MEASUREMENT OF WITHIN-BUSY-HOUR TRAFFIC VARIATIONS
Ole A. Pedersen
GTE Automatic Electric Laboratories
Northlake, Illinois USA

ABSTRACT

The following model is used to describe variations in traffic loading a group of \( m \) trunks observed at fixed intervals within a busy hour:

\[
X_s = \mu + c_1 s + c_2 s^2 + e_s
\]

where \( X_s \) is the number of trunks observed busy at observation \( s \); \( \mu \) is a measure of a constant component in the traffic; \( c_1 \) and \( c_2 \) are measures of linear and quadratic trends; and \( e_s \) is the random component. Least squares procedures are described for estimating the values of \( \mu, c_1 \) and \( c_2 \). The properties of \( e_s \) are studied using analysis of variance techniques. Procedures are presented for estimating the value of an intraclass correlation coefficient used to measure the degree of non-smoothness of the carried traffic. Procedures are also presented for estimating and accounting for the effects of correlation between observations due to the effects of call holding times.

The data base for these procedures consists of a record of the results of individual observations in a series of counts of busy trunks made at uniform intervals and a count of the number of calls handled in each period. These data are readily obtained in the field with modern traffic measurement equipment.

The procedures are illustrated by application to real traffic data.

Procedures for statistical significance testing and determining confidence intervals for estimates of parameters are not described or illustrated in the applications. They are adequately treated in statistical texts.

1. INTRODUCTION

A descriptive mathematical model of within-busy-hour traffic carried by a finite trunk group is presented. Terms in the model account for traffic variations due to the presence of linear and quadratic trends and a random component. Least squares procedures are used to estimate the magnitudes of the trends. Analysis of variance (ANOVA) techniques are used in studying the characteristics of the random component. Procedures are presented for estimating the value of an intraclass correlation coefficient used to measure the degree of non-smoothness of the carried traffic. The correlation coefficient is applicable to fitting a curve to the distribution of the number of trunks observed busy [3]. Procedures are also presented for estimating and accounting for the effects of correlation between observations on the analysis.

The data base consists of a record of the results of individual observations in a series of counts of busy trunks made at uniform intervals within a busy hour and a count of the number of calls handled in that period.

The procedures are illustrated by application to real traffic data.

Procedures for statistical significance testing and determining confidence intervals for estimates of parameters are not described or illustrated. They are adequately treated in statistical texts [4].

This paper supplements an earlier work on measurement of traffic variations [3]. Here we explicitly define a model of observed traffic and use it to make a more rigorous analysis of within-busy-hour traffic variations. As a result of the analysis, we obtain better defined conditions for testing the statistical significance of measured trends in the data in the presence of autocorrelation and a different procedure for adjusting estimates of the intraclass correlation coefficient for autocorrelation.

2. THE USAGE DATA

We shall study data collected in the following way:

A series of \( n \) observations are made at uniform intervals of time of the occupancy of a group of \( m \) trunks. A record is made of the number of trunks found busy at each observation. Thus a series of \( n \) numbers \( X_1, X_2, \ldots, X_n \) are obtained in which \( X_n = 0 \) is the number of trunks found busy at observation \( n \).

In addition to obtaining usage data, a count is made of the number of calls \( c \) that have been placed on the trunk group in the time between the first and last observations.

Since the values of \( X \) and \( X_n \) usually do not equal zero because of the existence of calls not beginning and ending within the observation...
period, the number of counted calls \( c^* \) needs adjustment as a measure of the number of calls \( c \) contributing to the observed usage of the trunk group. The number of calls \( c \) is estimated by letting

\[
c = c^* + (X_l - X_n)/2.
\]

This estimate is based on the assumption that calls incompletely handled within the study period contribute to the observed usage one-half of the average usage contributed per call completely handled within the period.

3. DESCRIPTIVE MODEL OF BUSY-HOUR TRAFFIC

The observed traffic carried by a group of \( m \) trunks shall be described by the following model of the observed state \( x_{is} \) of trunk \( i \) at the \( s \)th of \( n \) observations made within a busy hour

\[
x_{is} = \mu + \xi_1(s) + \alpha \xi_2(s) + e_{is}.
\]

In this model \( x_{is} \) is an indicator variable taking the value 1 when trunk \( i \) is found busy at observation \( s \) and equals 0 otherwise. This is a fixed effects model with deterministic constant component \( \mu \) and linear and quadratic trend components represented by the functions \( \xi_1(s) \) and \( \xi_2(s) \). The random component is represented by \( e_{is} \). According to the conditions of the study, we do not have data on the usage of individual trunks. Therefore, we shall use the model to account for differences in trunk usage.

The functions \( \xi_1(s) \) and \( \xi_2(s) \) are orthogonal polynomials. \( \xi_1(s) \) is a linear polynomial and \( \xi_2(s) \) is a quadratic polynomial in \( s \). They have the following properties:

\[
\sum_{s=1}^{n} \xi_1(s) = \sum_{s=1}^{n} \xi_2(s) = \sum_{s=1}^{n} \xi_1(s)\xi_2(s) = 0
\]

\[
\sum_{s=1}^{n} \xi_1(s)^2 = \sum_{s=1}^{n} \xi_2(s)^2 = n.
\]

In particular, \( \xi_1(s) \) and \( \xi_2(s) \) are as follows:

\[
\xi_1(s) = \lambda_1 (s - n(n + 1)/2)
\]

\[
\xi_2(s) = \lambda_2 ((s - n(n + 1)/2)^2 - (n^2 - 1)/12)
\]

where \( \lambda_1 \) and \( \lambda_2 \) are suitable scaling constants chosen to satisfy the above conditions. Tables of such orthogonal polynomials may be found in the references [2], [4]. They also have the property that

\[
\xi_1(s)^2 = \xi_1(n-s+1)^2.
\]

Let the expectations of the random variable \( e_{is} \) and its square be

\[
E e_{is} = 0 \quad E e_{is}^2 = \sigma^2
\]

where \( E \) is the expectation operator. Let \( e_{is} \) be defined by

\[
e_{is} = \frac{m}{n} e_{is}\]

and let the expectation of its square be

\[
E e_{s}^2 = m(1 + (m-1)\rho)\sigma^2
\]

where \( \rho \) is an intraclass correlation coefficient. If the value of \( \rho \) is zero, then we have "smooth traffic" and the distribution of the number of busy trunks is described by the binomial or Bernoulli distribution [3]. Furthermore, let

\[
E \sum_{s \neq t} e_{st} e_{st} = n(n-1)\rho m(1 + (m-1)\rho)\sigma^2
\]

where \( \rho \) is a mean autocorrelation coefficient. Finally, let \( e \) be defined by

\[
e = \frac{1}{s} \sum_{s} e_{is}.
\]

The expectation of its square is

\[
E e^2 = nm(1 + (n-1)\rho)(1 + (m-1)\rho)\sigma^2
\]

Let \( x_{s} \) be the number of trunks observed busy at observation \( s \) and \( X \) be the total number of busy trunk counts - i.e.,

\[
x_{s} = \frac{m}{n} x_{is} \quad X = \sum_{s} x_{s}.
\]

Let \( x \) be the average measured usage per trunk

\[
x = X/nm.
\]

The expectations of \( x_{s} \) and \( X \) are

\[
E x_{s} = m \mu + m \alpha \xi_1(s) + m \alpha \xi_2(s)
\]

\[
E X = nm \mu
\]

We also have

\[
E \sum_{s} x_{s}^2 = n m^2 \mu^2 + n m^2 \alpha^2 + n m(1 + (m-1)\rho)\sigma^2
\]

\[
E X^2 = n m^2 \mu^2 + n m(1 + (n-1)\rho)(1 + (m-1)\rho)\sigma^2
\]

where we have let

\[
\alpha^2 = \alpha_1^2 + \alpha_2^2.
\]

If we define \( \sigma^2 \) by

\[
\sigma^2 = \mu(1-\mu)
\]

then

\[
\sigma_0^2 = \sigma^2 - \sigma^2.
\]

The values of \( \mu \), \( \alpha_1 \) and \( \alpha_2 \) are unknown and must be estimated from the data. Using the mark \( \hat{\cdot} \) to identify estimates, we shall use the method of least squares to determine the values of \( \hat{\mu} \), \( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \) minimizing the quantity

\[
\sum_{s} e_{s}^2
\]

where \( e_{s} \) is an estimate of \( e_{is} \) given by

\[
\hat{e}_{s} = X_{s} - m(\hat{\mu} + \hat{\alpha}_1 \xi_1(s) + \hat{\alpha}_2 \xi_2(s))
\]

The least squares estimates of \( \mu \), \( \alpha_1 \) and \( \alpha_2 \) are

\[
\hat{\mu} = X/nm
\]
\[
\hat{\alpha}_1 = \frac{n}{S} \sum_{s} X_{s}^2 \hat{\xi}_1(s)/nm \quad \hat{\alpha}_0 = \frac{n}{S} \sum_{s} X_{s}^2 \hat{\xi}_0(s)/nm.
\]

We have
\[
E \hat{\alpha}_1 = \alpha_1 \quad E \hat{\alpha}_0 = \alpha_0
\]

so that \(\hat{\alpha}_1\) and \(\hat{\alpha}_0\) are unbiased estimates of \(\alpha_1\) and \(\alpha_0\).

Let \(\hat{\varepsilon}_{is}\) be an estimate of \(\varepsilon_{is}\) and \(\hat{\varepsilon}_s\) be an estimate of \(\varepsilon_{s}\) defined by
\[
\hat{\varepsilon}_{is} = x_{is} - (\hat{\mu} + \hat{\alpha}_1 \hat{\xi}_1(s) + \hat{\alpha}_0 \hat{\xi}_0(s)) \quad \hat{\varepsilon}_s = \sum_{i} \hat{\varepsilon}_{is}.
\]

Therefore
\[
E \hat{\varepsilon}_{is} = 0 \quad E \hat{\varepsilon}_s = 0.
\]

If \(\hat{\varepsilon}_s^2\) and \(S(e)\) are defined by
\[
\hat{\varepsilon}_s^2 = \hat{\varepsilon}_1^2 + \hat{\varepsilon}_2^2 \quad S(e) = \sum_{i} \hat{\varepsilon}_{is}^2
\]

then
\[
S(s) = nm \hat{\varepsilon}_s^2.
\]

Let \(R(\tau)\) be the autocorrelation function for the random variable \(e_s\),
\[
R(\tau) = E e_s e_{s+\tau}/m(1+m-1)\rho \sigma_0^2.
\]

\(R(\tau)\) is an even function of \(\tau\) and it has been shown in [3] that
\[
\sum_{\tau=1}^{(n-1)} (n-\tau) R(\tau) = 2 \sum_{s=1}^{(n-1)} (n-s) E \xi_1(s) \xi_1(s+\tau)/n.
\]

The expectations of \(nm \hat{\varepsilon}_s^2\) and \(nm \hat{\varepsilon}_s^2\) are
\[
nm E \hat{\varepsilon}_s^2 = nm \varepsilon_1^2 + (1+m-1)\rho \sigma_0^2
\]

\[
(n-1) R(\tau) \sum_{s=1}^{(n-1)} E \xi_1(s) \xi_1(s+\tau)(1+m-1)\rho \sigma_0^2/n.
\]

The effects of the last term in these expressions was studied by numerically examining functions \(\hat{\theta}_1(\tau)\) and \(\hat{\theta}_2(\tau)\) having the following form
\[
\hat{\theta}_1(\tau) = ((n-\tau) E \xi_1(s) \xi_1(s+\tau))/h. \quad (i = 1, 2)
\]

We have
\[
2 \sum_{\tau=1}^{(n-1)} (R(\tau) - \rho^\tau) \hat{\theta}_i(\tau)
\]

so that, if the left hand side of this equation equals zero, then
\[
2 \sum_{\tau=1}^{(n-1)} R(\tau) \sum_{s=1}^{(n-\tau)} \xi(s) \xi(s+\tau)/n = -\rho^\tau
\]

\[
E m \hat{\varepsilon}_s^2 = nm \varepsilon_1^2 + (1-\rho^\tau)(1+m-1)\rho \sigma_0^2.
\]

The values of \(\hat{\theta}_1(\tau)\) and \(\hat{\theta}_2(\tau)\) were calculated for \(n = 5\) and \(n = 10\).

![Figure 1](image)

**FIGURE 1.** Functions \(\hat{\theta}_1(\tau)\) and \(\hat{\theta}_2(\tau)\) for \(n = 5\) and \(n = 10\).

The results are plotted in Figure 1. Note that \(\hat{\theta}_1(\tau)\) is a function of \(\tau/n\) rather than of \(\tau\). \(\hat{\theta}_1\) and \(\hat{\theta}_2\) give relatively small weight to \((R(\tau) - \rho^\tau)\) for values of \(\tau/n\) less than 0.1. Therefore, if \((R(\tau) - \rho^\tau)\) has small values beyond \(\tau/n = 0.1\), which would occur for calls with average holding time of the order of 0.1 or less of the length of the observation period, then we have
\[
2 \sum_{\tau=1}^{(n-1)} (R(\tau) - \rho^\tau) \hat{\theta}_i(\tau) = 0 \quad (i = 1, 2).
\]

V. E. Benes has studied the autocovariance and autocorrelation of finite trunk groups [1].

4. ANALYSIS OF THE RANDOM COMPONENT
IN THE USAGE DATA

We have least squares estimates of the values of \(\mu\), \(\alpha_1\) and \(\alpha_0\) and still have to get estimates of \(\rho\) and \(\rho^\tau\). We shall use ANOVA techniques to do this [3].

We calculate the following sums of squared deviations (SS) in our usage data
\[
S_1(X) = \sum_{s} (X_s - \bar{X}_s)^2/m = \sum_{s} X_s^2/m - \bar{X}^2/nm
\]
\[
S_2(X) = \sum_{i} \sum_{s} (X_{is} - \bar{X}_s)^2 = \bar{X} - \sum_{s} X_s^2/m
\]
The quantity $nm\sigma^2$ in these expressions is eliminated as follows where we have assumed that the data have been collected over a period that is at least 10 times as long as the average holding time of the carried calls.

Form the SS's

\[
S_1(e) = \frac{n}{m} \sum_{s} (e_{s} - \bar{e}_{s} - \bar{e}_{s}/m)^2 = S(X) - nm\sigma^2
\]
\[
S_2(e) = \frac{m}{n} \sum_{s} (e_{s} - \bar{e}_{s})^2 = S_2(X)
\]
\[
S(e) = \frac{m}{n} \sum_{s} e_{s}^2 = S(X) - nm\sigma^2.
\]

Their expectations are

\[
E[S_1(e)] = (n-3)(1-\rho\tau)^2(1+(m-1)\rho)\sigma_0^2
\]
\[
E[S_2(e)] = n(m-1)(1-\rho)\sigma_0^2
\]
\[
E[S(e)] = (nm\rho - (3+(n-3)\rho\tau)(1+(m-1)\rho))\sigma_0^2.
\]

Note that, if $\sigma_1^2$ and $\sigma_2^2$ equal zero, then $nm\sigma^2$, $nm\sigma^2$ and the mean squared deviation (MS) of $S_1(e)$ are alternative estimates of $(1-\rho\tau)(1+(m-1)\rho)\sigma_0^2$.

We have not yet arrived at the stage where we can estimate the values of $\rho$ and $\rho\tau$. We must estimate the amount of autocorrelation $\rho\tau$ in the data and adjust $S_1(e)$ for its presence.

An alternative way to measure usage is to measure the holding times of calls handled by the trunks. Let $u$ be the observed holding time of call $u$ (note that $u$ is an integer), then

\[
X = \frac{c}{u} X_u.
\]

The average measured holding time per call $h$ is

\[
h = X/c.
\]

Furthermore, we have

\[
X^2 = \sum_{u} X_u^2 + \sum_{u} X_u X_v = \frac{n}{s} X_s^2 + \frac{n}{s/t} X_s^2.
\]

Let $X_u$ be a random variable with mean $\mu_h$ and variance $\sigma_h^2$ so that

\[
E[X_u] = \mu_h, \quad E[X_u^2] = \sigma_h^2 + \mu_h^2.
\]

The holding time $h$ is an unbiased estimate of $\mu_h$. Because of the measurement method, $\mu_h$ generally is less than the actual mean holding time per call, say $k$.

Let the holding times of calls $u$ and $v$ be stochastically independent so that

\[
E[X_u X_v] = \mu_h^2 (u^2 v).
\]

Therefore

\[
E[X^2] = c\sigma_h^2 + c^2 \mu_h^2 = nm(1+(n-1)\rho\tau)\sigma_0^2 + n^2 m^2 \mu_h^2
\]
and

\[
\sigma_h^2 = nm(1+(n-1)\rho\tau)(1+(m-1)\rho)\sigma_0^2.
\]

Using the holding time data, we form the SS's

\[
S_1(U) = \frac{c}{u} X_u^2 - X^2/nc
\]
\[
S_2(U) = X - \frac{c}{u} X_u^2/n
\]
\[
S(U) = X - X^2/nc.
\]

Their expectations are

\[
E[S_1(U)] = (c-1)\sigma_h^2/n
\]
\[
E[S_2(U)] = c(n-m\mu_h)\mu_h/n - c\sigma_h^2/n
\]
\[
E[S(U)] = c(n-m\mu_h)\mu_h/n - \sigma_h^2/n.
\]

We adjust $S_1(e)$ for autocorrelation by forming $S_1^*(e)$ where

\[
S_1^*(e) = S_1(e) + (n-3)(cS_1(U)/m(c-1) - S_1(e)/(n-3))/m.
\]

Its expectation is

\[
E[S_1^*(e)] = (n-3)(1+(m-1)\rho)\sigma_0^2.
\]

The appearance of the quantity $(n-3)$ instead of $(n-1)$ in these equations implies that both instead of none of $nm\sigma^2$ and $nm\sigma^2$ have been found to be statistically significant in obtaining $S_1(e)$ from $S_1(X)$. As a practical matter we do not contemplate calculating $S_1(U)$ from a measurement of the holding times of carried calls. However, we do assume that information is available from which we can obtain the value of $\sigma_h^2$. Since $\rho\tau$ is non-negative for carried traffic [1], we do not adjust $S_1(e)$ to obtain $S_1^*(e)$ unless the adjustment is non-negative.
If we form the variance ratio $F_T$ where

$$F_T = \frac{(n-3) \text{SS}_1/U}{c(n-1) \text{SS}_0}$$

then we can obtain estimates of $(1-\rho_T)$ and $\rho_T$ in the form

$$\hat{(1-\rho)} = \frac{S_1(e)}{S_1^*(e)} = \frac{n}{(n+F_T-1)}$$
$$\hat{\rho} = \frac{\text{SS}_0/n - S_1(e)}{(n-3) \text{SS}_0/n}$$

By equating $S_1(U)$ with its expectation, we get expressions for $S_1^*(e)$ and $\hat{\rho}$ in terms of $c_h^2$:

$$S_1^*(e) = S_1(e) + \frac{(n-3)(c_h^2/n - S_1(e)/(n-3))}{n}$$
$$\hat{\rho} = \frac{S_0/n - S_1(e)}{(n-3) c_h^2/n - S_1(e)}$$

Using the SS's $S_1^*(e)$ and $S_2(e)$, we form the variance ratio $F_0$ where

$$F_0 = \frac{(n-3) \text{SS}_1(e)/(n-3) \text{SS}_2(e)}{c_h^2/n}$$

Our estimates of $(1-\rho)$ and $\rho$ are

$$\hat{(1-\rho)} = \frac{m}{(m+F_0-1)}$$
$$\hat{\rho} = \frac{(F_0-1)/m}{(F_0-1)}$$

5. AN EXAMPLE

We illustrate the procedures with an example. Table 1 presents the result of 30 observations of busy-hour usage of a group of 18 inter-office trunks. The observations were made over a one hour period at 2-minute intervals during which a peg count register scored 140 calls for the group. The number of trunks observed busy at each observation is plotted in Figure 2.

<table>
<thead>
<tr>
<th>Source</th>
<th>f</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>S_1(X)</td>
<td>29</td>
<td>16.37222</td>
<td>0.56456</td>
<td>2.45507</td>
</tr>
<tr>
<td>S_2(X)</td>
<td>510</td>
<td>117.27780</td>
<td>0.22996</td>
<td>20.59887</td>
</tr>
<tr>
<td>S(e)</td>
<td>537</td>
<td>126.64700</td>
<td>0.23551</td>
<td>4.08608</td>
</tr>
</tbody>
</table>

Table 2. ANOVA table for data in Table 1.

No measurements were made of the holding times of any calls. We shall illustrate the autocorrelation effects of calls having exponentially distributed holding times. In order to estimate the effects of exponential holding times on our data, we need an expression for estimating the value of $c_h^2$. We have from the work of R. I. Wilkinson [5] that, for values of $n$ much larger than the average holding time per call $k$,

$$\mu_h \approx (1-e^{-1/k})^{-1}$$
$$c_h^2 \approx \mu_h/(\mu_h-1)$$

Therefore, for calls with measured average holding time $h$,

$$c_h^2 \approx h/(h-1)$$

Note that this approximation requires that $h$ is not less than 1.0.

The value of $c_h^2/nm$ is 0.61676. It is greater than the MS of $S_1(e)$ 0.34034 of Table 3. The value of $F_T$ is 1.81224 from which we estimate the value of $\hat{\rho}_T$ to be 0.02636. Table 4 is the ANOVA table obtained from Table 3 after adjustment for autocorrelation. The value of $S^*(e)$ in the table is

$$S^*(e) = S_1^*(e) + S_2(e).$$
Its expectation is

\[ E S^*(e) = (n - 3(1 + (n - 1)p)) \alpha^2 \]

<table>
<thead>
<tr>
<th>Source</th>
<th>( f )</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_1^*(e) )</td>
<td>27</td>
<td>9.43804</td>
<td>0.34956</td>
<td>1.52008</td>
</tr>
<tr>
<td>( S_2^*(e) )</td>
<td>510</td>
<td>117.27780</td>
<td>0.22996</td>
<td></td>
</tr>
<tr>
<td>( S^*(e) )</td>
<td>537</td>
<td>126.71584</td>
<td>0.23597</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 4.** ANOVA table for random component adjusted for autocorrelation.

The value of \( F_0 \) is given in the F-column of Table 4 from which we estimate the value of \( \rho \) to be 0.02808.

6. **STATISTICAL TESTS AND INFERENCES**

The discussion has dealt with procedures for processing usage data and obtaining estimates of parameters in a descriptive mathematical model of the traffic under study. The matters of making statistical tests and inferences have not been considered.

The assumption that the random variables \( e_{i} \) and, in particular \( e_{5} \) have, say, the normal distribution is the critical assumption in enabling the use of procedures based on the normal distribution in testing the statistical significance and developing confidence intervals for estimates of \( \mu \), the regression coefficients \( \alpha_{i} \) and \( \beta_{i} \), and the intraclass correlation coefficient \( \rho \). The assumption is weakened by the fact that we are studying binomial-like proportions - i.e. the degree of utilization of a group of trunks - . The assumption is strengthened by making a variance stabilizing transformation of the data and analyzing the transformed data. The angular or arcsin transformation is a suitable transformation for the kind of data we are dealing with. It is described in [3] and in standard statistical texts [4].

7. **ACKNOWLEDGEMENTS**

The Author wishes to thank N. L. Hart and M. H. Steinberg of the General Telephone Company of California for providing the traffic data used in this paper.

**REFERENCES**


