closer to the population mean, Stratified Ranked Set Sampling is recommended most especially when the population is small regardless of the variability of strata in the population. This implies that even if the stratification variable were not appropriately chosen, SRSS would more likely to give less biased estimate of the mean provided that the population is small. Furthermore, it was shown that the SRSS gives less biased estimates provided that the correlation of the concomitant variable and target variable is at least moderately strong.

On the other hand, if the researcher puts a heavier weight on the reliability of the estimates, then Stratified Random Sampling works better since SSRS has lower coefficient of variation (CV) in general. This holds true for large populations regardless of the stratum sizes, sampling rates, and the variability of elements in each stratum.

This paper has several limitations. The number of strata is fixed to 3. It might be of interest to determine the effect of the number of strata in the estimation procedure. Moreover, Ibrahim et al. (2010) studied different distributions of the population. It would be appealing to study the performance of SRSS and SSRS under different distribution assumptions other than normal and under different population sizes as well.

Furthermore, the study is limited in comparing the two sampling designs only. It would be better if the usual RSS were compared as well. In practice, it is possible that the auxiliary variable used has a low correlation with the target variable if it is the only information available to the researchers. Perhaps, it is also a good idea if this scenario would be added in the simulation study.

REFERENCES

CHEN, Z., BAI, Z., and SINHA, B., 2003, *Ranked Set Sampling: Theory and Application,* Springer.

CHEN, Z., 2007, Ranked Set Sampling: Its Essence and Some New Applications. *Environ. Ecol. Stat*., 14:355-363.

CHEN Z. and WANG Y., 2004, Efficient Regression Analysis with Ranked Set Sampling, *Biometrics* 60:997-1004.

COCHRAN, W., 1977, *Sampling Techniques*, 3rd ed., John Wiley and Sons, New York.

IBRAHIM, K. et al., 2010, Estimating the Population Mean Using Stratified Median Ranked Set Sampling, *Applied Mathematical Sciences*, 4(47):2341-2354.

KISH, L., 1965, *Survey Sampling*. John Wiley and Sons, New York.

LOHR, S., 1999, *Sampling: Design and Analysis*. Duxbury Press, Pacific Grove, CA.

MCINTYRE, G.A., 1952, A Method of Unbiased Selective Sampling using Ranked Sets, *Australian Journal of Agricultural Research,* 3, 385-390.

PATIL, G.P., 1995, Editorial: Ranked Set Sampling. *Environmental and Ecological Statistics,* 2: 271–285.

STOKES, S., 1977, Ranked Set Sampling with Concomitant Variables, *Communications in Statistics - Theory and Methods*, A6 (12): 1207-1211.

TAKAHASI, K. and WAKIMOTO, K., 1968, On Unbiased Estimates of the Population Mean Based on the Sample Stratified by Means of Ordering, *Annals of the Institute of Statistical Mathematics*, 20:1-31.

SAMAWI, H.M. & MUTTLAK, H.A., 1996, Estimation of Ratio using Rank Set Sampling, *Biomet J* 38:753-764.

WOLFE, D., 2004, Ranked Set Sampling: An Approach to More Efficient Data Collection, *Statistical Science,* 19(4):636-643.

ACKNOWLEDGMENT

The authors would like to express their sincerest gratitude to Joanne Guansing for sharing her expertise in programming and Dr. Erniel Barrios for giving insightful comments which improved this paper.

Length of a Time Series for Seasonal Adjustment: Some Empirical Experiments*

Lisa Grace S. Bersales, Ph.D.
University of the Philippines in Diliman

Use of 5 to 15 years of quarterly or monthly data is suggested when doing seasonal adjustment using X11 and its variants. This is meant to address changes in the structure of the time series. Philippine time series are good candidates for this practice since they usually exhibit frequent changes in patterns. Empirical validation of the suggested length of series is done for seasonal ARMA processes. Different quarterly series were simulated for the following situations and seasonal adjustment was done for various lengths of time series: (1) processes without any structural change; (2) processes with abrupt permanent change in structure; (3) processes with gradual permanent change in structure. For all types of processes, both weak and strong seasonality were considered. Regression models were used in testing the effect of length of series used in seasonal adjustment to the error in estimating the seasonal factor. Results show that the length of series used does not have significant effect on the seasonal adjustment for processes without structural change and with abrupt permanent structural change. On the other hand, for processes with gradual permanent change, use of longer lengths of series for seasonal adjustment is better.

Keywords: seasonal adjustment, seasonal factor, X11-ARIMA, seasonal ARMA processes

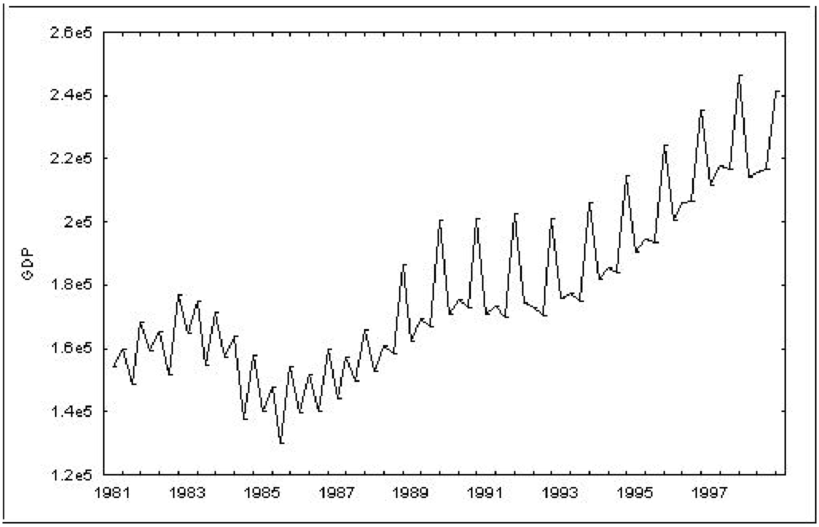
* Paper for the 2010 BSP Professorial Sterling Chair on Government and Official Statistics in the University of the Philippines Diliman, presented at the Bangko Sentral ng Pilipinas on February 22, 2011.

1. Introduction

Seasonal adjustment of series with changes in behavior usually results in a seasonally adjusted series that is not smoothened. Such situations may lead to misleading analyses (Ghysels, 1988). Specific illustrations of such situations are discussed by Castro and Osborn (2004) for periodic autoregressive processes whose periodicity remains, though in an altered form, after X11 seasonal adjustment. Thus, tests to detect deviations from deterministic seasonal patterns were developed (e.g.,Canova and Hansen, 1995; Busetti and Harvey, 2003). Of interest, however, is how series length impacts on the seasonal adjustment under non-deterministic seasonality. This issue of series length is a concern addressed by Findley and Martin (2003) who studied the performance of TRAMO SEATS and X11-ARIMA for short and moderate length series.

In the Philippines where official seasonal adjustment is done using X11-ARIMA, the problem of seasonally adjusting series with changing behavior is addressed by limiting the length of the series for seasonal adjustment to the suggested length of 5 to 15 years (Dagum, 1988; Bersales and Sarte, 1999). This practice has been done by agencies doing official seasonal adjustment since Philippine time series usually exhibit frequent changes in patterns. For example, quarterly Philippine Gross Domestic Product (GDP) from 1981 to 1998 as presented in Figure 1 shows changes in pattern with the new seasonal pattern starting in 1988.

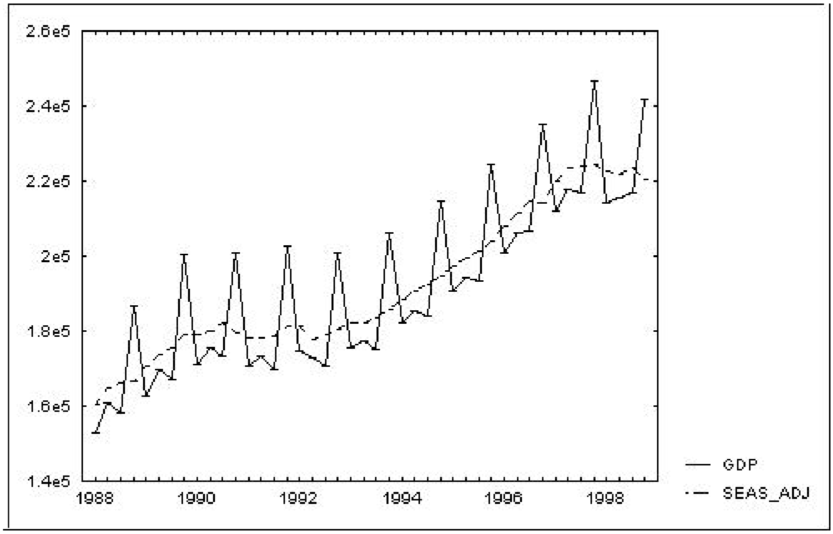
Figure 1. Philippine Gross Domestic ProductFirst Quarter 1981 – Fourth Quarter 1998



Source: Bersales and Sarte, 1999

The following figure, Figure 2, shows the historical plot of the original and seasonally adjusted quarterly GDP from the first quarter of 1988 to the fourth quarter of 1998. Seasonal adjustment was done for the time period 1988 to 1998, 11 years of data, instead of the whole period (1981-1998) for which data are available.

Figure 2. Original and Seasonally Adjusted GDP, 1988 – 1998



Source: Bersales and Sarte, 1999

Many guidelines used in seasonal adjustment resulted in users' experience with various types of data and may actually be considered rules of thumb. The current practice of using computational statistics in developing estimation procedures for complex statistical models may be used to evaluate some guidelines on seasonal adjustment. This paper aims to provide empirical validation of the use of 5 to 15 years of data for seasonal adjustment using simulated data from processes exhibiting the following behavior:

- (Type 1) processes without any structural change
- (Type 2) processes with abrupt permanent change in structure (shift in level)
- (Type 3) processes with gradual changes in structure.

For all three types of processes, realizations with both weak and strong seasonality were simulated.

The following plots illustrate the simulated series:

Figure A.1. Plot of Six Realizations for Type 1 Weak Seasonality

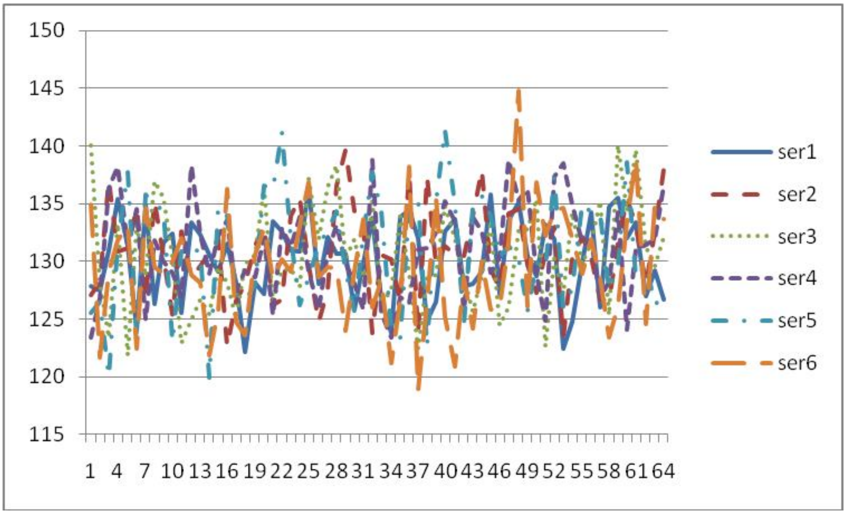


Figure A.2. Plot of Six Realizations for Type 1 Strong Seasonality

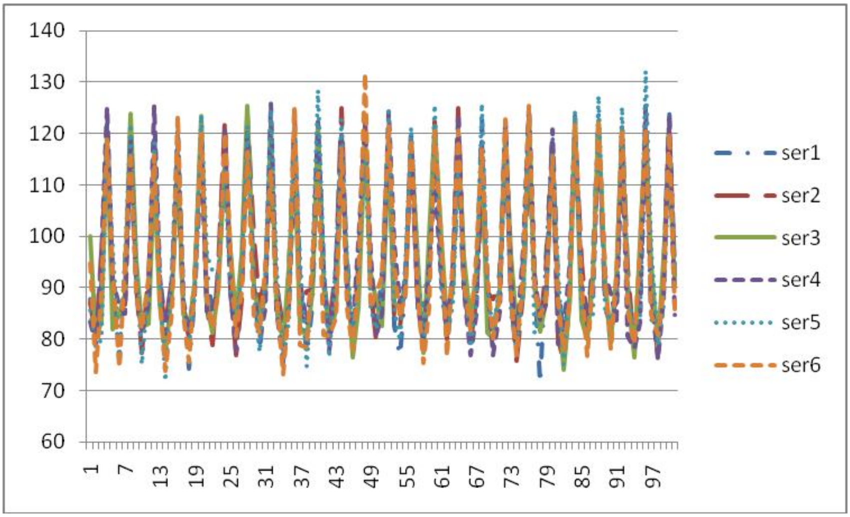


Figure A.3. Plot of Six Realizations for Type 2 Weak Seasonality

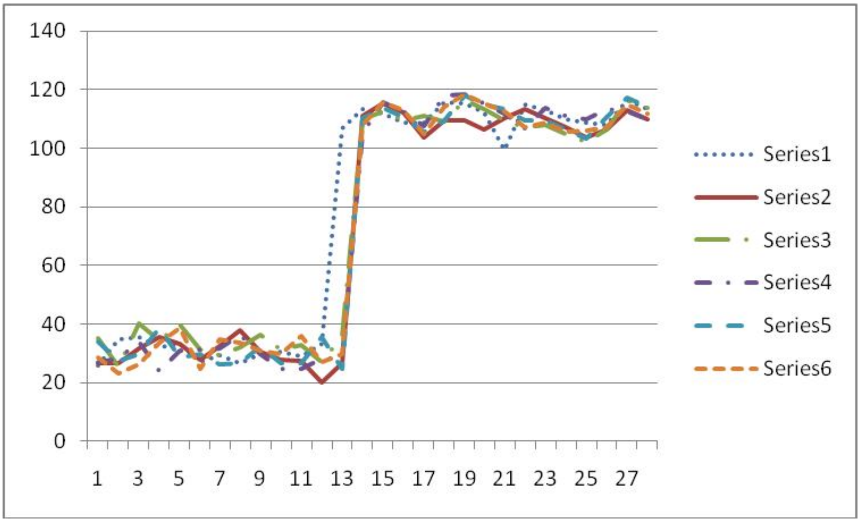


Figure A.4. Plot of Six Realizations for Type 2 Strong Seasonality

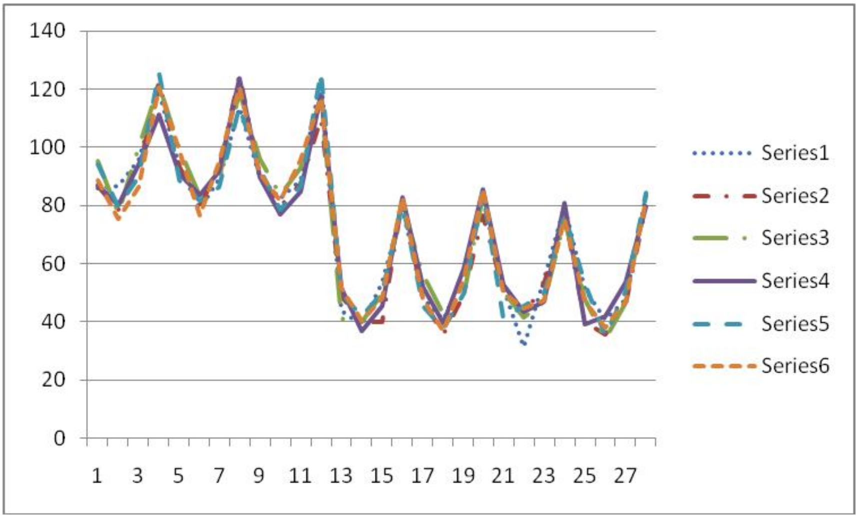


Figure A.5. Plot of Six Realizations for Type 3 Weak Seasonality

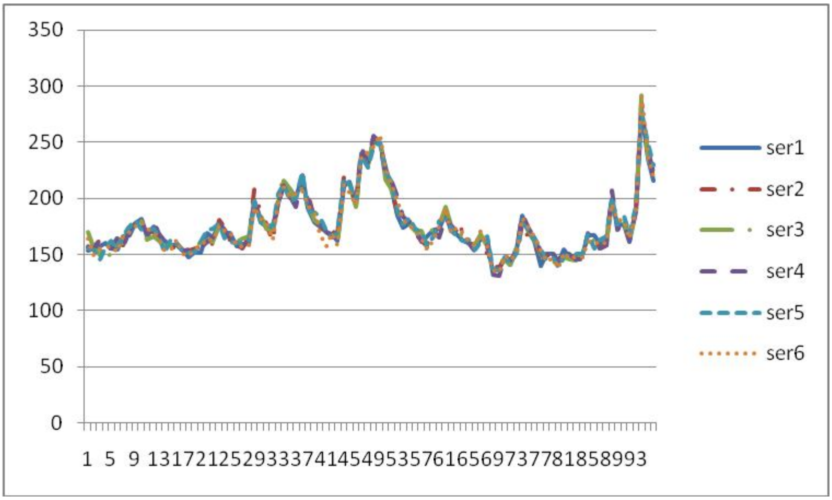
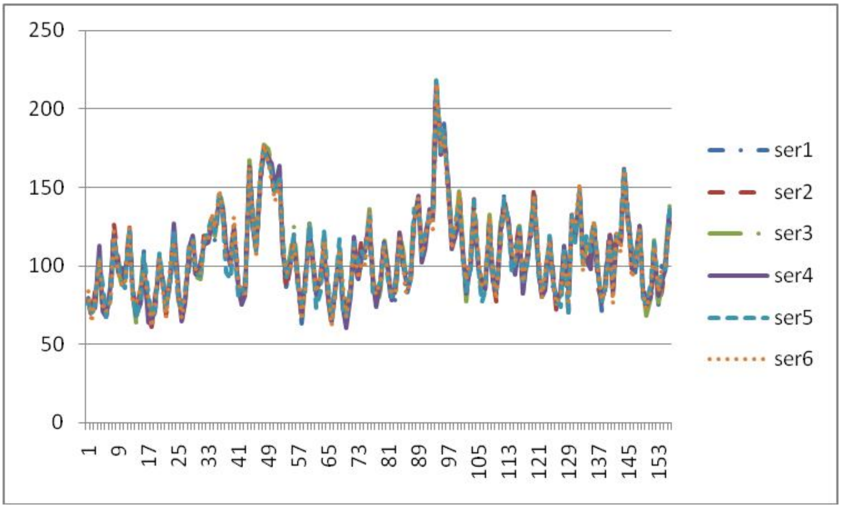


Figure A.6. Plot of Six Realizations for Type 3 Strong Seasonality



2. Methodology

The following procedures were used in achieving the objective of this study:

- *Step 1*.100 realizations with 30 years of quarterly data were generated for each of the following processes which can be modeled as purely Seasonal AR processes:

Type 1

Weak seasonality	Strong seasonality
Y(t)=130 +a(t), a(t)~N(0,4) for Quarter 1	Y(t)=90 +a(t), a(t)~N(0,4) for Quarter 1
Y(t)=128 +a(t), a(t)~N(0,4) for Quarter 2	Y(t)=80 +a(t), a(t)~N(0,4) for Quarter 2
Y(t)=130 +a(t), a(t)~N(0,4) for Quarter 3	Y(t)=90 +a(t), a(t)~N(0,4) for Quarter 3
Y(t)=133 +a(t), a(t)~N(0,4) for Quarter 4	Y(t)=120+a(t), a(t)~N(0,4) for Quarter 4

Type 2

Weak seasonality	Strong seasonality
Pattern 1	*Pattern 1*
Y(t)=30 +a(t), a(t)~N(0,4) for Quarter 1	Y(t)=90 +a(t), a(t)~N(0,4) for Quarter 1
Y(t)=28 +a(t), a(t)~N(0,4) for Quarter 2	Y(t)=80 +a(t), a(t)~N(0,4) for Quarter 2
Y(t)=30 +a(t), a(t)~N(0,4) for Quarter 3	Y(t)=90 +a(t), a(t)~N(0,4) for Quarter 3
Y(t)=33 +a(t), a(t)~N(0,4) for Quarter 4	Y(t)=120+a(t), a(t)~N(0,4) for Quarter 4
Pattern 2	*Pattern 2*
Y(t)=110 +a(t), a(t)~N(0,4) for Quarter 1	Y(t)=50 +a(t), a(t)~N(0,4) for Quarter 1
Y(t)=108 +a(t), a(t)~N(0,4) for Quarter 2	Y(t)=40 +a(t), a(t)~N(0,4) for Quarter 2
Y(t)=110 +a(t), a(t)~N(0,4) for Quarter 3	Y(t)=50 +a(t), a(t)~N(0,4) for Quarter 3
Y(t)=113 +a(t), a(t)~N(0,4) for Quarter 4	Y(t)=80 +a(t), a(t)~N(0,4) for Quarter 4

Type 3

Weak seasonality	Strong seasonality
Weak Type 1 series + GARCH(1.1) errors with ARCH parameter=.239 and GARCH parameter=.667	Strong Type 1 series + GARCH(1.1) errors with ARCH parameter=.239 and GARCH parameter=.667

- *Step 2*.Seasonal adjustment of all realizations were done using the multiplicative decomposition model of X11 with varying lengths of data (5 years to 15 years). For Type 2 processes, the seasonal adjustment concentrated on where the permanent change started (e.g., analysis focused on: 5 years of seasonal adjustment with the first year of data 5 years before the break, 5 years of seasonal adjustment with the first year of data 4 years before the break,…, 5 years of seasonal adjustment with the first year of data is the first year of the new behavior)

- *Step 3.* Seasonal factors of the processes were extracted using the multiplicative decomposition model of X11. The whole length of available data was used in the extraction. For realizations with abrupt permanent change, the seasonal factors for the old pattern and the new pattern were extracted separately. The mean seasonal factors from this step are assumed to be the actual seasonal factors for the process.
- *Step 4.* The mean seasonal factors from Step 2 were compared with the mean seasonal factors from Step 3. This produced the error series which was generated by getting the absolute value of the difference.
- *Step 5.* The following regression models were estimated and tests of significance were done to determine if length of the series used in seasonal adjustment has significant effect on the errors. For series with abrupt break, the regression included an independent variable reflecting first year of seasonal adjustment relative to the break in the series. This determines the quality of will is before or after the new behavior starts.
  - o For processes with abrupt permanent change:

    $$E(ERROR) = \theta_0+\theta_1Q1+\theta_2Q2+\theta_3Q3+ \theta_4LENGTH \quad \text{(model 1)}$$
  - o For other processes:

    $$E(ERROR) = \alpha_0+ \alpha_1Q1+\alpha_2Q2+\alpha_3Q3+\alpha_4YEAR +\alpha_5LENGTH \quad \text{(model 2)}$$

  where Q_k is an indicator variable representing quarter k; k=1,2,3

  YEAR= years before/after start of new behavior with value for 0 for the year where new behavior started, -1 year before new behavior, 1 year after new behavior started

  LENGTH= number of years of data used in the seasonal adjustment, with values from 5 to 15.

  Of interest is to test the significance of LENGTH.

Seasonal adjustment and regression analysis were done using Eviews6. Generation of simulated data was done using the following: normally distributed errors for Types 1 and 2 processes in Excel, GARCH (1,1) errors for Type 3 in Eviews6; and, all realizations in Excel.

3. Discussion of Results

Estimation and model 1 and model 2 using weighted least squares in the procedure used to answer the study resulted in Table 1. In the table, AE is ERROR in models 1 and 2.

$$E(ERROR) = \theta_0 + \theta_1Q1 + \theta_2Q2 + \theta_3Q3 + \theta_4LENGTH \quad \text{(model 1)}$$

$$E(ERROR) = \alpha_0+ \alpha_1Q1 + \alpha_2Q2 + \alpha_3Q3 + \alpha_4YEAR + \alpha_5LENGTH \quad \text{(model 2)}$$

The results in Table 1 show that the length of series used does not have significant effect on the errors in estimating the seasonal factor for processes without structural change and with abrupt shifts in level. On the other hand, for processes with gradual permanent change, use of longer lengths of series for seasonal adjustment is better.

It is further noted that for processes with abrupt shifts in level, what affects the seasonal adjustment is the start of the data being used for seasonal adjustment. The errors in estimation of the seasonal factor are higher when the start is just before the break starts. Once the seasonal adjustment starts with the new behavior, the estimates have lower errors. This result reinforces the practice of cutting the series for seasonal adjustment once a new pattern starts. Tables 2 and 3 clearly show the improvement in error as YEAR nears 0.

Table 1. Results of Regression of Error in Seasonal Factor Estimate versus Independent Variables

	Independent Variable	Coefficient	Std. Error	t-Statistic	Prob.
	C	0.006928	0.000888	7.802005	0.0000
Type 1	Q1	0.001537	0.000616	2.496280	0.0127
Weak Seasonality	Q2	-0.001412	0.000353	-3.997746	0.0001
	Q3	0.001355	0.000658	2.059321	0.0397
	LENGTH	-8.40E-05	7.35E-05	-1.142788	0.2534
	C	0.233896	0.028815	8.117034	0.0000
Type 1	Q1	-0.237455	0.028012	-8.477048	0.0000
Strong Seasonality	Q2	-0.238818	0.028011	-8.525991	0.0000
	Q3	-0.237864	0.028011	-8.491743	0.0000
	LENGTH	0.000627	0.002300	0.272629	0.7852
	C	0.054570	0.003606	15.13399	0.0000
Type 2	Q1	-0.029148	0.002715	-10.73395	0.0000
Weak Seasonality	Q2	-0.014551	0.002477	-5.873684	0.0000
	Q3	-0.024696	0.002160	-11.43114	0.0000
	YEAR	-0.013260	0.000533	-24.87590	0.0000
	LENGTH	-0.000128	0.000276	-0.461721	0.6444
	C	0.124382	0.007669	16.21886	0.0000
Type 2	Q1	-0.054010	0.004654	-11.60423	0.0000
Strong Seasonality	Q2	-0.042050	0.004673	-8.999352	0.0000
	Q3	-0.074169	0.004240	-17.49398	0.0000
	YEAR	0.021333	0.002457	8.682820	0.0000
	LENGTH	-0.000586	0.000514	-1.141329	0.2540
	C	0.243926	0.037478	6.508445	0.0000
Type 3	Q1	0.055877	0.022488	2.484713	0.0131
Weak Seasonality	Q2	-0.003505	0.020780	-0.168683	0.8661
	Q3	-0.001763	0.020904	-0.084328	0.9328
	LENGTH	-0.013670	0.003135	-4.360837	0.0000
	C	0.152922	0.016125	9.483651	0.0000
Type 3	Q1	0.055870	0.009412	5.935775	0.0000
Strong Seasonality	Q2	-0.007557	0.002514	-3.005926	0.0027
	Q3	-0.009601	0.002557	-3.755392	0.0002
	LENGTH	-0.011323	0.001303	-8.691081	0.0000

Table 2. Descriptive Statistics for AE for Type 1 Weak Seasonality

Categorized by values of YEAR
Sample: 1 1236
Included observations: 1236

YEAR	Mean	Std. Dev.	Obs.
-2	0.058404	0.035671	176
-1	0.072755	0.038330	220
0	0.014147	0.012380	260
1	0.011959	0.010359	296
2	0.010641	0.008583	284
All	0.029552	0.034249	1236

Table 3. Descriptive Statistics for AE for Type 2 Weak Seasonality

Categorized by values of YEARS
Sample: 2 1124
Included observations: 1123

YEARS	Mean	Std. Dev.	Obs.
-2	0.033068	0.037877	36
-1	0.054411	0.052092	380
0	0.028242	0.027535	380
All	0.038657	0.042030	1123

For Type 3 series, Tables 4 and 5 show that it is better to use longer series to seasonally adjust. This result reinforces the current practice of using long length for seasonally adjustment.

Table 4. Descriptive Statistics for AE, Type 3 Weak Seasonality

Categorized by values of LENGTH
Sample: 1 1968
Included observations: 1644

LENGTH	Mean	Std. Dev.	Obs.
5.00	0.362003	0.483999	100
6.00	0.261115	0.442540	112
7.00	0.205172	0.406549	128
8.00	0.168678	0.376878	144
9.00	0.136638	0.345669	156
10.00	0.105658	0.309267	164
11.00	0.048456	0.213433	168
12.00	0.048504	0.213422	168
13.00	0.048539	0.213415	168
14.00	0.048530	0.213417	168
15.00	0.119869	0.324546	168
All	0.126141	0.332062	1644

Table 5. Descriptive Statistics for AE, Type 3 Strong Seasonality

Categorized by values of LENGTH
Sample: 1 1608
Included observations: 1608

LENGTH	Mean	Std. Dev.	Obs.
5.00	0.220717	0.398762	64
6.00	0.110329	0.286633	112
7.00	0.073915	0.230748	128
8.00	0.071291	0.232702	144
9.00	0.070401	0.237083	156
10.00	0.066400	0.231688	164
11.00	0.009104	0.013501	168
12.00	0.008092	0.013655	168
13.00	0.007872	0.013757	168
14.00	0.007569	0.013815	168
15.00	0.007425	0.013868	168
All	0.046525	0.185411	1608

4. Conclusion

This study provides empirical evidence that validates the current practices of official seasonal adjustment of Philippine time series of using long lengths of series when doing seasonal adjustment and cutting the series where a new pattern starts. Results indicate that as long as the series do not have changes in behavior or the change is in an abrupt shift in level, the length of series, as long as within 5 to 15 years, has no significant effect on error of estimation of the seasonal factor. However, for series with more volatile changes, use of longer series is recommended. All these results hold for both weak and strong seasonality.

REFERENCES

BERSALES, L.G.S. and SARTE, G.F.,1999, Seasonal Adjustment of Time Series with Applications to Philippine Data (A Teaching Module on Seasonal Adjustment of Philippine Time Series).

BUSETTI, F. and HARVEY, A., 2003, Seasonality Tests, *Journal of Business and Economic Statistics*, 21:420-436.

CANOVA F. and HANSEN, B.E., 1995, Are Seasonal Patterns Constant over Time? A Test for Seasonal Stability, *Journal of Business and Economic Statistics*, 2:292-349.

CASTRO,T.B. and OSBORN, D., 2004, The Consequences of Seasonal Adjustment for Periodic Autoregressive Processes, *Econometrics Journal*, 7: 307-321.

DAGUM, E.B., 1988, The X11 ARIMA/88 Adjustment Method – Foundations and User's Manual, Statistics Canada.

GHYSELS, E., 1988, A Study Towards a Dynamic Theory of Seasonality for Economic Time Series, *Journal of the American Statistical Association*, 83: 168-172.

Food Inflation, Underemployment and Hunger Incidence: A Vector Autoregressive Analysis[1]

Dennis S. Mapa, Ph.D, Fatima C. Han
and Kristine Claire O. Estrada
University of the Philippines Diliman

The high level of hunger incidence in the country is perhaps one of the most pressing issues that need to be addressed by our policymakers. Official government statistics and data from self-rated hunger surveys show an increasing trend in hunger incidence among Filipino households. Data from the National Statistical Coordination Board (NSCB) show that the percentage of Filipinos experiencing hunger almost remained the same, decreasing only slightly from 11.1 percent in 2003 to 10.8 percent in 2009. The Social Weather Stations (SWS) quarterly surveys on hunger incidence also show an increasing trend in the percentage of families that experienced hunger, reaching an alarming level of 24 percent in December 2009, representing about 4.4 million households. One probable cause of the increasing trend in hunger is the rising food prices akin to what the country experienced in 2008. This paper aims to determine the impact of food inflation and underemployment on hunger incidence in the Philippines, using the hunger incidence data from the SWS. A vector autoregressive (VAR) model is used to determine the effect of a shock or increase to food inflation and underemployment on total involuntary hunger. Results show that an increase in food prices at the current quarter will increase hunger incidence for five quarters. Shocks to underemployment will also increase hunger incidence but the effects last for only two quarters. The results of this study provide relevant information that will be useful in crafting policies related to the Hunger Mitigation Program of the government.

Keywords: hunger, food inflation, underemployment, vector autoregressive

1. Introduction

Pope Benedict XVI, during a summit of the United Nations Food and Agriculture Organization (FAO) in November 2009 in Rome, referred to hunger as "the most cruel and concrete sign of poverty." The pontiff has reason to worry. World hunger reached a historic high in 2009 with 1.02 billion people experiencing hunger every day, according to estimates from the FAO. The number of individuals going hungry has reached the one billion mark for the first time in history. This represents about 15 percent of the world's population, estimated at 6.8 billion in 2009. The twin crises experienced in the past two years, the high cost of food in 2008 followed by the global financial crisis, increased the number of individuals who went hungry by about 100 million compared to the 2008 estimates of 915 million.

In the Philippines, hunger incidence in its various absolute dimensions, has been widespread and increasing in recent years, threatening to rip our social fabric. It is disturbingly high and embarrassing, in comparison to other countries in East and Southeast Asia. The food crisis, in 2008, resulting from high prices of basic commodities particularly rice, the global financial crisis and the impact of natural calamities (brought about by typhoons Ondoy and Peping) in 2009 are expected to raise the number of Filipinos who will join the ranks of those experiencing involuntary hunger. While these three shocks in the past two years will exacerbate further the poverty and hunger situation in the country, it will not fundamentally change the character of the poverty problem in the country (Balisacan and Mapa, 2010). Evidence from official statistics and national surveys of hunger by the Social Weather Stations (SWS) suggest that our country's hunger situation has already deteriorated during the period 2003 to 2008. What is disturbing is that the worsening problem of extreme poverty occurred against the backdrop of high growth rates as trumpeted by the past administration of Gloria Macapagal-Arroyo.[2] Undeniably, addressing the problem of hunger or extreme poverty is the single most important policy challenge facing the country today.

Our official statistics on the proportion of subsistence poor compiled by the National Statistical Coordination Board (NSCB) reveal a number of striking observations. First, the percentage of subsistence poor (or food poor) did not change much in recent years, only slightly decreasing from 11.1 percent in 2003 to 10.8 percent in 2009. However, in terms of magnitude, the number of food poor Filipinos has increased to about 9.44 million in 2009 from 8.8 million in 2003. This is primarily because of the relatively high population growth during the period.[3]

The results of the 7th National Nutrition Survey (NNS) of 2008 conducted by the Food Nutrition and Research Institute (FNRI) show there was a significant increase in the proportion of children aged 0-5 years who were underweight (indirect measure of hunger) from 24.6 percent in 2003 to 26.2 percent in 2008. Moreover, the same report shows that the proportion of children who were under height for age (stunted) also increased significantly to 27.9 percent in 2008 from 26.3 percent in 2003. The FNRI study also shows the same results in children between 6 to 10 years old: a significant increase in the prevalence of underweight from 22.8 percent in 2003 to 25.6 percent in 2008 and increase in the proportion of under height from 32.0 percent to 33.1 percent.

Given the inadequate progress in reducing the number of households living below the subsistence or food threshold and in minimizing the number of underweight children, the Philippines will most likely miss its Millennium Development Goal (MDG) target of halving the proportion of poor households living below the food threshold and halving the proportion of underweight children below 5 years old from 1991 to 2015.

The SWS national surveys on hunger also show that the hunger incidence in the country has deteriorated in the past years. The proportion of families experiencing involuntary hunger reached a record-high of 24 percent in December 2009, representing about 4.4 million households (SWS, 2010). The time series data on hunger incidence shows that the average hunger incidence from 2001 to 2009 (Arroyo administration) is 14.12 percent. Moreover, the average hunger incidence during this period increased by almost 8 percentage points, from 11.4 percent in 2001 to 19.2 percent in 2009. What is noticeable is that the trend of hunger incidence shifted and increased beginning the third quarter of 2003. In other words, the proportion of hunger incidence rapidly increased starting the 3rd quarter 2003 up to the 4th quarter of 2009, compared to the period before 2003. The data from the official statistics on hunger incidence (subsistence poor from NSCB), as well as other measures of hunger incidence from the FNRI and SWS, consistently show the same results: that hunger has worsened in the past years.

This paper examines the dynamic patterns of hunger incidence and the effects of the determinants of hunger using the quarterly time series data from the SWS national surveys on hunger. A vector autoregressive (VAR) model is used to analyze the impact of shocks on food prices and underemployment on the current and future hunger incidence. An important feature of this paper is the mainstreaming of the time series data on hunger incidence from the SWS into the econometric model through the VAR models. The organization of the paper is as follows: this section serves as the introduction, section 2 discusses the different methods of measuring hunger incidence in the Philippines as well as some of the government programs aimed at mitigating hunger incidence, section 3 discusses the trends in hunger incidence using the official statistics and the results from the self-rated hunger surveys, section 4 presents the results of the vector autoregressive (VAR) model for hunger incidence and section 5 concludes.

2. Measures of Hunger Incidence and Accelerated Hunger-Mitigation Program

2.1. National measures of hunger

Hunger is a complex phenomenon and a multi-dimensional concept. In the Philippines, there are several existing measures of hunger incidence. At the national level, Maligalig (2008) identifies four different measures of hunger: (1) the prevalence of food poor (or subsistence poor) computed by the National Statistical Coordination Board (NSCB); (2) the self-rated hunger incidence collected by the Social Weather Stations (SWS); (3) the hunger incidence compiled by the Bureau of Agricultural Statistics (BAS); and (4) the food security measures compiled by the Food and Nutrition Research Institute (FNRI).[4] The NSCB statistics on subsistence poor, measured from the Family Income and Expenditure Survey (FIES), and available every three years are also the official statistics on hunger in the country. The SWS, FNRI and the BAS measures of hunger incidence are referred to as the direct measures since these "were compiled on the basis of responses of individuals to questions about their experiences about hunger," while the proportion of subsistence poor is an indirect measure of hunger (Maligalig, 2008).

In addition to these four measures, Salud-Payuno (2009) cited other indicators of hunger incidence that are regularly reported by government agencies such as the percentage of pre-schoolers below six years old who are undernourished based on the annual survey collected by the National Nutrition Council (NNC), the percentage of underweight children between 0 to 5 year-olds and prevalence of thinness among 0 to 5 year-olds from the National Nutrition Survey of the FNRI and the hunger index developed by the NSCB.[5]

2.1.1. National Statistical Coordination Board (NSCB) measure of subsistence incidence

The official statistics on hunger incidence is the subsistence incidence or popularly called the food poor. The prevalence of subsistence poor refers to the proportion of families or individuals with per capita income/expenditure less than the per capita food threshold to the total number of families/individuals. The food threshold is determined using regional one-day menus priced at the provincial level. These menus are determined using low-cost nutritionally adequate food items satisfying basic food requirements of 2,000 calories which are 100% adequate for the recommended energy and nutrient intake (RENI) for energy and protein and 80% adequate for the RENI for vitamins, minerals and other nutrients (NSCB, 2010). The official statistics on subsistence incidence is determined using the food threshold and the income distribution derived from the Family Income and Expenditure Survey (FIES).

2.1.2 Food and Nutrition Research Institute (FNRI) measure of food insecurity

The Food and Nutrition Research Institute (FNRI), an agency affiliated with the Department of Science and Technology (DOST), conducts the National Nutrition Survey (NNS) to update the official statistics on the Philippines' food, nutrition and health situation (FNRI, 2010). The 2008 NNS is the seventh in a series of surveys undertaken by the FNRI every five years. The FNRI measure of hunger uses the Radimer-Cornell measures of food insecurity based on a set of 10 questions designed to evaluate food insecurity, adult's hunger and children's hunger.[6] In addition to the food insecurity measures, the NNS also provides information on the proportion of underweight and under height children, among other statistics.

2.1.3 Social Weather Stations (SWS) measure of hunger indicator

One criticism of the official statistics for measuring poverty and hunger incidence (from NSCB and FNRI) is that "being infrequently applied, (it) has fostered an illusion that poverty steadily declines" (Mahangas, 2009). On the one hand, the FIES is conducted only once every three years and the official hunger and poverty incidence statistics were reported only eight times from 1985 to 2006. The poverty and hunger incidence statistics from the 2009 FIES will only be released in 2011. On the other hand, the FNRI-NNS is conducted once every five years, the latest being the 2008 survey. If we are interested in measuring the impact of the recent global financial crisis on hunger and poverty incidence in the country, we will have to wait for NSCB's results in 2011or FNRI's results in 2013. Due to the lack of a frequent measure of hunger incidence (and also poverty incidence) in the country, government officials depend on the national quarterly surveys on hunger conducted by the Social Weather Stations (SWS), particularly during periods between the FIES years.[7] The SWS is a private, non-profit scientific institute established in 1985 to generate social survey data. The SWS hunger indicator is defined as the proportion of household heads reporting that their families have experienced hunger, without having anything to eat, at least once in the last three months (Mangahas, 2009). The SWS quarterly survey has 1,200 respondents from various parts of the country. The respondents are asked if they have experienced hunger in the past three months. If the respondent answers yes, a second question is then asked regarding the frequency of the experience. The SWS further classifies hunger into moderate if it happened "only once" or "a few times" and severe if it happened "often" or "always."[8] The SWS quarterly hunger indicator is reported beginning July 1998 and covers 46 quarters up to December 2009. Maligalig (2008), however, pointed out that the SWS hunger incidence figures may underestimate the true values because of potential sources of bias due to its design components. She argues that while the sampling error for all estimates from the quarterly survey is about 2.83 percent, non-sampling error due to potential problems with the sampling frame and sampling strategy can increase the over-all sampling and non-sampling error.[9]

2.2. Accelerated Hunger-Mitigation Program (AHMP)

To address the problem of hunger in the medium and long term, the administration of President Gloria Macapagal-Arroyo (GMA) initiated the Accelerated Hunger Mitigation Program, a strategy under the Medium Term Philippine Development Plan (MTPDP) of 2004-2010. The AHMP aims to address the causes of hunger, poverty, unavailability of food to eat, and a large family size. The AHMP aims for a holistic approach in addressing the problem of hunger and intervenes in both the supply side and the demand side.

On the supply side, the Department of Agriculture (DA), the Department of Health (DOH) and the Department of Social Work and Development (DSWD) take measures to produce more food and efficiently deliver this to those who need it. Some examples of these interventions are: (a) the Food for School Program of the DOH where a daily ration of one kilo of rice is provided for the families of grade 1, pre-school and day care centre children; (b) the *Tindahan Natin* (our store) Project of the National Food Authority (an attached agency of the DA) and the DSWD. A poor family can buy low-priced but good quality rice and noodles at the "*tindahan*"; (c) The *Gulayan ng Masa* (backyard gardening) and the Barangay Food Terminal programs of the DA which aim to provide alternative food sources. On the demand side, the AHMP aims to hire workers from poor areas to clean and maintain the country's roads and highways under the emergency public work and food for work programs of the Department of Public Works and Highways (DEVPULSE-NEDA, 2007). The National Nutrition Council, an agency affiliated with the DOH is given the oversight function to ensure the implementation of the programs and projects within the AHMP framework (NNC, 2010).

Perhaps the most successful government intervention program in terms of mitigating hunger is the Pantawid Pamilyang Pilipino Programs or 4Ps. The 4Ps is a poverty reduction and social development strategy of the national government that provides conditional cash grants to extremely poor households to improve the health, nutrition and education, particularly of children aged 0-14. The households were selected from the poorest provinces, cities and municipalities based on the 2006 Family Income and Expenditure Survey (FIES) and 2003 Small Area Estimates (SAE) of NSCB, respectively. The poorest households in the municipalities are identified through a Proxy-Means Test that determines the socio-economic variables such as asset ownership, type of housing, education of the household head, livelihood of the family and access to water and sanitation. A household-beneficiary with three children whose ages range from 0 to 14 years can receive a maximum of P1,400 per month (about US$30 at US$1= P46 exchange rate) or P15,000 per year (about

US$326) as long as they comply with the conditions related to the family's health and education. These conditions state that pregnant women must avail of pre- and post-natal care and be attended during childbirth by skilled attendant, that parents must attend responsible parenthood sessions (for family planning), that children aged 0 to 5 years old must receive regular preventive check-ups and vaccines, that children aged 3 to 5 years old must attend day care or pre-school classes at least 85 percent of the time and that children aged 6 to 14 years old must be enrolled in elementary and high school at least 85 percent of the time and receive de-worming pills twice a year (DSWD, 2010). As of June 2009, the 4Ps covered about 700,000 households from 255 municipalities and 15 cities in 45 provinces, out of the total of 80 provinces. Some economists, notably Balisacan (2009), point out that the 4Ps with an allotted budget of about P10 billion per year for the 700,000 families is a more efficient program for poverty alleviation compared to the expensive Comprehensive Agrarian Reform Program (CARP) which cost the government about P237 billion (in 1997 prices) to implement.

3. Trends in Hunger Incidence

Official hunger statistics from the NSCB, shown in Table 1, indicate that while the percentage of subsistence poor among households have decreased from 8.2 percent in 2003 to 7.9 percent in 2009, the number of families that are considered subsistence poor actually increased from about 1.36 million to 1.45 million during the same period, largely due to a higher population. Moreover, the figures from the same table show the actual number of subsistence poor households increased in the three major geographical areas, in Luzon (about 31,000 families), Visayas (about 29,000) and Mindanao (about 36,000) in the same period.

Table 1. Subsistence Incidence and Magnitude of Poor among Families

Major Island Group	Poverty Incidence among Families (%)			Magnitude of Poor Families		
	2003	2006	2009	2003	2006	2009
PHILIPPINES	8.2	8.7	7.9	1,357,833	1,511,579	1,453,843
Luzon	4.4	5.1	4.3	418,439	506,974	449,388
Visayas	12.1	12.6	11.4	392,854	439,348	421,494
Mindanao	14.5	14.1	13.8	546,540	565,257	582,961

Source: National Statistical Coordination Board (NSCB)

Subsistence incidence among the population follows the same trend. The numbers in Table 2 shows that while the percentage of Filipinos experiencing hunger almost remained the same, decreasing only slightly from 11.1 percent in 2003 to 10.8 percent in 2009, the number of Filipinos experiencing hunger increased to 9.44 million in 2009 from 8.80 million in 2003. Furthermore, hunger incidence in Mindanao rose by about 264,000, to 3.71 million in 2009 from 3.44 million in 2003. It is clear that hunger incidence has worsened in Mindanao.

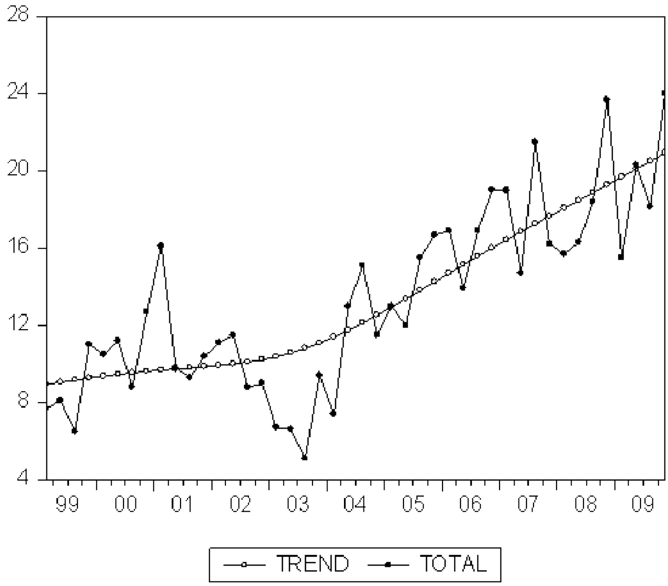
Table 2. Subsistence Incidence and Magnitude of Poor among the Population

Major Island Group	Subsistence Incidence among Population (%)			Magnitude of Poor Population		
	2003	2006	2009	2003	2006	2009
PHILIPPINES	11.1	11.7	10.8	8,802,918	9,851,362	9,440,397
Luzon	6.2	7.2	6.1	2,818,041	3,437,824	3,033,052
Visayas	16.2	16.8	15.3	2,540,826	2,806,891	2,699,031
Mindanao	18.7	18.2	18.2	3,444,051	3,606,647	3,708,314

Source: National Statistical Coordination Board (NSCB)

The time plot of the percentage of families experiencing hunger from the 1st quarter of 1999 to the 4th quarter of 2009 is shown in Figure 1 below, together with the estimate of the long-term trend of the percentage of hunger incidence computed using the Hodrick-Prescott (HP) filter.[10] The plot of the HP filter shows that the slope of the long-term trend component shifted during the 3rd quarter of 2003 and became steeper which indicates a relatively faster increase in the percentage of families that experienced hunger after the 3rd quarter of 2003 compared to the period before it.

Figure 1 Percentage of Families Experiencing Hunger (TOTAL) and the Long Term Trend (Hodrick-Prescott) from First Quarter 1999 to Fourth Quarter 2009



Source: Social Weather Stations (SWS) National Quarterly Surveys on Hunger and Authors' Computation of the Long Term Trend

4. Vector AutoRegressive (VAR) Model for Hunger Incidence

This paper examines the dynamic patterns of hunger incidence and the effects of the determinants of hunger, food prices and underemployment rate. A vector autoregressive (VAR) model is used to analyze the impact of shocks on food prices, measured using the food component of the Consumer Price Index (CPI), and underemployment rate on the current and future hunger incidence.[11] The food group, composed of rice, corn, dairy products, eggs, fish, meat, food and vegetables, among others, represents 46.58 percent of total weight of the CPI measured in 2000 (NSO, 2010). This group has the largest weight in the index and any change in the prices of the food group will have an impact on the overall inflation rate.

Underemployment rate is the proportion of underemployed persons to the total population 15 years old and up. Underemployed persons include all employed persons who express the desire to have additional hours of work in their present job or an additional job, or to have a new job with longer working hours. Visibly underemployed persons are those who work for less than 40 hours during the reference period and want additional hours of work (NSO, 2010). The summary statistics of the variables used in this study are given in Table 3 below.

Table 3. Summary Statistics for Hunger Incidence, Food Inflation and Underemployment Rate

	Hunger Incidence	Food Component of the CPI	Food Inflation	Underemployment Rate
Mean	13.29	121.90	1.20	19.56
Median	12.85	116.55	1.10	19.50
Maximum	24.00	165.50	6.91	26.10
Minimum	5.10	98.10	(1.62)	15.30
Std. Dev.	4.76	20.87	1.37	2.73
Skewness	0.36	0.73	1.71	0.45
Kurtosis	2.39	2.35	8.72	2.59

4.1 Augmented Dickey-Fuller (ADF) test

The time series on total hunger incidence, food component of the CPI and underemployment were tested for presence of unit roots using the Augmented Dickey-Fuller (ADF) test prior to building the VAR model. The results in Table 4 show that the time series hunger incidence and underemployment rates are stationary. However, the ADF test for the food component of the CPI showed that series has a unit root. The difference of the natural logarithm of the food component of the CPI was used in the VAR model.

Table 4. Results of the Augmented Dickey-Fuller (ADF) Tests

Variable	ADF test statistic	P-value	Conclusion
Underemployment	-3.913579	0.0042	Stationary
Total Hunger	-4.187364	0.0100	Stationary*
Food Inflation	2.825585	0.9985	Non-Stationary; I(1)

** Trend-Stationary series, the trend is deterministic*

4.2 Granger Causality test

One of the key questions that need to be addressed in regression is how useful some variables are for forecasting others. We need to investigate whether past values of a time series x_t can help forecast another series y_t. If it cannot, then we say that x_t does not Granger-cause y_t. The simplest and probably the best approach to test whether a particular observed series x_t Granger-causes y_t is through the use of the autoregressive (AR) specification. To implement this test, we assume a particular autoregressive lag p (usually selected using the AIC or SBC criterion) and estimate by OLS,

$$y_t = \alpha + \delta_1 y_{t-1} + \delta_2 y_{t-2} + \ldots + \delta_p y_{t-p} + \beta_1 x_{t-1} + \ldots + \beta_p x_{t-p} + \varepsilon_t$$

The results of the Granger causality tests show that Food Inflation *Granger Causes* Total Hunger Incidence at the 5 percent level of significance but not the reverse. Moreover, Underemployment also *Granger Causes* Total Hunger Incidence at the 10 percent level of significance but not the reverse.

Table 5. Results of the Granger-Causality Test

Null Hypothesis:	n	F-Statistic	P-Value *
Food Inflation does not Granger Cause Total Hunger	43	5.14096	0.0144 **
Total Hunger does not Granger Cause Food Inflation		1.00057	0.1616
Underemployment does not Granger Cause Total Hunger	43	2.81478	0.0506 ***
Total Hunger does not Granger Cause Underemployment		0.16594	0.3430
Underemployment does not Granger Cause Food Inflation	43	0.23455	0.3154
Food Inflation does not Granger Cause Underemployment		0.00583	0.4698

* *one-sided p-value;*
** *Food Inflation Granger Causes Total Hunger Incidence but not the reverse (at 5% level)*
*** *Underemployment Granger Causes Total Hunger Incidence but not the reverse (at 10% level)*

4.3 The VAR Model

The vector autoregressive (VAR) is commonly used for forecasting systems of interrelated time series and for analyzing the dynamic impact of random disturbances (or shocks) on the system of variables. The main distinction of the VAR approach, compared to the other econometric models, is that it treats every endogenous variable in the system as a function of the lagged values of all endogenous variables in the system. When we are not confident that a variable is actually exogenous, we can treat each variable symmetrically. In the three-variable case order one VAR (or VAR (1)) model we have,

$$
\begin{aligned}
y_t &= \beta_{10} - \beta_{12} z_t - \beta_{13} w_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \gamma_{13} w_{t-1} + \varepsilon_{yt} \\
z_t &= \beta_{20} - \beta_{21} y_t - \beta_{23} w_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \gamma_{23} w_{t-1} + \varepsilon_{zt} \\
w_t &= \beta_{30} - \beta_{31} y_t - \beta_{32} z_t + \gamma_{31} y_{t-1} + \gamma_{32} z_{t-1} + \gamma_{13} w_{t-1} + \varepsilon_{wt}
\end{aligned} \tag{1}
$$

where y_t is the total hunger incidence, z_t is the food inflation and w_t is the underemployment, all at quarter t. The ε_{yt}, ε_{zt} and ε_{wt} are white noise disturbance terms with means 0 and standard deviations σ_y, σ_z and σ_w, respectively. The equations in (1) are called the structural equations of the VAR. The parameters, β_{12}, β_{13}, β_{21}, β_{23}, β_{31} and β_{32} measure the contemporaneous effects while the γ's measure the lag 1 effects. The equations are not in reduced form since, for example, y_t has contemporaneous effect on z_t and w_t.

Isolating the time t variables on the left-hand side, we have,

$$
\begin{aligned}
y_t + \beta_{12} z_t + \beta_{13} w_t &= \beta_{10} + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \gamma_{13} w_{t-1} + \varepsilon_{yt} \\
\beta_{21} y_t + z_t + \beta_{23} w_t &= \beta_{20} + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \gamma_{23} w_{t-1} + \varepsilon_{zt} \\
\beta_{31} y_t + \beta_{32} z_t + w_t &= \beta_{30t} + \gamma_{31} y_{t-1} + \gamma_{32} z_{t-1} + \gamma_{33} w_{t-1} + \varepsilon_{wt}
\end{aligned} \tag{2}
$$

In matrix form,

$$
\begin{bmatrix} 1 & \beta_{12} & \beta_{13} \\ \beta_{21} & 1 & \beta_{23} \\ \beta_{31} & \beta_{32} & 1 \end{bmatrix}
\begin{bmatrix} y_t \\ z_t \\ w_t \end{bmatrix}
=
\begin{bmatrix} \beta_{10} \\ \beta_{20} \\ \beta_{30} \end{bmatrix}
+
\begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix}
\begin{bmatrix} y_{t-1} \\ z_{t-1} \\ w_{t-1} \end{bmatrix}
+
\begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \\ \varepsilon_{wt} \end{bmatrix}
$$

Simplifying, we have,

$$
\begin{aligned}
B\underline{x}_t &= \Gamma_0 + \Gamma_1 \underline{x}_{t-1} + \underline{\varepsilon}_t \\
\underline{x}_t &= B^{-1}\Gamma_0 + B^{-1}\Gamma_1 \underline{x}_{t-1} + B^{-1}\underline{\varepsilon}_t \\
\underline{x}_t &= \mathrm{A}_0 + \mathrm{A}_1 \underline{x}_{t-1} + \underline{e}_t
\end{aligned} \tag{3}
$$

where

$$
\underline{x}_t = \begin{bmatrix} y_t \\ z_t \\ w_t \end{bmatrix},\
B = \begin{bmatrix} 1 & \beta_{12} & \beta_{13} \\ \beta_{21} & 1 & \beta_{23} \\ \beta_{31} & \beta_{32} & 1 \end{bmatrix},\
\Gamma_0 = \begin{bmatrix} \beta_{10} \\ \beta_{20} \\ \beta_{30} \end{bmatrix},\
\Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix},\
\varepsilon_t = \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \\ \varepsilon_{wt} \end{bmatrix}
$$

The equations in (3) are called the reduced-form representation of a VAR (1) model. We can generalize the mathematical representation of the reduced-form VAR model as,

$$
\underline{x}_t = A_0 + A_1 \underline{x}_{t-1} + A_2 \underline{x}_{t-2} + \ldots + A_p \underline{x}_{t-p} + \underline{e}_t \tag{4}
$$

where $\underline{x}_t$ is a (k x 1) vector of endogenous variables, $\underline{A}_1$, $\underline{A}_2$,…, $\underline{A}_p$ are matrices of coefficients to be estimated, and $\underline{e}_t$ is a (k x 1) vector of forecast errors that may be contemporaneously correlated but are uncorrelated with their own lagged values and uncorrelated with all of the right-hand side variables. The error vector e_t is assumed to be normally distributed with mean $\underline{0}$ and covariance matrix $\underline{\boldsymbol{\Sigma}}$. The order of the VAR model (p) is determined using the information criteria (Akaike, Schwarz and the Hannan-Quinn).

The results of the VAR (1) model using the quarterly time series data on total hunger incidence, food inflation and underemployment are given in Table 6 below. The paper is interested in the first equation of the VAR where the dependent variable is total hunger incidence (under the column total hunger). The total hunger incidence at quarter t can be explained significantly by the lag 1 values of total hunger incidence, food inflation and underemployment. Lag 1 values of total hunger incidence and food inflation are significant at 1 percent level while lag 1 value of underemployment is significant at the 10 percent level.

While the VAR model in Table 6 can be used to forecast the future hunger incidence, the estimated parameters are not that useful in analyzing the dynamic relationships of food inflation and underemployment on total hunger incidence since the errors in equation (4) are not the original structural errors but the forecast errors. The dynamic relationship of the VAR model is derived using the Impulse Response Function (IRF).

Table 6. VAR (1) model Total Hunger Incidence, Food Inflation and Underemployment

	Total Hunger	Food Inflation	Underemployment
Total Hunger (lag 1)	0.69	0.05	0.04
	(0.11)	(0.05)	(0.09)
	[6.50]	[1.08]	[0.40]
Food Inflation (lag 1)	0.83	0.34	(0.00)
	(0.34)	(0.15)	(0.29)
	[2.42]	[2.31]	[-0.02]
Underemployment (lag 1)	0.32	(0.05)	0.46
	(0.17)	(0.07)	(0.14)
	[1.88]	[-0.64]	[3.25]
Constant	(2.75)	1.04	10.08
	(3.41)	(1.48)	(2.86)
	[-0.81]	[0.70]	[3.53]
R-squared	0.63	0.18	0.23
Adj. R-squared	0.60	0.11	0.17
Akaike Info. Criterion*	5.12	3.44	4.76
Schwarz Info. Criterion*	5.28	3.61	4.92

Standard errors are in ( ) and t-statistics in [ ];
** lag 1 is selected as the appropriate lag order using the AIC and SIC*

4.4 Impulse Response Function (IRF)

A shock to the i^{th} variable (e.g. increase in food prices or underemployment rate) not only directly affects the i^{th} variable but is also transmitted to all the other endogenous variables, in particular total hunger incidence, through the dynamic (lag) structure of the VAR. An impulse response function traces the effect of a one-time shock to one of the innovations (error terms) on the current and future values of the endogenous variables. If the error terms are contemporaneously uncorrelated, then the i^{th} innovation (ε_{it}) is simply a shock to y_{it} or what is referred to as "shock to itself."

4.5 Response of total hunger to a shock in food inflation

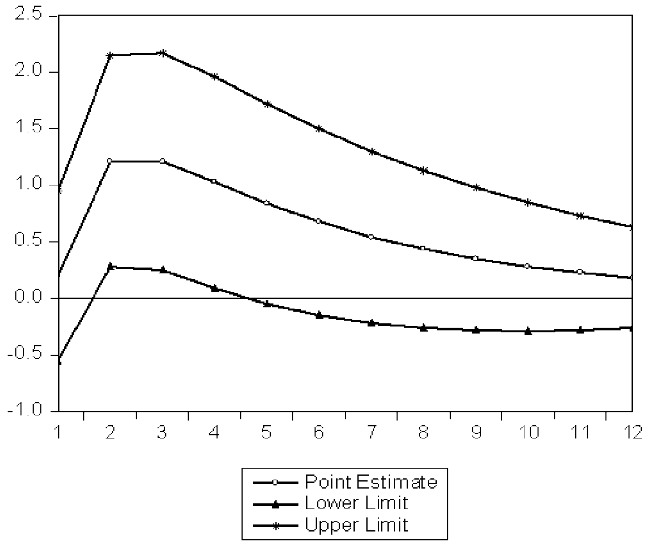
The response of total hunger to a shock in food prices is given in Table 7 below. The IRF shows that a one-time shock (or increase) to food prices at quarter t will have a significant effect on total hunger for the succeeding five periods, starting at quarters (t + 1) and ending at quarter (t +5). The effect of a shock to food prices is significant in increasing total hunger incidence at the 5% level for the first 3 quarters and significant at the 10% level for the last two quarters. After quarter (t + 5), the effect of the shock to food prices on total hunger is no longer significantly different from zero (or the effect decays to zero) as shown in Figure 2. In particular, a one standard deviation increase to food inflation (about 1.37 percentage points) at quarter 1 will increase total hunger by about 1.21 standard deviation or 5.76 percentage points in the next quarter, all things being the same. The increases in the next four quarters are: 5.77 percentage points (in quarter 3), 4.90 percentage points (in quarter 4), 3.99 percentage points (in quarter 5) and 3.22 percentage points (in quarter 6). The numbers mean that total hunger incidence is very sensitive to changes in food prices, a spike in food inflation equivalent to say one percentage point at the current quarter will increase hunger incidence by 4.21 percentage points in the next quarter or an additional 772,000 households that will experience hunger.

Table 7. Impulse Response Function – Response of Total Hunger Incidence to a one standard deviation increase in Food Inflation at Quarter 1

Quarter	Impact of an Increase in Inflation to Total Hunger	t-stat
1	0.20	0.44
2 *	1.21	2.13
3 *	1.21	2.08
4 *	1.03	1.81
5 **	0.84	1.56
6 **	0.68	1.35
7	0.54	1.17
8	0.44	1.03
9	0.35	0.91
10	0.28	0.82
11	0.23	0.74
12	0.18	0.67

** significant at the 5 percent level; ** significant at the 10% level (one-sided alternative)*
Cholesky Ordering: Food Inflation, Underemployment, Total Hunger

Figure 2. Increase in Hunger Incidencefrom Quarter 1 to Quarter 12 Resulting from One Standard Deviation Increase in Food Prices at Quarter 1



4.6 Response of total hunger to a shock in underemployment

While the one-time shock to food prices affects total hunger for a period of five quarters, the IRF results in Table 8 show that a one-time increase in underemployment rate at quarter t will have significant effects on total hunger for the succeeding two quarters: quarters (t +1) and (t + 2). Moreover, the effect of an increase in underemployment rate to total hunger is significant only at the 10% level. A one-standard deviation increase in underemployment rate, equivalent to about 2.37 percentage points at quarter 1 will increase total hunger incidence by about 0.76 standard deviation or 3.63 percentage points in the next quarter (quarter 2), all things being the same. The increase in quarter 3 is about 0.79 stanard deviation or 3.76 percentage points. After quarter 3, the impact of one-time increase in underemployment rate is no longer significant. An increase in underemployment rate of one percentage point at the current quarter will increase hunger incidence by 1.33 percentage points in the next quarter or an additional 244,000 households that will experience hunger.

The good news is that underemployment rate in April 2010 at 17.8 percent is lower compared to that of April 2009 at 18.9 percent and also lower than the average underemployment rate of 19.56 percent from the 1st quarter of 1999 to the 4th quarter of 2009 (although the unemployment rate spiked up to 8 percent in the same quarter).

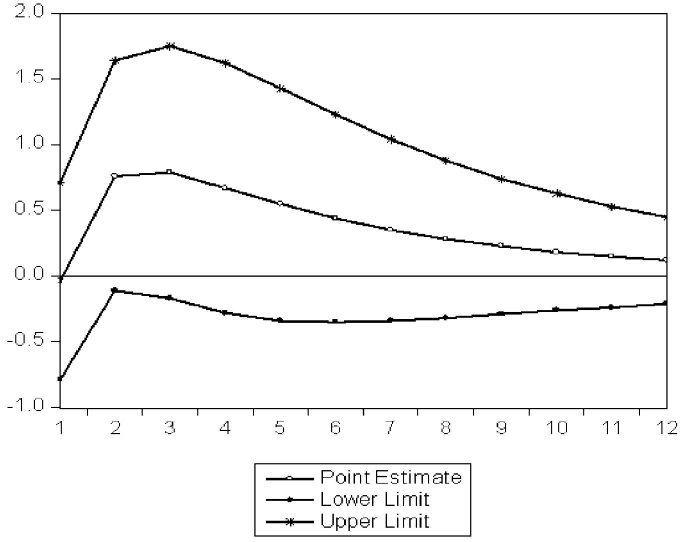
Moreover, of the 414,000 new jobs created from April 2009 to April 2010, about 210,000 of these came from the manufacturing sector. The bad news is that a significant number of new jobs created are still in the informal sector, such as trade-related jobs, and the government-created emergency jobs mainly as a response to the global financial crisis.

Table 8. Impulse Response Function – Response of Total Hunger Incidence to a One Standard Deviation Increase in Underemployment at Quarter 1

Quarter	Impact of an Increase in Inflation to Total Hunger	t-stat
1	-0.04	-0.09
2 **	0.76	1.43
3 **	0.79	1.36
4	0.67	1.16
5	0.55	1.02
6	0.44	0.92
7	0.35	0.84
8	0.28	0.78
9	0.23	0.72
10	0.18	0.67
11	0.15	0.63
12	0.12	0.59

*** significant at the 10% level (one-sided alternative)*
Cholesky Ordering: Food Inflation, Underemployment, Total Hunger

Figure 3 Increase in Hunger Incidence from Quarter 1 to Quarter 12 resulting from One Standard Deviation Increase in Underemployment Rate at Quarter 1



4.7 Forecast error variance decomposition

While the impulse response functions trace the effects of a shock to one endogenous variable on the other variables in the VAR model, the Forecast Error Variance Decomposition tells us the proportion of the movements in the series (e.g. total hunger) due to its "own" shocks versus the shocks to the other variables (food inflation and underemployment). In applied research it is typical for a variable to explain almost all of its forecast error variance at short horizons and smaller proportions at longer horizons. The variance decomposition provides information about the relative importance of each random innovation in affecting the variables in the VAR model. The forecast error variance decomposition of total hunger given in Table 9 below shows how much of the future error variance of total hunger can be explained by shocks to total hunger, food inflation and underemployment. The results show that shock to total hunger (or "own shock") can explain almost all, 99.53 percent, of the variance of the forecast error of total hunger at quarter (t + 1). The shocks to food inflation and underemployment have negligible effect to the forecast error variance of total hunger at the next quarter. However, at quarter (t + 2), about 14 percent of the forecast error variance of total hunger can now be explained by shocks to food inflation and underemployment. At quarter (t + 3), the total variance explained by food inflation and underemployment increased to about 21 percent. This value stabilizes at around 28 percent which implies that shocks to food inflation and underemployment explain about 28 percent of the future forecast error variance of total hunger, making these two variables important determinants of total hunger.

Table 9. Forecast Error Variance Decomposition of Total Hunger

Period	S.E.	Total Hunger	Food Inflation	Underemployment
1	2.988684	99.53245	0.446861	0.020694
		(4.44744)	(3.36775)	(2.72259)
2	3.898978	86.25213	9.909387	3.838486
		(10.6173)	(9.50938)	(5.48614)
3	4.445916	78.83321	15.06570	6.101084
		(13.1000)	(12.2656)	(8.47213)
4	4.777240	75.05440	17.68829	7.257312
		(14.2422)	(13.7084)	(10.4050)
5	4.981098	73.01823	19.10478	7.876993
		(14.8482)	(14.5297)	(11.6201)
6	5.108408	71.85578	19.91541	8.228810
		(15.2368)	(15.0431)	(12.4095)

5. Conclusions

This paper examines the dynamic patterns of hunger incidence in the Philippines using the quarterly survey data on hunger from the Social Weather Stations (SWS). The results of the econometric model based on the vector autoregressive (VAR) show that food inflation and underemployment are important determinants of hunger incidence in the Philippines. A one-time increase in food prices can lead to increases in hunger incidence that will last for five quarters, while a one-time increase in underemployment will lead to increases in hunger incidence for two quarters. An important contribution of this paper is the mainstreaming of the time series data on hunger incidence from the SWS into the econometric model through the VAR models. The results of the study are useful in crafting policies and programs that could help alleviate hunger in the country. For one, hunger incidence is very sensitive to changing food prices and thus the supply side strategies of the AHMP such as increasing food production and enhancing the efficiency of logistics and food delivery must be improved. Take for example the case of the *Tindahan Natin* (TN) stores that sells low-priced but quality goods. The SWS survey in June 2006 (no available data on access after this period) shows that only 6.6 percent of households said that there is a TN outlet in their locality and only 3.0 percent actually bought something from these outlets. Clearly, only a small percentage of the poor households have been reached by this program. Increasing the number of TN stores to increase the number of poor households that can access these outlets should be a priority of the DSWD.

However, Manasan and Cuenca (2007)[12] found that the two major programs under the hunger mitigation initiative of the Arroyo administration, namely: Tindahan Natin Program and Food-for-School Program, suffer from poor targeting. Exclusion and leakage rates are considerably high. Thus, in addition to increasing the number of TN stores, targeting system of the TN, as well as the FSP, program should be revisited.

In the case of the underemployment, increasing the number of new jobs that will be created and enhancing the quality of jobs are important factors that will decrease the hunger incidence in the country. Priorities should be made in the area of improving the investment climate for investors through stable and predictable government policies as well as battling corruption and red tape in government transactions.

The paper shows that hunger incidence is very dynamic and frequent monitoring, for example, quarterly, of hunger incidence through self-rated surveys, perhaps at the provincial level, is important in order to monitor and assess the effectiveness of the government programs (e.g. Tindahan Natin, Conditional Cash Transfer, Food for School, Comprehensive Livelihood and Emergency Employment Program) in mitigating hunger. These self-rated surveys can complement the official statistics on hunger incidence computed by the NSCB every three years from the FIES.

Finally, policies that address the hunger incidence in the country must include measures that will manage the country's bourgeoning population and bring down the fertility rate to a manageable level. Millions of Filipinos go through the vicious cycle of high fertility and poverty and hunger: a high fertility rate prolongs poverty/ hunger in households and poor households contribute to high fertility rates. Policy makers must address the country's rapid population growth head-on through proactive government policies.

Notes

1 Earlier version of this paper was presented at the Millennium Development Goals (MDG) Forum organized by the National Academy of Science and Technology (NAST) last February 2, 2010 at Trader's Hotel in Manila and at the Annual Seminar Series on Food Security and Sustainable Development organized by the Institute of Statistics (INSTAT) and the Southeast Asian Regional Center for Graduate Study and Research in Agriculture (SEARCA), University of the Philippines in Los Baños. The authors are grateful to the participants, particularly Dr. Arsenio M. Balisacan, Dr. Mahar Mangahas, Dr. Romulo A. Virola, for their comments and suggestions. The authors are also grateful to Administrator Carmelita N. Ericta of the National Statistics Office (NSO) and Ms. Rosie Sta. Ana, Chief of the Economic Indices and Indicators Division of the NSO, for providing us with the time series data on the Food Component of the Consumer Price Index, to Dr. Mahar Mangahas of the SWS for sharing the time series data on hunger incidence and to Dr. Jocelyn Juguan of the Food and Nutrition Research Institute (FNRI) for sharing the highlights of the 7th National Nutrition Survey (NNS). The authors would like to thank two anonymous referees for their helpful comments and suggestions. All errors are the authors' responsibility.

2 The average growth rate of the Gross Domestic Product (GDP) in constant prices from 2001 to 2009 is about 4.5 percent and with population growing at an average rate of 2.04 percent per year, the per capita GDP growth is only between 2.4 to 2.5 percent, a modest rate by East Asian standard.

3 The annual population growth from 2000 to 2007 is 2.04% based on the results of the 2007 Census of Population. In August 2007, the population of the Philippines is 88.57 million.

4 The Survey of Hunger Incidence in the Philippines (SHIP) was conducted by the Bureau of Agricultural Statistics (BAS), a service agency of the Department of Agriculture (DA) in August 2006. The SHIP covered more that 13,000 household-respondents. The SHIP used the same questions asked in the SWS quarterly survey. Unfortunately, no follow-up survey was made after 2006 and the results from the SHIP are not amenable for comparison across time.

5 Salud-Payumo (2009) also discussed the four measures discussed by Maligalig (2008) and referred to the NSCB measure of hunger incidence as the quantitative measure while the hunger measures from the SWS, BAS and FNRI as the qualitative measures. The NSCB's hunger index is measured as the average of three indicators: (a) the proportion of households with per capita energy consumption less than the requirement; (b) proportion of underweight children under 5 years and; (c) mortality rate of children under 5 years.

6 The 10 questions used are discussed in Maligalig (2008), pp. 120-121.

7 Government agencies involved in the Anti-Hunger Mitigation programs (AHMP), such as the Department of Social Work and Services (DSWD), National Nutrition Council (NNC) and the National Economic and Development Authority (NEDA) make use of the SWS hunger incidence indicator to gauge the effectiveness of the strategies.

8 While the SWS hunger indicator reports the total hunger incidence as well as the moderate and severe hunger incidence, this paper focus only on the total hunger incidence for its analysis.

9 The readers are referred to Maligalig's 2008 paper "Examining the Existing Direct Measures of Hunger in the Philippines" for an extensive discussion on the sampling and estimation issues. When one would like to measure the (partial) effect of a variable, X_t, but we can only observe an imperfect measure $\tilde{X}_t$, where $\tilde{X}_t = X_t + w_t$, one can show that the least squares estimator $\hat{\beta}_1$ has probability limit $\beta_{1=}(\sigma^2 x / (\sigma^2 x + \sigma^2_w))$ and is biased toward zero. However, if the error (w_t) is constant, the variance, σ^2_w is zero and the estimator is still consistent.

10 The HP filter, first proposed by Hodrick and Prescott (1997) uses a smooting method to obtain an estimate of the long-term trend component of a time series. The HP filter computes the permanent component (TR_t) of a time series y_t by minimizing the variance of y_t around TR_t, subject to a penalty that constrains the second difference of $TR_{t.}$

11 This paper uses underemployment rate as a measure of the quality of jobs, rather than unemployment rate.

12 PIDS Discussion Paper Series No. 2007-10 titled "Who benefits from the Food-for-School Program and Tindahan Natin Program: lessons in targeting"

REFERENCES

BALISACAN, A.M., 2009, Pathways Out of Rural Poverty: Is Agrarian Reform Passé?, A paper presented at the 6th Ayala Corporation-UPSE Economic Forum, at the University of the Philippines School of Economics Auditorium, July 10, 2009.

BALISACAN, A.M. and MAPA D.S., 2010, Beyond the 2010 Elections: Addressing the Key Constraints to Poverty Reduction, A paper presented at the League of Corporate Foundations (LCF) Forum on March 15, 2010 in Makati City.

DEPARTMENT OF SOCIAL WORK AND DEVELOPMENT (DSWD), Website: www.dswd.gov.ph.

DEVPULSE, 2007, Winning the War vs. Hunger, National Economic and Development Authority (NEDA) Development Advocacy Factsheet Vol. XI, No. 3, April 15, 2007.

__________, 2008, Nutrition in the Philippines: A Continuing Challenge, National Economic and Development Authority (NEDA) Development Advocacy Factsheet Vol. XII, No. 16, August 31, 2008.

FOOD AND AGRICULTURAL ORGANIZATION OF THE UNITED NATIONS, 2009, More People than Ever are Victims of Hunger, Website: www.foa.org.

FOOD NUTRITION RESEARCH INSTITUTE (FNRI), Website: www.fnri.dost.gov.ph.

HODRICK, R. and E. C. PRESCOTT, 1997, Postwar U.S. Business Cycles: An Empirical Investigation, *Journal of Money, Credit, and Banking*, 29:1-16.

MALIGALIG, D.S., 2008, Examining the Existing Direct Measures of Hunger in the Philippines, *The Philippine Statistician*, 57: 93-124.

MANGAHAS, M., 2009, The Role of Civil Society in Poverty Monitoring: The Case of the Philippines, A paper presented on The Impact of the Global Economic Situation on Poverty and Sustainable Development in Asia and the Pacific, September 28-30, 2009, Hanoi, Vietnam.

MANASAN, ROSARIO G. and CUENCA, J.S., 2007, Who Benefits from the Food-for-School Program and Tindahan Natin Program: Lessons in Targeting, PIDS Discussion Paper Series No. 2007-10.

NATIONAL STATISTICAL COORDINATION BOARD, Website: www.nscb.gov.ph.

NATIONAL STATISTICS OFFICE, Website: www.census.gov.ph.

POPE BENEDICT XVI, 2009, Address of His Holiness Benedict XVI to the Food and Agricultural Organization (FAO) on the Occasion of the World Summit on Food Security, November 16, 2009 at the FAO Headquarters in Rome.

SALUD-PAYUNO, C., 2009, Identifying Methodology for Measuring Hunger in the Philippines, July 24, 2009 Report to the Anti-Hunger Task Force, National Nutrition Council (NNC), Department of Health.

SOCIAL WEATHER STATIONS, 2010, Fourth Quarter 2009 Social Weather Survey: Hunger at new record-high 24.0%, SWS Media Release, January 12, 2010, Website: www.sws.org.ph.

Substance Use among Serious Adolescent Offenders Following Different Patterns of Antisocial Activity

Michelle Besana
University of the Philippines Visayas

Edward P. Mulvey, Ph. D.
University of Pittsburgh School of Medicine

The present study examines individual differences in the levels of substance use in a sample (n = 1,067) of male serious adolescent offenders following distinct trajectories of criminal offending over a three year period. The levels of substance use are compared for the different offender groups controlling the effects of age, ethnicity, and diagnosis of previous drug and alcohol abuse/ dependence. The association between antisocial activity and the level of substance use was also examined and compared for the different groups after controlling the effect of institutional placement. The growth or decline in substance use was investigated and compared for the different groups above and beyond the effects of antisocial activity and institutional confinement.

After fitting a series of hierarchical generalized linear models for repeated measurements data, results revealed that significant differences in the level of substance use exist among the different offender groups in the sample. Antisocial activity is associated with the level of substance use over time after controlling the effect of institutional placement in all offender groups. Above and beyond the effect of antisocial activity and institutional placement, substance use is increasing over the data collection period in all groups, but the rate of growth is highest in the lowest offending group.

Keywords: hierarchical generalized linear models, growth curve models, substance use, antisocial activity, delinquency, serious adolescent offenders

1. Introduction

Substance use is a significant problem among the youth, especially those involved with the juvenile justice system (Prinz & Kern, 2003). Adolescents in the juvenile justice system have rates of substance use several times that seen in the general adolescent population (Deschenes & Greenwood, 1994). In addition, the rate of diagnosable substance use disorders is higher among the more serious youth offenders (Huizinga & Jakob-Chien, 1998) and is estimated to be approximately one-half of this group of adolescents (Grisso, 2004). The co-occurrence of substance use problems and delinquency (Teplin et al., 2002; Young et al., 2007) as well as the link (Winters, 1998; Dawkins, 1997; D'Amico et al., 2008; Hammersley et al, 2003; Chassin, 2008; Elliot et al, 1985) between the two behaviors among adolescent offenders has received significant attention in previous research studies. Researchers believe that the association between the two behaviors is reciprocal in nature (D'Amico et al., 2008; Sullivan & Hamilton, 2007) and that they are predicted by similar risk factors (Elliot et al, 1985; Mason & Windle, 2002).

The current study examines the relationship between substance use and antisocial activity or delinquent criminal behavior in a sample of male serious adolescent offenders over a three year period. The analyses presented here addresses the question of whether or not the link between the two behaviors holds in groups of adolescent offenders that follow distinct patterns of offending. The different patterns of offending used in the study are those reflecting different offending trajectories derived in previous research (Mulvey et al., 2010). The level of substance use as well as its growth or decline over time in each offender group will be examined. The impact of antisocial activity on substance use over time for each group will also be explored after controlling the effect of institutional placement. Institutional placement is included in the analysis because spending time in a controlled residential environment curtails the opportunities for substance use, significantly affects the trajectories of criminal offending, and affects the relation between these two behaviors in a sample of serious adolescent offenders (Piquero et al., 2001; Mauricio et al., 2009; Mulvey et al., 2010). The effects of subject level variables such as age, ethnicity, and diagnosis of previous drug or alcohol abuse/dependence on the level of substance use are also included in the analysis. Previous studies revealed that these variables are significant predictive factors for illicit drug use in adolescent population (Young et al., 2002; Ljubotina et al., 2004; Howard & Jenson, 1999; Steinberg, 2002; Hofler et al., 1999; von Sydow et al., 2002; Chassin et al., 1996; Kandel et al., 1986).

In studying the relationship between delinquent behavior and its associated problems such as substance use in a sample of chronic youth offenders, it is important to distinguish between different offender groups (Eklund & Klinteberg, 2009) because patterns of substance use in this sample of high risk individuals may also be different. The knowledge of the levels of substance use and the relationship between substance use and criminal behavior in each offender group would allow policymakers to design early preventive programs and interventions that recognize the different needs of the individuals in the different groups.

2. Review of Literature

2.1 The unilateral relationship between substance use and antisocial behavior

Numerous studies support the link between substance use and antisocial behavior among adolescent offenders. On one hand, some studies assert that substance use predicts antisocial behavior among adolescent offenders; on the other hand, others support the reverse relationship. In a cross-sectional study of 293 highly delinquent offenders between 14 to 18 years old composed predominantly of males (81%), over a half of the group agreed that alcohol or drugs had been associated with getting upset or angry which had led eventually to offending (Hammersley et al., 2003). This study concludes that substance use predicted offending and that socially acceptable drugs (alcohol, tobacco, and cannabis) did so more than any other drugs. This same study, meanwhile, also found that almost half (44%) of the sample of highly delinquent adolescent offenders committed serious crimes to obtain money for drugs or alcohol (Hammersley et al., 2003), supporting the notion that substance use leads to offending.

Longitudinal studies on adolescent offenders have also examined this relationship over time. Studies confirm that previous substance use is a significant predictor of subsequent serious offending among adolescent offenders (D'Amico et al., 2008; Dembo et al., 2007) while other investigations indicate that conduct problems and aggression predict adolescent illicit substance use (Kellam et al., 1983) and that delinquent behavior predicts subsequent substance abuse and dependence (Chassin et al., 1999; Disney et al., 1999). Furthermore, studies also show that delinquent behavior among adolescent offenders eventually leads to accelerated levels of substance use over time (Hill et al., 2000; Hussong et al., 1998).

2.2 The link between substance use and antisocial behavior

The literature provides substantial evidence of the prevalence of substance use and substance use disorder among adolescents particularly among adolescent offenders. It has been recognized that very early substance use is a significant problem among the youth especially those who end up in the juvenile justice system (Prinz & Kern, 2003). Since the mid-1990s, the use of marijuana, stimulants, cocaine, and LSD is rising among the adolescent population (Steinberg, 2002), and evidence shows that the youth involved in the juvenile justice system are several times more likely to use alcohol and other drugs than adolescents in general (Deschenes & Greenwood, 1994). The rates of substance use and diagnosable substance use disorder are observed to be higher among more serious adolescent offenders (Huizinga & Jakob-Chien, 1998), and the proportion of adolescent offenders with diagnosable substance use disorder is approximately one-half (Grisso, 2004).

The comorbidity (co-occurrence) of substance use problems and delinquency among adolescent offenders is also supported in the literature. About half of the males and almost half of the females who had encounters with the juvenile justice system had substance use problems, with marijuana use disorder being the most common (Teplin et al., 2002). One study found that the prevalence of substance use disorders among adolescents aged 12-17 who had encounters with the juvenile justice system is almost three times that of the youth in the same age range who had never been jailed or detained (NSDUH Report, February 27, 2004). Another study indicates that adolescent offenders who continue to use drugs will more likely continue offending (Young et al., 2007).

The evidence of the link between antisocial behavior and substance use among adolescent offenders abound in the literature. On one hand, many adolescent conduct problems can result from alcohol and drug use; on the other hand, it is also possible that substance use may contribute to the maintenance of adolescent delinquency which may continue into adult antisocial behavior (Winters, 1998). Many studies support the link or positive association between substance use and delinquent criminal behavior among adolescents (Dawkins, 1997; D'Amico et al., 2008; Hammersley et al., 2003; Chassin, 2008; Elliot et al., 1985). Many researchers believe that this positive association is reciprocal in nature and that the relationship between the two behaviors is fairly stable over time (D'Amico et al., 2008; Sullivan & Hamilton, 2007), and that they are predicted by similar risk factors (Elliot et al., 1985; Mason & Windle, 2002).

2.3 The Pathways to Desistance Study

The Pathways to Desistance Study is a large-scale, two-site, prospective longitudinal study of a cohort of serious adolescent offenders. The rationale behind the study can be found in Mulvey et al. (2004). The study started in November, 2000, and enrollment of the subjects was completed in March, 2003. The project follows a sample N=1,354 serious juvenile offenders from adolescence to young adulthood in two metropolitan cities: Philadelphia, Pennsylvania and Phoenix, Arizona. The primary aims of the study are: (a) to describe the patterns by which serious adolescent offenders stop engaging in illegal activities; (b) to describe the role of social context and developmental changes in promoting these positive changes; and, (c) to compare the effects of sanctions and interventions in promoting these changes. The participants in the study are adolescent offenders between the ages of 14 and 18 at the time of adjudication, who have been found guilty of a serious offense (almost exclusively felony offenses, with a few serious misdemeanor charges included, e.g. weapons offense, sexual assault). Assessments were done at baseline and every six months thereafter for a period of three years and yearly assessments follow for a period of five years.

Table 1 shows the descriptive statistics of the enrolled participants in the Pathways to Desistance Study. The enrolled sample is 86% male and 14% female; it is 20% White, 41% African-American, 34% Hispanic, and 5% other. The average age of the participants at study index petition was 16.24 years old (S.D.= 1.10 years) with an average of approximately 2 prior petitions (S.D.= 2.14) at the time of adjudication. About 44% of the participants had been adjudicated of serious crimes against persons, 25% of property crime, 16% of drug offense, 10% of weapons offense, and 4% of other offenses (Schubert et al., 2004).

Table 1 Descriptive Statistics of Enrolled Research Participants and the Sample Used

Characteristic	Pathways sample		Sample used	
N	1354		1067	
Mean age at study index petition	16.24	(1.10)	16.21	(1.11)
Mean number of prior petitions [a]	1.92	(2.14)	2.03	(2.19)
Mean age at first prior petition	14.93	(1.64)	14.86	(1.64)
Race/Ethnicity				
Caucasian/White	20%		20%	
African American/Black	41%		43%	
Hispanic	34%		36%	
Other	5%		0%	
Most serious offense [b]				
Crime against person	44%		44%	
Property crime	25%		27%	
Drug offense	16%		14%	
Weapons offense	10%		10%	
Other	4%		4%	
Missing data	1%		1%	

Note. The values in parentheses are standard deviations
[a] *Average count of all prior petitions available in the subject's court records excluding probation violations*
[b] *Most serious charge on study index petition*

2.4 Selected findings from the Pathways study involving substance use and antisocial behavior

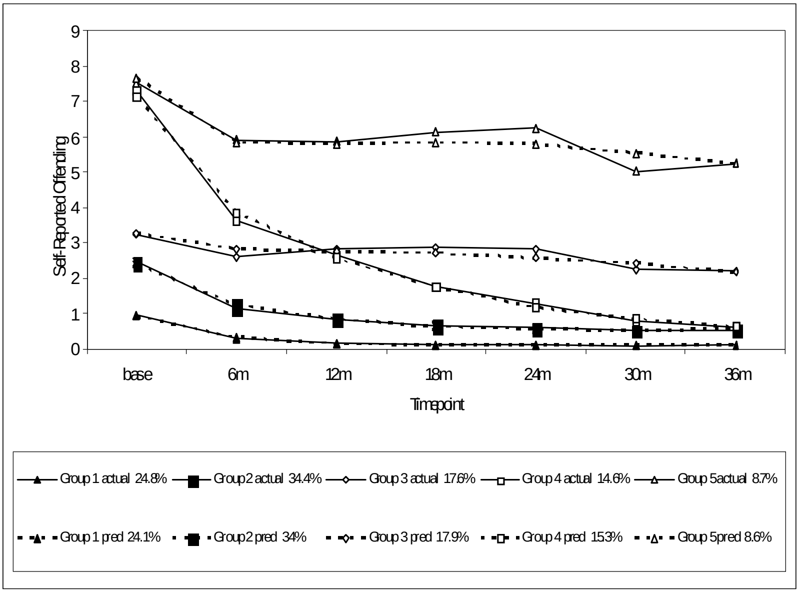
Several papers presented or published about the Pathways to Desistance Study have documented the relationship between substance use and antisocial behavior in this sample. One study found that the presence of a drug or alcohol disorder and level of substance use predict the level of self-reported offending and number of arrests (Losoya & Chassin, 2004). Moreover, the researchers found that baseline substance use and substance use disorders also predict continued delinquency involvement (after controlling for baseline offending levels), and that the relationship between substance use and substance use disorder and self-reported offending is strongest for those who had spent the least time in an institution over the one year follow-up period. Another study on the association between the two behaviors indicates that substance use goes hand-in-hand with criminal offending in this group of adolescent offenders across multiple waves of interviews (Mulvey et al., 2010). The study on the effects of risk and protective factors on alcohol and marijuana use across time on a sample of male participants revealed that time in supervised facility was shown to alter alcohol and marijuana trajectories over time (Mauricio et al., 2009).

A subsample of the Pathways to Desistance Study consisting of only males who completed at least four interviews (n=1,119) in a period of three years (including baseline assessment) were considered in another set of analyses which focused on finding distinct developmental offending trajectories by which the subjects behave after court adjudication and the different factors that differentiate these patterns. The purpose of these analyses was to give a better picture of the different pathways of criminal behavior among these adolescent offenders and to eventually provide explanations why adolescents desist from doing further criminal offending and why others continue (Mulvey, et al., 2010). After incorporating the time at risk for offending in the community in the analysis, the study identified five (5) developmental offending trajectory groups (see Figure 1), including two low offending groups (constituting about 59% of the sample; Groups 1 and 2), a moderate offending group (17.6%; Group 3), a "desister" group (14.6%; Group 4), and a "persister" group (8.7%; Group 5). The results revealed that the five trajectory offending groups differ significantly (but not dramatically) in terms of age, ethnicity, antisocial history, deviant peers, a criminal father, substance use, and psychosocial maturity. These factors, however, did not significantly differentiate the persister from the desister group.

3. Objectives of the Study

The general purpose of this paper is to study the nature of association between substance use and antisocial behavior across time among male serious adolescent offenders that follow distinct trajectories of offending after taking into account the effect of institutional placement. Specifically this paper aims to find out whether antisocial activity predicts substance use across time among the different groups of male serious adolescent offenders following distinct patterns of offending after controlling for institutional placement. The specific objectives of this paper are the following: (1) to compare the level of substance use among male serious adolescent offenders following different patterns of offending; (2) to determine the effect of age, ethnicity, previous drug or alcohol abuse/dependence problems, and institutional placement on the level of substance use; (3) to investigate the relationship between

Figure 1 Five-group trajectory solution using the Zero Inflated Poisson Model



antisocial activity and substance use across time among the different offender groups after controlling institutional placement; and, (4) to compare the rate of growth or decline in substance use among the different offender groups.

4. Method

This study is a secondary data analysis on a subsample of the participants in the Pathways to Desistance Study. Table 1 shows the descriptive characteristics of the complete Pathways sample and the sample used for this study. The sample (n=1,067) includes only the males who completed at least three follow-up interviews done approximately every six months over a period of three years who belong to three major ethnicity groups, namely: Caucasian/White, African American/Black, and Hispanic. The sample is composed of 20% White, 43% African American, and 36% Hispanic. The average age at study index petition of the subjects in the sample is 16.21 years old (S.D. =1.11). On the average, the subjects had 2.03 (S.D.=2.19) prior petitions and the average age of the subjects at first prior petition is 14.86 years old (S.D.=1.64). About 44% of the subjects in the sample had been convicted of serious crimes against persons, 27% of property crime, 14% of drug offense, 10% of weapons offense, and 4% of other offenses. In terms of the mean age at study index petition, mean number of prior petitions, mean age at first prior petition as well as the most serious charge on study index petition, this sample does not significantly differ from the over-all Pathways sample.

The outcome variable in this study is the level of substance use across the 36-month period excluding baseline measurement, namely: assessments at the 6-month, 12-month, 18-month, 24-month, 30-month, and the 36-month follow-up interview. The potential level 1 predictors are measurement occasions, antisocial activity, and institutional placement. Both antisocial activity and institutional placement are treated as time-varying covariates. The potential level 2 predictors include age at baseline interview, ethnicity, previous drug or alcohol abuse/ dependence problems, and membership in the different trajectory groups of offending.

Statistical analysis of the outcome variable and the various potential explanatory variables employs the hierarchical generalized linear models for repeated measurements data using the Negative Binomial distribution (Raudenbush & Bryk, 2002). The GLIMMIX procedure of SAS was used to derive the estimates of the different models considered in the model building procedure.

4.1 Measures

In the context of this study, substance use is measured by items adapted from the Alcohol and Health Study at the University of Missouri (Chassin et al., 1991). This self-report measure considers the adolescent's use of illegal drugs and alcohol over the course of his/her lifetime and in the past six-months. Substance use in this study is measured by aggregating the frequency of use (for the past 6 months) for alcohol, marijuana/hashish, sedatives/tranquilizers, stimulants/amphetamines, cocaine, opiates, ecstasy, hallucinogens, inhalants, and amyl nitrate/odorizers. The scale used for each substance is the following: 0: (never), 1: (1-2 times in the last 6 months), 2: (3-5 times in the last 6 months), 3: (1 time per month), 4: (2-3 times per month), 5: (1 time per week), 6: (2-3 times per week), 7: (4-5 times per week), and 8: (every day).

Antisocial activity is measured by the modified version of the Self-Report of Offending (Elliot, 1990; Huizinga et al., 1991) scale which measures the adolescent's involvement in antisocial and illegal activities at each interview. The scale is composed of 22-items listing different illegal activities, and the subject indicates whether or not he has done any of these activities over the "last six months." A sum of the number of items endorsed ranging from 0 – 22 is calculated and is used as a measure of antisocial activity for the purpose of this study. This over-all variety score has been shown to be a reliable and valid measure of the adolescent's overall involvement in illegal activities (Osgood et al., 2002; Thornberry & Krohn, 2000). The 22 items included in the self-report of offense measure are the following: (a) destroyed property, (b) set fire, (c) broke into building, (d) stole from store, (e) bought something stolen, (f) used credit card illegally, (g) stole car/motorcycle, (h) sold marijuana, (i) sold other illegal drugs, (j) carjacked someone, (k) drove while drunk/high, (l) was paid for sex, (m) forced someone for sex, (n) killed someone, (o) shot someone, (p) shot at someone (pulled trigger), (q) took something (with weapon), r) took something (no weapon), (s) beat up someone, (t) was in a fight, (u) beat up someone (with a gang), and (v) carried a gun.

The measure for previous drug or alcohol abuse/dependence diagnosis is taken from the Composite International Diagnostic Interview (CIDI, 1990), a comprehensive, fully structured interview used to assess a variety of mental disorders including alcohol abuse, alcohol dependence, drug abuse, and drug dependence. An indicator variable for this measure was used in the study to indicate the presence (1) or the absence (0) of the diagnosis.

For the purpose of this study, institutional placement is taken as the proportion of time in the recall period of approximately six months that the subject spent in a psychiatric hospital, detox/drug treatment facility, secure facility, or a residential treatment facility. This is determined from the subject self report at each follow-up interview.

The variable GROUP in this study is a categorical variable that represents membership in one of the five different developmental offending trajectory groups identified in previous research (Mulvey et al., 2010) that reflects the subject's pattern of criminal offending. Table 2 shows the number of subjects in the sample belonging to these five trajectories.

Table 2 Number of Subjects in the Sample Belonging to the Different Trajectory Groups of Offending

Trajectory Group	Number
Group 1	265
Group 2	367
Group 3	189
Group 4	155
Group 5	91
Total Sample	1067

4.2 The proposed model

The study uses hierarchical generalized linear models to describe individual change over time. In the context of this study, it is used to model the relationship between the dependent variable (frequency of substance use) and a number of independent variables over time. We let the dependent variable Y_{ti} represent the frequency of use for 10 substances measured at time t for subject i. The measurement occasions include assessments of approximately every six months for a period of three years. If we let μ_{ti} represent the average frequency of substance use for subject i at time t, we suppose that $Y_{ti} \mid \mu_{ti}$ follows a Negative Binomial distribution. The model assumes that the mean and variance of Y_{ti} are not equal. The variance function for this model is given by $\mu_{ti} + k\mu^2_{ti}$, where k is taken as a scale parameter (Schabenberger, 2005).

We model η_{ti} as the log average frequency of substance use for subject i at time t; thus, $\eta_{ti} = \log \mu_{ti}$. The level 1 and level 2 hierarchical structures are shown below.

Level 1 (Repeated measurements over time t for subject i)

$$\eta_{ti} = \pi_{0i} + \pi_{1i}\, Antisocial_{ti} + \pi_{2i}\, Lock_{ti} + \pi_{3i}\, Time_{ti} \tag{1}$$

Level 2 (Model for subject i)

$$\pi_{0i} = \beta_{00} + \beta_{01}\, Group_i + \beta_{02}\, Age_i + \beta_{03}\, Ethni_i + \beta_{04}\, Diag_i + r_{0i} \tag{2}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}\, Group_i + r_{1i} \tag{3}$$

$$\pi_{2i} = \beta_{20} + r_{2i} \tag{4}$$

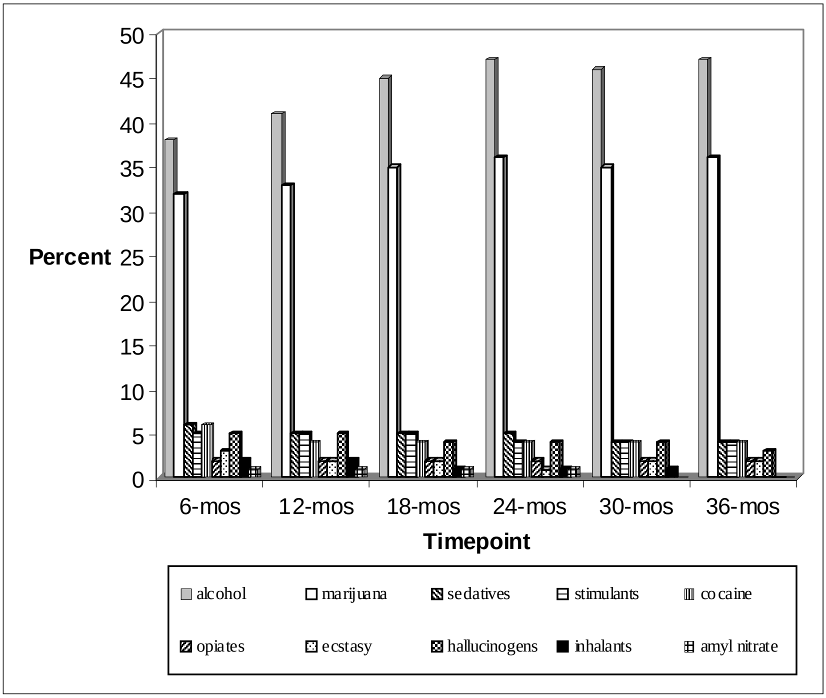
$$\pi_{3i} = \beta_{30} + \beta_{31}\, Group_i + r_{3i} \tag{5}$$

The predictors at level 1 are the level of antisocial activity (ANTISOCIAL), institutional placement (LOCK), and measurement occasions (TIME). For this study, antisocial activity and institutional placement are time-varying covariates and group mean centered. Measurement occasion is centered at the 18-month data collection period to facilitate easier and more meaningful interpretation of the growth parameters. At level 2, the predictors of the initial status (π_{0i}) are trajectory group membership (GROUP), age centered at 14 (AGE), ethnicity (ETHNI), and diagnosis of previous drug or alcohol abuse/dependence problems (DIAG). The model also takes into account the effect of trajectory group membership on the slope of antisocial activity and the rate of growth. The error terms at level 2 (r_{0i}, r_{1i}, r_{2i} and r_{3i}) are assumed to follow a multivariate normal distribution with a zero mean vector and a 4x4 unstructured variance-covariance matrix.

5. Results

The percentages of the sample that had alcohol to drink over the recall period are increasing across time with the following percentages at each assessment: 38%, 41%, 45%, 47%, 46%, and 47%, respectively (see Figure 2). The percentages of the sample that smoked marijuana over the recall period are 32%, 33%, 35%, 36%, 35%, and 36% respectively; thus, the rate of increase is minimal over time. The percentages of the sample that used other substances across time are approximately the following: sedatives (5%), stimulants (5%), cocaine (4%), opiates (2%), ecstasy (2%), hallucinogen (4%), inhalants (1%), and amyl nitrate (< 1%). Thus, it is clear that the level of use for these other substances is very low across time with the use of inhalants and amyl nitrate being the lowest.

***Figure 2.* Proportions of the sample using the different substances across time**



The frequency of use across the ten (10) substances over time (see Figure 3) does not only show the significantly large number of subjects that do not use any of these substances but also display a highly positively skewed distribution of the dependent variable over time.

Preliminary exploratory techniques using smooth non-parametric as well as the ordinary least squares summary of the individual change over time for selected subjects in the sample (see Figures 4 and 5) suggest differences in the level of substance use as well as in the rate of change over time. The graphs also suggest modeling individual change using linear or quadratic growth parameters. The smooth non-parametric and the ordinary least squares summary across all the subjects also reveal the same observation (see Figure 6). The graph suggests differences in the individual level of substance use and in the rate of change across time. Furthermore, the graph validates the use of models that are linear or quadratic in time to model individual change.

Figure 3. The Frequency of Use across 10 Substances over a Three Year Period for all Subjects

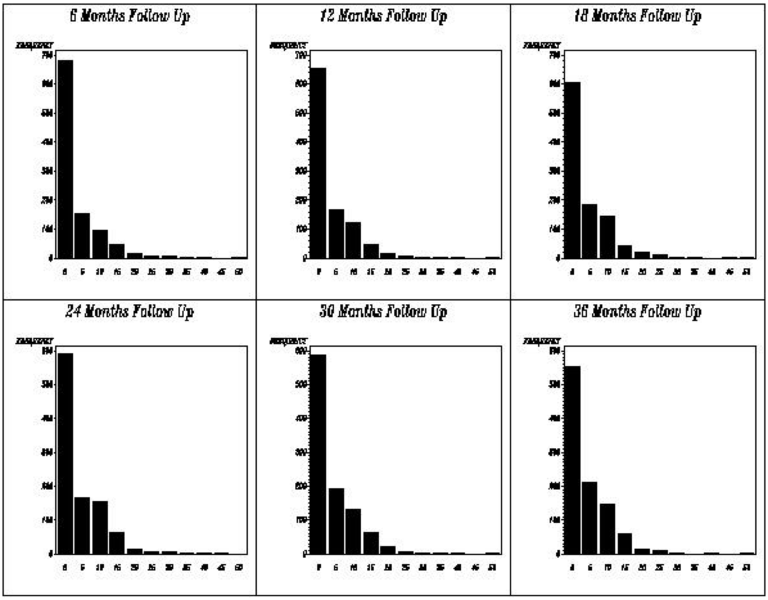


Figure 4. **Smooth non-parametric summaries of the individual change over time for selected subjects**

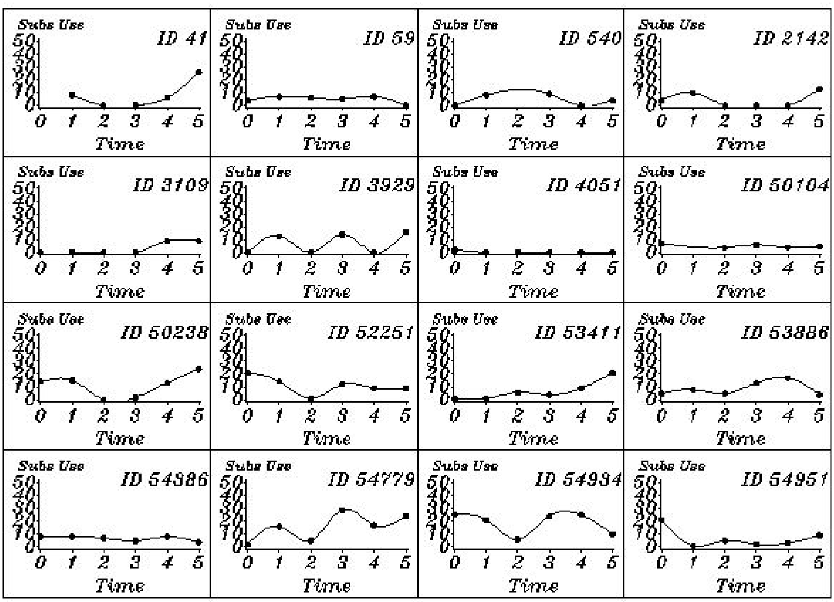


Figure 5. **OLS summaries of the individual change over time for selected subjects**

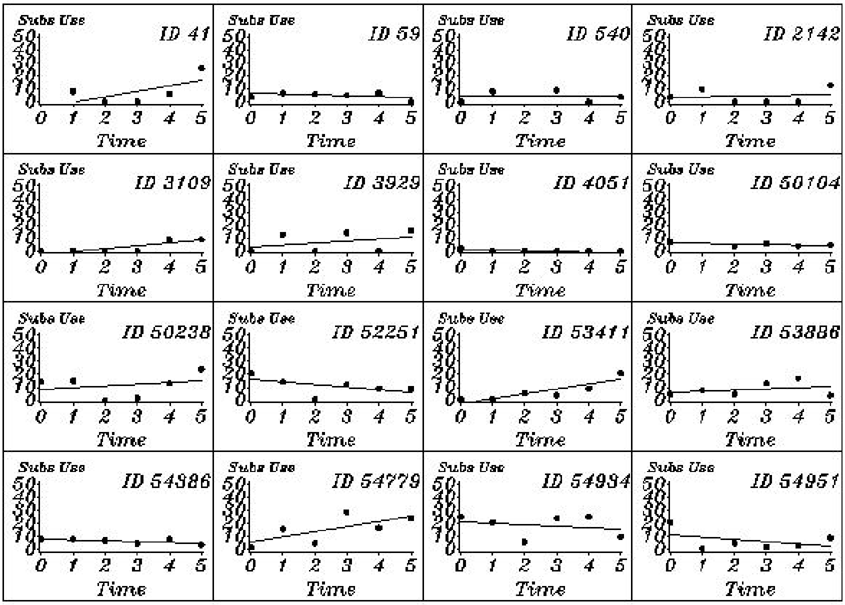
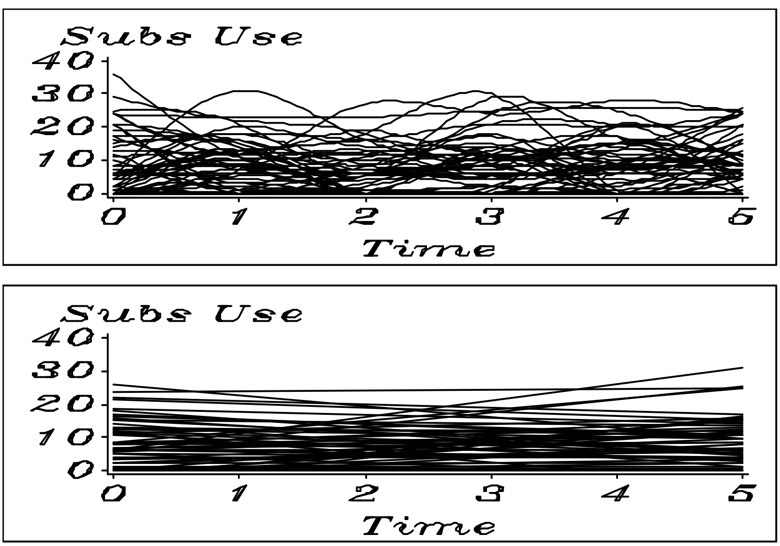


Figure 6. **Smooth non-parametric and OLS trajectories of substance use over time for all subjects**



Further graphical exploratory analysis using the OLS summaries of substance use across time for each trajectory offending group (see Figure 7) supports differences in the level of substance use as well as in the growth rate in substance use among the five offending groups. In addition, OLS summaries of substance use against the level of antisocial activity for each longitudinal offending group (see Figure 8) reveal differences in the strength of relationship between the two variables among the five developmental trajectories.

Exploratory techniques used to demonstrate the relationship between the dependent and the independent variables suggest that it is reasonable to use a hierarchical generalized model which is linear in the time variable. The link function used in the model was the log link function and the sampling model used at level 1 was the Negative Binomial distribution as opposed to the Poisson distribution since the data manifest overdispersion problems (Schabenberger, 2005). The model building procedure starts with the fully unconditional growth model and builds up by successively including all the potential level 1 and level 2 predictors in the proposed model until the final model is identified where all the model coefficients are significant and the fit statistics suggest a reasonable fit.

The result for fitting the fully unconditional growth model (model A) suggests that there is still unexplained variability in the dependent variable. We next fit a model (model B) using measurement occasions (TIME) as the sole level 1 predictor and the growth rate is constant while the level of substance use is random across the subjects. We then fit a model (model C) with measurement occasions, level of antisocial activity, and institutional placement as level 1 predictors and the growth rate and the slope of antisocial activity are constant across subjects but the slope of institutional confinement is random across subjects. In addition, we include the effect of membership to different trajectory offending groups (GROUP), age centered at 14 (AGE), ethnicity (ETHNI), and diagnosis of previous drug or alcohol abuse/ dependence (DIAG) on the level of substance use (π_{0i}). At the final stage of model fitting, we include the effect of membership to the different longitudinal offending groups (GROUP) on the growth rate and the slope of antisocial activity (ANTISOCIAL) to the previous model to form model D. The structural models for model D are as follows:

The Final Model (Model D)

Level 1

$$\eta_{ti} = \pi_{0i} + \pi_{1i}\, Antisocial_{ti} + \pi_{2i}\, Lock_{ti} + \pi_{3i}\, Time_{ti} \qquad (6)$$

Level 2

$$\pi_{0i} = \beta_{00} + \beta_{01}\, Group_i + \beta_{02}\, Age_i + \beta_{03}\, Ethni_i + \beta_{04}\, Diag_i + r_{0i} \qquad (7)$$

$$\pi_{1i} = \beta_{10} + \beta_{11}\, Group_i \qquad (8)$$

$$\pi_{2i} = \beta_{20} + r_{2i} \qquad (9)$$

$$\pi_{3i} = \beta_{30} + \beta_{31}\, Group_i \qquad (10)$$

Figure 7. **OLS summaries of substance use over time for each group for selected subjects**

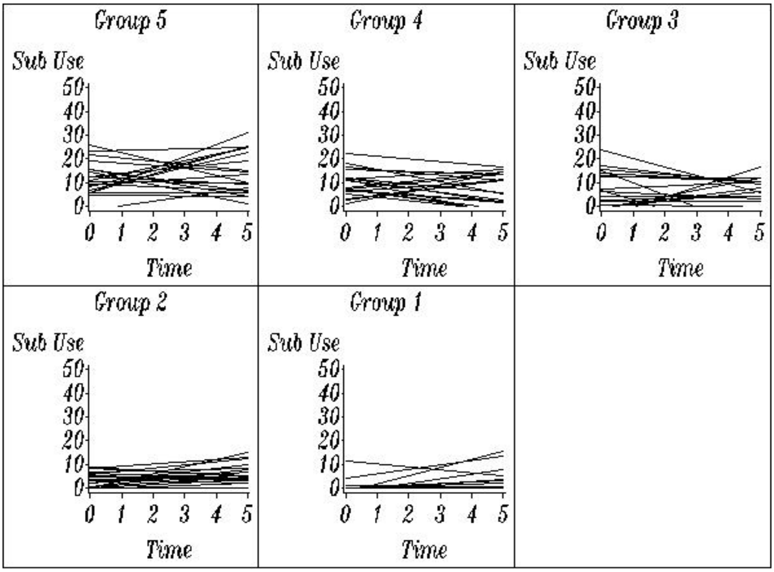
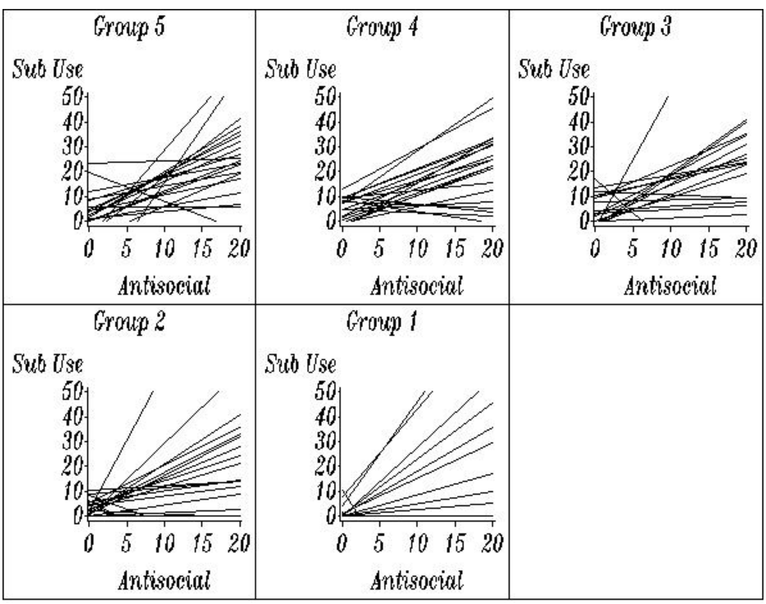


Figure 8**. OLS summaries of substance use against the level of antisocial activity for each group for selected**



The scale parameter for model D is highly significant which means that the use of the Negative Binomial distribution as a sampling model at level 1 is appropriate. Also, the ratio of the value of the generalized chi-square to the over-all degrees of freedom (DF) is 1.08 indicating no serious problems of overdispersion and that the model fits the data since this fit-statistic should ideally be around 1. This further supports the appropriateness of the use of Negative Binomial distribution at level 1. In addition, measurement occasions (TIME) was centered at the different assessment periods to check whether the model holds across the data collection period. The result indicates that model D fits the data appropriately and that the model indeed holds across the data collection period. Thus, model D is the final model selected and its result is shown in Table 3 together with the results from fitting the other models.

Table 3 Summary of Model Fitting Using Negative Binomial Distribution

	Model A	Model B	Model C	Model D
Fixed Effect				
Model for π_{0i}				
Intercept	1.38**	1.36**	1.26**	1.22**
Group 5			1.77**	1.85**
Group 4			1.16**	1.27**
Group 3			1.51**	1.59**
Group 2			0.90**	0.94**
Age14			0.15**	0.15**
Hispanic			-0.28**	-0.29**
Black			-0.56**	-0.58**
Diagnosis			0.25**	0.25**
Model for Antisocial Slope, π_{1i}				
Intercept			0.17**	0.36**
Group 5				-0.26*
Group 4				-0.21^{ns}
Group 3				-0.17^{ns}
Group 2				-0.04^{ns}
Model for Inst Placement Slope, π_{2i}				
Intercept			-1.28**	-1.31**
Model for Ind Level Growth Rate, π_{3i}				
Intercept		0.04**	0.07**	0.17**
Group 5				-0.12**
Group 4				-0.16**
Group 3				-0.12**
Group 2				-0.07*
Random Effect				
Intercept	0.84**	0.84**	0.54**	0.52**
Institutional Time			0.30**	0.34**
Scale	1.28**	1.31**	1.54**	1.54**
Fit Statistics				
Gener. Chi-sq/DF	1.00	1.00	1.06	1.08

** 0.01<p < 0.05 , ** p < 0.01, ^{ns} not significant*

The result of the final model indicates that substance use for typical subjects in groups 5, 4, 3, and 2 is significantly higher than those in group 1 after controlling for the effect of age, ethnicity, and diagnosis of previous drug or alcohol abuse/ dependence problems. Age has a strong significant positive effect on substance use for a typical subject after controlling for the effect of trajectory group membership, ethnicity, and diagnosis of previous substance abuse/dependence. Substance use for a typical Hispanic or African American subject is significantly lower than his White counterpart after controlling for the effect of trajectory group membership, age, and diagnosis of previous drug or alcohol abuse/dependence. Substance use for a typical subject who had previous drug or alcohol abuse/dependence problems is significantly higher than his counterpart with no previous drug or alcohol abuse/dependence problems after controlling for the effect of trajectory group membership, age, and ethnicity.

On the average, the level of antisocial activity has a strong significant positive relation with substance use for all offender groups. The estimated slope of antisocial activity in group 5 is significantly lower than those in the other trajectory offending groups (Groups 1, 2, 3, and 4) after controlling for the effect of institutional placement and assessment period. There is no significant difference in the estimated slope of antisocial activity among groups 1, 2, 3 and 4 holding constant the effect of institutional placement and assessment period. Institutional placement has a strong significant negative effect on substance use after controlling for the effect of the level of antisocial activity and assessment period. The estimated growth rates in substance use for groups 5, 4, 3 and 2 are significantly lower than group 1 after controlling for the effect of the level of antisocial activity and institutional placement.

6. Discussion

This paper models the differences in the individual level of substance use over a three year period of male adolescent offenders convicted of the most serious crimes taking into account the effect of the level of antisocial activity, institutional placement, age, ethnicity, the presence of previous drug or alcohol abuse/dependence problems, and membership in different trajectories of offending. The levels as well as the rates of growth in substance use over time are compared for the different offender groups in the sample. The paper further explores the nature and the degree of relationship that may exist between substance use and antisocial activity for the different groups.

The result from the final model reveals that, across time, substance use for typical subjects belonging to four of the longitudinal offending groups is significantly higher than those belonging to the lowest offending group (Group 1) whenever age, ethnicity, and diagnosis of previous drug or alcohol abuse/dependence are held constant. This indicates that above and beyond the effect of age, ethnicity, and diagnosis of previous drug or alcohol abuse/dependence, the average levels of substance use across time for typical (those with average levels of antisocial activity and institutional placement) subjects belonging to groups 2,3,4, and 5 are significantly higher than that in group 1. This finding is consistent with the result of a previous study of highly delinquent youth offenders with substance use involvement which supports that drug use was found to be highest among the most frequent offenders, lower in the medium offenders, and lowest in the less frequent offenders (Hammersley et al., 2003).

For a typical subject, age has a strong significant positive effect on substance use whenever the other subject level variables are held constant. This indicates that above and beyond the effect of ethnicity, presence of previous drug and alcohol abuse/dependence problems, and trajectory group membership, typical (those with average levels of antisocial activity and institutional placement) subjects who are older have significantly higher levels of substance use. This result supports findings of previous research that age (older adolescents) is a significant predictor of cannabis use or overall level of drug use among young people (Young et al., 2002; Ljubotina et al., 2004; Howard & Jenson, 1999).

Substance use for typical Hispanic or African American subjects is significantly lower than typical White subjects whenever the other subject level variables are held constant. This implies that above and beyond the effect of age, trajectory group membership, and diagnosis of previous drug or alcohol abuse/dependence problems, the level of substance use for White subjects (those with average levels of antisocial activity and institutional placement) is significantly higher than both African American and Hispanic subjects. This result replicated previous research findings that there exist ethnic differences in substance use and abuse (Steinberg, 2002), and that rate of substance use disorder is highest among White incarcerated or detained adolescent offenders and lowest among African American subjects (Teplin et al., 2006).

Substance use for a typical subject who had previous drug or alcohol abuse/dependence problems is significantly higher than a typical subject with no previous drug or alcohol abuse/dependence problems whenever the other subject level variables are held constant. This means that whenever age, ethnicity, and trajectory group membership are held fixed, the level of substance use for typical (those with average levels of antisocial activity and institutional placement) subjects with previous drug or alcohol abuse/dependence problems is significantly higher than his counterpart with no diagnosis of previous drug or alcohol abuse/dependence. This finding is consistent with the result of previous studies. Previous research findings indicate that early onset of substance use is a key predictor of illicit drug use among young people (Hofler et al., 1999; von Sydow et al., 2002; Howard & Jenson, 1999). Previous studies contend that diagnosis of substance use disorder is a strong predictor of adolescent's future alcohol and drug use which may even continue until adulthood (Chassin et al., 1996; Kandel et al., 1986).

On the average, the level of antisocial activity has a strong significant positive relation with substance use for all offender groups whenever the level of institutional placement and assessment period are held constant. The effect is significantly weaker among subjects belonging to the highest offending group (Group 5) compared to all other groups (Groups 1, 2, 3, and 4); in addition, there is no significant difference in the slope of antisocial activity over time among the four groups. It is worth noting that, in the marginal analysis (ignoring the effect of institutional placement) the correlation between substance use and antisocial activity across time is, on the average, observed to be highest in the persister group (Group 5) and lowest in the lowest offending group (Group 1). However, after controlling the effect of institutional placement, the reverse situation occurs. Although this may seem odd, this scenario occurs because institutional placement affects the level of offending in samples of active offenders (Piquero et al., 2001). Furthermore, the effect of institutional placement is different in different groups of adolescent offenders following different behavioral patterns (Mulvey et al., 2010). This finding is also supported by another result in this study that, on the average, institutional placement significantly lowers substance use whenever the level of antisocial activity and assessment period are controlled. According to the model, each additional delinquent act committed in the future is associated with an estimated increase in the average frequency of substance use by about 2.29 for group 1 and only 1.26 for group 5, for fixed levels of institutional placement and assessment period. Thus, above and beyond the effect of institutional confinement and assessment period, the level of antisocial activity, on the average, is associated with the level of substance use in different groups of adolescent offenders with different offending patterns. This result supports previous research findings that delinquent behavior predicts substance use among chronic young offenders (Kellam et al., 1983; Hill et al., 2000; Chassin et al., 1999; Disney et al., 1999). This study does not only confirm the strong link between delinquency and substance use among serious adolescent offenders (Dawkins, 1997; D'Amico et al., 2008; Mulvey et al., 2010; Chassin, 2008; Hammersley et al., 2003) but this also reveals that this strong link exists in different groups of serious adolescent offenders having different offending patterns across time even after the effect of institutional placement has been controlled.

Institutional placement has a strong significant negative effect on the level of substance use whenever the level of antisocial activity and assessment period are held constant. This research finding suggests that institutional placement, on the average, lowers the level of substance use across time above and beyond the effect of antisocial activity. For fixed levels of antisocial activity, longer periods of institutional confinement are associated with lower levels of substance use across time. This result is consistent with the findings of previous research that residence in a supervised setting suppressed age-related growth in alcohol and marijuana use (Mauricio et al., 2009). The result of the present study also confirms the significant effect of controlled residential environments on the trajectories of substance use and criminal offending of serious youth offenders (Piquero et al., 2001; Mauricio et al., 2009; Mulvey et al., 2010).

On the average, substance use is increasing over the data collection period but the rates of growth in substance use among subjects belonging to the four trajectory groups (Groups 5, 4, 3, and 2) of offending are, on the average, significantly lower than that in the lowest offending group (Group 1), whenever the level of antisocial activity and institutional placement are held constant. This implies that above and beyond the effect of antisocial activity and institutional placement, the rates of increase in substance use in the four offending groups are, on the average, significantly lower than that in the lowest offending group. The rate of increase in substance use in group 5 is mainly due to the increasing consumption of alcohol and marijuana while it is the increasing consumption of alcohol in group 1 which drives the increase in substance use. This finding reiterates previous results (Mauricio et al., 2009) on another subsample from the same Pathways to Desistance Study which supports that higher levels of initial substance use were associated with slower rate of growth in substance use and lower levels of initial use corresponds to higher growth rate. It is not known how much of these effects might be attributable to ceiling and floor effects on substance use.

In summary, the study supports that, across the three year data collection period, high levels of substance use in the sample are associated with subjects who are older, White, with previous substance abuse/dependence problems, and belong to longitudinal offending groups other than the lowest offending group (Group 1). Longer periods of institutional confinement are associated with lower levels of substance use. High levels of antisocial activity or delinquent behavior are associated with high levels of substance use. Furthermore, antisocial activity is still associated with the level of substance use over time, even after controlling the effect of institutional confinement, for all offender groups. And finally, above and beyond the effect of antisocial activity and institutional confinement, substance use is increasing over the data collection period in all trajectory offending groups, with the highest rate of increase in the lowest offending group.

The paper recognizes the importance of distinguishing between different offender groups when examining the relationship between delinquent behavior and associated problems (such as substance use in this case; Eklund & Klinteberg, 2009). The main contribution of this paper is its simultaneous analysis of the individual differences in substance use across time in a group of male serious youth offenders that follow different offending patterns. In addition, the levels of substance use are compared for different offender groups controlling the effects of other subject level variables and the association of antisocial activity with the level of substance use was examined and compared for the different groups after controlling for the different effects of institutional confinement. Finally, the growth in substance use was also investigated and compared for the different groups after controlling for the different effects of antisocial activity and institutional placement. The knowledge of the levels of substance use and the relationship between substance use and offending in each offender group suggests that preventive programs and interventions that recognize the different needs of the individuals in each group would be useful.

It is important to note that the sample used in this study was a purposive sample of serious adolescent offenders; these are males belonging to three major ethnicity groups, with female youth offenders and other ethnicity groups not represented in the analysis. The interpretation of the results may only work for the kind of population where the sample came from and may not reflect the actual situation in the general population of adolescent offenders. The self-report data on antisocial activity and substance use should also be interpreted with caution since juvenile offenders, particularly African Americans, may under-report their involvement in criminal activity and illegal substances and self-reports on substance use involvement may be more appropriate only for past use rather than the current use (McClelland et al., 2004).

REFERENCES

CHASSIN, L., 2008, Juvenile Justice and Substance Use, *Juvenile Justice*, 18(2):165-183.

CHASSIN, L., CURRAN, P.J., HUSSONG, A.M., & COLDER, C.R., 1996, The Relation of Parent Alcoholism to Adolescent Substance Use: A Longitudinal Follow-up Study, *Journal of Abnormal Psychology*, 105(1):70-80.

CHASSIN, L., PITTS, S., DELUCIA, C., & TODD, M., 1999, A Longitudinal Study of Children of Alcoholics: Predicting Young Adult Substance Use Disorders, Anxiety and Depression, *Journal of Abnormal Psychology*, 108:106-119.

CHASSIN, L., ROGOSCH, F., & BARRERA, M., 1991, Substance Use and Symptomatology among Adolescent Children of Alcoholics, *Journal of Abnormal Psychology*, 100(4):449-463.

D'AMICO, E.J., EDELEN, M.O., MILES, J., & MORRAL, A., 2008, The Longitudinal Association between Substance Use and Delinquency among High-risk Youth, *Drug and Alcohol Dependence*, 93:85-92.

DAWKINS, M.P., 1997, Drug Use and Violent Crime among Adolescents, *Adolescence*, 32:395-405.

DEMBO, R.,WAREHAM, J., & SCHMEIDLER, J., 2007, Drug Use and Delinquent Behavior: A Growth Model of Parallel Processes among High-risk Youths, *Criminal Justice and Behavior*, 34:680-696.

DESCHENES, E.P., & GREENWOOD, P.W., 1994, Treating the Juvenile Drug Offender, in D. MacKenzie & C. Uchida (eds.) *Drugs and Crime: Evaluating Public Policy Initiatives*, Thousand Oaks, CA, pp.253-280.

DISNEY, E., ELKINS, I., MCGUE, M., & IACONO, W., 1999, Effects of ADHD, Conduct Disorder, and Gender on Substance Use and Abuse in Adolescence, *American Journal of Psychiatry*, 156:1515-1521.

EKLUND, J.M. & KLINTEBERG, B., 2009, Alcohol Use and Patterns of Delinquent Behavior in Male and Female Adolescents, *Alcohol and Alcoholism*, 44(6):607-614.

ELLIOT, D.S., 1990. *National Youth Survey*, Denver, CO: University of Colorado, Institute of Behavioral Science.

ELLIOT, D.S., HUIZINGA, D., & AGETON, S.S., 1985, *Explaining Delinquency and Drug Use*, London: Sage.

GRISSO, T., 2004, *Double Jeopardy: Adolescent Offenders with Mental Disorders,* Chicago, IL: University of Chicago Press.

HAMMERSLEY, R., MARSLAND, L., & REID, M., 2003, Substance Use by Young Offenders: The Impact of the Normalization of Drug Use in the Early Years of the 21st Century, Home Office Research Study No.261.

HILL, K., WHITE, H. R., CHUNG, I.J., HAWKINS, J.D., & CATALANO, R.F., 2000, Early Adult Outcomes of Adolescent Binge Drinking: Person- and Variable-centered Analyses of Binge Drinking Trajectories, *Alcoholism: Clinical and Experimental Research*, 24:892-901.

HOFLER, M., LIEB, R., PERKONIGG, A, ET AL., 1999, Covariates of Cannabis Use Progression in a Representative Population Sample of Adolescents: A Prospective Examination of Vulnerability and Risk Factors, *Addiction*, 94:1679-1694.

HOWARD, M.O. & JENSON, J.M.,1999, Inhalant Use among Antisocial Youth: Prevalence and Correlates, *Addictive Behaviors,* 24:59-74.

HUIZINGA, D., ESBENSEN, F., & WEIHAR, A., 1991, Are There Multiple Paths to Delinquency?, *Journal of Criminal Law and Criminology*, 82:83-118.

HUIZINGA, D., & JAKOB-CHIEN, C., 1998, The Contemporaneous Co-occurrence of Serious and Violent Juvenile Offending and Other Problem Behaviors, in Loeber, R. & Farrington, D. (eds.) *Serious Violent Juvenile Offenders: Risk Factors and Successful Interventions*, pp. 47-67.

HUSSONG, A., CURRAN, P., & CHASSIN, L., 1998, Pathways of Risk for Accelerated Heavy Alcohol Use among Adolescent Children of Alcoholic Parents, *Journal of Abnormal Psychology,* 26:453-466.

KANDEL, D.B., DAVIES, M., KARUS, D., & YAMAGUCHI, K., 1986, The Consequences in Young Adulthood of Adolescent Drug Involvement: An Overview, *Archives of General Psychiatry,* 43(8):746-754.

KELLAM, S., BROWN, C., RUBIN, B., & ENSMINGER, M., 1983, Paths Leading to Teenage Psychiatric Symptoms and Substance Use: Developmental Epidemiological Studies in Woodlawn, in S. B. Guze, J. Earls & J. Barrett (eds.) *Childhood Psychopathology and Development*, pp.17-52.

LJUBOTINA, D., GALIC, J., & JUKIC, V., 2004, Prevalence and Risk Factors of Substance Use amongUrban Adolescents: Questionnaire Study, *Croatian Medical Journal*, 45:88-98.

LOSOYA, S., & CHASSIN, L., 2004, Substance Use and Serious Juvenile Offending. Paper presented as part of a symposium entitled “Sanctions and Services for Serious Juvenile Offenders: Findings from the Pathways to Desistance Study,” at the annual meeting of American Psychology-Law Society, Scottsdale, AZ.

MASON, W.A. & WINDLE, M., 2002, Reciprocal Relations between Adolescent Substance Use and Delinquency: A Longitudinal Latent Variable Analysis, *Journal of Abnormal Psychology*, 111(1): 63-76.

MAURICIO, A.M., LITTLE, M., CHASSIN, L., KNIGHT, G.P., PIQUERO, A.R., LOSOYA, S.H., & VARGAS-CHANES, D., 2009, Juvenile Offenders’ Alcohol and Marijuana Trajectories: Risk and Protective Factor Effects in the Context of Time in a Supervised Facility, *Journal of Youth & Adolescence*, 38:440-453.

MCCLELLAND, G., TEPLIN, L., & ABRAM, K., 2004, Detection and Prevalence of Substance Use amongJuvenile Detainees, *OJJDP Juvenile Justice Bulletin*.

MULVEY, E.P., SCHUBERT, C.A., & CHASSIN, L., 2010, Substance Use and Offending in Serious Adolescent Offenders. *OJJDDP Bulletin*, Washington, DC: United States Department of Justice, Office of the Juvenile Justice and Delinquency Prevention.

MULVEY, E., STEINBERG, L., FAGAN, J., CAUFFMAN, E., PIQUERO, A., CHASSIN, L., KNIGHT, G., BRAME,R., SCHUBERT, C., HECKER, T., & LOSOYA, S., 2004, Theory and Research on Desistance from Antisocial Activity among Serious Adolescent Offenders, *Youth Violence and Juvenile Justice: An Interdisciplinary Journal*, 2 (3):213-236.

MULVEY, E., STEINBERG, L., PIQUERO, A., BESANA, M., FAGAN, J., SCHUBERT, C., & CAUFFMAN, E., 2010, Trajectories of Desistance and Continuity in Antisocial Behavior Following Court Adjudication among Serious Adolescent Offenders, *Development and Psychopathology*, 22:453-475.

NSDUH REPORT, February 27, 2004, Substance Use, Abuse, and Dependence among Youths Who Have Been in Jail or a Detention Center, National Survey on Drug Use and Health Report, Office of Applied Studies, Substance Use and Mental Health Services Administration. Available at: (www.DrugAbuseStatistics .samhsa.gov).

OSGOOD, D.W., MCMORRIS, B.J., & POTENZA, M.T., 2002, Analyzing Multiple-item Measures of Crime and Deviance I: Item Response Theory Scaling, *Journal of Quantitative Criminology, 18* (3): 267-296.

PIQUERO, A.R., BLUMSTEIN, A., BRAME, R., HAAPANEN, R., MULVEY, E.P., & NAGIN, D.S., 2001, Assessing the Impact of Exposure Time and Incapacitation on Longitudinal Trajectories of Criminal Offending, *Journal of Adolescent Research*, 16:54-74.

PRINZ, R.J. & KERNS, S.E.U., 2003, Early Substance Use by Juvenile Offenders, *Child Psychiatry and Human Development*, 33 (4):263-277.

SCHABENBERGER, O., 2005, Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models, Cary, NC: SAS Institute.

SCHUBERT, C., MULVEY, E., STEINBERG, L., CAUFFMAN, E., LOSOYA, S., HECKER, T., CHASSIN, L., & KNIGHT, G., 2004, Operational Lessons from the Pathways to Desistance Project, *Youth Violence and Juvenile Justice: An Interdisciplinary Journal*, 2 (3):237-255.

STEINBERG, L., 2002, *Adolescence*, 6th ed., New York : McGraw-Hill.

SULLIVAN, C.J. & HAMILTON, Z.K., 2007, Exploring Careers in Deviance: A Joint Trajectory Analysis of Criminal Behavior and Substance Use in an Offender Population, *Deviant Behavior*, 28:497-523.

TEPLIN, L. ET AL., 2002, Psychiatric Disorders in Youth in Juvenile Detention, *Archives of General Psychiatry*, 59:1133-1143.

TEPLIN, L. ET AL., 2006, Psychiatric Disorders of Youth in Detention, Juvenile Justice Bulletin, Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice.

THORNBERRY, T. P., & KROHN, M.D., 2000, The Self-report Method for Measuring Delinquency and Crime, *Measurement and Analysis of Crime and Justice, 4:*33-83.

VON SYDOW, K., LIEB, R., PFISTER, H. ET AL., 2002, What Predicts Incident Use of Cannabis and Progression to Abuse and Dependence?: A Four (4) Year Prospective Examination of Risk Factors in a Community Sample of Adolescents and Young Adults, *Drug and Alcohol Dependence*, 68:49-64.

WINTERS, K.C., 1998, Kids and Drugs: Treatment Recognizes Link between Delinquency and Substance Abuse, Corrections Today.

WORLD HEALTH ORGANIZATION, 1990, Composite International Diagnostic Interview, Geneva, Switzerland, World Health Organization.

YOUNG, D., DEMBO, R., & HENDERSON, C., 2007, A National Survey of Substance Abuse Treatment for Juvenile Offenders, *Journal of Substance Abuse Treatment*, 32:255-266.

YOUNG, S.E., CORLEY, R.P., STALLINGS, M.C. ET AL, 2002, Substance Use, Abuse and Dependence in Adolescence: Prevalence, Symptom Profiles and Correlates, *Drug and Alcohol Dependence*, 68:309-322.

ACKNOWLEDGMENT

I would like to thank Dr. Edward P. Mulvey for his over-all guidance on the conceptualization of this paper and for allowing me to use the data set on a subsample of the participants of the Pathways to Desistance Study. The Pathways to Desistance Study was supported by funds from the following: Office of Juvenile Justice and Delinquency Prevention, National Institute of Justice, John D. and Catherine T. MacArthur Foundation, William T. Grant Foundation, Robert Wood Johnson Foundation, William Penn Foundation, Center for Disease Control, National Institute on Drug Abuse (R01DA019697), Pennsylvania Commission on Crime and Delinquency, and the Arizona Governor's Justice Commission. The content of this paper is the sole responsibility of the authors and does not necessarily represent the official views of these agencies. I would also like to thank the University of the Philippines Visayas through the UPV In-House Research for funding this project.

Nearest-Integer Response from Normally-Distributed Opinion Model for Likert Scale

Jonny B. Pornel, Vicente T. Balinas
and Giabelle A. Saldaña
University of the Philippines Visayas

This paper proposes that respondents' opinions on Likert Scale items are normally distributed around their latent ability although their observable responses will be integers in the scale nearest to those opinions. We tested the appropriateness of the model on actual data gathered by a Likert scale developed to measure attitude of teachers towards research undertaking. The soundness of common research practice of using mean and standard deviation to estimate the respondents' latent ability was tested. The results show that the NIRNDO model could be used appropriately to model responses on Likert scale. Also, the results show that using the mean response to a Likert scale, the resulting 95% confidence interval (mean $\pm$ 1.96 SEM) would be effective at least 90% of the time. This effectiveness is guaranteed for latent ability in the optimum range $[u + 0.8, v - 0.8]$ where u and v are the lowest and highest points in the scale, respectively.

Keywords: Likert Scale, NIRNDO Model, latent ability

1. Introduction

Likert scale (Likert, 1932) has become an important instrument in the fields of social sciences (Wu, 2007), education (Gay & Airasian, 2000), medicine, marketing research (Albaum, 1997) and others for measuring latent abilities such as attitude. Uebersax (2006) characterizes the Likert scale with the following main features: (1) it contains several items, (2) response levels are arranged horizontally, (3) response levels are anchored with consecutive integers, and (4) response levels are also anchored with verbal labels that are more-or-less evenly-spaced. Likert scale was intended to be a summated scale (Likert, 1932). A good Likert scale should have high internal consistency. The several items contained in a Likert scale should be designed to solicit opinions (agreement or disagreement, approval or disapproval, etc.) of the respondents on different situations related to the phenomenon being investigated. A summary measure of these opinions is intended to estimate the latent ability (e.g., attitude towards homeschooling) of the respondent.

Likert scale is a kind of polytomous test, that is, the scoring scheme is unlike the binary test where examinees have only two possible scores for each item, 1 or 0 (right or wrong). In polytomous test, the scoring scheme for each item have several score levels, such as (0, 1, 2 and 3), (1, 2, 3, 4, and 5), etc.

The popularity of Likert scale is mostly due to its ease of construction, administration, response (Albaum, 1997) and interpretation. These characteristics were important factors in the spread of its use in various fields whose practitioners cannot be rightfully assumed to be well-versed in psychometrics, and thus would gladly use tools that can be easily learned, applied and interpreted.

Other researchers are wary, however, of the common practice of using mean, standard deviation, and parametric tests on data gathered using Likert scale. Parametric statistical tools require at least interval data. The critics contend that Likert scale data are (a) of ordinal type and (b) coarse.

On the issue of the type of data from Likert scales, they contended that in Likert scale with choices such as (1) Strongly Disagree, (2) Disagree, (3) Agree, and (4) Strongly Agree, the choices can be ordered but the difference between 1 and 2 cannot be certainly claimed to be the same as the difference between 2 and 3, leading them to say that data gathered using Likert Scale are not of interval type. Thus, using parametric tools, for example, on this data are violations of sound statistical principles; they insisted that only nonparametric test should be used (Jamieson, 2004). Kuzon, Urbanchek & McCabe (1996) contended that the use of parametric analysis on ordinal data is one of seven deadly sins of statistical analysis.

On the issue of coarseness, there is a consensus that the latent variables, for which Likert Scales were devised to measure, are of an essence continuous. However, the scale forces the respondents to choose only from few choices or responses, and this results to different true scores being lumped together into the same category (Aguinis, Pierce and Culpepper, 2008). For Example, in a scale designed to measure attitude towards a topic and with possible responses of 1, 2, 3, 4, and 5, two respondents with opinions 2.7 and 3.2 (almost neutral attitude, though of varying degree and inclination) on a certain item of the scale, would most probably select response 3 as their response and will be lumped together as having the same opinion. In so doing, there is loss of information due to coarseness. This is not a trivial issue.

Russell and Bobko (1992), in their study on the effect of the response scale on the power of moderated regression, claimed that the loss of information due to Likert scale “greatly reduces the probability of detecting true interaction effects.”

Sung and Kang (2006) discussed several Item Response Theory (IRT) models for polytomous test, like Likert Scale, three of which were the rating scale model (Andrich, 1978), the partial credit model (Masters, 1982), and the generalized partial credit model (Muraki, 1992).

The rating scale model (RSM) proposes that the probability that examinee j chooses response r in a Likert scale is

$$P\left(R_{ji} = r \Big/ \theta_j, \beta_i\right) = e^{\sum_{c=u}^{r} \frac{(\theta_j - \beta_i - \tau_c)}{\sum_{y=u}^{v} e^{\sum_{c=u}^{y} (\theta_j - \beta_i - \tau_c)}}}$$

where θ_j is the latent ability of respondent j, β_i is the difficulty of item i, and τ_c is the location parameter for choice *c*. Further, *u* and *v* are the least and highest value of *c* respectively.

The partial credit model (PCM) is similar to RSM with the added parameter τ_{ci} instead of being constant for each choice across items, that is the probability that examinee j chooses response r in a Likert scale is

$$P\left(R_{ji} = r \Big/ \theta_j, \beta_i, \tau_{ci}\right) = e^{\sum_{c=u}^{r} \frac{(\theta_j - \beta_i - \tau_{ci})}{\sum_{y=u}^{v} e^{\sum_{c=u}^{y} (\theta_j - \beta_i - \tau_{ci})}}}$$

The graded partial credit model (GPCM) is an elaboration of PCM by adding the discrimination index α_i. That is, the probability that examinee j chooses response r in a Likert scale is

$$P\left(R_{ji} = r \Big/ \theta_j, \alpha_i, \beta_i, \tau_{ci}\right) = e^{\sum_{c=u}^{r} \frac{\alpha_i (\theta_j - \beta_i - \tau_{ci})}{\sum_{y=u}^{v} e^{\sum_{c=u}^{y} \alpha_i (\theta_j - \beta_i - \tau_{ci})}}}$$

However, Item Response models such as RSM, PCM and GPCM have stringent requirements on the items of the scales being considered. Using the models when the items do not satisfy those requirements could lead to some drastic outcome. These stringent requirements, as well as the mathematical rigor involved in IRT models made those models unpopular with researchers from fields not associated with good mathematical inclination.

In contrast to IRT models, the model being proposed in this paper does not take into account the parameters of each item in the scale. It assumed that the opinion ϑ of respondents on each item of a Likert Scale (that is, a Likert scale one that was designed to measure the respondent's latent variable θ such as attitude, self-concept, motivation, etc.) is normally distributed with mean and standard deviation . Using this assumption, the paper proceeded to determine whether the mean response $\bar{r}_j = \frac{\sum_{i=1}^{n} r_{ij}}{n}$ of respondent j on a Likert Scale with n items can be a good estimate of the respondent's latent ability being considered despite of the coarseness brought about by the response format inherent to the scale.

2. The Model

This paper proposes a new mathematical model for the responses to items of a Likert Scale. Such model was then used to explore the effect of coarseness on the soundness of common practice of using mean response in the data analysis.

The idea of this paper is modeled from the tendency of respondents, when confronted with several items of a Likert scale, to signify answers of varying degree from item to item. For example, a certain respondent who was asked to answer a Likert scale with 10 items, each item of which has five choices (1 – Strongly disagree, 2 – Disagree, 3 – Undecided, 4 – Agree, 5 – Strongly Agree), may have a response vector R= (3, 4, 3, 4, 2, 3, 3, 4, 3, 3). This score vector does not mean that the latent ability of the respondent varies. Instead, it signifies that the opinions of the respondent vary from item to item but these opinions hover around the latent ability. This paper assumes that, on any item of the scale, the distance of the respondent's opinion from the latent ability is normally distributed with mean θ and standard deviation σ.

Say, a respondent of a 4-point, 30-item Likert Scale on attitude towards compulsory sex education to grade school students have a slightly positive attitude towards the issue. Assuming that on the scale of 1 to 4 with 4 as the highest, his attitude is 2.75. Then, his opinion ϑ on any item of the scale, assuming it was excellently constructed, will be around 2.75 more or less. This paper proposed that $\vartheta \sim N(2.75, \sigma)$. However, since Likert Scale forces the respondent to choose only integral answers, then the respondent's answer to a certain item would be the integer nearest to his or her opinion for the item.

Formally, let *L(k,n)* denote a k-point n-item Likert scale. That is, the scale has *n* items and each item has *k* choices. It is important that these items should be unidimensional, or at least have high internal consistency, and focus on measuring opinions for a single phenomenon. There are many standard procedures to do this.

For any item, let *u* be the choice with the lowest value and *v* be the one with the highest value. It must be noted that $v = u + k - 1$. Let $I_1, I_2, \ldots, I_n$ be the items of *L(k,n)*. Also, let *L(k,n)* be a measure of a certain phenomenon (e.g. attitude, awareness, acceptability, etc.).

A certain respondent *j* of the Likert scale *L(k,n)* will have ability (i.e. level of attitude, awareness, acceptability, etc.) θ_j with regard to the phenomenon, where $u \leq \theta_j \leq v$. When confronted with item I_i, respondent *j* will then form opinion ϑ_{ji}. The model being advance here, the Nearest-Integer Response from Normally-Distributed Opinion (NIRNDO) model, contends that $\vartheta_{ji} \sim N(\theta_j, \sigma_j)$. It has to be pointed out that this θ_j may change due to some circumstances but is assumed to be stable for some length of time which makes it worth measuring. In this model, it is the opinion ϑ_{ji} that is assumed to vary with σ_j around θ_j.

Since a Likert scale records only integral response, then respondent *j* will need to convert opinion ϑ_{ji} to response R_{ji} where R_{ji} is an integer between *u* and *v* inclusive. For example, if *u* and *v* are 1 and 5 respectively, then R_{ji} can only be 1, 2, 3, 4, or 5. This model further claims that

$$R_{ji} = r \quad \text{where } |r - \vartheta_{ji}| = \min(|x - \vartheta_{ji}|) \quad (1)$$
$$\text{for all integer x from } u \text{ to } v \text{ inclusive}$$

Necessarily, $u \leq r \leq v$. For example, Consider $L(4, n)$ where $u = 1$ and $v = 4$. Then

$$\overline{R}'_{ji} = \begin{cases} 4 & \vartheta_{ji} \geq 3.5 \\ 3 & 2.5 \leq \vartheta_{ji} < 3.5 \\ 2 & 1.5 \leq \vartheta_{ji} < 2.5 \\ 1 & \vartheta_{ji} < 1.5 \end{cases} \quad (2)$$

Since the model contends that $\vartheta_{ji} \sim N(\theta_j, \sigma_j)$, it follows that to solve for probability that the observed response on item I_i is 1 given that the latent ability of respondent *j* is θ_j, the following equation will be used

$$P(R_{ji} = 1/\theta_j) = \int_{\infty}^{1.5} N(\theta_j, \sigma_j) \quad (3)$$

In general, $P(R_{ji} = r/\theta_j)$ of respondent *j* on item *i* the Likert Scale *L(k, n)* is

$$P(R_{ji} = r/\theta_j) = \int_a^b N(\theta_j, \sigma_j) \quad (4)$$

Where *a* and *b* are the lower and upper boundaries of r.

It must be pointed out that a and b are functions of r. In $L(4,n)$, the ordered pair (a,b) refers to $(-\infty, 1.5)$,$(1.5, 2.5)$, $(2.5, 3.5)$ and $(3.5, \infty)$ when $r = 1, 2, 3,$ and 4 respectively.

For notational convenience, let

$$\int_a^b N(\theta_j, \sigma_j) = \Phi_{a,b} \qquad (5)$$

Then,

$$P\,(r_{ji} = r/\theta_j) = \Phi_{a,b} \qquad (6)$$

From this, one may solve for the expected response $E(R_j)$ of respondent j,

$$E(R_j) = \sum_{r=u}^{v} r\Phi_{a,b} \qquad (7)$$

These results will be used in determining how good is the mean response on the items of Likert scale as an estimate of the latent variable being measured by the scale. The study will use simulated data using the model as well as actual data on teachers' attitude towards research undertakings (Pornel et al., 2010) measured using a Likert scale.

This paper aimed to introduce NIRNDO model and use it to explore the soundness of the popular practice of solving for the mean response for a Likert scale to determine the respondents' latent ability. Specifically, this study aimed to do the following:

1. Given actual response vector $R_j = (r_{j1}, r_{j2},..., r_{ji}, r_{jn})$ determine whether $\bar{r}_{ji}$ significantly differs from $\bar{r}'_{ji}$ where $\bar{r}'_{ji}$ is the response vector generated using $N(\hat{\theta}_j, \hat{\sigma}_j)$ and $\hat{\theta}_j = \bar{r}_{ji}$ and $\hat{\sigma}_j = S_{rji}$.
2. Given a respondent with ability level θ_j with σ_j, determine whether $\bar{r}'_{ji} \pm 1.96$SEM is a good estimator of θ_j, where $\bar{r}'_{ji}$ is the average of the generated responses r'_{ji}.
3. Determine the RMSE of $E(R)$ across different values of σ.

Simply speaking, the first objective is to verify whether the model could generate a response vector closely related to an observed response vector using the mean and standard deviation of the observed responses. This is equivalent to a test of normality of $\bar{r}_{ji}$. In some way, it will explore whether there is basis to the practice of using parametric test on data gathered using Likert scale. The second objective is to establish whether, in the context of the model, given a theoretical latent ability θ_j with σ_j, the mean response in a Likert scale is a good estimate of θ_j. The third objective is to explore how the error caused by the coarseness of the Likert scale vary with the σ_j.

3. The Simulations

In this paper, simulations one, two and three were conducted to achieve objectives one, two and three respectively. The results of simulation 1 determined whether NIRNDO model works well with actual data. On the other hand, simulation 2 explored whether the mean response in a Likert scale is useful in estimating the latent ability of the respondent. Lastly, the third simulation studied the parameter estimation error associated with Likert scale as affected by the variance of the respondent's opinions. The algorithms for these simulations were as follows:

<u>Simulation 1</u>

1. Given actual response vectors $R_j = (r_{j1}, r_{j2},..., r_{jn})$ of respondents in Pornel et al. (2010), solve for $\hat{\theta}_j = \bar{r}_{ji}$ and $\hat{\sigma}_j = S_{rji}$.
2. Generate the opinion vector by generating ϑ'_{ji} from $N(\hat{\theta}_j, \hat{\sigma}_j)$ for $n = 30$ times where n is the number of items of the Likert scale.
3. Determine the generated response vector $R'_j = (r'_{j1}, r'_{j2},..., r_{jn})$ using equation 1.
4. Do steps 1 to 3 for $m = 95$ times where m is the number of respondents.
5. Determine whether $\bar{r}'_{ji}$ significantly differs from $\bar{r}_{ji}$
6. Do steps 1 to 5 for $t = 100$, where t is the number of trials.

<u>Simulation 2</u>

1. Let $k = 4$
   a. Let $n = 30$
   b. Let $u = 1$ and $v = k$
   c. Let $\theta_j = u$
   d. Let $\sigma_j = 0.1$
   e. Generate the opinion vector by generating ϑ_{ji} from N (θ_j, σ_j) for n times.
   f. Determine the response vector $R'_j = (r'_{j1}, r'_{j2},..., r_{jn})$.
   g. Determine $\bar{r}'_{ji}$ and $SEM = {Sr'_{ji}}/{\sqrt{n}}$
   h. Determine the interval $[x,y]$ where $x = \bar{r}'_{ji} - 1.96SEM$ and $y = \bar{r}'_{ji} + 1.96SEM$
   i. Determine whether $\theta_j \in [x,y]$.

j. Do steps e to i for $m=10,000$ times.
k. Determine the proportion β_k wherein $\theta_j \in [x,y]$
l. Do steps d to k using σ_j with increasing value (increment is 0.1) until $\sigma_j = 1.0$
m. Do steps c to l using θ_j with increasing value (increment is 0.1) until v.

2. Do step 1 for $k = 5$ and 6.
3. Do steps 1 and 2 for $n = 25$ and 20.

Simulation 3

1. Let $k = 4$
2. Let $u = 1$ and $v = k$
3. Let $\sigma = .01$
   a. Determine $\theta - \hat{\theta}$ using equation 7 for $\theta = u+.1, u+.2, ..., v-.1$. Where $\hat{\theta} = E(R)$
   b. Solve for $[\theta - \hat{\theta}]^2$ for $\theta = u+.1, u+.2, ..., v-.1$
   c. Solve for the root mean square of error $\sqrt{\dfrac{\sum(\theta-\hat{\theta})^2}{(k-u)/.1}}$ for $\theta = u+.1, u+.2, ..., v-.1$
4. Do step 3 for $\sigma = 0.01, 0.02, ..., 1.50$
5. Do steps 1 to 4 for $k = 5$ and 6.
6. Determine Optimum range of σ that result to minimum RMSE.

4. Results and Discussion

First Simulation: Deviation of Generated Mean Response Using NIRNDO Model from the Observed Mean Response to a Likert Scale

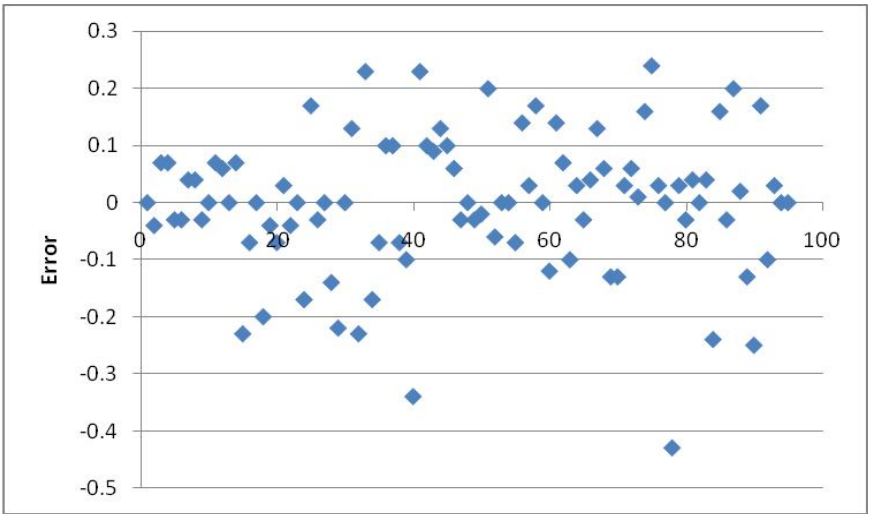
In the first simulation, the answers of 95 respondents to the Likert scale used by Pornel et al. (2010) to measure the teachers' attitude towards research undertakings were considered. The scale has originally 36 items. However, after item analysis and revision, it was reduced to 30 items. The revised instrument has a Cronbach Alpha reliability of 0.910. The least observed mean response of the respondents is $\bar{r}_{ji} = 2.17$, while the highest observed value is $\bar{r}_{ji} = 3.97$. For the standard deviation of the responses s_{rji}, the least observed value is 0.18, while the highest observed value is 1.49 as shown in Table 1.

Table 1 **Summary Measures of Responses to the Likert Scale used in Pornel et al. (2010)**

	Teachers' Response to Pornel et al. (2010)	
	Mean Response	SD
Maximum	3.97	1.49
Minimum	2.17	0.18
Mean	2.91	0.62

Using the mean, $\hat{\theta}_j = \bar{r}_{ji}$, and standard deviation of the responses, $\hat{\sigma}_j = s_{rji}$, the researchers generated opinions ϑ'_{ji} from $N(\hat{\theta}_j, \hat{\sigma}_j)$ for each respondent as predicted by NIRNDO model, then simulated the most probable responses, r'_{ji}, based on equation 1. Computing for the absolute error, $|\bar{r}'_{ji} - \bar{r}_{ji}|$, the researchers found that the maximum absolute error in one run of the simulation to be 0.43, the minimum is 0.00 and the average of the absolute errors is 0.09. When the errors were plotted they were found to hover randomly around zero as shown in Figure 1.

Figure 1 **Scatterplot Showing Errors of Estimate for each Respondent in One Trial**



Using the Wilcoxon Signed Rank test to determine the significance of the difference between the mean simulated response $\bar{r}'_{ji}$ and the mean observed response $\bar{r}_{ji}$ showed that there was no significant difference between the observed and projected mean 92% of the time. This test was performed 100 times.

Second Simulation: Efficiency of Mean Response on a Likert Scale to Estimate the Respondent's Latent Ability

In the second simulation, the researchers generated the responses of respondents given their latent ability θ_j (with the corresponding standard deviation σ_j) and then determined the average response $\bar{r}_{ji}$ and the standard error of the mean $SEM = s_{rji} / \sqrt{n}$. The researchers, then, determined how often the interval $\bar{r}_{ji} \pm 1.96SEM$ contains.

The result of the simulation shows that in $L(4,30)$**,** when σ is equal or less than 0.3, the interval $\bar{r}_{ji} \pm 1.96SEM$ is inconsistent in containing θ_j across different values of σ. That is, for some value of θ_j, the interval $\bar{r}_{ji} \pm 1.96SEM$ contains θ_j at least 90% of the time while in some values of $\theta_{j,}$ the interval contains θ_j, less than 90% of the time. However, when σ is greater than 0.3, the interval $\bar{r}_{ji} \pm 1.96SEM$ contained θ_j more than 90% of the time when the value of the ability θ_j is from 1.8 to 3.2 as shown in Table 2. The same trend was observed for $L(5,30)$ and $L(6,30)$ as indicated in Tables 3 and 4. That is, $\bar{r}_{ji} \pm 1.96SEM$ contains θ_j at least 90% of the time when θ_j is at least a distance of 0.8 from the edges and σ is greater than 0.3.

Similar patterns were found when n is either 25 or 20. The results for these simulations are found in Tables 7 to 12 of the Appendix.

Since, the interval $\bar{x} \pm 1.96SEM$ is supposed to be a 95% confidence interval for normally distributed variable, these results showed that the coarseness due to Likert scale lessen the effectiveness of the confidence interval. When θ_j is in the optimum range (0.8 from the edges of the scale), the confidence interval $\bar{r}_{ji} \pm 1.96SEM$ can be effective at least 90% of the time.

Checking the distribution of data in Pornel et al. (2010), it was found that 80% of the respondents to the Likert scale used in the study had mean responses between 1.8 and 3.2 inclusive as shown in Table 5. Also, 90% of the respondents had σ greater than 0.3 but not more than 1.0.

Table 2 Effectiveness of $\bar{r}_{ji} \pm 1.96SEM$ in containing θ_j in $L(4,30)$

	σ									
θ	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
1.0	100.00	100.00	94.54	60.23	27.94	11.76	5.17	2.40	1.28	0.69
1.1	0.00	14.89	76.45	87.23	69.76	49.16	35.37	24.99	18.09	13.36
1.2	0.00	32.72	87.42	**95.72**	**90.02**	80.60	68.17	54.82	44.25	36.12
1.3	0.06	52.06	90.49	95.57	93.36	87.73	80.10	71.18	63.02	54.59
1.4	9.35	74.98	91.08	93.97	93.70	**91.50**	87.51	82.58	76.46	69.73
1.5	95.80	95.94	96.01	95.75	94.27	92.19	**90.31**	87.29	83.55	79.33
1.6	8.89	75.09	90.38	93.76	93.85	93.51	91.55	**90.62**	88.07	84.51
1.7	0.03	52.64	89.45	93.50	94.02	93.29	93.51	91.78	**90.63**	87.81
1.8	0.00	32.46	83.41	93.15	94.66	94.23	93.53	93.01	91.85	**90.48**
1.9	0.00	14.93	77.06	92.29	93.28	93.70	94.53	93.56	93.12	92.14
2.0	100.00	99.99	97.05	93.01	92.79	92.84	93.25	93.12	92.87	92.98
2.1	0.00	14.65	76.93	92.13	93.32	93.62	94.28	94.03	94.01	93.53
2.2	0.00	31.89	83.45	92.91	94.17	93.89	93.52	93.56	93.78	93.97
2.3	0.08	53.11	89.13	93.79	93.75	93.54	93.49	93.56	94.07	94.02
2.4	9.09	75.20	90.39	93.59	94.07	93.79	94.09	93.90	94.49	93.87
2.5	95.94	96.24	95.79	94.80	94.30	93.67	93.52	94.04	93.85	94.32
2.6	8.84	75.57	90.51	93.77	94.67	93.88	94.55	93.88	94.06	93.72
2.7	0.04	52.50	89.12	93.55	94.13	93.41	93.90	93.94	93.83	93.95
2.8	0.00	32.22	83.02	92.92	94.22	94.26	94.13	93.55	93.80	93.43
2.9	0.00	14.59	77.04	92.73	93.24	93.77	93.80	94.19	94.32	93.77
3.0	100.00	100.00	100.00	93.44	92.49	93.26	93.62	93.41	93.25	92.67
3.1	0.00	14.15	77.29	92.17	93.38	93.61	94.17	93.62	93.52	92.75
3.2	0.00	32.58	82.57	92.91	94.21	94.70	94.06	93.80	**92.33**	**91.23**
3.3	0.09	52.53	88.56	93.59	93.45	93.56	92.92	92.22	89.97	88.21
3.4	8.99	74.92	90.91	93.83	93.94	93.04	91.83	**90.41**	88.08	85.34
3.5	95.65	95.90	95.73	95.12	94.02	92.34	**90.25**	87.54	83.88	79.27
3.6	9.30	75.03	90.71	93.34	93.54	**91.61**	87.64	82.74	77.02	69.83
3.7	0.06	53.26	90.24	95.39	**92.98**	87.53	81.26	72.03	62.64	54.63
3.8	0.00	32.42	87.03	**95.26**	89.96	79.31	68.09	56.20	44.70	35.94
3.9	0.00	15.25	76.42	87.10	69.71	50.09	34.95	25.66	18.46	13.77
4.0	100.00	100.00	94.50	61.14	27.77	12.02	5.08	2.49	1.63	0.75

A low σ may signal that the respondent exhibited the blind effect, a response bias wherein the respondent would choose the answers without reading each item. This happens when a disinterested respondent would decide to have a fix answer, say 3, in $L(4,30)$ scale, and decide to answer all or most of the items with the predetermined choice without reading any of them. With response vectors from this kind of respondents, the model will fail. However, it is also possible that the respondent has a very low *s*, especially when his latent ability θ is so near an integral choice in the scale (say θ = 2.9, this is so near the integral choice 3 of a 4-point Likert scale). In this situation, the respondent will signify answers that are mostly, if not all, 3. With this kind of respondent, the mean response would be a good estimate of the respondent's latent ability.

Table 3 Effectiveness of $\bar{r}_{ji} \pm 1.96SEM$ in containing θ_j in $L(5,30)$

	σ									
θ	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
1.0	100.00	100.00	95.19	61.05	27.35	11.24	5.23	2.71	1.27	0.68
1.1	0.00	14.62	76.56	87.16	69.41	50.22	35.63	25.42	18.30	14.46
1.2	0.00	32.58	86.91	**95.42**	**90.01**	80.74	68.52	55.16	44.73	35.38
1.3	0.03	52.69	90.27	95.63	93.51	88.24	80.37	72.08	62.42	55.18
1.4	8.78	73.86	90.96	93.95	93.94	**91.68**	87.94	82.45	75.84	69.52
1.5	96.54	95.32	95.44	95.82	94.52	92.83	**90.58**	87.08	84.00	79.10
1.6	9.23	76.03	90.51	93.65	93.81	93.12	91.95	**90.64**	87.87	84.94
1.7	0.08	53.07	89.26	93.81	93.76	93.74	93.18	91.97	**90.93**	88.12
1.8	0.00	31.83	83.94	93.06	94.33	94.35	93.50	92.99	92.18	**90.15**
1.9	0.00	14.88	77.07	92.24	94.06	93.47	94.22	94.01	93.34	91.61
2.0	100.00	100.00	96.84	92.62	93.20	93.37	93.45	93.25	93.05	92.41
2.1	0.00	14.61	77.31	92.33	93.70	94.17	93.94	94.76	93.93	92.98
2.2	0.00	31.98	83.22	93.13	94.34	94.08	93.52	94.00	94.21	93.87
2.3	0.07	52.67	89.44	93.32	94.19	93.74	93.70	93.60	93.58	94.01
2.4	8.90	75.17	90.46	93.63	94.03	93.84	94.13	93.86	94.42	93.80
2.5	95.85	95.85	95.54	95.31	95.75	94.14	93.53	94.17	93.85	94.17
2.6	9.52	74.88	90.93	93.64	94.12	93.74	93.92	94.14	94.17	94.19
2.7	0.07	53.10	88.44	93.93	93.88	93.49	94.31	93.95	94.08	93.76
2.8	0.00	33.14	83.01	93.10	94.71	93.91	94.03	93.76	94.36	93.58
2.9	0.00	14.82	77.37	92.49	94.05	94.20	94.15	94.52	93.97	94.06
3.0	100.00	100.00	96.40	92.83	93.03	93.64	93.81	93.44	93.50	94.26
3.1	0.00	15.34	76.63	92.76	93.33	94.16	94.35	94.55	93.84	94.46
3.2	0.00	32.24	83.05	92.81	94.14	94.57	93.68	94.09	93.59	93.92
3.3	0.05	52.09	89.22	93.74	93.90	94.12	94.15	93.72	94.21	93.72
3.4	8.80	74.57	91.00	93.98	94.30	93.68	97.75	93.87	93.89	94.18
3.5	95.94	95.95	96.22	94.80	93.88	93.67	93.84	93.84	93.93	93.81
3.6	9.31	74.70	90.59	93.50	94.02	93.92	94.09	93.63	93.71	94.08
3.7	0.05	52.38	89.17	93.65	94.03	94.08	93.89	93.88	93.80	93.38
3.8	0.00	31.65	83.15	93.02	93.96	94.26	93.71	93.75	93.78	93.70
3.9	0.01	14.49	77.26	92.29	93.50	93.99	94.25	94.13	94.03	93.51
4.0	100.00	99.99	97.09	93.09	92.88	93.25	93.54	93.12	93.03	92.72
4.1	0.00	15.14	77.54	92.32	93.53	93.67	93.35	93.83	92.97	92.40
4.2	0.00	32.07	82.94	93.20	94.61	94.20	94.26	92.75	**91.62**	**90.64**
4.3	0.04	52.36	89.65	93.74	93.88	93.60	93.53	91.92	89.91	87.12
4.4	8.83	75.12	90.88	93.31	93.44	92.95	91.56	**90.10**	88.04	85.22
4.5	95.55	95.93	95.56	95.12	93.78	92.29	**90.01**	87.17	84.02	79.18
4.6	9.12	75.29	90.43	93.46	93.72	**91.09**	88.00	82.58	75.93	70.28
4.7	0.11	52.43	89.94	95.26	**93.08**	87.79	79.77	71.85	63.04	55.39
4.8	0.00	32.71	87.84	**95.33**	90.40	80.56	67.27	56.05	45.54	36.71
4.9	0.00	14.37	76.37	86.81	69.78	49.36	36.04	24.88	18.20	13.93
5.0	100.00	100.00	94.70	61.47	28.08	11.65	5.47	2.75	1.40	0.81

Table 4 Effectiveness of $\bar{r}_{ji} \pm 1.96SEM$ in containing θ_j in $L(6,30)$

θ	σ 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1.0	100.00	100.00	94.54	60.23	27.94	11.76	5.17	2.40	1.31	0.71
1.1	0.00	14.89	76.45	87.23	69.76	49.16	35.37	24.99	18.09	13.39
1.2	0.00	32.72	87.42	**95.72**	**90.02**	80.60	68.17	54.83	44.27	36.19
1.3	0.06	52.06	90.49	95.57	93.36	87.73	80.10	71.18	63.02	54.60
1.4	90.35	74.98	91.08	93.97	93.70	**91.50**	87.51	82.58	76.49	69.80
1.5	95.8	95.94	96.01	95.75	94.27	92.19	**90.31**	87.29	83.55	79.34
1.6	8.85	75.09	90.38	93.76	93.85	93.51	91.55	**90.62**	88.06	84.47
1.7	0.03	52.69	89.45	93.50	94.02	93.29	93.51	91.78	**90.63**	87.81
1.8	0.00	32.46	83.41	93.15	94.66	94.23	93.53	93.01	91.84	**90.41**
1.9	0.00	14.93	77.06	92.29	93.28	93.70	94.53	93.56	93.09	92.07
2.0	100.00	99.99	97.05	93.01	92.79	92.84	93.25	93.13	92.88	92.88
2.1	0.00	14.65	76.93	92.13	93.32	93.62	94.28	94.00	93.96	93.40
2.2	0.00	31.89	83.45	92.91	94.17	93.89	93.52	93.56	93.73	93.84
2.3	0.08	53.11	89.13	93.79	93.75	93.54	93.49	93.57	93.93	94.16
2.4	9.09	75.20	90.39	93.59	94.07	93.80	94.07	93.95	94.41	93.80
2.5	95.94	96.24	95.79	94.80	94.30	93.67	93.60	93.97	93.69	94.27
2.6	8.84	75.57	90.51	93.77	94.67	93.91	94.59	93.92	94.13	93.82
2.7	0.04	52.50	89.12	93.55	94.13	93.41	93.94	93.93	93.80	94.00
2.8	0.00	32.22	83.02	92.92	94.21	94.38	94.18	93.70	94.18	93.95
2.9	0.00	14.59	77.04	94.73	93.25	93.85	94.00	94.22	94.85	94.24
3.0	100.00	100.00	96.71	93.44	92.52	93.31	93.60	93.65	93.89	93.90
3.1	0.00	14.15	77.29	92.17	93.47	93.76	94.43	94.07	94.27	94.34
3.2	0.00	31.48	83.13	92.96	94.33	94.58	93.73	93.83	94.03	94.08
3.3	0.04	52.57	89.20	93.70	93.80	94.09	93.70	94.08	93.80	94.10
3.4	9.07	74.89	91.15	94.12	93.84	94.14	94.10	94.09	94.03	94.41
3.5	95.8	95.78	95.22	95.13	94.16	93.74	93.84	93.60	94.26	94.10
3.6	8.85	75.34	90.88	93.75	93.87	94.03	93.90	94.02	94.02	94.29
3.7	0.07	53.91	89.16	93.63	93.87	93.8	93.64	94.05	93.77	93.66
3.8	0.00	32.31	83.87	93.03	94.40	94.28	93.86	94.11	94.36	94.14
3.9	0.00	15.31	77.50	92.59	93.38	93.96	94.85	94.42	94.16	94.19
4.0	100.00	100.00	96.77	93.26	92.97	93.32	93.48	93.60	93.45	94.12
4.1	0.00	15.69	76.65	92.06	93.57	94.04	94.05	93.90	94.45	94.21
4.2	0.00	31.88	83.28	93.18	94.13	94.22	93.60	93.81	94.08	94.33
4.3	0.04	53.73	89.06	93.90	94.04	93.89	93.99	93.47	94.40	93.75
4.4	8.65	74.68	90.64	93.75	93.82	93.93	93.74	93.72	94.28	94.17
4.5	95.73	95.93	95.74	94.93	93.53	93.38	93.67	93.76	94.24	94.18
4.6	8.80	74.57	90.70	93.75	94.41	93.85	94.43	93.58	93.90	94.34
4.7	0.04	52.58	88.2	93.23	93.48	94.33	93.73	93.62	94.17	93.68
4.8	0.00	32.72	83.76	93.06	94.86	94.32	94.14	93.96	94.24	94.01
4.9	0.00	15.26	76.88	92.75	93.53	93.52	94.45	93.93	93.95	93.76
5.0	100.00	100.00	97.1	92.94	92.71	94.02	93.29	93.37	92.86	92.49
5.1	0.00	15.46	76.42	92.90	93.1	93.59	94.40	93.98	93.30	92.34
5.2	0.00	32.31	83.02	93.32	94.47	94.04	94.12	92.56	92.33	**90.51**
5.3	0.05	53.34	89.42	93.99	93.92	93.68	93.06	92.19	**90.13**	87.84
5.4	8.90	75.50	90.37	93.34	93.89	92.75	92.19	**90.49**	87.62	85.84
5.5	95.98	96.11	95.48	95.37	94.36	92.56	**90.15**	86.90	83.61	79.47
5.6	8.56	74.88	91.09	93.63	93.80	**91.64**	88.38	82.47	76.39	69.97
5.7	0.06	52.83	90.52	95.44	93.57	88.04	80.09	72.62	63.45	55.36
5.8	0.00	32.83	87.34	**95.15**	**90.03**	79.84	68.04	55.61	45.40	35.92
5.9	0.00	14.77	77.26	86.98	69.98	50.60	36.62	29.84	18.68	13.41
6.0	100.00	100.00	95.00	60.01	27.87	11.87	4.97	2.32	1.43	0.86

Table 5 Distribution of Respondents Observed in the Study of Pornel et al. (2010)

Statistics	Range of values	% of respondents
θ	Below 1.8	0
	Between 1.8 & 3.2 inclusive	80
	Greater than 3.2	20
σ	Less than or equal to 0.3	5
	Greater than 0.3 but not more than 1.0	90
	Greater than 1.0	5

Third Simulation: Parameter Estimation Error Associated with Likert Scale as Affected by the Variance of the Respondent's Opinions

In this simulation, the researchers determine the difference between the theoretical theta θ and the estimated theta, $\hat{\theta}$, where $\hat{\theta}$ is taken to be equal to *E (R)*, given the value of standard deviation *s*. Results show that the root mean square error (RMSE) is high for s = 0.01but it decreases as s approaches 0.04. When the RMSEs were computed for different ranges of θ, it was found that in $L(4, n)$ the three ranges of θ differ in their rates of change as *s* increases as shown in Figure 2.

The RMSE of $L(4, n)$ for θ ranging from 1 to 4 is minimum at $\sigma = 0.35$, and the RMSE increases as σ increases. On the other hand, the RMSE of $L(4, n)$ for θ ranging from 1.5 to 3.5 is minimum at $\sigma = 0.45$. Lastly, the RMSE of $L(4, n)$ for θ ranging from 2 to 3 is minimum at $\sigma = 0.52$.

In $L(5, n)$, the RMSE is minimum for θ ranging from 1 to 5 at $\sigma = 0.38$, for θ ranging from 1.5 to 4.5 at $\sigma = 0.45$ and 0.46 and for θ ranging from 2 to 4 at $\sigma =$ 0.54 (see Fig. 3).

Lastly, when $L(6, n)$, was considered, the RMSE is minimum for θ ranging from 1 to 6 at $\sigma = 0.36$ and 0.37, for θ ranging from 1.5 to 5.5 at $\sigma = 0.46$ and for θ ranging from 2 to 4 at $\sigma = 0.53$ as shown Figure 4.

One may observe, that the optimum value of σ for the range *(u, v)* of θ is almost the same for $L(4, n)$, $L(5, n)$,and $L(6, n)$. The same is true for ranges $(u + .5, v - .5)$ and $(u + 1, v - 1)$**.** The relationship between σ and RMSE for $L\ (4, n)$ as presented in Figure 2 showed that the model is more accurate when the θ involved is near the middle of the spectrum than at the edges. As Figure 2 shows most of the errors are due to θ at the edges. The same trend was found for $L(5, n)$ and $L(6, n)$ as shown in Table 6 and Figures 3 and 4. Examining these results show that there is minimal difference in the RMSE among the three scales despite of the differences in the number of choices (4, 5 and 6). Thus, one may say that the error is more a function of the position of θ and the magnitude of σ and not of the number of choices or points of the Likert scale. Since *E(R)* do not vary with *n*, this simulation was not tested with different value of *n*.

Figure 2. Root Mean Square Error (RMSE) for $L(4, n)$ across Different Standard Deviation σ

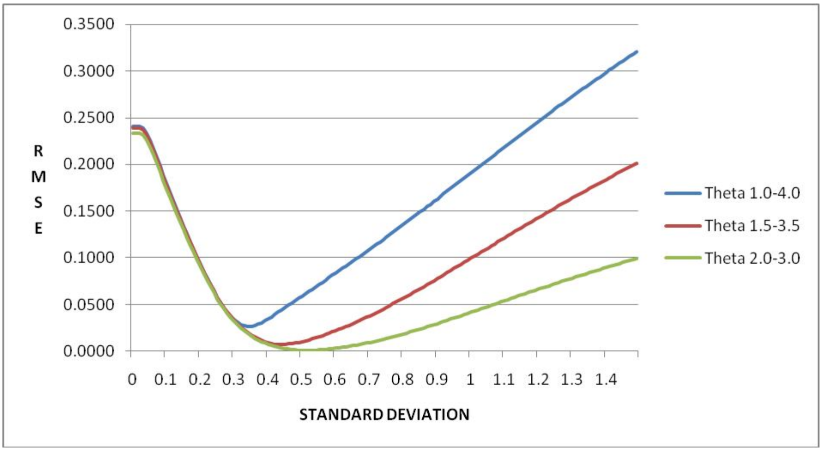


Figure 3. Root Mean Square Error (RMSE) for L (5, n) across Different Standard Deviation σ

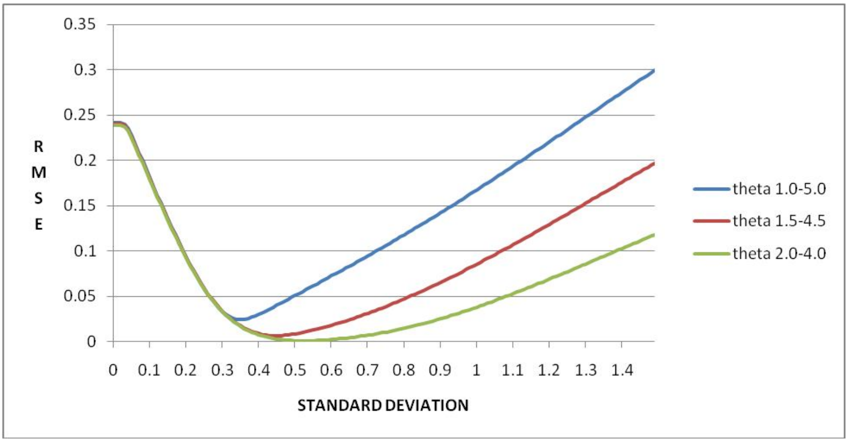


Figure 4. Root Mean Square Error (RMSE) for L (5, n) across Different Standard Deviation σ

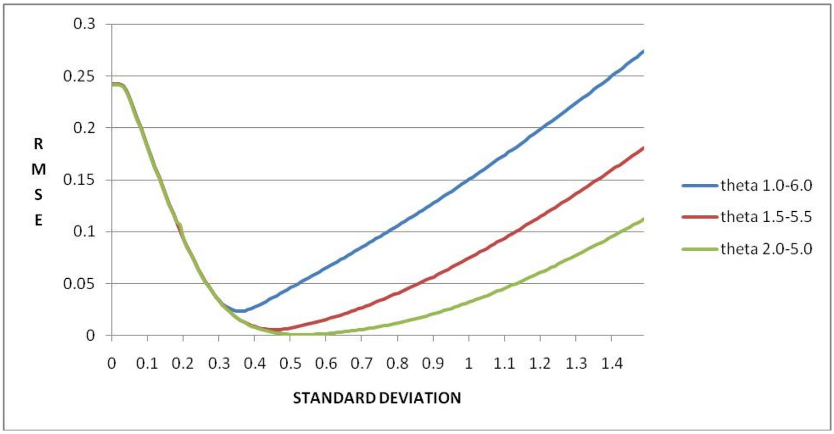


Table 6 Optimum Value of σ

Range of θ	Value of σ where RMSE is minimum		
	L (4, n)	L (5, n)	L (6, n)
(u, v)	0.35	0.38	0.36, 0.37
(u + .5, v - .5)	0.45	0.45, 0.46	0.46
(u + 1, v - 1)	0.52	0.54	0.53

5. Summary and Conclusions

The result of the first simulation showed that the model can generate a response vector with mean response not significantly different from the original response vector. Thus, for a Likert scale application that estimates the latent ability of the respondent using mean response, the NIRNDO is one good model to be considered.

The result of the second simulation showed that under the NIRNDO model, using the mean response to a Likert scale would work well in estimating the respondent's latent ability when the latent ability θ is at least 0.8 from the edges of the scale (that is, $u + 0.8 \leq \theta \leq v - 0.8$). Specifically, the interval $\bar{r}_{ji} \pm 1.96SEM$ contains θ at least 90% of the time when $u + 0.8 \leq \theta \leq v - 0.8$. This result, however, is consistent when $\sigma > 0.3$. It must be pointed out, however, that under the model, assuming $\bar{r}_{ji}$ is a normally distributed and continuous variable, the 95% confidence interval $\bar{r}_{ji} \pm 1.96SEM$ can be effective at least 90% of the time.

The result of third simulation supported the result of the second simulation. It shows that the optimum range of σ where the model has the minimum error is from 0.35 to 0.54 and that the model would not work well with θ near the edges.

The NIRNDO model could be a good model for Likert Scale, and that using mean response to estimate the latent variable being measured by the Likert Scale is feasible, although one need to be on lookout when faced with respondents who had extreme mean response, that is, mean response with 0.8 or less distance from one edge of the scale, or with very low or very high response variance. With this type of respondents, the model does not work well. Also, it is shown that Likert scale has less accuracy than a continuous scale. On the region where the model works well, an interval that would guarantee 95% accuracy for a continuous scale would guarantee 90% accuracy for Likert scale.

REFERENCES

AGUINIS, H., PIERCE, C., & CULPEPPER, S., 2008, Scale Coarseness as Methodological Artifact: Correcting Correlation Coefficients Attenuated from Using Coarse Scale. *Org. Research Methods,* 12(4):623-652.

ALBAUM, G., 1997, The Likert Scale Revisited: An Alternate Version (Product Preference Testing), *Journal of the Market Research Society,* 39(2):331-342.

ANDRICH, D., 1978, Application of a Psychometric Model to Ordered Categories which are Scored with Successive Integers, *Applied Psychological Measurement*, 2:581-594.

GAY, L. R., & AIRASIAN, P. W., 2000, *Educational Research* (6th ed.). NJ: Prentice-Hall.

JAMIESON, S., 2004, Likert Scales: How to (ab)use Them, *Medical Education, 38:*1212-1218.

KUZON, W.J., URBANCHEK,M.G., MCCABE,S., 1996, The Seven Deadly Sins of Statistical Analysis, *Annals of Plastic Surgery,* 37(3): 265-272.

LIKERT, R., 1932, A Technique for the Measurement of Attitudes, *Archives of Psychology,* 140:44-53.

MASTERS, G. N., 1982, A Rasch Model for Partial Credit Scoring. *Psychometrika,* 47:149-174.

MURAKI, E., 1992, A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, 16: 159-176.

PORNEL, J., BAYLOCON, E., MAQUIRAN, D., CAUSING, F., & TAN, M. J., 2010, A Study on Motivations, Deterrents, Needs, and Attitudes towards Research Undertakings of Public Secondary School Teachers, *Danyag: Journal of Humanities and Social Sciences,* 15(1): 36-44.

RUSSELL & BOBKO, 1992, Moderated Regression Analyses and Likert Scales: Too Coarse for Comfort, *Journal of Applied Psychology* 77(3):336-342.

SUNG, H.J. & KANG, T., 2006, Choosing a Polytomous IRT Model using Bayesian Model Selection Methods. A paper presented at the National Council on Measurement in Education annual meeting in San Francisco, April, 2006.

UEBERSAX, J., 2006, Likert Scales: Dispelling the Confusion. Statistical Methods for Rater Agreement website. [Accessed March 14, 2008]. Available from http://ourworld. compuserve.com/homepages/jsuebersax/likert2.htm.

WU, C.H., 2007, An Empirical Study on the Transformation of Likert-Scale Data to Numerical Scores, *Applied Mathematical Sciences*, 1(58): 2851-2862.

Copula-Based Vector Autogressive Models for Bivariate Cointegrated Data

Hideaki Taima and Ana Maria L. Tabunda, Ph.D.
University of the Philippines Diliman

The copula method is well applied in finance and actuarial science but its application in economic studies is limited and its use in the cointegration framework virtually nil. This paper explores the use of copula method to analyze the remaining dependence after a cointegration relationship is modeled. Specifically, simulated data is used to characterize the behavior of the dependence parameter estimates of several copulas fitted to the distribution of the residuals after cointegrated Vector Autoregressive (VAR) and Vector Error-Correction Mechanism (VECM) models are fitted, as well as evaluate the forecasting ability of the copula-based models. The Clayton, Frank, Gaussian, Gumbel and Plackett copulas are used and are compared on the basis of bias, root mean square error (RMSE) and maximum likelihood. The density forecasting ability of the copula-based VAR and VECM is then compared with that of standard models via conditional Kullback-Leibler Information Criterion (KLIC) divergence measure using simulated and empirical data. The simulation results indicate that the copula-based models generally have better density forecasting ability than standard VAR and VECM models, a finding that is supported in the application of a copula-based VAR to empirical data.

Keywords: Copula, Cointegration, VAR, VECM

1. Introduction

It is well known that economic variables have certain relationships among themselves, for example comovements. Cointegration is one of the breakthrough approaches to characterizing several series through the imposition of long-run co-movement restriction. Substantial work has since been introduced to model multivariate time series following the approach of cointegration. However, satisfying the assumption of joint normality of the series is usually difficult in applications. In this paper we explore the use of copulas to model the dependence between series that are not necessarily normal.

The copula is a function that links multivariate distribution functions to their one-dimensional marginal distribution functions (Nelson, 2006). Specifically, the joint distribution function of a random vector can be represented via Sklar's theorem in terms of the copula function with the marginal distribution functions of the components of the vector as the arguments of the copula function. The application of the copula helps separate the parameters of the marginal distributions from their intrinsic association as captured by the dependence parameters. An attractive feature of this approach is that the copula and the association parameter are invariant under continuous and monotonically increasing transformations of the marginal variables. Hence copulas have an advantage when the interest centers on the intrinsic association among the marginals (Joe, 1997; Kim et al., 2007).

The copula method has been used to analyze multivariate data with flexible functional forms rather extensively in finance and in actuarial science. For example, the use of the copula in conjunction with the generalized autoregressive conditional heteroskedastic model (GARCH), or the copula-based GARCH, facilitates the analysis of several series of returns that are non-normal (Jondeau and Rockinger, 2006) and allows for flexibility in modeling the conditional dependence structure between the Deutsche mark and the Japanese yen relative to the US dollar (Patton, 2005). Various analyses in finance show better results from the use of copulas in multivariate GARCH.

The application of the copula method to economic studies, however, is limited and its use in the cointegration framework is virtually nil. The copula-based studies with applications in econometrics include Granger et al. (2006), which looks into the multivariate GARCH model, and Mitchell (2007), which models dependence between the survey of professional forecasters (SPF)'s inflation and output growth density forecasts of the US economy. Bianchi et al. (2009), on the other hand, uses the copula-based vector autoregression (VAR) approach to forecast industrial production series in the core European Monetary Union (EMU) countries and provide evidence that the copula-VAR model outperforms or, at worst, is at par with standard VAR models.

This study explores the use of copula-based vector autoregressive models under a cointegration framework to yield insights on the effect of any remaining dependence on forecast models for bivariate cointegrated variables. Specifically, we use simulated data to evaluate the performance of copula-based cointegrated VAR and vector error correction mechanism model (VECM) against standard models in the presence of dependence. Following Vuong (1989) and Bao et al. (2007), we employ the Kullback-

Leibler information criterion (KLIC) divergence measure between two conditional densities, described in Section 3.3, to evaluate the density forecasting ability of the copula-based models.

One limitation of this study is that it considers only two symmetric distributions for the error terms used in the data generating model in the simulations – the bivariate normal distribution and the bivariate student's t distribution. Thus the conclusions drawn in the paper will be applicable only for such situations.

This paper is organized as follows: Section 2 introduces the framework of the model and reviews the definition of copula. Section 3 presents simulation results under different scenarios depending on the assumptions on the form of the error distribution, the strength of the dependence between the component error terms, and the model fitting method. In particular, the simulation section compares five forms of copula that were used in estimating the distribution of the residuals obtained from fitting bivariate VAR and VECM models under three dependence scenarios for the model error terms. The copulas are compared on the basis of accuracy, precision as well as fit to the distribution of the data, while the copula-based VECM and VAR models are compared with standard VECM and VAR models in terms of density forecasting ability. The copula-based models are then fitted to Japan consumption and income and compared with standard VAR models. Section 4 gives a summary of the results and presents recommendations.

2. Copula-Based Bivariate Cointegrated Vector Autoregressive Model

The VAR and VECM models are useful in studying linear dynamic relationships among several time series variables. The VAR is the generalization of the univariate autoregression model and can be used whether the vector variables are cointegrated or not.

Assuming the y_t's are I(1), a VAR process could be modeled as follows:

$$\Delta y_{1,t} = \sum_{i=1}^{p} A_{11i}\Delta y_{1,t-i} + \sum_{i=1}^{p} A_{12i}\Delta y_{2,t-i} + \sqrt{v_{1,t}}\, e_{1,t}$$

$$\Delta y_{2,t} = \sum_{i=1}^{p} A_{21i}\Delta y_{1,t-i} + \sum_{i=1}^{p} A_{22i}\Delta y_{2,t-i} + \sqrt{v_{2,t}}\, e_{2,t} \qquad (1)$$

where $e_{1,t}$ and $e_{1,t}$ have mean 0 and variance σ^2_{1et} and σ^2_{2et}, respectively.

The bivariate Vector Error Correction Mechanism (VECM), on the other hand, is given by

$$\Delta y_{1,t} = \alpha_1 (y_{1,t-1} + \beta_1 y_{2,t-1}) + \sum_{i=1}^{p-1} \Gamma_{11i} \Delta y_{1,t-i} + \sum_{i=1}^{p-1} \Gamma_{12i} \Delta y_{2,t-i} + \sqrt{u_{1,t}}\, \varepsilon_{1,t}$$

$$\Delta y_{2,t} = \alpha_2 (y_{1,t-1} + \beta_1 y_{2,t-1}) + \sum_{i=1}^{p-1} \Gamma_{21i} \Delta y_{1,t-i} + \sum_{i=1}^{p-1} \Gamma_{22i} \Delta y_{2,t-i} + \sqrt{u_{2,t}}\, \varepsilon_{2,t} \qquad (2)$$

where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ have 0 and mean variance σ_{1t}^2 and σ_{2t}^2 respectively.

The VECM can be viewed as a restricted VAR. It incorporates the cointegrating relationship(s) in a VAR model equation.

Now an expression for the conditional joint distribution of the error terms of VECM and VAR models can be obtained through Sklar's theorem (1959). For example, the conditional joint distribution of $U(\varepsilon_{1,t}, \varepsilon_{2,t})$ in equation (2) can be expressed in the following manner:

$$(\varepsilon_{1,t}, \varepsilon_{2,t}) \sim U(\varepsilon_{1,t}, \varepsilon_{2,t}; \theta) = C_t\,(F_{1,t}\,(\varepsilon_{1,t}; \delta_1), \mathrm{F}_{2,\mathrm{t}}\,(\varepsilon_{2,t}; \delta_2); \varphi) \qquad (3)$$

where $\delta_{1,}\ \delta_2$ and θ are marginal parameters and copula, respectively.

The use of copulas allows one to model different types of dependence in a flexible way and allows for various marginal distributions. Through the use of copulas, it is possible for the marginal distributions to have different degrees of freedom. For instance distribution F_1 may have a student's t-distribution with v_1 degree of freedom, while distribution F_2 has a student's t-distribution with v_2 degree of freedom, or F_2 has a normal distribution.

2.1 Copula

The definition of a bivariate copula is given in Nelson (2006) as follows:

Definition. A 2-dimensional copula is a function C whose domain is $\mathbf{I}^2$, where the unit square $\mathbf{I}^2$ is the Cartesian product $\mathbf{I}\times\mathbf{I}$, $\mathbf{I}=[0,1]$. The function C has the following properties:

1. For every u, v in $\mathbf{I}$, where $\mathbf{I}=[0,1]$,

   $$C(u,0)=C(0,v)=0, \qquad C(u,1)=1 \text{ and } C(1,v)=v,$$

2. For every u_1,u_2, v_1,v_2 in $\mathbf{I}$ such that $u_1 \leq, u_2$, *and* $v_1 \leq v_{2,}$

   $$C(u_2,v_2)-C(u_2,v_1)-C(u_1,v_2)+C(u_1,v_1) \geq 0.$$

The following theorem provides the framework for the application of copulas.

Theorem (Sklar's Theorem). Let H be a joint distribution function with margins F and G. Then there exists a copula C such that for all x, y in $\overline{R}$

$$H(x,y) = C(F(x), G(y)) \tag{4}$$

If F and G are continuous, then C is unique; otherwise C is uniquely determined on Range $F \times$ Range G. Conversely, if C is a copula and F and G are distribution functions, then the function H defined by (4) is a joint distribution function with margins F and G.

To obtain the best-fitting model, Granger et al. (2006) tested several forms of copulas. As mentioned in their paper, there is no guidance from economic theory regarding the choice of copula, thereby necessitating modeling and comparing several copulas and choosing the best in terms of maximum log-likelihood. In this paper we use the one-parameter copulas commonly used in finance and actuarial science, which are: Gaussian copula, Clayton copula, Gumbel copula, Frank copula and Plackett copula. Each copula allows for different dependence properties.

The *Gaussian copula* is the copula derived from the multivariate Gaussian distribution. The bivariate Gaussian Copula is obtained by the inversion method as follows:

$$\begin{aligned} C_G(u,v) &= \Phi_{pxy}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) \\ &= \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-p_{xy}^2)^{1/2}} \exp\left[\frac{-(s^2 - 2p_{xy}st + t^2)}{2(p_{xy}^2 - 1)}\right] dsdt \end{aligned} \tag{5}$$

where Φ_{pxy} is the CDF of standard bivariate distribution with linear correlation parameter *pxy* and Φ is CDF of standard normal distribution.

The *Clayton, Frank and Gumbel copulas* belong to the Archimedean family of copulas. The Archimedean copula with generator φ is given by

$$\mathrm{C}\,(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)) \tag{6}$$

where φ is a continuous, convex, strictly decreasing function from [0,1] to $[0,\infty]$ such that $\varphi(1)=0$, and $\varphi^{[-1]}$ is the pseudo-inverse function of given by

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \le t \le \varphi(0) \\ 0, & \phi(0) \le t \le +\infty \end{cases} \tag{7}$$

The specific copula functions and generators corresponding to the Clayton, Frank and Gumbel copulas are shown in Table 1.

Table 1. Forms of Selected Archimedean Copulas

	$C_\theta(u,v)$	$\varphi_\theta(t)$	$\varphi_\theta(t)$
Clayton family	$\max\left(\left[u^{-\theta}+v^{-\theta}-1\right]^{-1/\theta},0\right)$	$\frac{1}{\theta}(t^{-\theta}-1)$	$[-1,0)\setminus(0,\infty)$
Frank family	$-\frac{1}{\theta}\left(1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{e^{-\theta}-1}\right)$	$-\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$(-\infty,\infty)\setminus\{0\}$
Gumbel family	$\exp\left(-\left[(-\ln u)^{\theta}+(-\ln v)^{\theta}\right]^{-1/\theta}\right)$	$(-\ln t)^{\theta}$	$[1,\infty)$

The Placket copula, on the other hand, has the following form,

$$C(u,v)=\frac{\left[1+(\theta-1)(u+v)\right]-\sqrt{\left[1+(\theta-1)(u+v)\right]^2-4uv\theta(\theta-1)}}{2(\theta-1)} \qquad (8)$$

for $\theta>0$, $\theta\neq1$.

2.2 Estimation

Let (Z_{1t},Z_{2t}) denote a continuous bivariate random vector where t=1,…,T. Let $F_i(Z_{it})$ and $f_i(z_{it})$ denote the cumulative distribution function (cdf) and probability density function (pdf) of Z_{it}, respectively. Let $U_i= F_i(Z_{it})$ and let $C(U_{1t},U_{2t})$ denote the joint cdf of (U_{1t},U_{2t}), $c(u_{1t},u_{2t})$ denote the pdf corresponding to $C(u_{1t},u_{2t})$, and $H(Z_{it},Z_{2t})$ and $h(z_{1t},z_{2t})$ denote the cdf and pdf of (Z_1,Z_2), respectively. Then the joint density function $h(z_{1t},z_{2t})$of (Z_{1t},Z_{2t}) can be expressed via Sklar's theorem in the form of (4) as follows:

$$h(z_{1t}, z_{2t}) = c\{F_1(z_{1t}), F_2(z_{2t})\}\, f_1(z_{1t})\, f_2(z_{2t}) \qquad (9)$$

The log-likelihood function is given by:

$$L(\theta)=\sum_{t=1}^{T}\log\left[c\left(F_1(z_{1t}),F_2(z_{2t})\right)f_1(z_{1t})f_2(z_{2t})\right] \qquad (10)$$

where θ is the set of all parameters of both the marginal distributions and the copula.

Hence, given a set of marginal pdf's and copula, the log-likelihood may be written as in (10), and the maximum likelihood estimator obtained, where $\hat{\theta}_{MLE} = \max_{\theta \in \Theta} L(\theta)$.

In the simulation study, we employ multi-stage maximum log likelihood (MSML) because the computational burden is lighter especially when there are a lot of parameters to be estimated. MSML was proposed for estimating copula by Joe and Xu (1996) under the name inference functions for margins (IFM). The estimation procedure is performed in two steps. In the first step, parameter estimates are obtained separately by optimizing the univariate likelihoods based on the margins. This is then followed by optimizing the multivariate likelihood treated as a function of the copula parameter. In addition the canonical maximum likelihood (CML) method is employed in the empirical study. The CML method estimates each marginal distribution nonparametrically by the empirical distribution. This method is used to check the consistency of the dependence parameter with its IFM since it is difficult to determine the true distributions in empirical data.

3. Application

3.1.1 Simulation Design

The data employed in this study is generated from a first order cointegrated bivariate VAR model which has the form:

$$\Delta Y_t = \begin{bmatrix} -1 \\ -0.5 \end{bmatrix} \begin{bmatrix} 0.5 & 1 \end{bmatrix} Y_{t-1} + \varepsilon_t \tag{11}$$

where $Y_t=(y_{1t},y_{2t})'$ and error term $\varepsilon_t = \left(e_{1t}, e_{2t}\right)$.

We consider two scenarios for the distribution of the error term, i.e., bivariate normal distribution with the marginal distributions having mean 0 and variance 1, and bivariate student's t- distribution with 3 degrees of freedom (d.f). In each scenario, three values of the dependence measure are considered corresponding to Kendall's tau of 0.3, 0.5 and 0.9, or equivalently, a linear correlation coefficient r of 0.454, 0.707 and 0.908, given the relation $\tau = \frac{2}{\pi} \arcsin \rho$ that holds for (essentially) all elliptical distributions (Lindskog, 2000). Thus estimates of the copula parameter of normally- and student's t-distributed variables can be obtained given the linear correlation coefficient (Kendall's tau).

The scenarios and cases considered in the simulations are as follows:

Scenario I: The error term has a bivariate normal distribution

Case (i): $\varepsilon_t = (e_{1t}, e_{2t}) \sim N(0,1)$ with $\tau=0.3$

Case (ii): $\varepsilon_t = (e_{1t}, e_{2t}) \sim N(0,1)$ with $\tau=0.5$

Case (iii): $\varepsilon_t = (e_{1t}, e_{2t}) \sim N(0,1)$ with $\tau=0.9$

Scenario II: The error term has a bivariate student's t-distribution

Case (iv): $\varepsilon_t = (e_{1t}, e_{2t}) \sim t_3$ with $\tau=0.3$

Case (v): $\varepsilon_t = (e_{1t}, e_{2t}) \sim t_3$ with $\tau=0.5$

Case (vi): $\varepsilon_t = (e_{1t}, e_{2t}) \sim t_3$ with $\tau=0.9$

In each case under each scenario, 200 samples each consisting of 200 data points are generated using equation (11) with the error terms following the aforementioned distributions and corresponding Kendall's tau. A sample size of 200 for each data set is considered with a view to applying the procedure to an actual data set that has less than 200 observations. Each generated sample is estimated via VECM and VAR model. Copula parameters are estimated from the error terms of marginal models to observe the performance of the copulas after filtering by VECM(1) and VAR(1).

For given marginals and method of copula estimation, let θ_i be the one-parameter copula estimator of q for the *i*th generated sample, i=1,…,200, and θ_0 represent the true value of the copula parameter given Kendall's tau and let N denote the number of generated samples. We consider two criteria, root mean squared error (RMSE) and estimated bias where $RMSE = \left[N^{-1}\sum\{\theta_i - \theta_0\}^2\right]^{1/2}$ and estimated bias = $N^{-1}\sum\theta_i - \theta_0$ in evaluating the five copulas tried.

3.1.2 Bias and RMSE and Log-Likelihood of Models of Copula

Tables 2 and 3 present the bias and RMSE of each copula model measured assuming the error terms follow a bivariate normal distribution and a bivariate student's t-distribution, respectively. It is seen from these tables that, with the exception of the Clayton copula in some cases and the Plackett copula in many cases, the biases obtained in estimating the dependence parameter are relatively small compared to the value of the true parameter irrespective of whether the true distribution of the error term is normal or Student's t-distribution.

In addition, irrespective of the distribution of the error term and copula used, the bias and RMSE of the copula parameter in the VECM model are lower than that produced by VAR when the dependence coefficient is relatively low($\tau=0.3$). But when the dependence coefficient is moderate ($\tau=0.5$), the VAR model produces lower bias and RMSE when the error distribution is normal for all copulas except the Clayton copula. However, the results are reversed when the error terms follow a Student's t-distribution and τ=0.5; the VECM model this time has generally lower bias and RMSE for all copulas except the Clayton copula. On the other hand, when the dependence coefficient is high (τ=0.9), in contrast to the results when the coefficient is low, the RMSE and bias of the estimates obtained from the VAR are generally lower compared to the corresponding VECM-based values.

Table 2. Bias and MSE of Copulas in Bivariate Normal Distribution*

Copula	Model		τ=0.3	τ=0.5	τ=0.9
Gaussian		**TRUE θ_0**	0.45	0.71	0.99
	VECM	**Bias**	0.09	0.05	0.03
		RMSE	0.14	0.07	0.03
	VAR	**Bias**	0.14	0.04	0.01
		RMSE	0.16	0.06	0.01
Gumbel		**TRUE θ_0**	1.43	2.00	10.00
	VECM	**Bias**	0.17	0.07	4.95
		RMSE	0.63	0.23	5.00
	VAR	**Bias**	0.30	0.04	3.22
		RMSE	0.02	0.17	3.30
Clayton		**TRUE θ_0**	0.86	2.00	18.00
	VECM	**Bias**	0.08	0.41	12.45
		RMSE	0.79	0.53	12.51
	VAR	**Bias**	0.25	0.46	10.06
		RMSE	1.19	0.53	10.13
Frank		**TRUE θ_0**	2.92	5.74	20.90
	VECM	**Bias**	1.10	0.73	1.68
		RMSE	3.01	1.27	3.23
	VAR	**Bias**	1.76	0.62	4.74
		RMSE	4.27	1.04	5.23
Plackett		**TRUE θ_0**	4.00	11.40	530.00
	VECM	**Bias**	3.77	1.39	431.24
		RMSE	15.76	3.99	432.66
	VAR	**Bias**	6.84	0.94	341.70
		RMSE	28.42	2.90	344.32

* *Bias and MSE were computed on 200 random samples of length 200 from bivariate cointegrated autoregressive process of order one.*

The results in Tables 2 and 3 suggest that, when the dependence between the error terms is high and therefore the departure from the long-run equilibrium relationship in the previous period does not sufficiently explain the variation in the two variables, simply fitting a copula-based VAR model on the data tends to yield more accurate and more precise dependence estimates. But when the dependence in the error terms not accounted for by the cointegrating relationship is low, then there is an advantage to incorporating the cointegrating relationship in the model.

However, when the dependence between the error terms is moderate, the choice of copula and the model fitting method is not as clear cut. Thus greater care has to be exercised in choosing the model fitting method as well as the copula when the dependence between the error terms is suspected to be moderate and accurate and precise estimates of the dependence parameter are desired.

Table 3. Bias and MSE of Copulas in Bivariate Student's-t Distribution*

Copula	Model		τ=0.3	τ=0.5	τ=0.9
Gaussian		**TRUE θ_0**	0.09	0.04	0.03
	VECM	**Bias**	0.09	0.04	0.03
		RMSE	0.13	0.08	0.03
	VAR	**Bias**	0.15	0.05	0.01
		RMSE	0.16	0.07	0.02
Gumbel		**TRUE θ_0**	1.43	2	10
	VECM	**Bias**	0.17	0.21	4.97
		RMSE	0.25	0.32	5.02
	VAR	**Bias**	0.28	0.23	3.28
		RMSE	0.33	0.33	3.38
Clayton		**TRUE θ_0**	0.06	0.14	1.27
	VECM	**Bias**	0.09	0.21	12.47
		RMSE	0.27	0.45	12.52
	VAR	**Bias**	0.24	0.21	10.31
		RMSE	0.34	0.41	10.38
Frank		**TRUE θ_0**	2.92	5.74	20.9
	VECM	**Bias**	1.05	1.19	2.12
		RMSE	1.38	1.65	3.41
	VAR	**Bias**	1.66	1.31	3.45
		RMSE	1.84	1.69	4.40
Plackett		**TRUE θ_0**	4	11.4	530
	VECM	**Bias**	2.75	5.31	430.96
		RMSE	3.50	7.05	432.13
	VAR	**Bias**	4.22	5.72	337.99
		RMSE	4.72	7.46	341.13

* *Bias and MSE were computed on 200 random samples of length 200 from bivariate cointegrated autoregressive processes of order one*

Tables 4 and 5 present the number and percentage of models which attained the maximum log-likelihood among the 200 samples of 200 observations, under assumptions of a normal distribution and a student's t distribution, respectively.

As may be expected, the Gaussian copula fits the marginal model better when the error term is normally distributed irrespective of whether the marginal models are estimated by VECM or VAR (Table 4). Nevertheless when τ=0.3, there is still about 30% to one-third chance that other copulas fit better than the Gaussian copula whether VAR or VECM is fitted. However, when the dependence coefficient is higher (τ=0.5 and τ=0.9), the Gaussian copula is far superior to the other copulas.

Table 4. Number and Percentage which Attained Maximum Log-likelihood in Bivariate Normal Distribution*

Copula	Model		τ=0.3	τ=0.5	τ=0.9
	VECM	Number	134	162	177
Gaussian		percentage	67%	81%	89%
	VAR	Number	144	160	161
		percentage	72%	80%	81%
	VECM	Number	20	21	18
Gumbel		percentage	10%	11%	9%
	VAR	Number	20	21	17
		percentage	10%	11%	9%
	VECM	Number	16	3	0
Clayton		percentage	8%	2%	0%
	VAR	Number	12	5	0
		percentage	6%	3%	0%
	VECM	Number	23	9	5
Frank		percentage	12%	5%	3%
	VAR	Number	19	11	20
		percentage	10%	6%	10%
	VECM	Number	7	5	0
Plackett		percentage	4%	3%	0%
	VAR	Number	5	3	2
		percentage	3%	2%	1%

* *log-likelihood were computed on 200 random samples of length 200 from bivariate cointegrated autoregressive processes of order one*

When the error term has a student's t distribution, the Gumbel copula and the Plackett copula appear to be good candidates for estimating the marginal distributions when τ=0.3, irrespective of whether VECM or VAR models are used (Table 5). This outcome corroborates the results of Granger et al. (2006) where the Gumbel copula fitted best with parameter coefficient 1.0977 even as the said study used the skewed t-distribution for the marginal density.[1] When τ=0.5 the Gaussian copula is added to these two candidates. When τ=0.9, the Gaussian copula unexpectedly outperforms the other copulas, yielding the highest number of resamples with maximum log-likelihood, especially when used in conjunction with VECM.

Thus, the results in Tables 4 and 5 indicate that the Gaussian copula is, as may be expected, a very good choice when the error terms can be assumed to follow a normal distribution, and its performance is even better the error terms are moderately or highly correlated. But when the correlation between the error terms is low, however, there is a 30% to 33% chance that another copula can model the distribution of the data better than the Gaussian copula. But in the absence of guidance as to which copula will perform better, the Gaussian copula is a safe choice.

Table 5. Number and Percentage which Attained Maximum Log-likelihood in Bivariate Student's-t Distribution*

Copula	Model		τ=0.3	τ=0.5	τ=0.9
Gaussian	VECM	Number	18	53	156
		percentage	9%	27%	78%
	VAR	Number	24	48	97
		percentage	12%	24%	49%
Gumbel	VECM	Number	57	59	36
		percentage	29%	30%	18%
	VAR	Number	60	60	58
		percentage	30%	30%	29%
Clayton	VECM	Number	25	7	0
		percentage	13%	4%	0%
	VAR	Number	12	2	1
		percentage	6%	1%	1%
Frank	VECM	Number	3	5	5
		percentage	2%	3%	3%
	VAR	Number	1	4	13
		percentage	1%	2%	7%
Plackett	VECM	Number	87	76	3
		percentage	44%	38%	2%
	VAR	Number	103	86	3
		percentage	52%	43%	16%

* *log-likelihood were computed on 200 random samples of length 200 from bivariate cointegrated autoregressive processes of order one*

Interestingly, the Gaussian copula also performs satisfactorily when the error distribution is student's t, provided the dependence parameter is high and its performance is further improved if the copula-based VECM model is fitted. Thus, the slight loss in accuracy and precision in the dependence estimate can be offset by the better fit afforded by a Gaussian copula even if a VECM model is used in the presence of high correlation among the error terms. For low to moderate dependence, however, the Plackett and Gumbel copulas perform better when the error distribution is student's t.

3.1.3 Comparison of Models

The standard VAR and VECM models are compared with the copula-based models on the basis of the multivariate conditional KLIC divergence measure. This measure is defined as the distance between the true multivariate conditional density $g_t\left(x_{1,t},\ldots,x_{m,t}\middle|F_{t-1};\lambda_m\right)$ and the model based density $f_t\left(x_{1,t},\ldots,x_{m,t}\middle|F_{t-1}\right)$:

$$L^m\left(f_t : g_t, \lambda_m\right) = E\left[\ln f_t(x_{1,t},\ldots,x_{m,t}) - \ln g_t(x_{1,t},\ldots,x_{m,t};\lambda_m)\right]$$

Low values of the KLIC indicate that the model is close to the true density. In this paper, we compare the density forecasts ability of the models by measuring the distances between the measured distribution density and the true distribution density through conditional KLIC divergence measure (Bao et al., 2007). In addition, we approximate the true multivariate distribution density using the non-parametric multivariate product kernel estimator suggested by Scott (1992) and Li and Racine (2007) which is given by:

$$\hat{f} = \frac{1}{nh_1 \cdots h_m} \sum_{t=1}^{n} \left\{ \prod_{j=1}^{m} K\left(\frac{x_j - x_{jt}}{h_j} \right) \right\}$$

where $K(\cdot)$ denotes the Gaussian kernel and h_j denotes kernel bandwidth. The kernel bandwidth is calculated as $\hat{h} = [4/(m+2)]^{1/(m+4)} \cdot \hat{\sigma}_j \cdot n^{1/(m+4)}$ where $\hat{\sigma}_j$ is the standard deviation of x_{jt}.

Table 6 below presents the number of models which attained the lowest KLIC and the total KLIC of models when the error terms are jointly normally distributed. In general, at least one copula-based model attains the most number of fitted models with lowest KLIC in any given dependence scenario, indicating better density forecasting ability compared to standard models.

Table 6. Number of Models which Attained the Lowest KLIC out of 200 Samples and Total KLIC of the Models when the Error Term is Normally Distributed

Models		τ=0.3 No. with. min KLIC	τ=0.3 Total KLIC	τ=0.5 No. with min.KLIC	τ=0.5 Total KLIC	τ=0.9 No. with min. KLIC	τ=0.9 Total KLIC
Standard	VECM	11	14.35	29	13.4	8	24.7
Copula-Based	Gauss	24	4.83	15	5.69	0	23.11
VECM	Gumbel	20	5.12	18	5.42	0	22.6
	Clayton	2	17.15	0	26.37	0	37.32
	Frank	29	5.4	64	4.48	0	20.82
	Plackett	11	5.37	13	6.51	0	24.65
Standard	VAR	3	72.17	1	52.28	15	19.35
Copula-Based	Gauss	30	3.27	4	5.68	13	2.96
VAR	Gumbel	16	3.39	12	5.32	60	2.68
	Clayton	0	14.78	0	20.07	0	9.62
	Frank	30	3.04	41	4.07	98	2.51
	Plackett	24	3.58	3	6.15	6	3.8

The simulation results indicate that the Frank copula-based models perform rather well across the three dependence scenarios considered (low, moderate, and high) when the error term is normally distributed, and not the Gaussian-based models as one might expect. When the correlation between the error terms is low, the Gaussian copula-based model is nearly as good as the Frank-based model irrespective of whether the VECM or VAR model is used. However, when the correlation is moderate, the Frank copula-based VECM is best. When the correlation is high, the Frank copula-based VAR is the best choice.

As in the case of a jointly normally distributed error term, at least one copula-based model outperforms the corresponding standard VECM or VAR model when the error term follows a student's t-distribution (Table 7). And as in the normal case, the results suggest that the use of copula-based VECM models is not advisable in the presence of high correlation. However, unlike in the normal case, no single copula is the best choice across the dependence scenarios. The Clayton-based VAR performs best under low correlation, the Gaussian-based VECM under moderate correlation and the Frank-based VAR under high correlation.

Table 7. Number of Models which Attained the Lowest KLIC out of 200 Samples and Total KLIC of the Models when the Error Term is Student's t-distribution

Models		τ=0.3		τ=0.5		τ=0.9	
		No. with. min KLIC	Total KLIC	No. with min.KLIC	Total KLIC	No. with min. KLIC	Total KLIC
Standard	VECM	5	41.2	15	30.51	10	12.81
Copula-Based	Gauss	41	6.34	61	4.25	0	6.72
VECM	Gumbel	8	8.45	7	5.83	3	5.92
	Clayton	41	5.27	21	9.35	0	16.18
	Frank	6	8.6	5	6.37	4	5.2
	Plackett	6	9.41	7	5.95	0	7.23
Standard	VAR	2	77.43	6	31.99	6	24.65
Copula- Based	Gauss	8	8.36	51	4.11	1	3.1
VAR0	Gumbel	2	11.05	6	6	68	2.02
	Clayton	79	3.69	10	9.46	0	10.15
	Frank	2	11.52	6	6.55	106	1.96
	Plackett	0	12.24	5	6.1	2	2.9

3.2 Empirical Study

3.2.1 Data

Copula-based models are tried on the quarterly time series data on the real consumption expenditure and gross national income of Japan from 1957Q1 to 2004Q1. The data, obtained from International Financial Statistics (IFS) published by the International Monetary Fund, is in constant prices in billions of yen.

3.2.2 Tests for Statandionarity and Cointegration

The results of the Augmented Dickey Fuller (ADF) test, Phillip-Perron (PP) test the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test and the Johansen Cointegration test are presented in Tables 8 and 9. Results from the three unit roots tests indicate that the variables are integrated order of one.[2] Moreover, the results of the cointegration test indicate that there is a cointegration relationship between income and consumption.

Table 8. Results of Unit Root Tests

	Test	Level	1st differenced
Logged consumption	ADF	-2.7339	-4.1132**
	PP	-4.0699	319.8787**
	KPSS	0.3494**	0.264239
Loged income	ADF	-2.6091	-5.1188**
	PP	-2.2892	-245.2109
	KPSS	0.4338**	0.1102

** *stands for significant at one percent significance level*

Table 9. Johansen Cointegration Test Results

	Number of Cointegration	Maximal Eigenvalue	Trace
Test Statistics	At most one	0.89	0.89
	None	44.78**	45.66

** *stands for significant at one percent level*

3.2.3 Estimation of Econometric Models and Copulas

We employed Akaike information criterion (AIC) and Schwarz's information criterion (SIC) to determine the appropriate lag length of the VECM and VAR models. These measures indicated a lag length of four periods for both the VECM and VAR models.

Table 10 presents the parameter estimates for the copula-based VAR econometric model. Copula-based VECM models for the data were not pursued further as the error correction terms in the VECM model were not significant, indicating that both consumption and income are not responding to adjustments from disequilibrium. The VECM and VAR models return almost the same estimated coefficients. Thus only the copula-based VAR is compared to the standard VAR model.

Table 10. Estimates of Copula-based VECM and VAR

		VAR			
		Consumption		Income	
intercept		0.007**	(0.002)	0.003^	(0.001)
Lag1	Consumption	-0.466**	(0.062)	-0.0002	(0.047)
	Income	0.397**	(0.125)	-0.015	(0.081)
Lag2	Consumption	-0.250**	(0.077)	0.064	(0.050)
	Income	0.565**	(0.126)	0.224**	(0.082)
Lag3	Consumption	-0.321**	(0.075)	0.020	(0.049)
	Income	0.377**	(0.131)	0.225**	(0.085)
Lag4	Consumption	0.414**	(0.125)	0.027	(0.081)
	Income	0.109	(0.125)	0.138^	(0.081)
ECM			-		-
Residual s.e		0.024		0.015	
d.f (residual s.e.)		175		175	
Multiple R-squared		0.76		0.26	
Adjusted R-squared		0.75		0.23	
F-statistics		69.91		8.016	
d.f (F-statistics)		8;175		8;175	
p-value		0.000		0.000	

Numbers inside parenthesis are standard errors of estimates
*** and * stand for significant at 1 and 0.5 percent significant level, respectively, while ^ denotes significant at 10 percent significant level*

The same copulas used in the simulation study were used to model the dependence between the marginals, except for the Gumbel copula. Optimization problems were encountered when the Gumbel copula was tried for both VECM and VAR models and using both IFM and CML methods. The standard normal distribution was found to fit the data for the margins and the residual distribution passed the goodness-of-fit test.

Table 11 reports estimates of the dependence parameter based on the IFM and CML methods that were applied after fitting the VAR models. The estimated coefficients of the dependence parameter yielded by the IFM and CML methods show small differences for the model residuals. This implies that the marginal models are correctly specified (Kim, et al., 2007b). The dependence estimates obtained imply relatively low dependence. For instance, the estimated parameter coefficients of the Gaussian and Frank copulas imply τ values of 0.2 to 0.3 approximately, that of the Clayton copula indicates a τ value in the 0.1 to 0.3 range, while that of the Plackett copula yields a τ of approximately 0.2 to 0.4. However the calculated value of Kendall's tau on the empirical distribution of the error terms is 0.27, which is not inconsistent with the above results. Among the copulas employed, the Gaussian copula registers the best fit to the data, implying symmetry in the dependence structure.

Table 11. Estimates of the Dependence Parameter of Copula *

	VAR			
	IFM			
	Gauss	Clayton	Frank	Plackett
Parameter estimates	0.406	0.329	2.897	4.501
	(0.056)	(0.109)	(0.535)	(1.0196)
Maximized log-likelihood	16.663	11.284	13.920	16.142
	CML			
	Gauss	Clayton	Frank	Plackett
Parameter estimates	0.379	0.548	2.407	3.541
	(0.059)	(0.123)	(0.477)	(0.772)
Maximized log-likelihood	14.146	13.784	12.849	14.355

** All estimates are significant at a=0.01. Numbers inside parentheses are standard errors of estimates.*

3.2.4 Model Comparison and Evaluation

Table 12 presents the out-of-sample conditional KLIC. To calibrate the out-of-sample conditional KLIC, the time period is divided into the in-sample period t_1, where $t_1 = 1,\ldots,R$ and the out-of-sample period t_2, where $t_2 = R+1,\ldots,T$. In this study, R is set as half period of time T, i.e., R=T/2. The out-of-sample KLIC symmetry is calculated using the parameters (mean and variance of the empirical distribution and dependent parameter of copula) obtained from the in-sample period.

Although the Gaussian copula shows best fit to the data, the Frank copula model attains the lowest conditional KLIC in the out-of-sample period. The estimated Kendall's tau for the empirical data is 0.27 and is, therefore, relatively low. Thus the results of the empirical model are not inconsistent with the simulation results.

Table 12. Out-of-sample Conditional KLIC measures

		VAR
Standard		3.185
Copula-based models	Gauss	3.109
	Clayton	2.992
	Frank	2.947
	Plackett	2.986

4. Summary and Recommendations

The simulation results suggest that the choice of copula and model to be fitted, i.e., whether copula-based VECM or copula-based VAR, depends on the underlying distribution of the model error terms, the strength of the correlation between the error terms as well as focus of the study – i.e., whether estimation of the dependence parameter of the joint distribution is of interest, the fit of the copula-based models to the data, or the forecasting ability of the copula-based VAR or copula-based VECM model. Initially, this general finding may come as a surprise until it is recalled that different criteria are used to compare the competing copulas and models. A case in point is the ability of the multiple regression model to generate good forecasts in the presence of multicollinearity despite ill-behaved parameter estimates.

When the focus of the investigation is estimation of the dependence parameter, the simulations results suggest the use of a copula-based VECM when the correlation is low and the use of a copula-based VAR when the correlation is high irrespective of the underlying error distribution and choice of copula. However, when the correlation is moderate, the results suggest the use of copula-based VAR when the error term is normal, and the use of a copula-based VECM when the error term is student's t except, again, for the Clayton copula.

However, when the interest is in fitting the copula-based models to the data, the choice of copula is apparently more crucial than the choice between copula-based VAR and copula-based VECM models. The Gaussian copula expectedly outperforms the other copulas when the underlying error distribution is normal, particularly when the correlation between the error terms is moderate or high. It also performs satisfactorily when the error distribution is student's t, provided the dependence parameter is high and the copula-based VECM model is used. For low to moderate dependence between error components following a student's t distribution, however, the Plackett and Gumbel copulas outperform the Gaussian copula, with the copula-based VAR tending to outperform its VECM counterpart when the Plackett copula is used.

When forecasting ability is the main concern, copula-based models generally outperform standard models. The simulation results further indicate the superiority of copula-based VAR models over VECM-based models when the correlation in the error terms is high, irrespective of whether the error term follows a normal or a student's t distribution. When the error term is normally-distributed, the Frank copula, and not the Gaussian copula, is the copula of choice. When the error term follows a student's t-distribution, on the other hand, no single copula or model-fitting method is shown to be the best choice across the dependence scenarios. Results indicate the use of Clayton-based VAR, Gaussian-based VECM and the Frank-based VAR for the low, moderate and high correlation scenarios, respectively.

Results obtained from applying copula-based VAR on real consumption expenditure and real gross national income in Japan are consistent with the simulation findings. In particular, the copula-based VAR models display better density forecasting performance than the standard VAR model, with the Frank copula-based VAR providing the best forecasting performance.

Future simulation studies can focus on alternative data generating mechanisms as well as on a finer grid of values for the dependence parameter t (say, in increments of 0.1 from 0.1 to 0.9). Other types of copula such as the two-parameter copula can also be tried under different dependence scenarios such as asymmetrically distributed data.

NOTES

1 Granger et. al. did not employ cointegration analysis, but application of the Johansen test to the data used in the paper (which was taken from St. Louis Federal Reserve web page) indicated the existence of a cointegration relationship.

2 The null hypothesis of the KPSS test is that there is no unit root, while for the ADF and PP tests the null hypothesis is that there is a unit root.

REFERENCES

BAO, Y., LEE, T.H. and BURAK, S., 2007, Comparing Density Forecast Models, *Journal of Forecasting*, 26:203-225.

BIANCHI, C., CARTA, A., FANTAZZINI, D., DE GIULI, M.E. and MAGGI, M.A., 2009, Copula-VAR-X Approach for Industrial Production Modelling and Forecasting, *Applied Economics*, 24.

GRANGER, C.W.J., TERÄSVIRTA, T. and PATTON A.J., 2006, Common Factor in Conditional Distributions for Bivariate Time Series, *Journal of Econometrics*, 132: 43-57.

JOE H., 1997, *Multivariate Models and Dependence Concepts*, Chapman & Hall/CRC.

JONDEAU, E. and ROCKINGER, M., 2006, The Copula-GARCH model of Conditional Dependencies: An International Stock Market Application, *Journal of International Money and Finance,* 25:827-853.

KIM, G., SILVAPULLE, M.J. and SILVAPULLE, P., 2007, Estimating the Error Distribution in Multivariate Heteroscedastic Time Series Models, *Monash Econometrics and Business Statistics*, No 8/07.

LINDSKOG, F., 2000, Linear Correlation Estimation, Working Paper, Zürich.

MITCHELL, J., 2007, Constructing Bivariate Density Forecasts of Inflation and Output Growth Using Copulas: Modelling Dependence Using the Survey of Professional Forecasters, National Institute of Economic and Social Research.

NELSON, R.B., 2006, *An Introduction to Copulas*, Springer-Verlag New York.

PATTON, A.J., 2005, Modelling Time-Varying Exchange Rate Dependence, *Intentional Economic Review*, 47(2):527-556.

SKLAR, A., 1959, *Fonctions de repartition à n dimensions et leurs marges*, publ. Inst. Statist. Univ. Paris,8,229-231.

VUONG, Q.H., 1989, Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses. *Econometrica*, 57: 307-334.

A Dose of Business Intelligence: Data Mining

Joseph Ryan G. Lansangan
University of the Philippines Diliman

With the advancement of data warehousing and processing, large volume and/or different sources of data are now available, such as data on customer profile, transaction details, business processes, and even marketing efforts. Data is processed and summarized into useful information for business strategies. Given the appropriate techniques and tools, companies become proactive on their decisions and/or actions, with insights made for the future using present and/or historical information. Companies value such processes, and hence they continue to gather data, formulate strategies, and make actions, which in turn, become new information and/or yield new business directions. Such cycle of three components—data, question, and decision—remains at the core of today's business intelligence. With the continuous influx of data, questions arise, and hence actions are to be made. Equivalently, with the new directives, information is desired to arrive at certain decisions. And, when actions are made, data comes in and possibly new directions and/or objectives are created.

As an aid to decision-making, crucial to the business intelligence framework (or cycle) is data mining. Data mining is the process of extracting useful (hidden) information or knowledge from large volume of data, commonly implemented on an automated, timely and quick manner as solutions to or support for different analytical queries and/or business problems. Data mining is used to uncover inherent patterns based on historical information, allowing for statistical predictions, characterization and/or classifications of data. Information is then presented in meaningful ways, be it for exploratory reasons (e.g., deep-dive or drill-down analyses) or for modeling purposes. Thus companies with data mining capacity become more forward-looking based on what can be inferred from what information is available. Since data is built from the customers themselves, companies tend to be "customer-centered," and since the processes are carried out to meet certain objectives, companies become "tactically-driven."

Data Mining Techniques

Data mining techniques can be classified into five general areas. First, visual representations techniques are graphical interpretations of complex (and even simple) relationships, which are commonly the “front-end” of other data mining techniques but are also used as “post-hoc” procedures. Data is accessed via specialized views and/or drill-down processes for deeper analyses. Second, variable/feature selection methods are dimension-reduction techniques to summarize data into “relatively fewer” features, commonly used to identify the “more important” information. These are often conducted as data pre-processing, but are also used for index-derivation objectives. Third, segmentation and clustering techniques are used to find groups of “similar” characteristics based on relevant dimensions. Segments or clusters are made based on different similarity (or dissimilarity) measures, the objective of grouping often for profiling purposes, for “targeting” specific segments, or for classifying (of “new” units). Fourth, association rules are used to look for significant relationships and/or sequences among transactions (or events), with the rules based on frequent patterns. Common applications are collaborative filtering, market basket analysis and sequence analysis. Fifth, predictive modeling looks into developing a “model” based on discovered patterns or trends in the data, with the “model” being used to predict future outcome and/or identify impacts of changes in behaviors or activities. Predictive models are commonly used for robust customer valuation (or scoring) and identification (e.g., customers who are most likely to respond to an offer).

The different data mining techniques may address specific objectives, but their essence for a particular company remains the same – to identify and/or understand their customers, gain insights on the company’s products and/or services, and take action based on what is presented by or inferred from the data. Visual representations are the most straight-forward, giving deeper perspectives of what the data/information conveys more than what is obvious, using 3-dimensional plots or interactive charts. Feature selection techniques yield interpretable and/or actionable information based on the “best” set/s of variables (relatively fewer than the original set of variables, or combined at fewer dimensions) that capture/s the most from the data. Segments and clusters derived from grouping techniques give deeper comprehension of latent or data-based affinities. Association rules may yield both inexplicable and interpretable rules, but still give knowledge on who the customers are or why customers make transactions (or participate in certain events). Predictive and/or forecasting models are best used to anticipate or forecast patterns or movements, thus the company can decide in a statistical sense (or at calculated risks) using available data.

The Data Mining Process

Different sources in the literature and different data mining software provide different frameworks of the data mining process. But somehow, the data mining process (or any analytical procedure for business intelligence, in this case) can be summarized in three stages – (1) objective and/or data setting (2) data processing and/or analysis, and (3) documentation and execution. These can be further classified as follows – under objective and/or data setting, the company must (a) know the business directives and/or identify specific objectives or queries, (b) then translate the business objectives into analytical objectives, and (c) prepare the data and map out the methodology (if data requirements and/or methods do not suffice to meet the objectives, then the objectives must be re-aligned or the data must be gathered and/or methods must be modified); for data processing, activities include (d) extraction, transformation and loading of data, and (e) analytics proper which includes validation and/or assessment procedures; and finally, activities under documentation and execution include (f) report writing and (g) implementation of decisions/actions.

Note that the discussed stages and/or processes above can be both simplistic and complicated. In the case of setting the business directives, it can be as simple as the Business Intelligence (BI) unit identifying the specific objectives; it can be as not-that-simple as the top management giving general company goals and thus the BI unit works in collaboration with other units (e.g., Marketing unit, Contact center) to come up with specific goals that meet and/or are parallel with the company goals. Translation of the specific objectives into analytical objectives together with data preparation and methodology-sketching are relatively easy tasks, but these become difficult when the company has limited resources (e.g., data sources, software to be used, statistician/s or analyst/s to be engaged, knowledge of methods). The analytics proper has its own simple and complex issues, which basically depend on both the tool/technique and the user. As there are no fixed steps to follow (but standards or best practices remain), analyses are never permanent for a given problem.

How Good Data Mining can Be

To best apply the different data mining techniques, one should not only know what technique is appropriate for a given data, but should always be guided by what the business objective/s is/are. Though it seems that data mining is driven by data, what remains fundamental are (1) the company's motivation or directive – what the company desires to do or needs to address (prior to data mining); and (2) the company's understanding of the results – how the company reacts with the results (during and/or after data mining).

Since data mining entails uncovering hidden information, the discovery process can be complicated, but the effective use of data mining first and foremost lies on the reason/motivation for the conduct of such. Questions and/or objectives must be in place, since these remain as the foundation of all analyses to be made – these questions/objectives essentially define the analyses, taking into account what data or resource is available. As data mining methods can be subjective in nature, approaches may vary but must always be within the scope of the objectives.

Similarly, since the results of data mining are sometimes difficult to understand and appreciate, companies must be able to translate the (mostly quantitative) results into solutions to its business problems or as action-items to meet the business objectives. Levels of interpretability of the results range from easy (e.g., Decision tree models) to difficult (e.g., Neural network models), but the results must always be taken "as is," on the assumption that the data used and the processes made prior to generation of the results are accurate and/or acceptable. Data mining may yield non- or counter-intuitive results, but for as long as the results are extracted from a data using statistically sound processes, the results are (empirically) valid. Rather than challenging such results based on other non-empirical evidences (or, on "similar" studies), such results must be accepted and interpreted in the context the company understands.

Data mining techniques will forever be present – once new data comes in, there will always be something to work on, and hence, innovations are possible. But the importance of an old or new technique must be paralleled with the importance of when/how the technique is used and how the results are interpreted. More often than not, the success of a data mining technique depends not only on the tool or the algorithm/technique, but also on the user's statistical and analytical sense and sensibility.

Bootstrap Methods

Erniel B. Barrios, Ph.D
University of the Philippines Diliman

The classical framework of statistical inference relies heavily on the sampling distribution as a link between the information provided by the sample and the generalizations it provides about the population. Samples are drawn independently and assumed to come from some special distributions so that a closed-form sampling distribution is easily derived. In the more complicated scenario, desirable properties are derived invoking large sample sizes in cases where the sampling distribution is not mathematically tractable. Many statistics are usually analyzed dealing with small samples, often resulting to more complicated standard errors.

Developments in statistical inference had been influenced tremendously by access to efficient computing facilities that allow verification of properties of complicated statistics or those without closed-form. The Bootstrap is a resampling method involving large amount of computations that facilitates small sample inference on a variety of estimation and hypothesis testing problems.

Efron (1979) introduced the bootstrap as a special case of the jackknife, a resampling method already known much earlier. The method was originally intended for independent set of observations, say a random sample $x = (x_1, x_2, ...x_n)$ from F. With the aim of understanding the sampling distribution of the random variable $R(x, F)$ from *x*, Efron (1979) provided the following bootstrap algorithm:

1. Construct the sample probability distribution $\hat{F}$, putting mass $1/n$ at each of the mass point $x_1, x_2,...,x_n$.
2. With $\hat{F}$ fixed, draw random sample of size n from $\hat{F}$, say $X_i^* = x_i^*, X_i^* \sim iid\, \hat{F},\ i = 1,2,...,n$ with replacement. The bootstrap sample is composed of the set $x^* = (x_1^*, x_2^*,..., x_n^*)$. m replicates of x^* is generated where m is reasonably large.
3. Approximate the sampling distribution of by the distribution of computed from the m replicates.

The bootstrap estimate of the mean and variance of the sampling distribution of R are:

$$R_B = \frac{\sum_{j=1}^{m} R_j^*}{m} \qquad V(R_B) = \frac{\sum_{j=1}^{m} \left(R_j^* - R_B\right)^2}{m}$$

Resampling is intended to smoothen the estimate by trimming off the bias that the nuisance of the sample selection might create. The method was illustrated to be applicable in variance estimation for statistics with complicated sampling distributions and for modeling problems like regression analysis.

Starting from the computationally attractive value of the bootstrap method, analytical properties were established, e.g., consistency provided by the Glivenko-Cantelli Theorem [Given $x_1, x_2, ..., x_n$ *iid* F let $F_n(x) = \frac{1}{n}\sum_{m=1}^{n} I_{(x_m \le x)}$. As $n \to \infty$ $\sup_x |F_n(x) - F(x)| \to 0$ a.s.]. Davison and Hinkley (1997: 31-38) observed that error in the bootstrap are classified into statistical error [the small difference between the true distribution F and the estimate $F_n(x)$] and the simulation error [since properties of statistics are approximated by empirical properties in simulation, influenced by factors like resample size, replication size, etc.]. Resampling size is an important factor that could influence the error in the bootstrap, for example, (Bickel et al., 1997) resampled with m<n observations and concluded that this approach can be expected to exhibit relative advantage on small samples.

The theoretical basis of the bootstrap has continued to be provided by over 1,000 papers since 1979 (Efron, 2000). While the bootstrap continued to define the landscape on the interplay between computing and statistical inference, there are also reminders that this is not the ultimate solution to all statistical problems. Beran (1997) recognized the viability of convergence of the bootstrap to the correct limiting distribution, but noted that convergence fails at superefficiency points in the parameter space. Furthermore, superefficiency is only a sufficient condition for bootstrap failure. Andrews (2000) further cautioned the bootstrap is not a universal solution to statistical inference problems, also provided counterexample illustrating that bootstrap is inconsistent when the parameter is on the boundary of the parameter space.

Efron also introduced the bootstrap in the context of model-based inference, where instead of the random sample, resampling is performed on the empirical distribution of the residuals. For the regression model $X_i = g_i(\beta) + \varepsilon_i$ $i=1,2,..., n$, $\varepsilon_i \sim iid\ F$ the parameter β is estimates by $\hat{\beta}$ (e.g., ordinary least squares). The sampling distribution is defined as: $\hat{F}: mass \frac{1}{n}$ at $\hat{\varepsilon}_i = x_i - g_i(\hat{\beta}),\ i=1,2,...,n$. From the pair $(\hat{\beta},\hat{F})$, the bootstrap sample is computed from the fitted model as $X_i^* = g_i(\hat{\beta}) + \varepsilon_i^*,\ \ \varepsilon_i^* \sim iid\ \ \hat{F}$. Each of the bootstrap samples can provide an estimate of β following the same estimation procedure used (e.g., ordinary least squares). From all the bootstrap replicates, we get $\hat{\beta}*^1, \hat{\beta}*^2,..., \hat{\beta}*^m$ and used to estimate the distribution of $\hat{\beta}*$. In model-based inference, Paparoditis and Politis (2005) underscored the importance of the choice of residuals. For example, to maximize power in bootstrap-based hypothesis testing, residuals are obtained using a sequence of parameter estimators that converge to the true parameter value both under the null and alternative hypothesis.

The bootstrap was initially introduced for independent cross-section data, but recently, it has been defined for time series data and other dependent observations as well. There are many theoretical justifications of time series bootstrap, example, Politis and Romano (1994) established convergence of certain sums of stationary time series that can facilitate bootstrap resampling. The block bootstrap was among the early proposal for time series data. While the method is very straightforward, there are associated problems like independence of block to maintain the dependence structure within the block. The size of the block is a crucial quantity that should be determined to assure success in block bootstrap. The AR-sieve was also introduced as a residual-based method similar to the model-based approach. Local bootstrap was also introduced but in the context of local regression framework (nonparametric) and to account for the nonparametric model, resampling allows the empirical distribution to vary locally in the time series. Bühlman (2002) compared different methods for time series bootstrap. The block bootstrap is recognized as the most general and simple generalization of the original independent resamples but is criticized for the possible artifacts it may exhibit when blocks are linked together. Blocking can potentially introduce some dependence structure in addition to those naturally existing in the data. The AR-sieve is less sensitive to selection of a model than the block length. The local bootstrap for nonparametric estimation is observed to yield slower rate of convergence. Generally, the AR-sieve is advantageous among the bootstrap approaches for time series data.

Recently, the bootstrap has been introduced to more complex situations and in more complicated models. In modeling nonstationary volatility, Xu (2008) used autoregression around a polynomial trend with stable autoregressive roots to illustrate how nonstationary volatility affects the consistency, convergence rates and asymptotic distributions of the estimators. Westerlund and Edgerton (2007) proposed a bootstrap test for the null hypothesis of cointegration in panel data. Dumanjug, et al. (2010)

developed a block bootstrap method in a spatial-temporal model. Chernick, et al. (2010) also provide a comprehensive summary of the development of the bootstrap method as it identifies a good range of literature on the subject matter.

REFERENCES

ANDREWS, D., 2000, Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space, *Econometrica*, 68(2): 399-405.

BERAN, R., 1997, Diagnosing Bootstrap Success, *Annals of the Institute of Statistical Mathematics*, 49(1): 1-24.

BICKEL, P., GÖTZE, F., AND VAN ZWET, W., 1997, Resampling Fewer Than Observations: Gains, Losses, and Remedies for Losses, *Statistica Sinica*, 7: 1-31.

BÜHLMAN, P., 1997, Sieve Bootstrap for Time Series, *Bernoulli*, 3(2): 123-148.

BÜHLMAN, P., 2002, Bootstrap for Time Series, S*tatistical Science*, 17(1): 52-72.

CHERNICK, M., GONZALEZ-MANTEIGA, W., CRUJEIRAS, R., AND BARRIOS, E., 2010, Bootstrap Methods, in M. Lovric, ed., *International Encyclopedia of Statistical Sciences*, N.J: Springer.

DAVISON, A. AND HINKLEY, D., 1997, *Bootstrap Methods and Their Applications*, Cambridge:Cambridge University Press.

DUMANJUG, C., BARRIOS, E., AND LANSANGAN, R., 2010, Bootstrap Procedures in a Spatial-Temporal Model, *Journal of Statistical Computing and Simulation*, 80(7):809-822.

EFRON, B., 1979, Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics*, 7(1): 1-26.

EFRON, B., 2000, The Bootstrap and Modern Statistics, *Journal of the American Statistical Association*, 95(452): 1293-1296.

PAPARODITIS, E., AND POLITIS, D., 2005, Bootstrap Hypothesis Testing in Regression Models, *Statistics and Probability Letters*, 74: 356-365.

POLITIS, D., AND ROMANO, J., 1994, Limit Theorems for Weakly Dependent Hilbert Space Valued Random Variables with Application to the Stationary Bootstrap, *Statistica Sinica*, 4: 461-476.

WESTERLUND, J., AND EDGERTON, D., 2007, A Panel Bootstrap Cointegration Test, *Economic Letters*, 97: 185-190.

XU, K., 2008, Bootstrapping Autoregression Under Nonstationary Volatility, *Econometrics Journal*, 11: 1-26.

Guidelines for Authors

The Philippine Statistician (TPS) is the official scientific journal of the Philippine Statistical Association. Inc. (PSA). It considers papers resulting from original research in statistics and its applications. Papers will be sent for review on the assumption that this has not been published elsewhere nor is submitted in another journal.

Aims and Scope

The Journal aims to provide a media for the dissemination of research by statisticians and researchers using statistical method in resolving their research problems. While a broad spectrum of topics will be entertained, those with original contribution to the statistical science or those that illustrates novel applications of statistics in solving real-life problems will be prioritized. The scope includes, but is not limited to the following topics:

- Official Statistics
- Simulation Studies
- Survey Sampling
- Time Series Analysis
- Nonparametric Methods
- Econometric Theory and Applications
- Other Applications
- Computational Statistics
- Mathematical Statistics
- Statistics Education
- Biostatistics
- Experimental Designs and Analysis

In addition to research articles, the Journal will have the following sections that may appear in some of its issues (but not necessarily in all):

- *Letters to the Editor.* This section will provide a forum for the airing of opinions on issues pertinent to the statistical community or offers commentaries on articles that have appeared in the journal.
- *Notes* section will include notices and announcements of upcoming events, conferences, calls for papers.
- *Review* section will present reviews on statistics books and software.
- *Teacher's Corner* shares experience of some teachers related to statistical education.

Articles submitted for the four special sections above will be reviewed only by the Editor and/or Associate Editors.

Submission of Manuscript

Only unpublished manuscripts will be considered. They will be refereed and evaluated on content, language and presentation. The article in MS word format, without author's identification should be sent as email attachment to Erniel B. Barrios, Editor, T*he Philippine Statistician*: ernielb@yahoo.com / barrios.erniel@gmail.com. A separate file containing the title of the paper, authors(s) (corresponding author identified) and their affiliations and complete address should be included in a separate file to be emailed along with the main article. To avoid delays and difficulties in submission, authors should follow instructions on style prescribed below.

A printed page in the Journal will have a maximum amount of space of 4.5" by 8.5."

Organization of Manuscript

The manuscript should be written in 8.5"x11" page using Times New Roman font size 12 with 1" margin in all sides.

Manuscripts must be organized in the following manner:

- ***Title***: This should be brief and concise.
- ***Abstract and Key Words***: An abstract of at most 250 words must be submitted with the manuscript. It precedes the article text. The abstract should summarize objectives, results, and main conclusions, but it should not contain any graph or complex mathematical notation and no references. Three to six keywords should be identified.
- ***Article Text***: Sections should be concise and numbered in a decimal system. Tables, Figures, and Artwork may be used within the body of the article, should be numbered consecutively. Tables, Figure Titles and Legends, Figure Artwork must be strategically placed in the article text. The original files for the tables, figures, and artwork should also be submitted to facilitate typesetting. Authors must obtain written permission to reproduce or adapt all or part of a figure from a copyrighted source. Mathematical equations cited in the text should be numbered consecutively. Numbers should be placed at the rightmost margin of the equation line in parenthesis. Matrices should appear in bold and vectors in italics. All other symbols should appear in italics. The preferred software for equations are Mathtype and MS Equation Editor, which are add-ins of MS Word.
- ***Acknowledgments***: An acknowledgment section may be included at the end of the article. This section should acknowledge financial assistance in the form of grants or university funding, assistance by individual colleagues, and any other pertinent information. This section will be inserted by the author only upon acceptance of the paper.
- ***References***: All references included in the list at the end of an article must be cited in the text. References are cited in the text in the following format: (Author, Year). Up to two authors can be cited in the text. If there are three or more authors, only the first author will be cited in the text, e.g. (Author et al, Year). The following format will be followed in the listing references:

Journal Article

LANDAGAN, O. and BARRIOS, E., 2007, An Estimation Procedure for a Spatial-Temporal Model, *Statistics and Probability Letter,* 77(4):401-406.

Book

KOTTAK, C., 1991, *Anthropology: The Exploration of Human Diversity*, 5th ed., New York: McGraw-Hill, Inc.

Book Chapter

FINK, E. and PRATT, K., 2008, Indexing Compressed Time Series, in Last, M., Kandel, A. and Bunke, H., eds., *Data Mining in Time Series Databases*, Singapore: World Scientific, pp. 43-66.

Internet Document

MUNDLAK, Y., LARSON, D. and BUTZER, R., 2002, Determinants of Agricultural Growth in Indonesia, the Philippines, and Thailand, V. 1, *World Bank Working Paper 2803*, The World Bank. Available at: http://econ.worldbank.org/external/default/main?pagePK=64165259&piPK=64165421&theSitePK=469372&menuPK=64216926&entityID=000094946_02032604542948

For particulars about the style, please download the Philippine Statistician Style Guide at www.philstat.org.ph.

- ***Appendices*:** A single appendix is headed, "APPENDIX: FOLLOWED BY A DESCRIPTIVE TITLE". If there are two or more appendices, they should be labeled, "APPENDIX A," "APPENDIX B," and so on.
- ***Editorial Style***: In addition to content, manuscripts are evaluated on their conciseness and clarity. Thus, the Journal gives premium to well-written and well-structured papers that will be of interest to a wide segment of the readership. Manuscripts and reviews that have been accepted for publication will be copy edited in accordance to accepted rules of correct grammar, usage, spelling, and punctuation. To avoid common problems of style, for guidelines on style, usage, and the preparation of technical manuscripts for publication, the following reference may be consulted:

  The Chicago Manual of Style (14th ed.) (1993),
  Chicago: University of Chicago Press.

Editorial Notes

Use quotation marks only when a standard term is used in a nonstandard way and to indicate the beginning and ending of a direct quotation.

1. Hyphens are used when two or more adjectives or an adjective and a noun together modify another noun; for example, *goodness-of-fit test* is the equivalent of *test for goodness of fit*. Most words with prefixes such as sub and non are not hyphenated, for example, *subtable*, *nonnormal.*

2. Italics are used to introduce important terms, when appropriate; they are to be used sparingly to indicate emphasis.
3. Abbreviations and acronyms should be minimized; those that are used are spelled out on their first appearances in the manuscript with the shortened form given in parentheses, for example, *best linear unbiased estimate (BLUE).*
4. Numbers under 10 are spelled out when they are not part of an equation or an expression containing symbols.
5. The sign % is always used when giving a specific percentage, for example, 23%, not 23 percent. Otherwise use the word *percent.*

Copyright Transfer Form

Authors of accepted papers will be required to submit an author copyright transfer form before the final release of the journal.

PHILIPPINE POSTAL CORPORATION
NATIONAL CAPITAL REGION
Liwasang Bonifacio
1000 Manila

Permit No.
Date Granted

SWORN STATEMENT

The undersigned ERNIEL B. BARRIOS editor / managing editor / business manager / owner publisher of EDITOR (title of publication) THE PHILIPPINE STATISTICIAN frequency of issue ANNUALLY language ENGLISH in which printed at PSA OFFICE office of publication ________ entered as THIRD CLASS (first, second and /or third class) office of entry UP POST OFFICE after having duly sworn to in accordance with law, hereby submit the following statement of ownership, management, circulation, etc, which is required by act 2580 as Amended by Commonwealth Act 201

ADDRESS

Editor	SCHOOL OF STATISTICS, UNIVERSITY OF THE PHILIPPINES OLD SCWCD BLDG., MAGSAYSAY AVE., DILIMAN
Associate Editors	LISA GRACE S. BERSALES, UP DILIMAN; ZITA VJ ALBACEA, UP LOS BAÑOS
Business Manager	________
Owner	PHILIPPINE STATISTICAL ASSOCIATION, INC
Publisher	PHILIPPINE STATISTICAL ASSOCIATION, INC
Office of Publication	RM 214 PSSC BLDG., COMMONWEALTH AVE., DILIMAN, QUEZON CITY

If publication is owned by a corporation, stockholders owning one percent or more of the amount of stocks ________ not applicable ________

Bond holders, mortgager or other security holders owning one percent or more of the total amount of Security ________ not applicable ________

In case of daily publication average number of copies printed and circulated of each issue during the preceding month of ________ not applicable ________

1. Sent to paid subscribers ________
2. Sent to others than paid subscribers ________
Total --

In case of publication other than daily total number of copies printed and circulated at the last issue date ________ 2011 ________

1. Sent to paid subscribers 400
2. Sent to others than paid subscribers 200
Total -- 600

For the Editor:

TERESITA G. ABAD, Managing Director
Designation and Signature

MAR 31 2011

SUBSCRIBED AND SWORN TO, before me this ________ day of ________ 200__ in The City of MANILA, Philippines. Affiant exhibiting to me his/her Community Certificate No. CCI2010 15828460 issued at Quezon City on January 6, 2011.

____
Person Administer Oath

ATTY. RONALD SEGUNDINO C. CHIN
NOTARY PUBLIC-CITY OF MANILA
ADMIN. NO. 2011-099-UNTIL DEC. 31, 2011
ROLL NO. 54895
IBP NO. 805532/11-15-2010 MANILA
PTR NO. MLA. 0269771/12-29-2010 MLA
MCLE Compliance No. III-0016300

DOC. NO. 97
PAGE NO. 16
BOOK NO. XXIII
SERIES OF 2011

THE BOARD OF DIRECTORS and OFFICERS for 2011

Members of the Board

Individual Members

Zita V.J. Albacea	Estela T. De Guzman	Imelda E. De Mesa
Chona P. Guatlo	Dennis S. Mapa	Romeo S. Recide
Gervacio G. Selda	Vicente B. Valdepeñas, Jr.	

Institutional Members

ABS-CBN Broadcasting Corporation	Represented by	Ruthie A. Floresta
Bangko Sentral ng Pilipinas		Rosabel B. Guerrero
Globe Telecom		Raul M. Macatangay
Government Service Insurance System		Maria Fe S. Dayco
National Statistical Coordination Board		Lina V. Castro
Social Weather Stations		Gerardo A. Sandoval

PSA Regional Chapters	Represented by	Vicente T. Balinas *(PSA Region 6)*

Executive Committee and Officers

President:	Vicente B. Valdepeñas, Jr.
Vice President:	Estela T. De Guzman
Secretary:	Lina V. Castro
Treasurer:	Imelda E. De Mesa
Ex officio Member (Immediate Past President):	Nieves L. Osorio

Working Committees in 2011

Annual Conference
Chair: Romeo S. Recide

Annual Meeting and Christmas Party
Chair: Gerardo A. Sandoval
Co-Chair: Ruthie A. Floresta

Finance
Chair: Imelda E. De Mesa

Institutional Development
Chair: Estela T. De Guzman

Membership
Chair: Cynthia C. Vallesteros
Co-Chair: Chona P. Guatlo

Nomination and Election
Chair: Ludivinia D. Gador *(Sept - Dec)*
Chair: Angeles R. Buenaventura† *(Jan - Aug)*

Publications and Information
Chair: Lina V. Castro

Search and Awards
Chair: Lisa Grace S. Bersales
Co-Chair: Rosabel B. Guerrero

Statistics Education
Chair: Gervacio G. Selda, Jr.
Co-Chair: Zita V.J. Albacea

Training
Chair: Dennis S. Mapa

PSA Secretariat

Managing Director: *Teresita G. Abad*

Secretariat Assistants: *Stephanie Rae C. Andres*
Reynaldo B. Belluga, Jr.