Forecasting Demand for Rural Electric Cooperative Call Center

Taeyoon Kim
Research Assistant
Department of Agricultural Economics
Oklahoma State University
Stillwater, OK 74078
taeyoon.kim@osktate.edu

Philip Kenkel
Professor, and Bill Fitzwater Cooperative Chair
Department of Agricultural Economics
Oklahoma State University
516 Agricultural Hall
Stillwater, OK 74078
phil.kenkel@okstate.edu

B. Wade Brorsen
Regents Professor and Jean & Patsy Neustadt Chair
Department of Agricultural Economics
Oklahoma State University
414 Agricultural Hall
Stillwater, OK 74078
wade.brorsen@okstate.edu

Selected Paper prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, Atlanta, Geogia, January 31 – February 3, 2009.

Forecasting Demand for Rural Electric Cooperative Call Center

Abstact

This research forecasts peak call volume to allow a centralized call center to minimize staffing costs. A Gaussian copula is used to capture the dependence among nonnormal distributions. Peak call volume can be easily and more accurately predicted using the marginal probability distribution with the copula function than without using a copula. The modeling approach allows simulating adding another cooperative. Ignoring the dependence that the copula includes, causes peak values to be underestimated.

Introduction

KAMO, an Oklahoma based rural electric cooperative (REC) established an after hour call center operation for 7 member cooperatives in 2006. The center provided significant cost advantage with some member cooperatives saving over $150,000 annually. By 2008 the center had expanded to 18 RECs and KAMO was considering expansion into other states. KAMO contacted Oklahoma State University for assistance in investigating the feasibility of the call center expansion.

The call centers' goal is to answer almost all calls and so it has staff to handle peak calls. Also, call volume is affected by number of customers, season, geographic location and individual REC characteristics such as line maintenance. One of the major challenges is to accurately forecast peak and average call volume which determine staffing and equipment needs for the design of the fee structure. In forecasting peak call volume, which is mainly related to the events of ice storms and other disasters, extreme value theory can be applied (Haan and Ferreira, 2006). Modeling total call volume of 18 RECs also raises issues of correlation and dependency. Copula functions, which represent a multivariate joint distribution in the form of dependency structure, should be considered (Cherubini, Luciano, and Vecchiato, 2004). Another challenge presented in the REC call center is the need to forecast the extreme call volume. Peak call volume dictates personnel needs and is a key issue in business expansion. A cooperative call center would be expected to generate efficiencies in managing peak call volume because severe weather events which generate the majority of power outage calls do not occur simultaneously across a wide geographic region. Adding additional RECs into KAMO's call center could conceivably reduce peak call volume per member.

The objective of this research is to model call volume for a centralized REC call center and forecast the impact of adding the additional RECs. One contribution of this research over previous research is the use of copula functions in modeling call center volume. These methods were not used in previous research on various industries such as telephone (Avramidis, Deslauriers and L'Ecuyer, 2004), banking (Taylor, 2008), postal service (Xu, 2000) and other marketing company (Andrews and Cunningham, 1995). The study also varies from past efforts by using several continuous distributions rather than discrete distributions. Previous call center data research has modeled call volume with Poission distribution or modifications of the Poisson such as negative binomial, Hurdle Poisson and zero-inflated Poisson (Liu and Cela, 2008). Finally, the study illustrates the use of empirical distribution for predicting extreme values. This technique, which also comes from extreme value theory is appropriate for modeling hourly call volume because the hourly call data has a high degree of skewness.

Theory

The objective for operating call center is to minimize its costs which are subject to satisfactory customer service. Staffing which typically depends on the number of calls received from customers is a main part of total costs for call center so that forecasting peak and average call volume accurately is essential in order to minimize total costs. Besides staffing strategies suggested by Atlason, Epelman, and Henderson (2008), the focus in this paper is on precise forecasting which depends on multivariate joint distributions considering random variables' correlation. Within agricultural economics, there were several research on multivariate nonnormal probability distributions. Taylor (1990) provided two procedures for empirically estimating correlated nonnormal joint probability density function (p.d.f.), and Richardson, Klose, and Gray (2000) also showed how to simulate a multivariate empirical distribution using correlated error terms. These past efforts are heuristic to approach copula functions.

A copula function which allows specific dependency structure between multivariate joint distibutions is used in our model because it represents the complex relationship of calls from individual REC customers entering a centralized call center. In addition, graphical explanation is made at the end of copula concept. Extreme value theory is also incorporated because it helps forecast extreme values created by short term peaks in call volume.

A two-dimensional copula $C(u,v)$ is defined as

(1) $$C:\ [0,1]^2 \rightarrow [0,1].$$

As shown in equation (1), a two-dimensional copula has a two-dimensional domain ranged from 0 and 1 and a one-dimensional codomain ranged from 0 and 1.

A copula has the following properties:

$$C(u,0) = C(0,v) = 0 \text{ for all } u,v \in [0,1]$$

(2) $$C(u,1) = u \text{ and } C(1,v) = v \text{ for all } u,v \in [0,1]$$

$$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0 \text{ for all } u_1 \leq u_2,\ v_1 \leq v_2.$$

Equation one implies that if one of the random variables is 0 then the copula is 0. The second equation implies that if one of random variables is 1 then the copula has the same value as the other variable. The last equation indicates that copula is a non-decreasing function.

According to the Sklar's Theorem, any joint distribution $H(x,y)$ with cumulative density functions (c.d.f.) of $F(x)$ and $G(y)$ can be expressed as

(3) $$H(x,y) = C\big(F(x),G(y)\big),$$

where $C(\cdot,\cdot)$ is a uniquely determined copula function.

Equation (3) implies that the copula function represents a bivariate joint distribution. If the distribution functions ($F(x), G(y)$) and the copula ($C(\cdot,\cdot)$) are continuous, then equation (3) can be restated in terms of the p.d.f. as

(4) $$h(x, y) = c\big(F(x), G(y)\big)\cdot f(x)\cdot g(y) ,$$

where $h(x, y) = \partial^2 H(x, y)/\partial x \partial y$, $f(x) = \partial F(x)/\partial x$, $g(x) = \partial G(y)/\partial y$, and $c\big(F(x), G(y)\big) = \partial^2 C\big(F(x), G(y)\big)/\partial F(x)\partial G(y)$ is the copula's density.

Equation (4) shows that a marginal distribution ($h(x, y)$) consists of a second-differentiated copula ($c(\cdot,\cdot)$) which has parameters to indicate the dependency structure between two variables (x, y) and two probability density functions. Then, canonical representation (Cherubini, Luciano, and Vecchiato, 2004, p.154) for the n dimensions can be expressed as

(5) $$h(x_1, x_2, \ldots, x_n) = c\big(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)\big)\cdot \prod_{i=1}^{n} f_i(x_i) ,$$

where $c\big(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)\big) = \partial^n C\big(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)\big)/\partial F_1(x_1)\partial F_2(x_2)\ldots\partial F_n(x_n)$.

Let $\boldsymbol{X} = \{x_{1t}, x_{2t}, \ldots, x_{nt}\}_{t=1}^{T}$ be the sample data matrix, and then the log-likelihood function can be expressed as

(6) $$l(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln c\big(F_1(x_{1t}), F_2(x_{2t}), \ldots, F_n(x_{nt})\big) + \sum_{t=1}^{T}\sum_{i=1}^{n} \ln f_i(x_{it}) ,$$

where $\boldsymbol{\theta}$ is the set of all parameters of both the marginals and the copula.

Thus, maximum likelihood estimation (MLE) is possible using equation (6). Let correlation matrix, $\boldsymbol{R}$, be a symmetric and positive definite matrix with $\mathrm{diag}(\boldsymbol{R}) = (1, 1, \ldots, 1)'$ then Gaussian copula can be defined as

(7) $$c(u_1, u_2, \ldots, u_n) = \frac{1}{|\boldsymbol{R}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \varsigma' (\boldsymbol{R}^{-1} - \boldsymbol{I}) \varsigma \right),$$

where $u_n = F_n(x_{nt})$, $\boldsymbol{R}$ is $n \times n$ correlation matrix, $\varsigma = \left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_3)\right)'$, and Φ is the standard normal c.d.f.

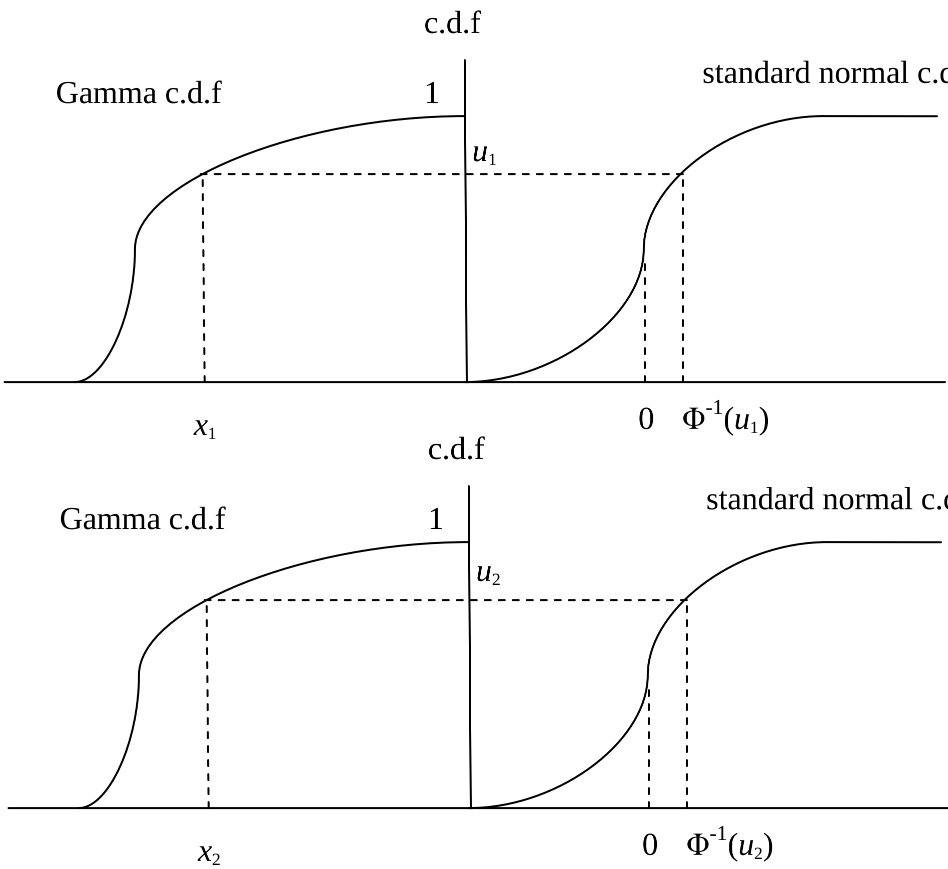

Figure 1. Graphical depiction of the transformation of gamma random variables into standard normal random variables

Figure 1 illustrates the transformation of gamma random variables into standard normal random variables. Let's assume that random variables x_1 and x_2 follow a Gamma distibution. Then, u_1 and u_2 are values for the gamma c.d.f. of x_1 and x_2, respectively, and both range from 0 to 1. Also,

$\Phi^{-1}(u_1)$ and $\Phi^{-1}(u_2)$ are inverse of standard normal c.d.f. of u_1 and u_2 , respectively, and both are the transformed data of gamma random variables (x_1 and x_2) into a standard normal random variables ($\Phi^{-1}(u_1)$ and $\Phi^{-1}(u_2)$) using Gaussian copula. The correlation between $\Phi^{-1}(u_1)$ and $\Phi^{-1}(u_2)$ is a copula parameter.

Regarding MLE in equation (6), it could be computationally difficult to estimate jointly the parameters of the marginal distributions and the parameters of the copula in the case of a high dimension so that Joe and Xu (1996) proposed two steps to estimate $\boldsymbol{\theta}$, which provides a good starting point for obtaining an MLE estimator and is called the inference for margins (IFM). As a first step, the margins' parameters $\boldsymbol{\theta}_1$ are estimated by usual MLE such as:

$$(8) \qquad \hat{\boldsymbol{\theta}}_1 = \operatorname{ArgMax}_{\theta_1} \sum_{t=1}^{T} \sum_{i=1}^{n} \ln f_i(x_{it}; \boldsymbol{\theta}_1) .$$

As a second step, the copula parameters $\boldsymbol{\theta}_2$ are estimated given $\hat{\boldsymbol{\theta}}_1$:

$$(9) \qquad \hat{\boldsymbol{\theta}}_2 = \operatorname{ArgMax}_{\theta_2} \sum_{t=1}^{T} \ln c\left(F_1(x_{1t}), F_2(x_{2t}), \ldots, F_n(x_{nt}); \boldsymbol{\theta}_2, \hat{\boldsymbol{\theta}}_1\right).$$

In terms of efficiency for the IFM estimator, Joe (2005) points out that the IFM has efficiency loss with strong dependence, and suggests to estimate equations with a combination of univariate and bivariate log-likelihoods or with low-dimensional margins.

In terms of fitting marginal distributions, extreme value theory can be applied if modeling the extreme values is the chief concern. Typically, the generalized Pareto distribution (GPD) resulted from the Fisher-Tippett theorem in the generalized extreme value distribution (GEV) is considered (McNeil, 1997) but the empirical distribution is also another possibility (Ghoudi and Rives, 1995). This empirical distribution is used here because of high skewness in call volume per hour. More specifically, the empirical distribution is estimated as the marginal distributions and then copula parameters are estimated by MLE like this:

$$(10) \qquad \hat{\boldsymbol{\theta}}_2 = \operatorname{ArgMax}_{\theta_2} \sum_{t=1}^{T} \ln c\left(\hat{F}_1(x_{1t}), \hat{F}_2(x_{2t}), \ldots, \hat{F}_n(x_{nt}); \boldsymbol{\theta}_2\right),$$

where $\hat{F}_i(x_{it})$ is the cumulative distribution calculated from the estimated empirical distribution. This is called the canonical maximum likelihood (CML) (Cherubini, Luciano, and Vecchiato, 2004, p.160).

Simulation can be used to look at the effects from adding an additional REC as well as to predict peak call volume more precisely. Here are procedures using Gaussian copula (Cherubini, Luciano, and Vecchiato, 2004, p.181):

- Find the Cholesky decomposition $\boldsymbol{A}$ of $\boldsymbol{R}$
- Simulate n independent random variates $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)'$ from $N(0, 1)$
- Set $\boldsymbol{x} = \boldsymbol{Az}$
- Set $u_i = \Phi(x_i)$ with $i = 1, 2, \ldots, n$, where Φ denotes the univariate standard normal c.d.f.
- $(u_1, \ldots, u_n)' = \left(F_1(y_1), \ldots, F_n(y_n)\right)'$, where $F_i(\cdot)$ is the i^{th} margin
- $F_i^{-1}(u_i) = y_i$, where y_i is i^{th} simulated data from Gaussian copula and its margin

Data and Procedures

Monthly data from January 2006 to June 2008 and hourly data from Apr. 18, 2008 to Jun. 30, 2008 were obtained from KAMO's call center. Call volume data from 14 RECs which had data for the entire study period were used. In order to capture regional effects which are a main concern for KAMO, call volume data was aggregated over similar regions such as eastern Oklahoma with 7 RECs (Eastern), western Oklahoma with 4 RECs (Western), and south western Missouri with 3 RECs (Missouri). This grouping simplified the use of copula functions because it reduces joint variables into 3 from 14. Because each REC has a different member of members (meters) call volume per member rather than total call volume was used. This removed the differences in call volume only to the REC size.

Table 1 and table 2 shows summary statistics of data. There is no truncation below zero in case of call volume per month because each month has at least 360 calls from Western, 766 calls from Missouri, and 2,757 calls from Eastern while there is a truncation below zero in case of call volume per hour.

The @Risk add in program in Microsoft Excel was used to select the best fitting distibution for the marginal distributions of monthly call volume data. The Gamma distibution was shown to have the best fit among several continuous distibutions. Therefore, Gamma distributions and Gaussian copula with IFM method are used.

In terms of marginal distributions for hourly call volume, several distributions like Parato distribution, Gamma distibution, and etc. are tested and they were rejected. Thus, empirical distributions and Gaussian copula with CML method are selected.

Consider the following Gamma p.d.f. with explanatory variables:

$$(11) \qquad f(y_t \mid \boldsymbol{x}_t) = \begin{cases} \dfrac{1}{\Gamma(\alpha_t)\,\beta^{\alpha_t}}\, y_i^{\alpha_t - 1} \exp(-\dfrac{y_t}{\beta}), & 0 < y_t < \infty \\ 0, & \text{elsewhere,} \end{cases}$$

where $\alpha_t = \boldsymbol{x}_t'\boldsymbol{\delta}$, $\Gamma(\alpha_t) = \int_0^\infty y_t^{\alpha_t - 1} \exp(-y_t)\ dy_t$, y_t is call volume per month for the t^{th} observation, α_t is the shape parameter which is determined by $\boldsymbol{x}_t'\boldsymbol{\delta}$, $\boldsymbol{x}_t$ is a vector of explanatory variables, here dummy variables for season, $\boldsymbol{\delta}$ is a vector of unknown parameters to be estimated, and β is the scale parameter to be estimated.

Gaussian copula with 3 random variables can be defined as

$$(12) \qquad c(u_1, u_2, u_3) = \frac{1}{|\boldsymbol{R}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\boldsymbol{\varsigma}'(\boldsymbol{R}^{-1} - \boldsymbol{I})\boldsymbol{\varsigma}\right),$$

where $\boldsymbol{R}$ is 3×3 correlation matrix, $\varsigma = \left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \Phi^{-1}(u_3)\right)'$, and Φ is the standard normal c.d.f.

Since there are 3 series of sample data, the log-likelihood function can be expressed as

$$(13) \qquad l(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln c\left(u_{1t}, u_{2t}, u_{3t}\right)\right) + \sum_{t=1}^{T} \sum_{i=1}^{3} \ln f_i(x_{it}) .$$

Thus, equation (13) enables to use MLE but it is computationally difficult to estimate jointly the parameters of the marginal distributions and the parameters of the copula so that IFM method is used here, i.e., the parameters for gamma distribution such as α, β, and $\boldsymbol{\delta}$ are estimated via MLE using the log-likelihood function from equation (11) and then copula parameters in the $\boldsymbol{R}$ matrix given $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\boldsymbol{\delta}}$ are estimated with MLE using the log-likelihood function from equation (12).

Consider the following empirical distribution.

$$(14) \qquad f(y_t) = \begin{cases} \dfrac{1}{T}, & t = 1, \ldots, T \\ 0, & \text{elsewhere} \end{cases} ,$$

where T is the number of observations.

In regard to the CML method for houly call volume data, the empirical distribution is easily calculated assuming equal probability within observations in equation (14) and then copula parameters are estimated via MLE from equation (15).

$$(15) \qquad \hat{\boldsymbol{\theta}}_2 = \operatorname{ArgMax}_{\theta_2} \sum_{t=1}^{T} \ln c\left(\hat{u}_{1t}, \hat{u}_{2t}, \hat{u}_{3t}; \boldsymbol{\theta}_2\right).$$

In terms of simulation for call volume, procedures from Cherubini, Luciano, and Vecchiato (2004, p 181) are used to simulate the 99th percentile and effects from adding an additional RECs on total number of call volume to the call center. For instance, let’s assume that an additional

REC with 100,000 members in the Western area is added. Then, the Gamma distribution for the added REC is assumed to have the same parameters as the Gamma distribution for existing RECs in the Western region. Copula parameters for added REC are assumed to be as same as the existing Western and the correlation between added REC and existing Western is assumed to be 1.

Results

As mentioned in data and procedure, all call volume data are transformed as per meter basis. Total meters are 262,552. The Eastern region had the most meters and the Missouri had the fewest meters. Descriptive statistics for monthly and hourly call volume are shown in Tables 1 and 2 respectively. The houly call volume had the highest degree of positive skewness reflecting the fact that the peak hourly call volume was far above the mean value.

Figure 2 shows the average monthly call volume from Jan. 2006 to Jun. 2008 for the three regions. Overall, June is the highest month and October is the lowest. Figure 3 also indicates that summer is the highest season and fall is the lowest.

Table 3 and table 4 provide copula parameters and average call volume per month with IFM method from equations (11) and (12). The dependence between the Eastern and Western region is the highest while the dependence between the Missouri and Western region is the lowest. However the difference in dependence is not large. In terms of average monthly calls per meter, summer is higher than the other seasons in all three regions and the Eastern region is higher than other regions.

Table 5 and figure 4 indicate the simulation results for the 99^{th} percentile of total monthly call volume before and after adding an additional REC with 100,000 meters. Table 5 and figure 4 use the results of parameters for the Gamma distribution and Gaussian copula in table 3 and 4. Overall, the summer season has the highest 99^{th} percentile while fall has the lowest. The total numbers of calls from adding an additional REC are shown in parentheses in table 5. Not surprisingly, total calls increase with the increase in total meters for all seasons. But on a call volume per meter basis, adding an additional REC from the Western or Missouri region results in a decline in the 99th percentile while adding an additional REC from the Eastern region results in a slight increase in the 99th percentile. If the call center expands their business by adding RECs in the Western and Missouri region, the 99th percentile per member call volume for each existing REC member would be expected to decline. This would imply a decrease in staffing costs per member since the call center staffing is designed to meet the 99th percentile call volume. If the call center expands their business by adding RECs in the Eastern region, the 99th percentile would be expected to increase so that its staffing costs per member would increase. Thus, the advantage of business expansion depends on the regions where call center are expanded.

It is interesting to look at the differences between the simulated 99th percentile of monthly total call volume calculated with copula and without the copula. This comparison is shown in table 6 and figure 5. The simulated 99th percentile with the copula is larger than the 99th percentile simulated without using the copula approach because all copula correlations are positive. Both simulated 99th percentiles are not perfectly consistent with the 99th percentile observed in the data because the copula correlations were estimated using yearly call volume per meter data rather than separated data according to the season.

Table 7 and table 8 provide copula parameters and average call volume per hour using the CML method from equation (14) and equation (15), respectively. Data for each seasonal time period were separated and then estimated for the parameters of the empirical distribution and Gaussian copula. Overall, copula correlations for the day time period from 9:00 to 16:00 (military time format) are consistently higher than other time periods. The highest hourly average call volume occurred during the 7:00-9:00 and 17:00-19:00 time periods. These results are also shown in figure 6.

Figure 7 shows that the simulated 99th percentile of average call volume per hour from the empirical distribution and Gaussian copula is similar to the maximum value while far above the average. This reflects high degree of positive skewness in the hourly call volume data.

Table 9 and figure 8 indicate the simulation results for the 99th percentile of total houly call volume before and after adding an additional REC with 100,000 meters. The simulation results use the parameters for the empirical distribution and Gaussian copula reflected in tables 7 and 8. Overall, adding an additional REC from Western or Missouri region causes the 99th percentile to decrease. There are some exceptions in the Missouri region. Adding an additional REC from Eastern results in an increase in the 99th percentile. These results are similar to the results for the monthly call volume.

Table 10 and figure 9 show that the simulated 99th percentile of total houly call volume created with the copula is slightly higher than the 99th percentile without copula. Again, these results are similar to results for the monthly call volume. However, the simulated 99th percentile calculated with the copula is closer to the 99th percentile of the observed data. This result is different from the results for monthly data. The improved correspondence with the observed data was observed because the copula correlations were estimated separately for each time period.

Conclusions

This research focused on forecasting peak call volume to allow a centralized call center to minimize staffing costs. A copula was estimated to capture the correlation among nonnormal distributions. The Gamma distribution was found to provide the best for monthly data. The empirical distribution rather than generalized Pareto distibution was selected to present the houly data based on the ability to represent the 99th percentile in the observed data which had a high degree of positive skewness.

Estimating average and peak call volume, simulating data to forecast the 99th percentile, and examining the effects of adding additional cooperative are all important questions for call center managers. These estimates can be easily and more accurately performed using the marginal probability distribution with the copula function. The impact of adding an additional cooperative on the peak (99th percentile) per meter call volume affects the cost structure of call center. The results indicated that the impact of an additional call center customer depended on the regional location of the cooperative. Another important result demonstrated by the call center example is that when positive dependence among data series exists, ignoring their dependence can cause their peak values to be underestimated.

Table 1. Summary Statistics of Mothly Call Volume per Meter from January 2006 to June 2008

Statistics	Total	Missouri	Eastern	Western
Meters	262,552	35,552	158,747	68,253
Mean	0.04064	0.04040	0.04756	0.02452
	(10,670)	(1,436)	(7,549)	(1,674)
Min	0.01503	0.02154	0.01737	0.00527
	(3,945)	(766)	(2,757)	(360)
Max	0.09468	0.09280	0.09888	0.08585
	(24,858)	(3,299)	(15,697)	(5,859)
S D	0.01626	0.01540	0.01832	0.01637
Skewness	1.44382	1.50094	0.87316	2.29085

Note: The number of calls is reported in parentheses.

Table 2. Summary Statistics of Hourly Call Volume per Meter from April 18 to June 30, 2008

Statistics	Total	Missouri	Eastern	Western
Meters	262,552	35,552	158,747	68,253
Mean	0.000104	0.000095	0.000113	0.000089
	(27)	(3)	(18)	(6)
Min	0	0	0	0
	(0)	(0)	(0)	(0)
Max	0.002331	0.006385	0.003691	0.002842
	(612)	(227)	(586)	(194)
S D	0.00022	0.00034	0.00029	0.00024
Skewness	4.77100	9.99587	6.02059	5.78903

Note: The number of calls is reported in parentheses.

Table 3. Copula Correlation Matrix with Call Volume per Month

Areas	Missouri	Eastern	Western
Missouri	1	0.6903	0.5646
Eastern	0.6903	1	0.7310
Western	0.5646	0.7310	1

Table 4. Monthly Average Call Volume for Each Season with Gamma Distribution

Season	Missouri	Eastern	Western
Spring	0.0416	0.0485	0.0278
Summer	0.0526	0.0627	0.0279
Fall	0.0326	0.0368	0.0195
Winter	0.0342	0.0413	0.0217

Table 5. The Simulated 99th Percentile of Total Call Volume per Month Before and After Adding Additional REC of Each Region with 100,000 meters

Season	Before Adding additional Cooperative	After Adding additional Cooperative from Each Region		
		Missouri	Eastern	Western
Spring	0.0791	0.0755	0.0814	0.0738
	(20,759)	(27,360)	(29,504)	(26,760)
Summer	0.0925	0.0890	0.0965	0.0831
	(24,284)	(32,284)	(34,978)	(30,112)
Fall	0.0645	0.0618	0.0664	0.0597
	(16,923)	(22,388)	(24,081)	(30,112)
Winter	0.0694	0.0659	0.0717	0.0642
	(18,225)	(23,882)	(26,012)	(21,659)

Note: Total calls are reported in parentheses.

Table 6. The Simulated 99th Percentile of Total Call Volume per Month w/ and w/o copula

Season	99th Percentile from Data	w/ Copula	w/o Copula
Spring	0.0720	0.0791	0.0675
Summer	0.0947	0.0925	0.0801
Fall	0.0373	0.0645	0.0540
Winter	0.0603	0.0694	0.0586

Table 7. Copula Correlations with Hourly Data

Time Interval	Missouri-Eastern	Missouri-Western	Eastern-Western
0:00-1:00	0.13178	0.26914	0.49720
1:00-2:00	0.37434	0.34901	0.42042
2:00-3:00	0.49103	0.18939	0.36169
3:00-4:00	0.34563	0.17957	0.35679
4:00-5:00	0.15410	0.11435	0.41036
5:00-6:00	0.18165	0.25380	0.35449
6:00-7:00	0.39200	0.22285	0.20904
7:00-8:00	0.45009	0.05212	0.36747
8:00-9:00	0.17990	0.01533	0.66560
9:00-10:00	0.79666	0.61793	0.58243
10:00-11:00	0.67941	0.35134	0.64749
11:00-12:00	0.69031	0.67080	0.64922
12:00-13:00	0.84017	0.75538	0.64731
13:00-14:00	0.50869	0.30617	0.47935
14:00-15:00	0.39858	0.28631	0.41241
15:00-16:00	0.43136	0.47946	0.53843
16:00-17:00	0.10890	-0.31448	0.14443
17:00-18:00	0.42807	0.40596	0.40917
18:00-19:00	0.26342	0.22551	0.23891
19:00-20:00	0.14820	0.05896	0.12693
20:00-21:00	0.18341	0.21110	0.36966
21:00-22:00	0.21152	0.31290	0.26793
22:00-23:00	0.34803	0.41538	0.28892
23:00-24:00	0.42241	0.12696	0.25707

Note: Copula parameters of diagonal are 1's and copula parameters of off-diagonal are symmetric. The format containing copula parameters is actually as same as Table 3 for a certain time interval.

Table 8. Average Call Volume per Hour from Empirical Distribution and Gaussian Copula

Hour	Missouri	Eastern	Western
0:00-1:00	0.000136	0.000084	0.000083
1:00-2:00	0.000109	0.000073	0.000070
2:00-3:00	0.000063	0.000082	0.000091
3:00-4:00	0.000067	0.000079	0.000070
4:00-5:00	0.000060	0.000053	0.000035
5:00-6:00	0.000075	0.000090	0.000041
6:00-7:00	0.000192	0.000158	0.000097
7:00-8:00	0.000207	0.000179	0.000137
8:00-9:00	0.000054	0.000070	0.000049
9:00-10:00	0.000035	0.000098	0.000076
10:00-11:00	0.000034	0.000043	0.000071
11:00-12:00	0.000032	0.000103	0.000076
12:00-13:00	0.000060	0.000095	0.000047
13:00-14:00	0.000058	0.000072	0.000033
14:00-15:00	0.000128	0.000062	0.000031
15:00-16:00	0.000094	0.000074	0.000050
16:00-17:00	0.000219	0.000163	0.000072
17:00-18:00	0.000136	0.000258	0.000205
18:00-19:00	0.000082	0.000212	0.000150
19:00-20:00	0.000077	0.000158	0.000131
20:00-21:00	0.000100	0.000145	0.000111
21:00-22:00	0.000078	0.000102	0.000124
22:00-23:00	0.000056	0.000093	0.000130
23:00-24:00	0.000137	0.000082	0.000148

Table 9. The Simulated 99th Percentile of Total Call Volume per Month Before and After Adding Additional REC of Each Region with 100,000 meters

Hour	Before Adding additional Cooperative	After Adding additional Cooperative from Each Region		
		Missouri	Eastern	Western
0:00-1:00	0.001034	0.002390	0.001161	0.000881
1:00-2:00	0.000773	0.001654	0.000746	0.000642
2:00-3:00	0.001099	0.000950	0.001187	0.000868
3:00-4:00	0.001108	0.000869	0.001295	0.000869
4:00-5:00	0.000647	0.000717	0.000762	0.000481
5:00-6:00	0.000994	0.000772	0.001167	0.000727
6:00-7:00	0.001007	0.001452	0.001088	0.000826
7:00-8:00	0.000950	0.000934	0.001042	0.000821
8:00-9:00	0.000625	0.000499	0.000728	0.000499
9:00-10:00	0.002049	0.001549	0.002355	0.001493
10:00-11:00	0.002314	0.001715	0.002665	0.001704
11:00-12:00	0.001569	0.001164	0.001802	0.001189
12:00-13:00	0.001847	0.001456	0.002129	0.001348
13:00-14:00	0.001495	0.001105	0.001744	0.001083
14:00-15:00	0.000941	0.001951	0.001082	0.000692
15:00-16:00	0.001125	0.000899	0.001284	0.001034
16:00-17:00	0.001139	0.000877	0.001328	0.000842
17:00-18:00	0.001002	0.000806	0.001114	0.000886
18:00-19:00	0.002112	0.001549	0.002470	0.001557
19:00-20:00	0.001497	0.001098	0.001757	0.001336
20:00-21:00	0.001430	0.001065	0.001677	0.001090
21:00-22:00	0.000787	0.000680	0.000856	0.000937
22:00-23:00	0.000748	0.000619	0.000637	0.001264
23:00-24:00	0.000846	0.001520	0.000815	0.001050
Average	0.001214	0.001179	0.001370	0.001005

Table 10. The Simulated 99th Percentile of Total Call Volume per Hour w/ and w/o copula

Time Interval	99th Percentile from Data	w/ copula	w/o copula
0:00-1:00	0.00132545	0.001034080	0.000978854
1:00-2:00	0.00075795	0.000773180	0.000618925
2:00-3:00	0.00100552	0.001098830	0.000978854
3:00-4:00	0.00120738	0.001108350	0.001096930
4:00-5:00	0.00090649	0.000647491	0.000643682
5:00-6:00	0.00129117	0.000994089	0.000986471
6:00-7:00	0.00113501	0.001007420	0.000891252
7:00-8:00	0.00100171	0.000950288	0.000864591
8:00-9:00	0.00075033	0.000624638	0.000609403
9:00-10:00	0.00223575	0.002049120	0.001969130
10:00-11:00	0.00233097	0.002313830	0.002231940
11:00-12:00	0.00176727	0.001569210	0.001504460
12:00-13:00	0.00206816	0.001847250	0.001782500
13:00-14:00	0.00150446	0.001494940	0.001477800
14:00-15:00	0.00100171	0.000940766	0.000914105
15:00-16:00	0.00114263	0.001125491	0.001081690
16:00-17:00	0.00112359	0.001138820	0.001129300
17:00-18:00	0.00107788	0.001001710	0.000927435
18:00-19:00	0.00227003	0.002111960	0.002094820
19:00-20:00	0.00187772	0.001496850	0.001491130
20:00-21:00	0.00156921	0.001430190	0.001416860
21:00-22:00	0.00071986	0.000786511	0.000723666
22:00-23:00	0.00076937	0.000748423	0.000708431
23:00-24:00	0.00131022	0.000845547	0.000697005

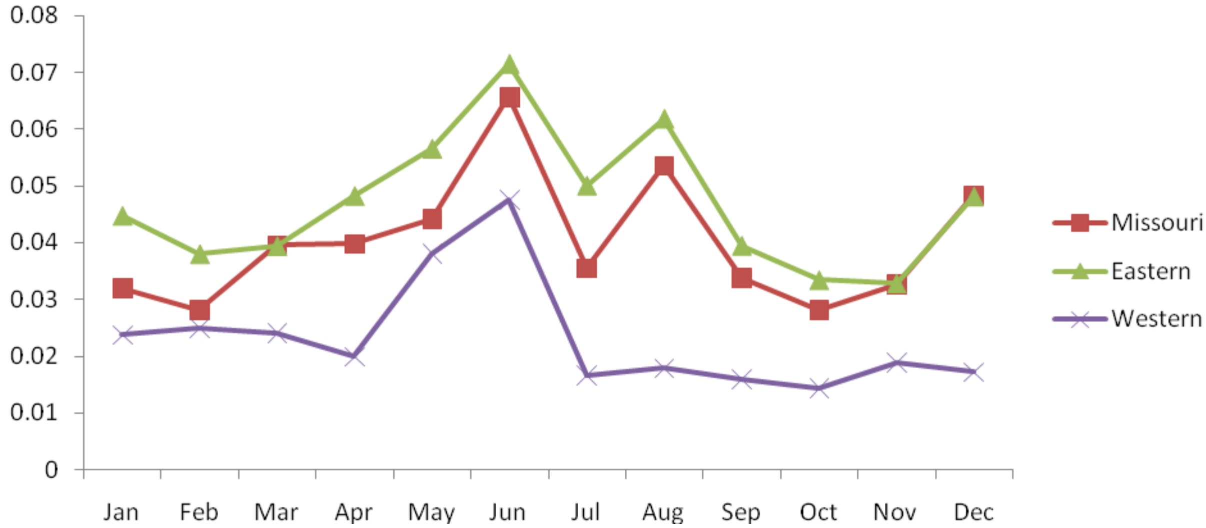

Figure 2. Average call volume per month

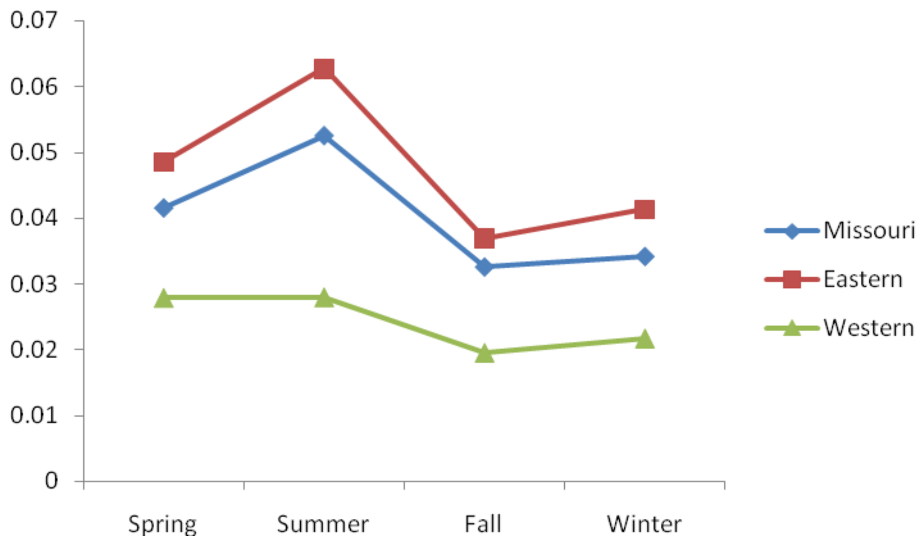

Figure 3. Average call volume per month for each season

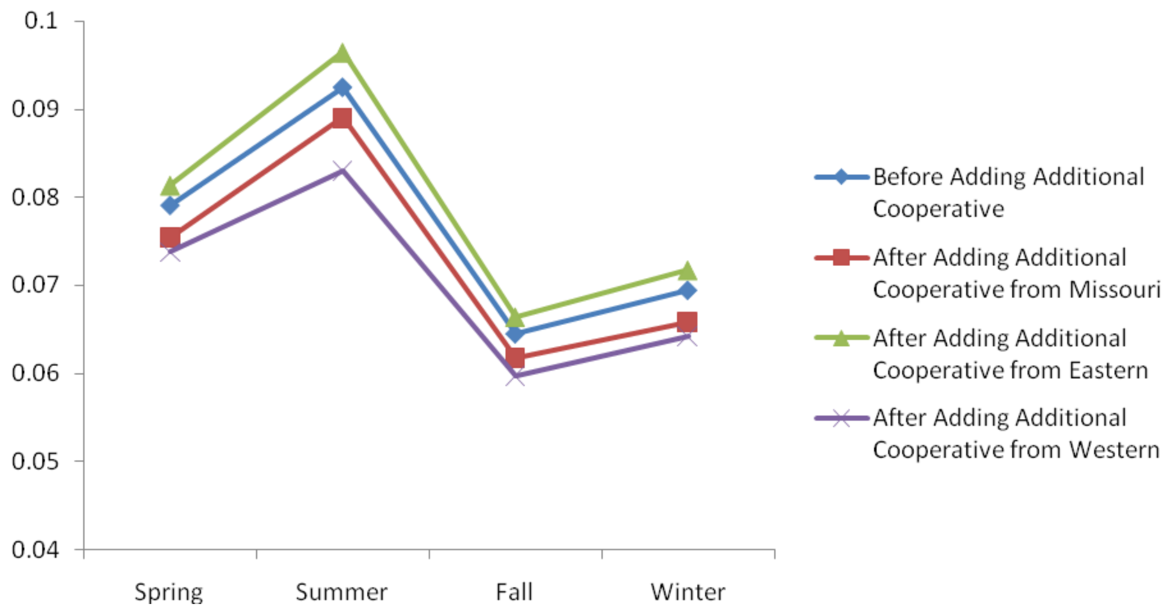

Figure 4. The 99th percentile of total call volume per month before and after adding additional cooperative from each region with 100,000 meters

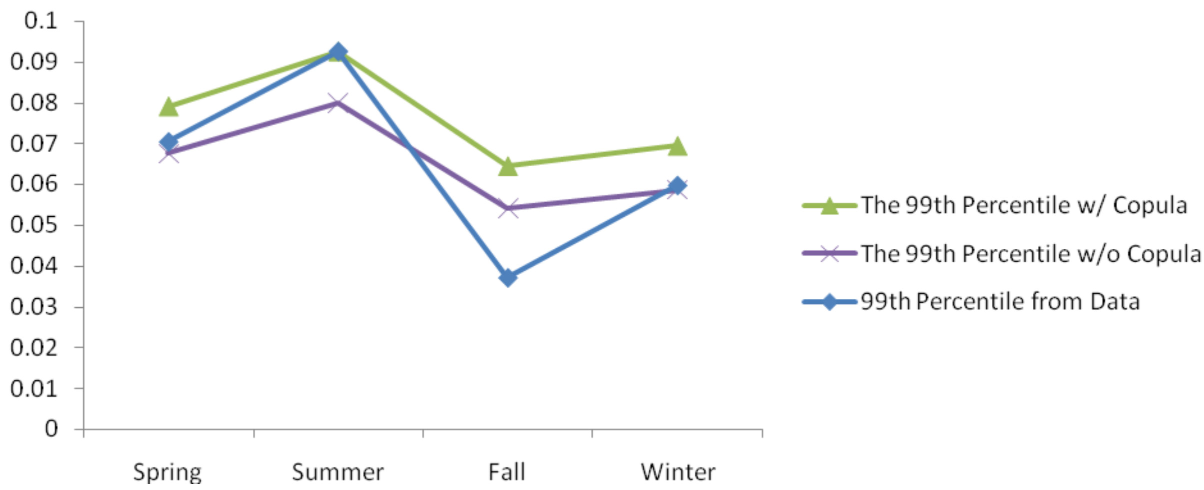

Figure 5. The 99th percentile from data, and the simulated 99th percentile with copula and without copula for call volume per month

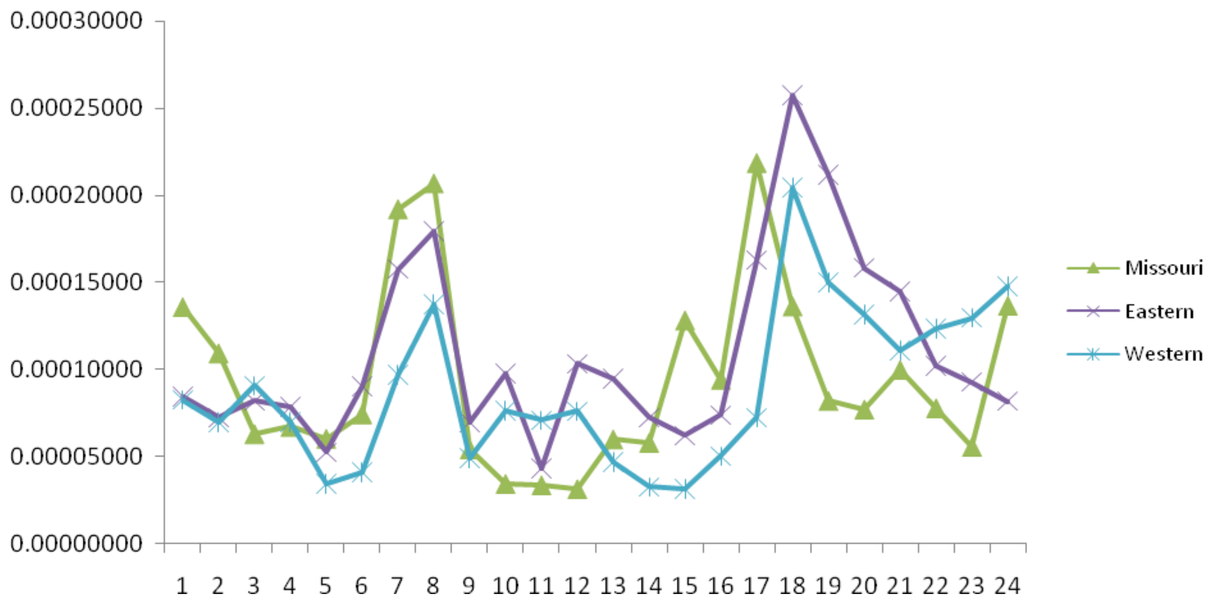

Figure 6. Average call volume per hour

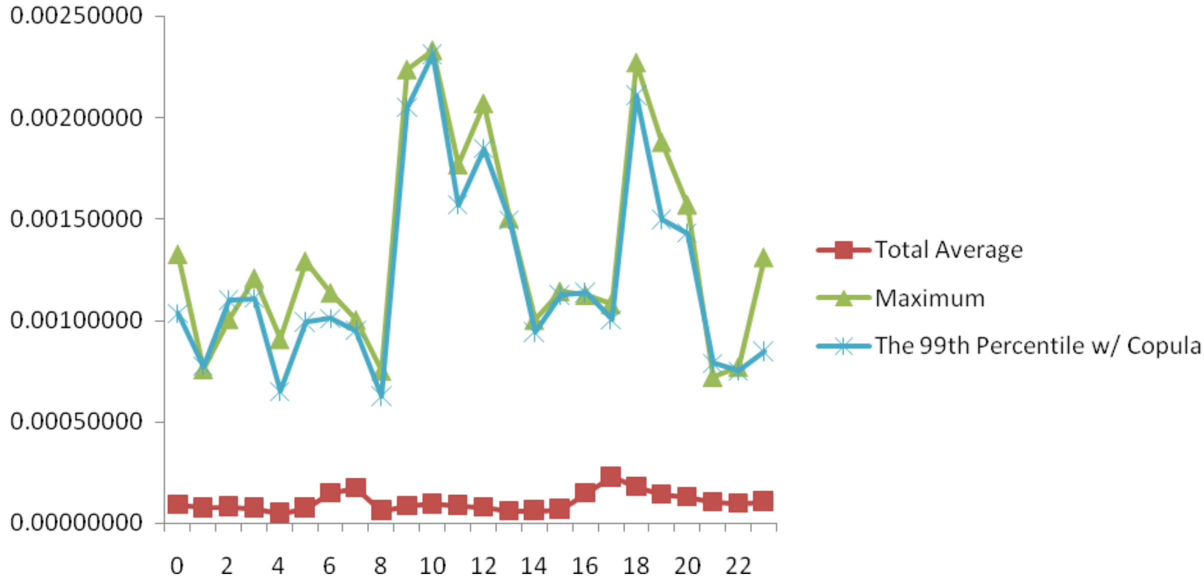

Figure 7. Average, maximum, and the simulated 99th percentile from empirical distribution with gaussian copula for total call volume per hour

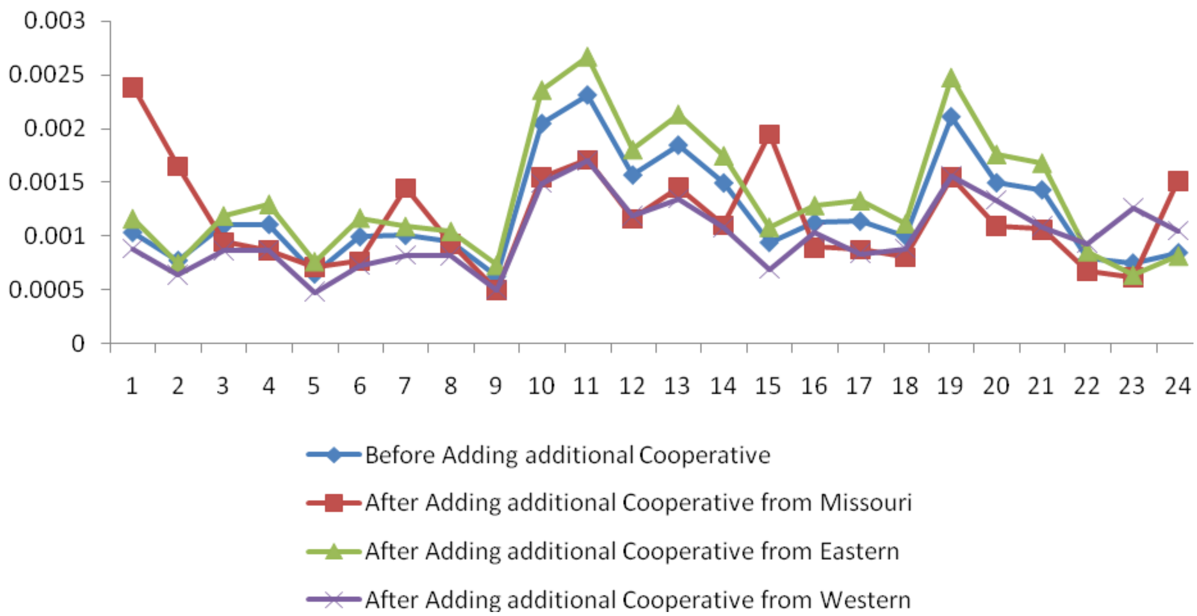

Figure 8. The 99th percentile of total call volume per hour before and after adding additional cooperative from each region with 100,000 meters

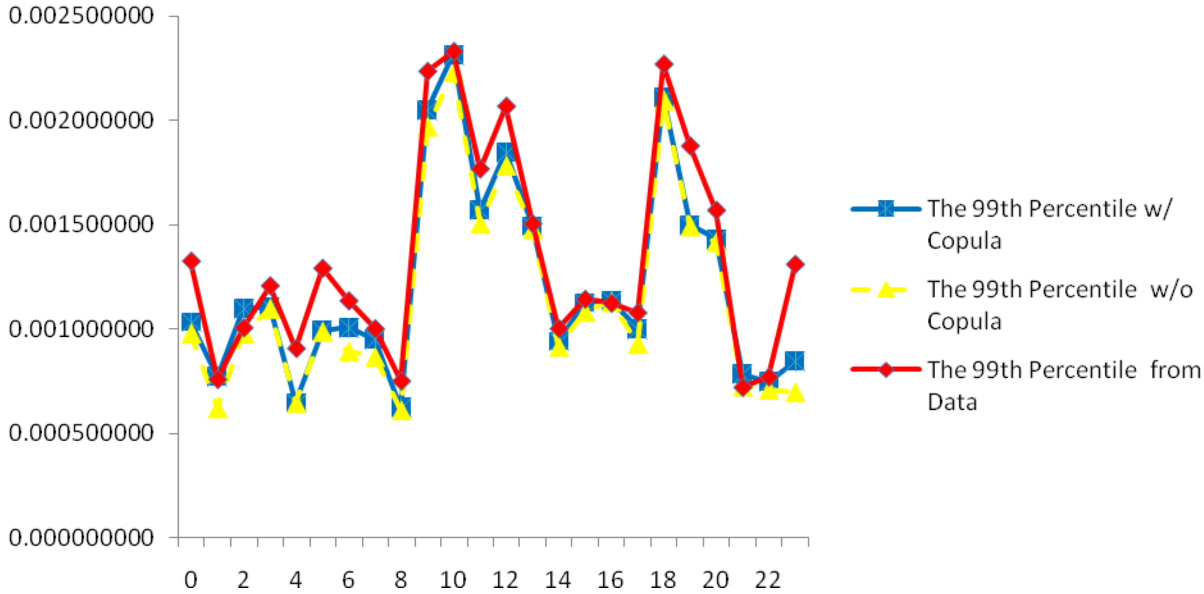

Figure 9. The 99th percentile from data, and the simulated 99th percentile with copula and without copula for total call volume per hour

References

Andrews, B. H., and S. M. Cunningham. 1995. "L.L. bean improves Call-Center Forecasting." *Interfaces* 25(6): 1-13.

Atlason, J., M. A. Epelman, and S.G. Henderson. 2008. "Optimizing Call Staffing Using Simulation and Analytic Center Cutting-Plane Methods." *Management Science* 54(2): 295-309.

Avramidis, A. N., A. Deslauriers, and P. L'Ecuyer. 2004. "Modeling Daily Arrivals to a Telephone Call Center." *Management Science* 50(7): 896-908.

Cherubini, U., E. Luciano, and W. Vecchiato. 2004. *Copula Methods in Finance*. West Sussex: John Wiley & Sons Ltd.

Genest, C., K. Ghoudi, and L. -P. Rives. 1995. "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions." *Biometrika* 82: 543-552.

Hann, L. D., and A. Ferreira. 2006. *Extreme Value Theory: An Introduction*. New York: Springer Science+Business Media LLC.

Joe, H., and J.J. Xu. 1996. *The Estimation Method of Inference Functions for Margins for Multivariate Models*. Dept. of Statistics University of British Columbia, Technical Report 166.

Joe, H. 2005. "Asymptotic Efficiency of the Two-Stage Estimation Method for Copula-Based Models." *Journal of Mutivariate Analysis* 94: 401-419.

Liu, W., and J. Cela. 2008. "Count Data Models in SAS." Paper presented at SAS Global Forum 2008, San Antonio TX, 16-19 March.

McNeil A. J. 1997. "Estimating the Tails of Loss Severity Distributions using Extreme Value Thoery." *ASTIN Bulletin International Actuarial Association* 12(1): 117-137.

Richardson, J. W., S. L. Klose, and A. W. Gray. 2000. "An Applied Procedure for Estimating and Simulating Multivariate Empirical (MVE) Probability Distributions In Farm-Level Risk Assessment and Policy Analysis." *Journal of Agricultural and Applied Economics* 32(2): 299-315.

Taylor, C. R. 1990. "Two Practical Procedures for Estimating Multivariate Nonnormal Probability Density Functions." *American Journal of Agricultural Economics* 72: 210-217.

Taylor, J. W. 2008. "A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center." *Management Science* 54(2): 253-265.

Xu, W. 2000. "Long Range Planning for Call Centers at Fedex." *The Journal of Bussiness Forecasting Methods and Systems* 18(4): 7-11.

Appendix

1. IFM method with Gamma distribution and Gaussian copula

```
/*First step for IFM method with three variables*/

/*Missouri - gamma distribution, shape=r, scale=lambda*/
proc nlmixed data=month1 hess;
parms r=.1 lambda=.1 a1=.1 a2=.1 a3=.1;
r1=r+a1*d1+a2*d2+a3*d3;
model mi~gamma(r1,lambda/1000);
run;

/*Eastern - gamma distribution, shape=r, scale=lambda*/
proc nlmixed data=month1 hess;
parms r=.1 lambda=.1 a1=.1 a2=.1 a3=.1;
r1=r+a1*d1+a2*d2+a3*d3;
model ea~gamma(r1,lambda/1000);
run;

/*Western - gamma distribution, shape=r, scale=lambda*/
proc nlmixed data=month1 hess;
parms r=.1 lambda=.1 a1=.1 a2=.1 a3=.1;
r1=r+a1*d1+a2*d2+a3*d3;
model we~gamma(r1,lambda/1000);
run;

/*Second step for IFM method */
Data test;set month1;
r1=10.1038;a11=2.1768;a21=5.4154;a31=-0.4710;r11=r1+a11*d1+a21*d2+a31*d3;lambda1=.0033875;
r2=8.8642;a12=1.5386;a22=4.6013;a32=-0.9597;r22=r2+a12*d1+a22*d2+a32*d3;lambda2=.0046589;
r3=3.1391;a13=0.8853;a23=0.8987;a33=-0.3123;r33=r3+a13*d1+a23*d2+a33*d3;lambda3=.0069037;
lpdf1=logpdf('gamma',mi,r11,lambda1);
lpdf2=logpdf('gamma',ea,r22,lambda2);
lpdf3=logpdf('gamma',we,r33,lambda3);
u=cdf('gamma',mi,r11,lambda1);
v=cdf('gamma',ea,r22,lambda2);
w=cdf('gamma',we,r33,lambda3);
x=probit(u);y=probit(v);z=probit(w);
proc univariate data=test normal;
var x y z;
run;
proc corr data=test;
var x y z;
run;
```

```
/*Alternative second step for IFM method */
proc nlmixed tech=trureg itdetails hess data=month1;
parms r12=0.8 r23=0.75 r13=0.66;
r1=10.1038;a11=2.1768;a21=5.4154;a31=-0.4710;r11=r1+a11*d1+a21*d2+a31*d3;lambda1=.0033875;
r2=8.8642;a12=1.5386;a22=4.6013;a32=-0.9597;r22=r2+a12*d1+a22*d2+a32*d3;lambda2=.0046589;
r3=3.1391;a13=0.8853;a23=0.8987;a33=-0.3123;r33=r3+a13*d1+a23*d2+a33*d3;lambda3=.0069037;
lpdf1=logpdf('gamma',mi,r11,lambda1);
lpdf2=logpdf('gamma',ea,r22,lambda2);
lpdf3=logpdf('gamma',we,r33,lambda3);
u=cdf('gamma',mi,r11,lambda1);
v=cdf('gamma',ea,r22,lambda2);
w=cdf('gamma',we,r33,lambda3);
x=probit(u);y=probit(v);z=probit(w);
lnc=-(1/2)*log(2*r12*r23*r13-r13**2-r12**2+1-r23**2)-(1/2)*x*(-x*(-
2*r12*r23*r13+r13**2+r12**2)/(-1+r23**2-2*r12*r23*r13+r12**2+r13**2)
+y*(r12-r13*r23)/(-1+r23**2-2*r12*r23*r13+r12**2+r13**2)-z*(r12*r23-r13)/(-1+r23**2
-2*r12*r23*r13+r12**2+r13**2))-(1/2)*y*(x*(r12-r13*r23)/(-1+r23**2-
2*r12*r23*r13+r12**2+r13**2)
-y*(-2*r12*r23*r13+r12**2+r23**2)/(-1+r23**2-2*r12*r23*r13+r12**2+r13**2)-z*(-r23+r12*r13)/(-1
+r23**2-2*r12*r23*r13+r12**2+r13**2))-(1/2)*z*(-x*(r12*r23-r13)/(-1+r23**2-
2*r12*r23*r13+r12**2
+r13**2)-y*(-r23+r12*r13)/(-1+r23**2-2*r12*r23*r13+r12**2+r13**2)+z*(2*r12*r23*r13-r13**2
-r23**2)/(-1+r23**2-2*r12*r23*r13+r12**2+r13**2));
model mi~general(lnc+lpdf1+lpdf2+lpdf3);
RUN;
```

2. MLE with Gamma distibution and Gaussian copula

```
proc nlmixed tech=trureg itdetails qp=100 hess data=month1;
parms r12=0.9263 r23=0.7200 r13=0.5432
    r1=10.1038 a11=2.1768 a21=5.4154 a31=-0.4710 lambda1=0.0033875
    r2=8.8642  a12=1.5386 a22=4.6013 a32=-0.9597 lambda2=0.0046589
    r3=3.1391  a13=0.8853 a23=0.8987 a33=-0.3123 lambda3=0.0069037;
r11=r1+a11*d1+a21*d2+a31*d3;
r22=r2+a12*d1+a22*d2+a32*d3;
r33=r3+a13*d1+a23*d2+a33*d3;
lpdf1=logpdf('gamma',mi,r11,lambda1/1000);
lpdf2=logpdf('gamma',ea,r22,lambda2/1000);
lpdf3=logpdf('gamma',we,r33,lambda3/1000);
u=cdf('gamma',mi,r11,lambda1/1000);
v=cdf('gamma',ea,r22,lambda2/1000);
w=cdf('gamma',we,r33,lambda3/1000);
x=probit(u);y=probit(v);z=probit(w);
lnc=-(1/2)*log(2*r12*r23*r13-r13**2-r12**2+1-r23**2)
-(1/2)*x*(-x*(-2*r12*r23*r13+r13**2+r12**2)/(-1+r23**2-2*r12*r23*r13+r12**2+r13**2)
+y*(r12-r13*r23)/(-1+r23**2-2*r12*r23*r13+r12**2+r13**2)-z*(r12*r23-r13)/(-1+r23**2
-2*r12*r23*r13+r12**2+r13**2))-(1/2)*y*(x*(r12-r13*r23)/(-1+r23**2-
2*r12*r23*r13+r12**2+r13**2)
-y*(-2*r12*r23*r13+r12**2+r23**2)/(-1+r23**2-2*r12*r23*r13+r12**2+r13**2)-z*(-r23+r12*r13)/(-1
+r23**2-2*r12*r23*r13+r12**2+r13**2))-(1/2)*z*(-x*(r12*r23-r13)/(-1+r23**2-
2*r12*r23*r13+r12**2
+r13**2)-y*(-r23+r12*r13)/(-1+r23**2-2*r12*r23*r13+r12**2+r13**2)+z*(2*r12*r23*r13-r13**2
-r23**2)/(-1+r23**2-2*r12*r23*r13+r12**2+r13**2));
model ea~general(lnc+lpdf1+lpdf2+lpdf3);
RUN;QUIT;
```

3. CML method with empirical distribution and Gaussian copula

```
proc rank data=hour1 out=epdf fraction;
ranks u v w;
var mi ea we;
data epdf1;set epdf;
x=probit(u);y=probit(v);z=probit(w);
proc corr data=epdf1;
var x y z;
run;
```

4. Simulation with Gamma distibution and Gaussian copula

```
data one;
d1=0;d2=0;d3=0;/*winter*/
r12=0.6903;r13=0.5646;r23=0.7310;
p21=r12;p22=sqrt(1-r12**2);
p31=r13;p32=-(-r23+r12*r13)/sqrt(1-r12**2);
p33=sqrt((r23**2-1+r12**2+r13**2-2*r12*r13*r23)/(-1+r12**2));
r1=10.1038;a11=2.1768;a21=5.4154;a31=-0.4710;r11=r1+a11*d1+a21*d2+a31*d3;lambda1=.0033875;
r2=8.8642;a12=1.5386;a22=4.6013;a32=-0.9597;r22=r2+a12*d1+a22*d2+a32*d3;lambda2=.0046589;
r3=3.1391;a13=0.8853;a23=0.8987;a33=-0.3123;r33=r3+a13*d1+a23*d2+a33*d3;lambda3=.0069037;
do ii=1 to 10000;
z1=rannor(12345);
z2=rannor(23565);
z3=rannor(87975);
x1=z1;
x2=p21*z1+p22*z2;
x3=p31*z1+p32*z2+p33*z3;
u1=probnorm(x1);
u2=probnorm(x2);
u3=probnorm(x3);
asim=gaminv(u1,r11)*lambda1;
bsim=gaminv(u2,r22)*lambda2;
csim=gaminv(u3,r33)*lambda3;
output;
end;
run;
data two;set one;
asim1=asim*35552;
bsim1=bsim*158747;
csim1=csim*68253;
total=(asim1+bsim1+csim1)/(35552+158747+68253);
proc means data=two;
var asim bsim csim total;
run;
proc univariate data=two;
var asim bsim csim total;
run;
```

5. Simulation with empirical distribution and Gaussian copula

```
data hour11;set hour1;
array aa{74} a1-a74;
array bb{74} b1-b74;
array cc{74} c1-c74;

r12=0.37434;r13=0.34901;r23=0.42042;
p21=r12;p22=sqrt(1-r12**2);
p31=r13;p32=-(-r23+r12*r13)/sqrt(1-r12**2);
p33=sqrt((r23**2-1+r12**2+r13**2-2*r12*r13*r23)/(-1+r12**2));

do ii=1 to 10000;
z1=rannor(12345);
z2=rannor(23565);
z3=rannor(87975);
x1=z1;
x2=p21*z1+p22*z2;
x3=p31*z1+p32*z2+p33*z3;
u1=probnorm(x1);
u2=probnorm(x2);
u3=probnorm(x3);
I1=int(u1*74+1);
I2=int(u2*74+1);
I3=int(u3*74+1);
asim=aa(I1);
bsim=bb(I2);
csim=cc(I3);
output;
end;
run;

data hour12;set hour11;
asim1=asim*35552;
bsim1=bsim*158747;
csim1=csim*68253;
total=(asim1+bsim1+csim1)/(35552+158747+68253);
proc means data=hour12;
var asim bsim csim total;
run;
proc univariate data=hour12;
var asim bsim csim total;
run;
```