

Efficient Confidence Intervals for the Defective Totals using SUR
Framework

<http://www.doi.org/10.62341/istj-vol39-1-NA68>

Received	2026/09/10	تم استلام الورقة العلمية في
Accepted	2026/09/29	تم قبول الورقة العلمية في
Published	2026/09/30	تم نشر الورقة العلمية في

Efficient Confidence Intervals for the Defective Totals using SUR Framework

Nuri Eltabit Almadni

University of Wadi Shati, Brak, Libya

n.almadni@wau.edu.ly

Abstract

Estimating the total number of defective units is an important problem in quality control and industrial applications. This study develops approximate confidence intervals for the defective total within the framework of the Seemingly Unrelated Regression (SUR) model, where the total is expressed as a linear combination of reliable estimates which yields sum of the estimated percentage of the various components. Using the multivariate central limit theorem, the asymptotic distribution of the estimator vector is derived, leading to the construction of confidence intervals based on both the standard normal (z) and Student's t quantiles. A simulation study is conducted under different sample sizes and parameter settings to evaluate the coverage performance of the proposed intervals. The results indicate that confidence intervals based on z-quantiles provide satisfactory coverage for moderate and large sample sizes, whereas those based on t-quantiles are generally more conservative, particularly for small samples and low parameter values. The proposed approach offers a practical method for inference on defective totals in correlated multivariate settings.

Keywords: SUR model, binomial model, multivariate central limit theorem, asymptotic inference, confidence intervals.

فترات الثقة الفعالة لإجمالي الوحدات المعيبة باستخدام إطار الانحدار غير المرتبط ظاهرياً (SUR)

نوري التابت محمد المدني

قسم علوم الحاسوب، كلية العلوم، جامعة وادي الشاطئ، براك، ليبيا

n.almadni@wau.edu.ly

الملخص

يُعد تقدير العدد الإجمالي للوحدات المعيبة مشكلة مهمة في ضبط الجودة والتطبيقات الصناعية. تطوّر هذه الدراسة فترات ثقة تقريبية للإجمالي المعيب ضمن إطار نموذج الانحدار غير المرتبط ظاهرياً (SUR)، حيث يُعبّر عن الإجمالي بوصفه تركيبة خطية من تقديرات موثوقة، مما ينتج عنه تقدير لمجموع موزون للتقديرات المتعلقة بنسبة المكونات المختلفة. وباستخدام نظرية الحد المركزي متعددة المتغيرات، يُشتق التوزيع التقاربي لمتجه المقدّرات، مما يؤدي إلى بناء فترات ثقة تستند إلى كميات التوزيع الطبيعي القياسي (Z) وكميات توزيع t لستودنت. وأُجريت دراسة محاكاة باستخدام أحجام عينات وإعدادات مختلفة للمعاملات لتقييم أداء التغطية للفترات المقترحة. وتشير النتائج إلى أن فترات الثقة القائمة على كميات Z تحقق تغطية مرضية لأحجام العينات المتوسطة والكبيرة، في حين تكون الفترات القائمة على كميات t أكثر تحفظاً بصورة عامة، ولا سيما عند أحجام العينات الصغيرة وقيم المعلمات المنخفضة. ويوفر النهج المقترح طريقة عملية للاستدلال على إجمالي الوحدات المعيبة في البيئات متعددة المتغيرات ذات الارتباط.

الكلمات المفتاحية: نموذج SUR، النموذج ذي الحدين، نظرية الحد المركزي متعددة المتغيرات، الاستدلال التقاربي، فترات الثقة.

1. Introduction

There are two methods for constructing confidence intervals for the proportion p of the binomial distribution (see Sec. 2) such that their coverage is closed to the nominal level. These are the interval based on the normal quantiles and the interval based on Student's t -

Efficient Confidence Intervals for the Defective Totals using SUR
Framework

<http://www.doi.org/10.62341/istj-vol39-1-NA68>

quantile. We observed that the t -quantile method gives better coverage than the z -quantile method. [1], [2], [3] and others show that the slandered interval perform poorly and erratically, more so than is generally appreciated, furthermore, their coverage is known to be poor for p near 0 or 1. [3] "Asymptotic confidence intervals by the z -quantiles" which produced intervals guarantee achieving at least the nominal value of the true coverage, and also, it's accurate even at smallest sample sizes.

In this research, we use both quantiles for the total of a certain product (using the z -quantile and using the t -quantile with $n - 1$ degrees of freedom), where n is the number of observations of each component. The model used was introduced by [4]. This model analyses a system of ' r ' equations with error terms correlated cross equations, while these errors are uncorrelated between individuals $i = 1, \dots, n$, but they are correlated across the equations for the j th. The r regression equations are unrelated or 'seemingly unrelated'. Suppose that the Gauss-Markov assumptions hold for each the r equations; then the ordinary least square (OLS) estimator of each equation taken separately are the Best Linear Unbiased Estimator (BLUE). However, because the errors are correlated across equations is inefficient compared with joint estimation (Aitken's generalised least square [10]).

2. Description of the model

Suppose, we have n observations of $2r$ components of data (taking in account any vertical dependencies between the components, i.e.,

(U_{ij}, X_{ij}) , $i = 1, \dots, n, j = 1, \dots, r$ are $n \times$

$2r$ dimensional data such that each observational unit indexed by the subscript i , as well as $U_{ij}, X_{ij} \geq 0, \forall i = 1, \dots, n, j = 1, \dots, r$. The i^{th} observation represents unit, for instance:

X_{i1}, X_{i2} is the number of units respectively for a product $i, i = 1, \dots, n$.

U_{i1}, U_{i2} is the number of defective units for a product $i, i = 1, \dots, n$.
Next we introduce to the models used in the simulation.

2.1. Poisson and Binomial Models

Each of the pairs random variables of (U_{i1}, U_{i2}) , (X_{i1}, X_{i2}) follows a bivariate Poisson model with parameters $(\lambda_0 + \lambda_1, \lambda_0 + \lambda_2)$, $(\lambda_0 + \lambda_1 + \mu_1, \lambda_0 + \lambda_2 + \mu_2)$ respectively, where

$$X_{ij} = U_{ij} + V_{ij}, U_{ij} \sim \text{Poiss}(\lambda_0 + \lambda_j), V_{ij} \sim \text{Poiss}(\mu_j),$$

$X_{ij} \sim \text{Poiss}(\lambda_0 + \lambda_j + \mu_j)$, $\text{Cov}(U_{i1}, U_{i1}) = \text{Cov}(X_{i1}, X_{i1}) = \lambda_0$,
as well as, all

$$(U_j, V_j), \text{ and } (U_j, V_j), j \neq j^l = 1, 2$$

are independent (for more details on the bivariate Poisson model see [5] and [6]). Further, the conditional distribution

$$P(U_{ij} | X_{ij}) \sim \text{Bin}(X_{ij}, p_j),$$

is a binomial distribution with parameter $p_j = \frac{\lambda_0 + \lambda_j}{\lambda_0 + \lambda_j + \mu_j}$, and the

sample observations $X_{ij} > 0, \forall i = 1, \dots, n, j = 1, 2$, we

also denote $\hat{p}_j = \frac{\sum_{i=1}^n U_{ij}}{\sum_{i=1}^n X_{ij}}$ as a ratio estimator of p_j , $E(\hat{p}_j) = p_j =$

$$\frac{U_j}{X_j}, j = 1, \dots, r,$$

where U_{ij}, X_{ij} are *i.i.d*, $i = 1, \dots, n$.

3. The Seemingly Unrelated Regression (SUR) Model

Consider the r – variate linear models:

$$U_{ij} = X_{ij} p_j + e_{ij}, j = 1, \dots, r$$

with the following assumptions:

$$E(e_{ij} | X_{ij}) = 0, p_j = \frac{U_j}{X_j}, \text{ with the variance depend on } X_{ij}$$

$$\text{Var}(e_{ij} | X_{ij}) = \sigma_j^2 X_{ij}, \text{ Cov}(e_{ij} e_{i^l j} | X_{ij} X_{i^l j}) = 0,$$

$$\forall i \neq i^l, i, i^l = 1, \dots, n.$$

We stake these equations in to linear model

$$U_i = X_i p + e_i, i = 1, \dots, n \text{ and}$$

$$U_i = (U_{i1}, U_{i2}, \dots, U_{ir})^T, \text{ the design matrix}$$

$$X_i = \text{diag}(X_{i1}, X_{i2}, \dots, X_{ir})^T$$

Efficient Confidence Intervals for the Defective Totals using SUR
Framework

<http://www.doi.org/10.62341/istj-vol39-1-NA68>

and $p = (p_1, p_2, \dots, p_r)^T$, as well as the error vector $e_i = (e_{i1}, e_{i2}, \dots, e_{ir})^T$ has the covariance matrix

$$\Sigma_i = \text{Cov}(e_i) = \begin{pmatrix} \sigma_1^2 X_{i1} & \sigma_{12} \sqrt{X_{i1} X_{i2}} & \dots & \sigma_{1r} \sqrt{X_{i1} X_{ir}} \\ \sigma_{12} \sqrt{X_{i1} X_{i2}} & \sigma_2^2 X_{i2} & \dots & \sigma_{2r} \sqrt{X_{i2} X_{ir}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1r} \sqrt{X_{i1} X_{ir}} & \sigma_{2r} \sqrt{X_{i2} X_{ir}} & \dots & \sigma_r^2 X_{ir} \end{pmatrix}$$

$\text{Cov}(e_i e_i) = 0_{r \times r}, i \neq i^l = 1, \dots, n,$

The model will be weighted by the linear transformation

$$A_i U_i = A_i X_i p + A_i e_i, \quad i, e, \tilde{U}_i = \tilde{X}_i p + \tilde{e}_i$$

where, the transformation matrix is

$$A_i = \text{diag}(\tilde{X}_{i1}^{-1}, \tilde{X}_{i2}^{-1}, \dots, \tilde{X}_{ir}^{-1}), \tilde{X}_{i1} = \sqrt{X_{ij}}, j = 1, \dots, r,$$

the weighed response variable

$$\tilde{U}_i = (\tilde{U}_{i1}, \tilde{U}_{i2}, \dots, \tilde{U}_{ir})^T,$$

the weighted error vector

$$\tilde{e}_i = (\tilde{e}_{i1}, \tilde{e}_{i2}, \dots, \tilde{e}_{ir})^T,$$

and the weighted design matrix

$$A_i = \text{diag}(\tilde{X}_{i1}^{-1}, \tilde{X}_{i2}^{-1}, \dots, \tilde{X}_{ir}^{-1}) \tilde{X}_{i1} = \sqrt{X_{ij}}, j = 1, \dots, r,$$

$$\tilde{U}_{ij} = \frac{U_{ij}}{\sqrt{X_{ij}}}, \tilde{e}_{ij} = \frac{e_{ij}}{\sqrt{X_{ij}}}, X_{ij} > 0,$$

$$\forall i = 1, \dots, n, j = 1, 2, \dots, r,$$

the $\tilde{e}_i$ has a covariance matrix

Efficient Confidence Intervals for the Defective Totals using SUR
Framework

<http://www.doi.org/10.62341/istj-vol39-1-NA68>

$$\Sigma = \text{Cov}(\tilde{e}_i) = \tilde{A}_i \text{Cov}(e_i) \tilde{A}_i^T = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1r} \\ \sigma_{12} & \sigma_2^2 & & \sigma_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1r} & \sigma_{2r} & \cdots & \sigma_r^2 \end{pmatrix}$$

where

$$\sigma_j^2 = E(\tilde{U}_{ij} - X_{ij}p_j)^T,$$

$$\sigma_{jj^l} = E\left((\tilde{U}_{ij} - \tilde{X}_{ij}p_j)(\tilde{U}_{ij^l} - \tilde{X}_{ij^l}p_{j^l})\right), \quad j < j^l = 1, \dots, r.$$

Thus, the staked SUR model (see also [7]) will be

$$\tilde{U} = \tilde{X}p + \tilde{e}$$

$$\tilde{X} = (\tilde{X}_1^T, \tilde{X}_2^T, \dots, \tilde{X}_n^T)^T$$

, with dimention $rn \times n$

$$p = (p_1, p_2, \dots, p_r)^T$$

$$\tilde{e} = (\tilde{e}_1^T, \tilde{e}_2^T, \dots, \tilde{e}_n^T)^T$$

, with dimention $rn \times 1$

$$\tilde{\Sigma} = \text{Cov}(\tilde{e}) = \begin{pmatrix} \tilde{\Sigma} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{\Sigma} \end{pmatrix}_{rn \times rn} = I_n \otimes \tilde{\Sigma} \text{ where}$$

$$\text{Var}(\tilde{e}_i) = \tilde{\Sigma}_{r \times r}, \text{ and } \text{Cov}(\tilde{e}_i \tilde{e}_{i^l}) = \mathbf{0}_{r \times r}$$

($\mathbf{0}_{r \times r}$ matrix of zeros with dimention $r \times r$)

$$\forall i \neq i^l, i, i^l = 1, \dots, n.$$

The ordinary least square estimator of the model ratio p is then obtained from the SUR model as

$$\begin{aligned}
 \mathbf{p} &= (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{U}} = \\
 &= \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{X}}_i \right)^{-1} \sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{U}}_i = \left(\sum_{i=1}^n \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{U}}_i \\
 &= \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i^2 \right)^{-1} \sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{U}}_i \\
 &= \left(\frac{\sum_{i=1}^n \sqrt{\mathbf{X}_{i1}} \tilde{\mathbf{U}}_{i1}}{\sum_{i=1}^n \mathbf{X}_{i1}}, \dots, \frac{\sum_{i=1}^n \sqrt{\mathbf{X}_{ir}} \tilde{\mathbf{U}}_{ir}}{\sum_{i=1}^n \mathbf{X}_{ir}} \right)^T, \\
 \sqrt{\mathbf{X}_{ij}} \tilde{\mathbf{U}}_{ij} &= \mathbf{U}_{ij}, \quad j = 1, \dots, r.
 \end{aligned}$$

4. Asymptotic normal of the estimator vector $\hat{\mathbf{p}}$

Assume that the random error vectors $\tilde{\mathbf{e}}_i$ are not necessary normally distributed but *i.i.d* random vectors, $i = 1, \dots, n$, i.e., $E(\tilde{\mathbf{e}}_i) = 0$, and $\text{Cov}(\tilde{\mathbf{e}}_i) = \tilde{\Sigma}$. Moreover, under certain conditions on the design matrix $\mathbf{X}_i$ one can show that in large sample sizes, $\hat{\mathbf{p}}$ follows a multivariate asymptotic normal distribution. Since the pairs $(\mathbf{X}_i, \mathbf{U}_i)$ are *i.i.d* the transformed $(\tilde{\mathbf{X}}_i, \tilde{\mathbf{U}}_i)$ are also *i.i.d*.

The OLS estimator of the SUR model can be written as

$$\begin{aligned}
 \mathbf{p} &= (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{U}} \\
 &= \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{X}}_i \right)^{-1} \left(\sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{X}}_i \mathbf{p} + \sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{e}}_i \right)
 \end{aligned}$$

So,

$$\begin{aligned}
 \sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}) &= \left(\frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{X}}_i \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{e}}_i \\
 &= \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{e}}_i \right) \quad (1)
 \end{aligned}$$

Efficient Confidence Intervals for the Defective Totals using SUR
Framework

<http://www.doi.org/10.62341/istj-vol39-1-NA68>

Implies that

$$\left(\frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{X}}_i\right)^{-1} \rightarrow^P \left(E(\tilde{\mathbf{X}}_i^T \tilde{\mathbf{X}}_i)\right)^{-1},$$

by consistency, since the matrix $E(\tilde{\mathbf{X}}_i^T \tilde{\mathbf{X}}_i)$ non singular (by [8], p. 612), thus

$$\left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i\right)^{-1} \rightarrow^P \text{diag}(E(X_1), \dots, E(X_r))^{-1} > 0$$

From (1) and by the multivariate central limit theorem

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{e}}_i \rightarrow^D N_r(\mathbf{0}, E(\tilde{\mathbf{X}}_i^T \tilde{\Sigma} \tilde{\mathbf{X}}_i)), \text{ where}$$

$$\begin{aligned} \tilde{\Sigma} &= \text{Cov}(\tilde{\mathbf{X}}_i^T \tilde{\mathbf{e}}_i) \\ &= E(\text{Cov}(\tilde{\mathbf{X}}_i^T \tilde{\mathbf{e}}_i | \tilde{\mathbf{X}}_i^*)) + \text{Cov}(E(\tilde{\mathbf{X}}_i^T \tilde{\mathbf{e}}_i | \tilde{\mathbf{X}}_i^*)) \end{aligned}$$

$$= E(\tilde{\mathbf{X}}_i^T \text{Cov}(\tilde{\mathbf{e}}_i | \tilde{\mathbf{X}}_i^*) \tilde{\mathbf{X}}_i) = E(\tilde{\mathbf{X}}_i^T \tilde{\Sigma} \tilde{\mathbf{X}}_i)$$

$$\begin{aligned} &= \begin{pmatrix} \sigma_1^2 E(X_{i1}) & \sigma_{12} E(\sqrt{X_{i1} X_{i2}}) & \dots & \sigma_{1r} E(\sqrt{X_{i1} X_{ir}}) \\ \sigma_{12} E(\sqrt{X_{i1} X_{i2}}) & \sigma_2^2 E(X_{i2}) & \dots & \sigma_{2r} E(\sqrt{X_{i2} X_{ir}}) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1r} E(\sqrt{X_{i1} X_{ir}}) & \sigma_{2r} E(\sqrt{X_{i2} X_{ir}}) & \dots & \sigma_r^2 E(X_{ir}) \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1^2 E(X_1) & \sigma_{12} E(\sqrt{X_1 X_2}) & \dots & \sigma_{1r} E(\sqrt{X_1 X_r}) \\ \sigma_{12} E(\sqrt{X_1 X_2}) & \sigma_2^2 E(X_2) & \dots & \sigma_{2r} E(\sqrt{X_2 X_r}) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1r} E(\sqrt{X_1 X_r}) & \sigma_{2r} E(\sqrt{X_2 X_r}) & \dots & \sigma_r^2 E(X_r) \end{pmatrix} \end{aligned}$$

where,

$$E(\tilde{\mathbf{X}}_i^T \tilde{\mathbf{e}}_i | \tilde{\mathbf{X}}_i^*) = \tilde{\mathbf{X}}_i^T E(\tilde{\mathbf{e}}_i | \tilde{\mathbf{X}}_i^*) = \mathbf{0},$$

Efficient Confidence Intervals for the Defective Totals using SUR
Framework

<http://www.doi.org/10.62341/istj-vol39-1-NA68>

$$\tilde{\mathbf{X}}_i^* = (X_{i1}, X_{i2}, \dots, X_{ir})^T$$

It follows that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\mathbf{X}}_i^T \tilde{\mathbf{e}}_i \rightarrow^D N_r(\mathbf{0}, \tilde{\Sigma}^*),$$

hence, by Slutsky's lemma ([9], pp 119-120), Eq. (1) can be formulated as

$$\sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}) \rightarrow^D N_r(\mathbf{0}, \text{diag}(E(X_1), \dots, E(X_r))^{-1} \times \tilde{\Sigma}^* \times \text{diag}(E(X_1), \dots, E(X_r))^{-1})$$

$$\begin{aligned} &\equiv N_r \left(\mathbf{0}, \begin{pmatrix} \frac{\sigma_1^2}{E(X_1)} & \frac{\sigma_{12}E(\sqrt{X_1 X_2})}{E(X_1)E(X_2)} & \dots & \frac{\sigma_{1r}E(\sqrt{X_1 X_r})}{E(X_1)E(X_r)} \\ \frac{\sigma_{12}E(\sqrt{X_1 X_2})}{E(X_1)E(X_2)} & \frac{\sigma_2^2}{E(X_2)} & \dots & \frac{\sigma_{2r}E(\sqrt{X_2 X_r})}{E(X_2)E(X_r)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1r}E(\sqrt{X_1 X_r})}{E(X_1)E(X_r)} & \frac{\sigma_{2r}E(\sqrt{X_2 X_r})}{E(X_2)E(X_r)} & \dots & \frac{\sigma_r^2}{E(X_r)} \end{pmatrix} \right) \\ &\equiv N_r(\mathbf{0}, \Sigma^{**}) \end{aligned} \quad (2)$$

where,

$$\Sigma^{**} = \begin{pmatrix} \frac{\sigma_1^2}{E(X_1)} & \frac{\sigma_{12}E(\sqrt{X_1 X_2})}{E(X_1)E(X_2)} & \dots & \frac{\sigma_{1r}E(\sqrt{X_1 X_r})}{E(X_1)E(X_r)} \\ \frac{\sigma_{12}E(\sqrt{X_1 X_2})}{E(X_1)E(X_2)} & \frac{\sigma_2^2}{E(X_2)} & \dots & \frac{\sigma_{2r}E(\sqrt{X_2 X_r})}{E(X_2)E(X_r)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1r}E(\sqrt{X_1 X_r})}{E(X_1)E(X_r)} & \frac{\sigma_{2r}E(\sqrt{X_2 X_r})}{E(X_2)E(X_r)} & \dots & \frac{\sigma_r^2}{E(X_r)} \end{pmatrix}$$

5. Approximate confidence intervals for the linear combination of $\hat{p}$

By the Cramer-Wold device ([9], p.147), Eq . (2) holds iff, for any vector

$\mathbf{a} = (a_1, \dots, a_r)^T \in \mathbf{R}$, we have

$$\begin{aligned} \sqrt{n}(\mathbf{a}^T \hat{\mathbf{p}} - \mathbf{a}^T \mathbf{p}) &\rightarrow^D N(0, \mathbf{a}^T \Sigma^{**} \mathbf{a}) \\ &\rightarrow^D N\left(0, a_j^2 \frac{\sigma_j^2}{E(X_j)} + 2 \sum_{j^l=2}^r \sum_{j=1}^{j^l-1} a_j a_{j^l} \frac{\sigma_{jj^l} E(\sqrt{X_j X_{j^l}})}{E(X_j)E(X_{j^l})}\right) \end{aligned} \quad (3)$$

Now, consider the consistent estimators

$$\bar{X}_{.j} = \frac{1}{n} \sum_{i=1}^n X_{ij}, \bar{X}_{.jj^l} = \frac{1}{n} \sum_{i=1}^n \sqrt{X_{ij} X_{ij^l}}, \quad j \neq j^l, \quad j, j^l = 1, \dots, r.$$

And

$$\begin{aligned} S_j^2 &= \frac{1}{n-1} \sum_{i=1}^n \frac{(U_{ij} - X_{ij} \hat{p}_j)^2}{X_{ij}} = \frac{1}{n-1} \sum_{i=1}^n (\tilde{U}_{ij} - \hat{p}_j \sqrt{X_{ij}})^2 \\ S_{jj^l}^2 &= \frac{1}{n-1} \sum_{i=1}^n \frac{(U_{ij} - X_{ij} \hat{p}_j)(U_{ij^l} - X_{ij^l} \hat{p}_{j^l})}{\sqrt{X_{ij}} \sqrt{X_{ij^l}}} \\ &= \frac{1}{n-1} \sum_{i=1}^n (\tilde{U}_{ij} - \sqrt{X_{ij}} \hat{p}_j)(\tilde{U}_{ij^l} - \sqrt{X_{ij^l}} \hat{p}_{j^l}), \\ \tilde{U}_{ij} &= \frac{U_{ij}}{\sqrt{X_{ij}}}, \quad \tilde{U}_{ij^l} = \frac{U_{ij^l}}{\sqrt{X_{ij^l}}}, \quad X_{ij}, X_{ij^l} > 0, \forall_{j \neq j^l}, \quad j, j^l = \\ &1, \dots, r, i = 1, \dots, n. \end{aligned}$$

Efficient Confidence Intervals for the Defective Totals using SUR
Framework

<http://www.doi.org/10.62341/istj-vol39-1-NA68>

Substituting these entries for the corresponding parameters in the asymptotic variance in Eq. (3), we obtain the standard error of $\mathbf{a}^T \hat{\mathbf{p}}$

$$s.e(\mathbf{a}^T \hat{\mathbf{p}}) = \sqrt{\frac{1}{n} \left(\sum_{j=1}^r a_j^2 \frac{s_j^2}{\bar{X}_j} + 2 \sum_{j^l=2}^r \sum_j^{j^l-1} a_j a_{j^l} \frac{s_j \bar{X}_j \bar{X}_{j^l}}{\bar{X}_j \bar{X}_{j^l}} \right)}$$

Therefore, the asymptotic confidence interval for the defective total U is obtained using the z-quantile

$$\left[\hat{U} \pm Z_{1-\frac{\beta}{2}} s.e(\hat{U}) \right]$$

Where

$$U = \mathbf{a}^T \mathbf{p} = \sum_{j=1}^r a_j p_j = \sum_{j=1}^r X_j p_j ,$$

$$\hat{U} = \mathbf{a}^T \hat{\mathbf{p}} = \sum_{j=1}^r a_j \hat{p}_j = \sum_{j=1}^r X_j \hat{p}_j , \quad a_j = X_j , \quad X_j = \sum_{i=1}^n X_{ij} , \quad j = 1, \dots, r.$$

And, the conservative confidence interval by the t – quantile
Is given by

$$\left[\hat{U} \pm t_{(n-1, 1-\frac{\beta}{2})} s.e(\hat{U}) \right]$$

where $t_{(n-1, 1-\frac{\beta}{2})}$ is the $\left(1 - \frac{\beta}{2}\right)$ quantile

of the t – distribution with $(n - 1)$ degrees of freedom, under the approximation that

$$\frac{\hat{U} - U}{s.e(\hat{U})} \sim t_{n-1}$$

Efficient Confidence Intervals for the Defective Totals using SUR
Framework

<http://www.doi.org/10.62341/istj-vol39-1-NA68>

6. Simulation Study

For approximating the true coverage of the total, we considered two confidence intervals, the asymptotic (with the z -quantile) and the conservative (with the t -quantile). i.e, approximating the true coverage of the total U . We simulated the case $r = 2$, (using R program, R is a free software environment for statistical computing and graphics and can be downloaded from CRAN Mirrors on <http://www.r-project.org>).

The sample points U_{ij}, X_{ij} , $i = 1, \dots, n, j = 1, 2$ for sample sizes $n = 10, 15, 20, 50, 100$ from the relevant Poisson distribution given in Sec. 2 with different positive values of the model parameter λ with 10000 replications, and $a_1 = 1, a_2 = 2, a_j \geq 0$ (arbitrary).

We use the sample variances and covariance s_j^2, s_{12} , $j = 1, 2$, for the standard error of the linear combination given Sec.5. The coverage is displayed graphically in the figures and in the corresponding tables with horizontal lines indicating the nominal level. As a coverage of the confidence intervals demonstrates a similar pattern for all the confidence levels, it is enough to present the results at the nominal level 0.95.

Invalid intervals are excluded (that the intervals with invalid standard errors arising from non-positive observations or non-positive variances) these intervals are not counted and the percentage of such exclusions is reported. Note that, taking only the informative observations reduced the actual sample size n to the reduced random sample size n_j . Hence we adjust the degrees of freedom for the t -quantile on the reduced sample size $\min(n_j)$, where n_j is the number of the informative observations of X_j , $j = 1, 2$. Finally, due to space constraints we present only the most interesting results.

For very small values of the Poisson parameter, the procedure yields results many non-informative observations.

- 1) For small Poisson parameter values (< 1), the coverage of confidence intervals based on the z -quantile is closer to the nominal level especially at $n > 20$, while the confidence intervals based on the adjusted t -quantile lie above the nominal level for all sample sizes, Fig. 1 (left and right subfigures).

Efficient Confidence Intervals for the Defective Totals using SUR
Framework

<http://www.doi.org/10.62341/istj-vol39-1-NA68>

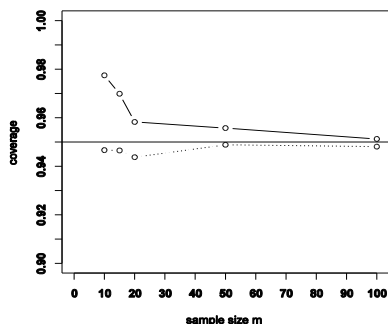


Fig. 1a

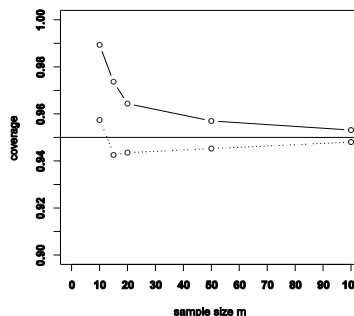


Fig. 1b

Fig. 1 Coverage of confidence intervals based on the z -quantile (dashed lines) and the adjusted t -quantiles (solid lines) for $\lambda_0 = 0.2$, $\lambda_1 = 0.3$, $\lambda_2 = 0.3$, $\mu_1 = 0.4$, $\mu_2 = 0.3$ (left subfigure). And for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = 0.5$ and $a_1 = 1$, $a_2 = 2$ (right subfigure).

Tables 1, 2 show that the percentage of the excluded intervals which decreases as the sample size becomes larger by both quantiles. Moreover, it is specially high for small n and λ , because the simulation produced results many non-informative observations, that is invalid intervals that are excluded, where λ denotes the chosen model parameter values.

Table 1 The percentage of excluded confidence intervals for $\lambda_0 = 0.2$, $\lambda_1 = 0.3$, $\lambda_2 = 0.3$, $\mu_1 = 0.4$, $\mu_2 = 0.3$ and $a_1 = 1$, $a_2 = 2$.

	sample size n				
	10	15	20	50	100
The percentage of the excluded intervals	0.9249	0.3090	0.2651	0.1376	0.0592

Table 2 The percentage of excluded confidence intervals for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = 0.5$ and $a_1 = 1$, $a_2 = 2$.

	sample size n			
	10	15	20	50
The percentage of the excluded intervals	0.3570	0.2935	0.2552	0.1283
	0.0504			

3) For larger parameter values (≥ 1)

Efficient Confidence Intervals for the Defective Totals using SUR
Framework

<http://www.doi.org/10.62341/istj-vol39-1-NA68>

The coverage of the t -quantiles method is similar to that for small parameter values and the coverage of the z -quantile is better for ($n \geq 20$) but falls below the nominal for ($n < 20$), for both parameter setting e as seen in Fig. 2 (Fig. 2a, Fig. 2b). The corresponding Table 3 shows that the exclusions decrease as the sample sizes increase.

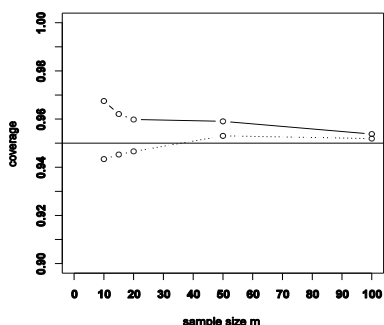


Fig. 2a

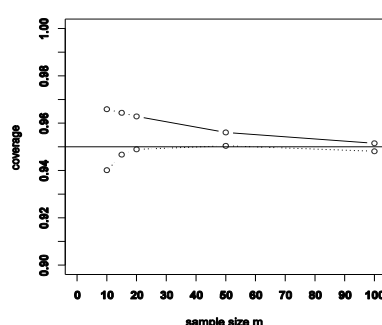


Fig. 2b

Fig. 2 Coverage of confidence intervals based on the z -quantiles (dashed lines) and adjusted t -quantiles (solid lines) for $\lambda_0 = 2$, $\lambda_1 = 3$, $\lambda_2 = 3$, $\mu_1 = 4$, $\mu_2 = 3$ (left subfigure). And for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = 5$ and $a_1 = 1$, $a_2 = 2$ (right subfigure).

4) On the other side, Fig. 3 (left and right subfigures) shows that, the proposed t -quantile method gives reasonable coverage (recommended for the given parameters) for both sample sizes and parameter values $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda = 0.5, 0.8, 1, 2, 5$, while the z -quantile method produced dramatically lower coverage specially for small sample size, for all λ as seen in Fig. 3a (not recommended for the given parameters).

Efficient Confidence Intervals for the Defective Totals using SUR
Framework

<http://www.doi.org/10.62341/istj-vol39-1-NA68>

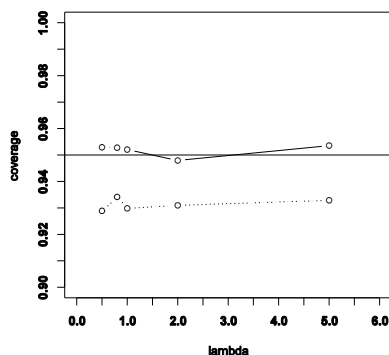


Fig. 3a

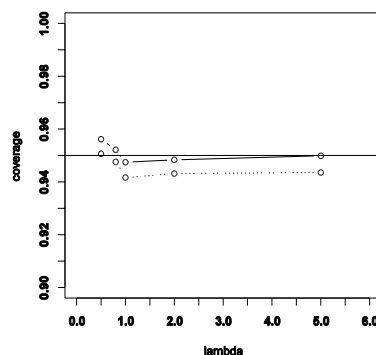


Fig. 3b

Fig. 3 Coverage for $n = 15$ and for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$ (left subfigure). And for $n = 50$, $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$ (right subfigure).

Table 3 The percentage of excluded confidence intervals for $\lambda_0 = 2$, $\lambda_1 = 3$, $\lambda_2 = 3$, $\mu_1 = 4$, $\mu_2 = 3$ and $a_1 = 1$, $a_2 = 2$.

	sample sizes n				
	10	15	20	50	100
The percentage of the excluded intervals	0.3287	0.2814	0.2462	0.1348	0.0557

Table 4 The percentage of excluded confidence intervals for the sample size $n = 15$ and for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$ and $a_1 = 1$, $a_2 = 2$.

	λ				
	0.5	0.8	1	2	5
The percentage of the excluded intervals	0.2884	0.2802	0.2803	0.2742	0.2700

Table 5 The percentage of excluded confidence intervals for the sample size $n = 50$ and for $\lambda_0 = \lambda_1 = \lambda_2 = \mu_1 = \mu_2 = \lambda$ and $a_1 = 1$, $a_2 = 2$.

	λ				
	0.5	0.8	1	2	5
The percentage of the excluded intervals	0.1290	0.1343	0.1382	0.1308	0.1302

Efficient Confidence Intervals for the Defective Totals using SUR
Framework

<http://www.doi.org/10.62341/istj-vol39-1-NA68>

From the Tables 4 and 5, we observe that: the exclusions tend to decrease slightly as the parameter values λ becomes larger (although not monotonically in Table 5), and that the exclusions in Table 5 are lower than in Table 4 as the sample sizes increases ($n = 15$ to $n = 50$).

7. Conclusion

For small parameter values the procedure as produced more non-informative observations with high probability. On the other side, as n or λ become larger, the percentage of excluded intervals becomes smaller.

The proposed confidence intervals based on the t -quantile highly conservative specially at small sample sizes and small λ (not recommended); there coverage exceeds that of the z -quantile intervals, whereas the z -quantile intervals are recommended when λ is larger and the sample sizes is sufficiently large.

We recommend that the conservative confidence intervals not be used for small parameter values and small sample sizes, while the asymptotic confidence intervals are preferable, as they require only a sufficiently large sample size.

Finally, when the ratio is near to 0.5, one expect better coverage; however, when its close to 0 or 1 the confidence intervals produce larger coverage.

References

- [1]. Agresti A., and Coull, B.A., 1998, "Approximate is better than exact for interval estimation of binomial proportions", The American Statistician, 52, pp 119-126.
- [2]. Brown, L.D., Cai, T., and DasGupta, A., 1999, "Interval estimation for a binomial proportion", Technical Report, Department of Statistics, Purdue University.
- [3]. Santner, T. J., 1998, "A note on teaching binomial confidence intervals", Teaching Statistics, 20, pp 20-23.
- [4]. Zellner, A., 1962, "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias", J. Amer. Statist. Ass., 57, pp 348-368.
- [5]. Johnson, N., Kotz, S., and Balakrishnan, N., 1997, "Discrete Multivariate Distributions", Wiley, New York, pp 124-138.
- [6]. Karlis, D., and Tsiamirtzis, P., 2007, "Exact Bayesian modeling

Efficient Confidence Intervals for the Defective Totals using SUR
Framework

<http://www.doi.org/10.62341/istj-vol39-1-NA68>

- for bivariate Poisson data and extensions", Stat. Comput., 18, pp 27-40.
- [7]. Timm, N.H., 2002, "Applied Multivariate Analysis", Springer Verlag, Inc, New York.
- [8]. Stock, J .H., and Waston, M.W., 2003, "Introduction to econometrics", Pearson Education, Inc, pp 588-591.
- [9]. Knight, K.C., 2000, "Mathematical Statistics", Chapman & Hall/CRC, Texts in Statistical Science, pp 120-149.
- [10]. Aitken, A.C., 1936, "A further note on multivariate selection", Proceedings of the Edinburgh Mathematical Society, 5, 37-40.
- [11]. Simon, Gary, 1996, "z or t for Binomial Confidence Intervals" Teaching Statistics, 18(3), pp 90-91.