

## Estimating the parameters of a nonhomogeneous Poisson process with linear rate

William A. Massey<sup>1</sup>, Geraldine A. Parker<sup>2</sup> and Ward Whitt<sup>3</sup>  
*AT&T Bell Laboratories, Murray Hill, NJ 07974-0636, USA*

Received November 1993; revised January 1995

Motivated by telecommunication applications, we investigate ways to estimate the parameters of a nonhomogeneous Poisson process with linear rate over a finite interval, based on the number of counts in measurement subintervals. Such a linear arrival-rate function can serve as a component of a piecewise-linear approximation to a general arrival-rate function. We consider ordinary least squares (OLS), iterative weighted least squares (IWLS) and maximum likelihood (ML), all constrained to yield a nonnegative rate function. We prove that ML coincides with IWLS. As a reference point, we also consider the theoretically optimal weighted least squares (TWLS), which is least squares with weights inversely proportional to the variances (which would not be known with data). Overall, ML performs almost as well as TWLS. We describe computer simulations conducted to evaluate these estimation procedures. None of the procedures differ greatly when the rate function is not near 0 at either end, but when the rate function is near 0 at one end, TWLS and ML are significantly more effective than OLS. The number of measurement subintervals (with fixed total interval) makes surprisingly little difference when the rate function is not near 0 at either end. The variances are higher with only two or three subintervals, but there usually is little benefit from going above ten. In contrast, more measurement intervals help TWLS and ML when the rate function is near 0 at one end. We derive explicit formulas for the OLS variances and the asymptotic TWLS variances (as the number of measurement intervals increases), assuming the nonnegativity constraints are not violated. These formulas reveal the statistical precision of the estimators and the influence of the parameters and the method. Knowing how the variance depends on the interval length can help determine how to approximate general arrival-rate functions by piecewise-linear ones. We also develop statistical tests to determine whether the linear Poisson model is appropriate.

<sup>1</sup> AT&T Bell Laboratories, Room 2C-120, Murray Hill NJ 07974-0636, USA. E-mail: will@research.att.com.

<sup>2</sup> Educational Testing service, Princeton, NJ 08554, USA. (This work was done at AT&T Bell Laboratories as part of an AT&T Summer Research Program fellowship.)

<sup>3</sup> AT&T Bell Laboratories, Room 2C-178, Murray Hill, NJ 07974-0636, USA. E-mail: wow@research.att.com.

## 1. Introduction and summary

Traffic measurements and traffic models have long played an important role in telecommunications systems, e.g., see Cole [5] and Rahko [17]. However, the emergence of new services and new technologies has led to new kinds of traffic and new traffic models, e.g., see Frost and Melamed [8], Leland, Taqqu, Willinger and Wilson [1], Meier-Hellstern, Wirth, Yan and Hoeflin [15] and Part I of Roberts [18]. Just as with overflow traffic associated with alternative routing, in many new situations the classical Poisson process traffic model is not nearly appropriate. On the other hand, in other situations it evidently is still appropriate.

We want to be able to determine if a Poisson process traffic model is appropriate and, when it is, we want to be able to estimate its parameters from measurements. These issues are relatively well understood in the familiar setting of *homogeneous* arrival processes, i.e., when the arrival rate is constant. However, in reality the arrival process is typically *nonhomogeneous*; i.e., the arrival rate typically varies significantly in time, e.g., see pp. 258–260 and chapter 6 of Hall [9]. It is much easier to detect departures from a homogeneous Poisson process than from a non-homogeneous Poisson process. Indeed, we may actually have a Poisson process when we think we do not, if we do not properly account for the nonhomogeneity; i.e., predictions of higher variability associated with non-Poisson homogeneous processes can often be explained by fluctuations in the deterministic arrival-rate function of a nonhomogeneous Poisson process; e.g., see Holtzman and Jagerman [10]. This led us in [13] to consider non-Poisson homogeneous processes as approximations for non-homogeneous Poisson processes.

A major difficulty with the nonhomogeneous Poisson process model is that it has infinitely many parameters. In particular, it is parameterized by its arrival-rate function  $\lambda(t)$ . A natural first step is to restrict attention to special parametric families of arrival-rate functions. In many cases it is reasonable to regard the arrival-rate function as *linear over appropriate subintervals*, i.e., as piecewise linear. For example, data show that the arrival-rate function for telephone calls over a day is very nonlinear, but it often is reasonable to approximate it by a linear function within single hours, especially when the available data are summaries over subintervals (e.g., five minutes). The linearity is a helpful simplification because it reduces the number of parameters.

A specific motivating application for us is the AT&T long distance network. Summaries of telephone calls by origin and destination (for the more than 100 major switches) are currently collected every five minutes. Fig. 1 displays these summaries for three origin-destination pairs in one hour. Since the time is between 8:00 am and 9:00 am, it is not surprising that there is significant increase over the hour. This example shows that the rates might not nearly be constant, but that they might be approximately linear over one hour.

Hence, in this paper we consider the case of a linear rate over a subinterval. In particular, we assume that we have a nonhomogeneous Poisson process over the

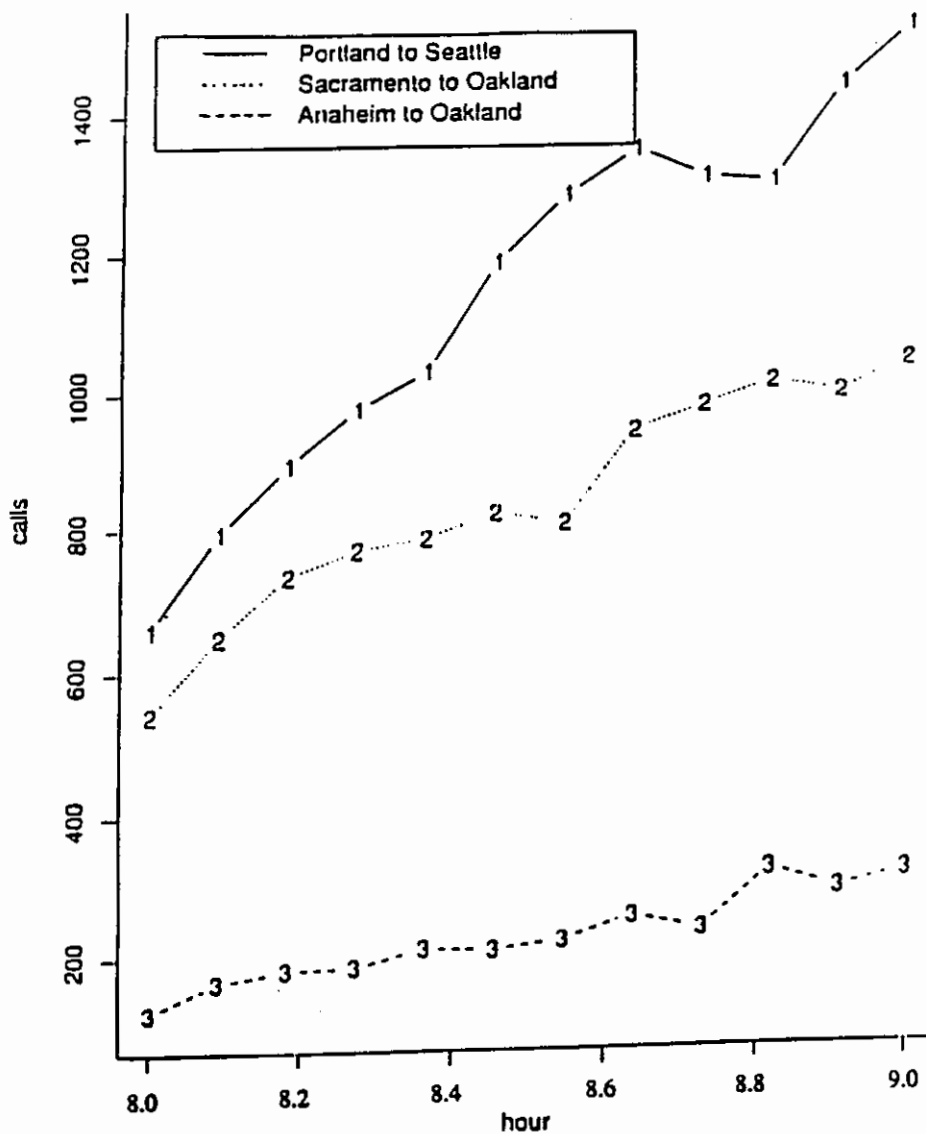


Fig. 1. Traffic count summaries for three origin-destination pairs in the AT&T long distance network during one hour.

interval  $[0, T]$  with arrival-rate function

$$\lambda(t) = a + bt, \quad 0 \leq t \leq T, \tag{1}$$

and we investigate how to estimate the two parameters  $a$  and  $b$  in (1) and the arrival rates  $\lambda(t)$  based on arrival process data. We assume that the overall time interval  $(0, T]$  is divided into  $N$  measurement subintervals  $((k - 1)T/N, kT/N]$ ,  $1 \leq k \leq N$ , and we observe the number of points in each. In addition to finding estimators  $\hat{a}$  and  $\hat{b}$  for the parameters  $a$  and  $b$  in (1), we want to provide insight into the choice of the interval length  $T$  and the number  $N$  of measurement subintervals. Having a larger  $T$  will yield better estimates if the arrival-rate function is indeed nearly linear,

but it can yield worse estimates if the arrival-rate function is only nearly linear over short intervals. Having a larger  $N$  may be more costly, but may improve the quality of the estimate.

We primarily focus on the variance of the estimators as a function of the parameters and the method. Explicit expressions for the variances enable us to see how the parameters  $a$ ,  $b$ ,  $T$  and  $N$  affect the estimation. We do not specifically discuss how to choose subintervals to approximate a general arrival-rate function by a piecewise-linear arrival-rate function, but our variance formulas can help make the choice.

Our estimation might be based on a single realization of an arrival process or multiple independent samples. For example, we might have data from the same hour on the same day of the week for  $n$  different weeks. If we consider the  $N$  sample means obtained from the  $Nn$  observations, then the variances of all the estimators we consider are inversely proportional to  $n$ . Since the effect of  $n$  is predictable, we henceforth only consider the case  $n = 1$ .

It is important to emphasize that we are interested in this linear Poisson process model from *two points of view*. First, we obviously want to understand how different estimation procedures perform when the model is approximately valid. Second, we want to develop ways to determine whether or not the linear Poisson model is appropriate. The model is not valid if the arrival process is not Poisson or if the arrival-rate function over the designated subinterval is nonlinear. We develop ways to test whether the linear Poisson model is appropriate. Such a test might be based on a single realization of the process or upon multiple realizations.

There is a substantial history of statistical inference for nonhomogeneous Poisson processes, e.g., see Basawa [1], Brown [2] and Snyder and Miller [19], but we are unaware of any work closely related to what we do here. In the physical sciences it is more natural to consider Poisson processes with nonlinear rates; see [1] and [19].

We are ultimately interested in characterizing arrival processes so that we can analyze the performance of telecommunications and other service systems with these arrival processes. There is a growing literature on ways to analyze service system models with nonhomogeneous Poisson arrival processes; e.g., see Choudhury, Lucantoni and Whitt [4], Davis, Massey and Whitt [6], Hall [9], Massey and Whitt [12], Taaffe and Ong [20] and references cited in these sources. In this paper we investigate ways to select nonhomogeneous Poisson arrival process models given arrival process data. Less attention has been given to fitting arrival process models to data; see Basawa [1] and Hall [9].

In our study we consider three different estimation procedures: (1) *ordinary least squares* (OLS), (2) *iterative weighted least squares* (IWLS) and (3) *maximum likelihood* (ML). These estimation procedures are described in detail in sections 2–5. OLS is the natural simple estimator; the other two are attempts to do better. The ML approach is discussed by Basawa [1] and Snyder and Miller [19]. We are led to consider weighted least squares, because the observations have unequal

variances. To achieve the best linear unbiased estimator, we would want weights inversely proportional to the variances (see section 2). Our IWLS procedure estimates the desired weights from estimates of the variances, which in turn depend on estimates of the parameters  $a$  and  $b$ , successively improving the estimates for both  $(a, b)$  and the weights by iteration. When our estimates of  $(ab)$  produce weights, which in turn reproduce these same estimates of  $(a, b)$ , the iteration has converged.

We prove a fundamental result linking these estimators; the proof appears in section 8. This is reminiscent of the classical result for linear models with i.i.d. normally distributed residuals; then ML coincides with OLS. Indeed, it can be deduced from the theory of generalized linear models; see p. 31 of McCullagh and Nelder [14].

#### THEOREM 1

The ML estimators coincide with the solution of IWLS.

We also study these estimators by computer simulation. Given the parameters  $a, b, T$  and  $N$ , we randomly generate the resulting Poisson random variables and study how the estimation procedures perform. Since we know the parameters with the simulation, we also can consider the *theoretical optimal weighted least squares* (TWLS), obtained by using weights inversely proportional to the known variances (depending on  $a$  and  $b$ ).

From our numerical experience, we find that the three estimation procedures (OLS, ML and TWLS) do not differ much from each other when the rate function is not nearly 0 at either end; i.e., they all tend to be unbiased and they all have nearly the same variances. The estimators fail to be precisely unbiased because of adjustments to account for nonnegativity constraints. Even though ML and IWLS coincide by Theorem 1, differences can occur due to numerical accuracy of the algorithms. We found that the ML procedure developed here was more robust for very large  $N$ .

We will show that OLS tends to be as good as TWLS when the slope  $b$  is relatively small, while TWLS tend to be significantly better than OLS when  $b$  is large and the rate function is near 0 at one endpoint. Indeed, we show that when the rate function is 0 at one endpoint, as occurs when the nonnegativity constraint is violated, then ML and TWLS coincide and that these are significantly more effective than OLS, primarily because they predict the intercept ( $a$  when  $a = 0$  and  $a + bT$  when  $a + bT = 0$ ) much better; i.e., the bias created by adjustments to the nonnegativity constraints is substantially less with ML and TWLS. The IWLS estimator usually converges very quickly (e.g., 2–5 iterations) to an estimator very close to the TWLS estimator.

As a basis for understanding the performance of these estimation procedures, we develop analytical formulas for the variances of the OLS and TWLS estimators, assuming that the nonnegativity conditions are not violated. Since the numerical results show that ML is very close to TWLS, these formulas give a clear quantitative

description of the statistical precision of both the OLS and ML procedures and the advantage of ML over OLS, provided that the rate function is indeed not near 0 at either end.

As should be anticipated, we find that the total length of the interval,  $T$ , is very important, so we should strongly prefer long intervals if linearity is not sacrificed. However, we find that the number  $N$  of measurement subintervals makes surprisingly little difference. When the rate function is not near 0 at either end, the variances decrease as  $N \rightarrow \infty$ , but rapidly converge to limits. The variance values are noticeably larger for  $N = 2, 3$  or  $4$ , but hardly at all by  $N = 10$ . The insignificance of  $N$  might be anticipated because the total number of Poisson counts in  $[0, T]$  is independent of  $N$ . The insignificance of  $N$  is established by extensive simulation results and explicit formulas for OLS. For OLS we have simple formulas for the variances for all  $N$ , but for TWLS we only have simple formulas for the asymptotic variances as  $N \rightarrow \infty$ ; we develop an algorithm to compute the TWLS variances for all finite  $N$ . Our numerical experience indicates that the dependence on  $N$  depicted by the OLS formulas applies approximately (but not exactly) to the dependence on  $N$  for the other estimators.

The role of  $N$  is different, however, when the rate function is near 0 at one end. Then increasing the number of measurement subintervals still does not help OLS, but it improves the performance of the ML and TWLS methods, with there being greater improvement for TWLS. For example, when  $a = 0$ , both the bias and the sample standard deviation are decreasing in  $N$  for TWLS. For  $a \approx 0$  and large  $N$ , ML does not perform as well as TWLS, but ML is significantly better than OLS in this setting.

Here is how the rest of the paper is organized. In section 2 we describe the weighted least squares procedures. In section 3 we present the formulas for the OLS variances. In section 4 we develop the asymptotic formulas for the TWLS variances. In section 5 we develop the ML estimation procedure. In section 6 we describe a simulation experiment to evaluate the estimation procedures and present some of the numerical results. In section 7 we develop statistical tests to see if the linear Poisson model is appropriate. In section 8 we prove theorem 1.

## 2. The least squares estimators

There are four parameters:  $a$ ,  $b$ ,  $T$  and  $N$ . We assume that we have a nonhomogeneous Poisson process with rate function  $\lambda(t) = a + bt$  over  $[0, T]$  as in (1). We count the number of points in the  $N$  subintervals  $((k-1)T/N, kT/N]$ ,  $1 \leq k \leq N$ . This sampling procedure from a single realization of the nonhomogeneous Poisson process over  $[0, T]$  produces  $N$  mutually independent Poisson random variables  $Y_k$  with means

$$\lambda_k = \frac{T}{N}(a + bx_k), \quad (2)$$

where

$$x_k = \left(k - \frac{1}{2}\right) \frac{T}{N}, \quad 1 \leq k \leq N. \quad (3)$$

If we form the linear model  $Y = \alpha + \beta x + \epsilon$ , i.e., if we assume that  $Y_k = \alpha + \beta x_k + \epsilon_k$  for each  $k$ , then we can estimate the parameters  $\alpha$  and  $\beta$  by *weighted least squares*; i.e., using positive weights  $w_k$ ,  $1 \leq k \leq N$ , with  $\sum_{k=1}^N w_k = N$ , we can choose estimators  $\hat{\alpha}$  and  $\hat{\beta}$  to minimize the weighted sum of the squared errors, i.e., to find

$$\min_{\alpha, \beta} \sum_{k=1}^N w_k (Y_k - [\alpha + \beta x_k])^2. \quad (4)$$

Applying calculus with (4) in the usual way, we obtain

$$\hat{\beta} = \frac{\sum_{k=1}^N w_k (x_k - \bar{x})(Y_k - \bar{Y})}{\sum_{k=1}^N w_k (x_k - \bar{x})^2} = \frac{\sum_{k=1}^N w_k (x_k - \bar{x}) Y_k}{\sum_{k=1}^N w_k (x_k - \bar{x})^2} \quad (5)$$

and

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{x}, \quad (6)$$

where

$$\bar{x} = N^{-1} \sum_{k=1}^N w_k x_k \quad \text{and} \quad \bar{Y} = N^{-1} \sum_{k=1}^N w_k Y_k. \quad (7)$$

From (2), we see that  $\alpha = aT/N$  and  $\beta = bT/N$ . Hence, associated estimators for  $a$  and  $b$  are

$$\hat{b} = \frac{N}{T} \hat{\beta} \quad \text{and} \quad \hat{a} = \frac{N}{T} \hat{\alpha} \quad (8)$$

for  $\hat{\beta}$  in (5) and  $\hat{\alpha}$  in (6). The resulting estimators  $\hat{a}$  and  $\hat{b}$  are linear functions of the observations  $Y_k$  and are unbiased; i.e.,  $E\hat{a} = a$  and  $E\hat{b} = b$ .

However, it is natural to impose the *constraint* that the estimated rate function  $\hat{a} + \hat{b}t$  be nonnegative throughout  $[0, T]$ . This constraint reduces to

$$\hat{a} \geq 0 \quad \text{and} \quad \hat{b} \geq -\hat{a}/T. \quad (9)$$

Because of the sum of squares in (4) is a convex function of  $(\alpha, \beta)$ , it is easy to determine the constrained minimum. If the constraints in (9) are satisfied by the solution

in (5)–(8), then we are done. If not, then either (i)  $\hat{a} < 0$  or (ii)  $\hat{a} \geq 0$  and  $\hat{b} < -\hat{a}/T$ . In the first case, the minimum is attained with  $\hat{a} = 0$  and  $\hat{b} = N\hat{\beta}/T$ , where  $\hat{\beta}$  is the solution to

$$\min_{\beta} \sum_{k=1}^N w_k (Y_k - \beta x_k)^2, \quad (10)$$

i.e.,

$$\hat{b} = \frac{N \sum_{k=1}^N w_k x_k Y_k}{\sum_{k=1}^N w_k x_k^2}. \quad (11)$$

In the second case the minimum is attained with  $\hat{a} = -\hat{b}T$  and  $\hat{b} = N\hat{\beta}/T$ , where  $\hat{\beta}$  is the solution to

$$\min_{\alpha} \sum_{k=1}^N w_k (Y_k - \beta(T - x_k))^2, \quad (12)$$

i.e.,

$$\hat{b} = -\frac{N \sum_{k=1}^N w_k (T - x_k) Y_k}{\sum_{k=1}^N w_k (T - x_k)^2}. \quad (13)$$

In this framework, the simplest procedure is *ordinary least squares* (OLS) in which all the weights are the same, i.e.,  $w_k = 1$ ,  $1 \leq k \leq N$ . However, we are motivated to consider unequal weights because, if  $b \neq 0$ , then the variances of the successive  $Y_k$  variables are unequal. Since the variable  $Y_k$  has a Poisson distribution, its variance equals its mean. Thus, from (2), we see that the variances are

$$\text{Var } Y_k = \lambda_k \equiv \frac{T}{N}(a + bx_k), \quad 1 \leq k \leq N. \quad (14)$$

Of course, since we do not know  $a$  and  $b$  in advance, we do not know these variances in advance. However, if we did know the variances in advance, then we might want to use the weights

$$w_k = \frac{N}{\lambda_k} \bigg/ \sum_{k=1}^N \left( \frac{1}{\lambda_k} \right), \quad 1 \leq k \leq N, \quad (15)$$

because these weights produce the minimum variance estimator among linear functions of the observations  $Y_k$  that are unbiased (assuming that we do not consider the



constraints in (9)). This can be verified by a direct calculation using Lagrange multipliers as for the Gauss-Markov theorem, as on p. 341 of Mood and Graybill [16], or by transforming the problem to the equal-variance case; see p. 78 of Draper and Smith [7] or p. 81 of Weisberg [21]. For the transformation, we replace  $Y_k = \alpha + \beta x_k + \epsilon_k$  with  $\sigma_k^{-1} Y_k = \alpha \sigma_k^{-1} + \beta \sigma_k^{-1} x_k + \sigma_k^{-1} \epsilon_k$ . Then  $\sigma_k^{-1} \epsilon_k = (\sigma_k^{-1} Y_k - \alpha \sigma_k^{-1} - \beta \sigma_k^{-1} x_k)$  are independent random variables with mean 0 and variance 1. Hence, the standard linear regression theory can indeed be applied to the transformed problem.

Since we do not actually know  $a$  and  $b$  in advance with data, we cannot use the weights in (15). However, in our simulation studies evaluating different estimators we can use the weights (15). We call this case *theoretically optimal weighted least squares* (TWLS).

We can approach TWLS above with data by using *iterative weighted least squares* (IWLS). We first perform ordinary least squares to obtain the estimates  $\hat{a}$  and  $\hat{b}$ . We then form the associated estimate  $\hat{\lambda}_k = (T/N)(\hat{a} + \hat{b}x_k)$  and use this in (15) to obtain weight estimates  $\hat{w}_k$ . We then successively obtain new estimates of  $(\hat{a}, \hat{b})$  and  $\hat{w}_k$  until insignificant change results. Our numerical experience indicates that IWLS consistently converges very quickly (e.g., 2–5 iterations) and that the resulting estimates are very close to the TWLS estimates. See Carroll [3], McCullagh and Nelder [14] and p. 87 of Weisberg [21] for further discussion about IWLS.

It is easy to see how the estimators are related when we make adjustments to satisfy the nonnegativity constraints in (9). Note that (11) becomes  $\hat{b} = N \sum_{k=1}^N x_k Y_k / T \sum_{k=1}^N x_k^2$  for OLS, but (11) becomes  $\hat{b} = N \sum_{k=1}^N Y_k / T \sum_{k=1}^N x_k = 2 \sum_{k=1}^N Y_k / T^2$  for both TWLS and IWLS, and similarly for (13). Hence, OLS coincides with TWLS when  $b = 0$ , while IWLS coincides with TWLS whenever the constraints (9) are violated.

We conclude this section by noting that, while it is natural to apply the adjustments to satisfy the nonnegativity constraints when we have a single data set  $\{(x_k, Y_k) : 1 \leq k \leq N\}$ , it is natural to combine the data before making adjustments, when we have multiple samples. It is easy to see that with a single data set, the adjustment minimizes mean square error.

### 3. Analytical formulas for ordinary least squares

If we ignore the nonlinear adjustment in the OLS estimators in order to satisfy the constraints in (9), then the estimators  $\hat{a}$  and  $\hat{b}$  in (5)–(8) are linear functions of the random variables  $Y_k$ , so that it is not difficult to calculate their means and variances. Since the random variables  $Y_k$  are independent and since a Poisson distribution approaches a normal distribution as the mean increases, it is natural to regard the joint distribution of the estimators as approximately multivariate normal as well, but we do not exploit this property. (Either large  $N$  or large means would imply normality, the first by the central limit theorem and the second by a direct normal approximation for the Poisson distribution.)

If we ignore the constraints in (9), then the estimators are all *unbiased*, i.e.,  $E\hat{a} = a$  and  $E\hat{b} = b$  for any weights. However, the adjustment to meet the constraint introduces some bias. For example, if  $b$  is positive but not too large, then  $E\hat{a} > a$  while  $E\hat{b} < b$ . Henceforth in this section we ignore the constraints in (9).

We can also calculate the variances for OLS. We use a superscript  $O$  to indicate OLS and a subscript  $N$  to indicate the dependence on  $N$ , but we omit them from  $x_k$  and  $Y_k$ . From (2), (3), (5) and (8),

$$\text{Var } \hat{b}_N^O = \left(\frac{N}{T}\right)^2 \frac{\sum_{k=1}^N (x_k - \bar{x})^2 \left(\frac{T}{N}\right) (a + bx_k)}{\left(\sum_{k=1}^N (x_k - \bar{x})^2\right)^2} = \left(\frac{N^2}{N^2 - 1}\right) \frac{6(2a + bT)}{T^3}. \quad (16)$$

(Part of the calculation in (16) is based on  $\sum_{k=1}^N (x_k - \bar{x})^2 = T^2(N^2 - 1)/12N$ .) The dependence upon  $a$ ,  $b$ ,  $T$  and  $N$  is quite clear from (16). To see the impact of  $N$  in the standard deviation, we can expand the  $N$ -term in inverse powers of  $N$ , i.e.,

$$\frac{N}{\sqrt{N^2 - 1}} = 1 + \frac{1}{2N^2} + \frac{3}{8N^4} + \frac{5}{16N^6} + O\left(\frac{1}{N^8}\right) \quad \text{as } N \rightarrow \infty. \quad (17)$$

From (16) and (17), we see that the standard deviation  $SDV(\hat{b})$  for  $N = 10$  exceeds the standard deviation  $N = \infty$  by only about 0.5%.

Next it is easy to focus on  $m \equiv a + bT/2$ . The OLS estimator for  $m$  is

$$\text{Var } \hat{m}_N^O = \left(\frac{N}{T}\right)^2 \text{Var } \bar{Y} = \frac{1}{T^2} \sum_{k=1}^N \left(\frac{T}{N}\right) \left(a + b\left(k - \frac{1}{2}\right) \frac{T}{N}\right) = \frac{2a + bT}{2T}. \quad (18)$$

Formula (18) is elementary, since  $m = Z/T$  where  $Z \equiv \sum_{k=1}^N Y_k$  is Poisson, because it is the sum of  $n$  independent Poisson random variables. Note that (18) is independent of  $N$ .

Next,  $a = m - bT/2$ , so that the variance of  $\hat{a}_N^O$  is

$$\text{Var } \hat{a}_N^O = \text{Var } \hat{m}_N^O + \left(\frac{T^2}{4}\right) \text{Var } \hat{b}_N^O - T \text{Cov}(\hat{m}_N^O, \hat{b}_N^O), \quad (19)$$

where

$$\begin{aligned} -T \text{Cov}(\hat{m}_N^O, \hat{b}_N^O) &= -T \text{Cov}\left(\frac{N}{T} \frac{1}{N} \sum_{k=1}^N Y_k, \left(\frac{N}{T}\right) \frac{\sum_{k=1}^N (x_k - \bar{x}) Y_k}{\sum_{k=1}^N (x_k - \bar{x})^2}\right) \\ &= -\frac{12}{T(N^2 - 1)} \sum_{k=1}^N (x_k - \bar{x}) \text{Var } Y_k = -b. \end{aligned} \quad (20)$$

From (16), (18)–(20),

$$\text{Var } \hat{a}_N^0 = \left( \frac{N^2}{N^2 - 1} \right) \frac{6(2a + bT)}{4T} + \frac{2a + bT}{2T} - b \rightarrow b + \frac{4a}{T} \text{ as } N \rightarrow \infty. \quad (21)$$

Next, for any  $t$ , let  $\hat{\lambda}_N^0(t) = \hat{a}_N^0 + \hat{b}_N^0 t$  be the estimator of the arrival rate at time  $t$ . Then

$$\text{Var } \hat{\lambda}_N^0(t) = \text{Var } \hat{a}_N^0 + t^2 \text{Var } \hat{b}_N^0 + 2t \text{Cov}(\hat{a}_N^0, \hat{b}_N^0), \quad (22)$$

where

$$\text{Cov}(\hat{a}_N^0, \hat{b}_N^0) = \text{Cov}(\hat{m}_N^0, \hat{b}_N^0) - \frac{T}{2} \text{Var}(\hat{b}_N^0).$$

Finally, we give approximate formulas for the mean and the standard deviation of the intercept estimator  $\tilde{a}_N^0$  assuming that the adjustment to satisfy the nonnegativity condition (9) is made. This shows the bias caused by the nonnegativity constraint in (9). We assume that the estimator  $\hat{a}_N^0$  before adjustment is normally distributed with mean 0 and variance  $\text{Var } \hat{a}_N^0$  in (21). Then  $\tilde{a}_N^0 = 0$  when  $\hat{a}_N^0 \leq 0$ , while  $\tilde{a}_N^0 = \hat{a}_N^0$  when  $\hat{a}_N^0 > 0$ . Hence, using basic properties of the positive normal distribution, we see that the mean is

$$E\tilde{a}_N^0 = \left( \frac{\text{Var } \hat{a}_N^0}{2\pi} \right)^{1/2} \approx 0.4 \text{SD}(\hat{a}_N^0) \rightarrow 0.4\sqrt{b} \text{ as } N \rightarrow \infty. \quad (23)$$

and the standard deviation is

$$\text{SD } \tilde{a}_N^0 = \left( \frac{1}{2} - \frac{1}{2\pi} \right)^{1/2} \text{SD } \hat{a}_N^0 \approx 0.584 \text{SD } \hat{a}_N^0 \rightarrow 0.584\sqrt{b} \text{ as } N \rightarrow \infty. \quad (24)$$

#### 4. Asymptotic formulas for TWLS

In this section we develop asymptotic formulas for the sample variances associated with TWLS, i.e., using the weights in (15). We first develop formulas for the limits as  $N \rightarrow \infty$ . Then we consider the asymptotic form of these limiting formulas as  $\epsilon \equiv bT/a$  gets small. The parameter  $\epsilon$  is a relevant measure of departure from homogeneity;  $|bT|$  is the difference between the left and right end points of  $\lambda(t)$  over  $[0, T]$ , whereas  $a$  is the left end point. At the outset, we assume that  $a \neq 0$  and  $b \neq 0$ ; we consider the cases  $a = 0$  and  $b = 0$  at the end of this section. We

use a superscript  $T$  and a subscript  $N$  for the estimators to indicate the dependence on TWLS and  $N$ , but we omit them in  $\lambda_k$ ,  $x_k$  and  $w_k$  in (2), (3) and (15).

As  $N \rightarrow \infty$ , the sums approach integrals which can be directly integrated. First,

$$\begin{aligned}\bar{x}_N^T &\equiv \frac{1}{N} \sum_{k=1}^N w_k x_k = \sum_{k=1}^N \lambda_k^{-1} x_k / \sum_{k=1}^N \lambda_k^{-1} \\ \rightarrow \bar{x}_\infty^T &\equiv \int_0^T \frac{tdt}{a+bt} / \int_0^T \frac{dt}{a+bt} = \frac{T}{\ln(1+[bT/a])} - \frac{a}{b} \quad \text{as } N \rightarrow \infty,\end{aligned}\quad (25)$$

where  $\ln(x)$  is the natural logarithm. By the same reasoning,

$$\begin{aligned}\text{Var } \hat{b}_N^T &= \left(\frac{N}{T}\right)^2 \frac{\sum_{k=1}^N w_k^2 (x_k - \bar{x}_N^T)^2 \lambda_k}{\left(\sum_{k=1}^N w_k (x_k - \bar{x}_N^T)^2\right)^2} \\ \rightarrow \text{Var } \hat{b}_\infty^T &\equiv \left(\int_0^T \frac{(t - \bar{x}_\infty^T)^2}{a+bt} dt\right)^{-1} = \frac{2b}{T(T - 2\bar{x}_\infty^T)} \quad \text{as } N \rightarrow \infty,\end{aligned}\quad (26)$$

$$\left(\frac{N}{T}\right)^2 \text{Var } \bar{Y}_N^T \rightarrow \frac{b}{\ln(1+[bT/a])} \quad \text{as } N \rightarrow \infty,\quad (27)$$

$$\text{Cov}\left(\frac{N}{T} \hat{b}_N^T, \frac{N}{T} \bar{Y}_N^T\right) \rightarrow 0 \quad \text{as } N \rightarrow \infty,\quad (28)$$

and

$$\begin{aligned}\text{Var } \hat{a}_N^T &= \left(\frac{N}{T}\right)^2 \text{Var } \bar{Y}_N^T + (\bar{x}_N^T)^2 \text{Var } \hat{b}_N^T - 2\bar{x}_N^T \text{Cov}\left(\frac{N}{T} \hat{b}_N^T, \frac{N}{T} \bar{Y}_N^T\right) \\ &\rightarrow \frac{b}{\ln(1+[bT/a])} + \frac{2(\bar{x}_\infty^T)^2 b}{T(T - 2\bar{x}_\infty^T)} \quad \text{as } N \rightarrow \infty.\end{aligned}\quad (29)$$

Indeed, our numerical experience indicates that  $\text{Cov}(\hat{b}_N^T, \hat{Y}_N^T) = 0$  for all  $N$ . For  $\hat{\lambda}_N^T(t) = \hat{a}_N^T + \hat{b}_N^T t$ ,

$$\text{Var } \hat{\lambda}_N^T(t) = \text{Var } \hat{a}_N^T + t^2 \text{Var } \hat{b}_N^T + 2t \text{Cov}(\hat{a}_N^T, \hat{b}_N^T),\quad (30)$$

where

$$\text{Cov}(\hat{a}_N^T, \hat{b}_N^T) = \text{Cov}(\bar{Y}_N^T - \bar{x}_N^T \hat{b}_N^T, \hat{b}_N^T) = \text{Cov}(\bar{Y}_N^T, \hat{b}_N^T) - \bar{x}_N^T \text{Var } \hat{b}_N^T.\quad (31)$$

The formulas for  $\bar{x}_\infty^T$ ,  $\text{Var } \hat{b}_\infty^T$  and  $\text{Var } \hat{a}_\infty^T$  in (25), (26) and (29) do not look much like their counterparts for OLS. However, a connection can be seen when we develop expressions for the TWLS asymptotic formulas as  $\epsilon \equiv bT/a$  gets small. We will show that the asymptotic relative difference between OLS and TWLS is of order  $O(\epsilon^2)$  as  $\epsilon \rightarrow 0$ .

First, applying the familiar asymptotic expansion for the logarithm,  $\ln(1+z) = z - z^2/2 + \dots$ , to (25), we obtain

$$\bar{x}_\infty^T = \frac{T}{2} \left( 1 - \frac{\epsilon}{6} + \frac{\epsilon^2}{12} - \frac{19}{360} \epsilon^2 + O(\epsilon^4) \right) \quad \text{as } \epsilon \rightarrow 0. \tag{32}$$

Next, from (25), (26) and (32),

$$\begin{aligned} \text{Var } \hat{b}_\infty^T &= \frac{12a}{T^3 \left( 1 - \frac{\epsilon}{2} + \frac{19}{60} \epsilon^2 + O(\epsilon^3) \right)} \\ &= \frac{12a}{T^3} \left( 1 + \frac{\epsilon}{2} - \frac{\epsilon^2}{15} + O(\epsilon^3) \right) \quad \text{as } \epsilon \rightarrow 0. \end{aligned} \tag{33}$$

Thus, the asymptotic relative difference between the asymptotic variances  $\text{Var } \hat{b}_\infty^T$  of the OLS and TWLS methods is

$$\frac{\text{Var } b_\infty^O - \text{Var } \hat{b}_\infty^T}{\text{Var } \hat{b}_\infty^T} = \frac{\epsilon^2}{15} + O(\epsilon^3) \quad \text{as } \epsilon \rightarrow 0. \tag{34}$$

Next, to treat  $\text{Var } \hat{a}_\infty^T$ , note that

$$(\bar{x}_\infty^T)^2 \text{Var } \hat{b}_\infty^T = \frac{3a}{T} \left( 1 + \frac{\epsilon}{6} - \frac{7}{180} \epsilon^2 + O(\epsilon^3) \right) \quad \text{as } \epsilon \rightarrow 0$$

and

$$\frac{b}{\ln(1+\epsilon)} = \frac{a}{T} \left( 1 + \frac{\epsilon}{2} - \frac{\epsilon^2}{12} + O(\epsilon^2) \right) \quad \text{as } \epsilon \rightarrow 0,$$

so that

$$\text{Var } \hat{a}_\infty^T = \frac{4a}{T} + b - \frac{a\epsilon^2}{5T} + O(\epsilon^3) \quad \text{as } \epsilon \rightarrow 0 \tag{35}$$

and

$$\frac{\text{Var } \hat{a}_\infty^0 - \text{Var } \hat{a}_\infty^T}{\text{Var } \hat{a}_\infty^T} = \frac{\epsilon^2}{20} + O(\epsilon^3) \quad \text{as } \epsilon \rightarrow 0. \quad (36)$$

So far in this section, we have excluded the cases  $a = 0$  and  $b = 0$ . The case  $b = 0$  is obtained as the limit as  $b \rightarrow 0$  as described above; then TWLS coincides with OLS.

The cases  $a = 0$  without adjustments is more complicated, so we do not treat it analytically. To see the complications, note that the limiting integral  $\int_0^T t^{-1} dt$  diverges. We now analyze the case in which we assume that  $a = 0$ , as we do if we make adjustments after finding  $\hat{a} < 0$ . For  $a = 0$  and finite  $N$ , the TWLS slope estimator is

$$\hat{b}_N^T = \frac{2}{T^2} \sum_{k=1}^N Y_k. \quad (37)$$

The associated variance is

$$\text{Var } \hat{b}_N^T = \frac{2b}{T^2}, \quad (38)$$

which is independent of  $N$ . By (16) and (38), when  $a = 0$ ,

$$\frac{\text{Var } \hat{b}_N^O}{\text{Var } \hat{b}_N^T} = \frac{3N^2}{N^2 - 1} \rightarrow 3 \quad \text{as } N \rightarrow \infty. \quad (39)$$

Since ML coincides with TWLS when  $a = 0$  (see section 5 below), formula (39) roughly indicates how much more efficient is the ML slope estimator than the OLS slope estimator when  $a$  or  $a + bT$  is very small and an adjustment is made to satisfy the constraints in (9). As shown in section 6, the overall relative efficiency of ML compared to OLS for the slope is actually only about 1.5.

We also obtain (38) asymptotically from (26) as  $a \rightarrow 0$ . Note that  $\bar{x}_\infty \rightarrow 0$  as  $a \rightarrow 0$  for  $\bar{x}_\infty$  in (25). Thus  $\text{Var } \hat{b}_\infty^T \rightarrow 2b/T^2$  and  $\text{Var } \hat{a}_\infty^T \rightarrow 0$  as  $a \rightarrow 0$ . In contrast, by (23),

$$\text{Var } \hat{a}_N^O = \left( \frac{N^2}{N^2 - 1} \right) \frac{6b}{4} - \frac{b}{2} \rightarrow b \quad \text{as } N \rightarrow \infty. \quad (40)$$

In section 6 we will see that TWLS is indeed significantly better than OLS for estimating the intercept in this region. However, ML does not share the advantage of large  $N$ .

**5. Maximum likelihood**

Let  $p(\mathbf{n}; a, b)$  be the probability that the count vector  $\mathbf{Y} \equiv (Y_1, \dots, Y_N)$  for the  $N$  subintervals is  $\mathbf{n} \equiv (n_1, \dots, n_N)$  when the parameter pair is  $(a, b)$ . The ML estimator is the pair  $(\hat{a}, \hat{b})$  that maximizes  $p(\mathbf{Y}; a, b)$  or, equivalently,

$$\begin{aligned} \ln p(\mathbf{Y}; a, b) &= -\sum_{k=1}^N \lambda_k + \sum_{k=1}^N Y_k \ln \lambda_k = \sum_{k=1}^N \ln(Y_k!) \\ &= -aT - \frac{bT^2}{2} + \sum_{k=1}^N Y_k \ln((a + bx_k)(T/N)) + C \end{aligned} \tag{41}$$

for a constant  $C$ . Note that  $\ln p(\mathbf{Y}; a, b)$  is a strictly concave function of  $(a, b)$  provided that  $Y_k > 0$  for some  $k$ , and  $a + bx_k > 0$  for all  $k$ , which we assume is the case. (If  $Y_k = 0$  for all  $k$ , then (41) is obviously maximized by  $a = b = 0$ .)

By differentiating with respect to  $a$  and  $b$  in (41), we find that the pair  $(\hat{a}, \hat{b})$  is the solution to the two equations

$$\sum_{k=1}^N \frac{Y_k}{a + bx_k} = T \tag{42}$$

and

$$\sum_{k=1}^N \frac{x_k Y_k}{a + bx_k} = \frac{T^2}{2} \tag{43}$$

provided that the constraints in (9) are satisfied.

If  $\hat{a} < 0$  by (42) and (43), then we set  $\hat{a}_N^M = 0$  and minimize  $\ln p(\mathbf{Y}, 0, b)$ , obtaining

$$\hat{b}_N^M = 2 \sum_{k=1}^N Y_k / T^2, \tag{44}$$

just as in (37). If we knew that  $a = 0$ , then we would also use (44) and obtain

$$\text{Var } \hat{b}_N^M = 2b/T^2, \tag{45}$$

just as in (38). Similarly, if  $\hat{a} + \hat{b}T < 0$ , then we set  $\hat{a}_N^M = -bT$  and maximize  $\ln p(\mathbf{Y}, -bT, b)$  to obtain

$$\hat{b}_N^M = -2 \sum_{k=1}^N Y_k / T^2. \tag{46}$$

We now consider how to solve (42) and (43). Multiplying eqs. (42) and (43) by  $a$  and  $b$ , respectively, and adding, we obtain the linear equation

$$S \equiv \sum_{k=1}^N Y_k = aT + \frac{bT^2}{2} = \sum_{k=1}^N EY_k. \quad (47)$$

(We include the last expression to emphasize that  $aT + bT^2/2$  is the expected total number of points in the interval  $[0, T]$ .)

By virtue of (47), we can substitute  $2(S - aT)/T^2$  for  $b$  in (42), we obtain the single equation

$$g(a) \equiv \sum_{k=1}^N \frac{Y_k}{a\bar{x} + x_k \left( \frac{S}{T} - a \right)} = 2. \quad (48)$$

First note that  $a = S/T$  and  $b = 0$  always solves (47) and (48), but there may be other solutions. Next note that  $g(a)$  in (48) has singularities for each  $x_k > \bar{x}$  for which  $Y_k > 0$ . In particular, the singularities are at

$$a_k^\# = \frac{x_k}{x_k - \bar{x}} \frac{S}{T} \quad \text{for } k > N/2,$$

all of which are greater than  $S/T$ .

By looking at the second derivative of  $g$ , we see  $g$  is infinitely differentiable and convex in the open interval  $(0, S/T)$ . Hence, there is at most one root to equation (48) in the interval  $(0, S/T)$ . If  $g'(S/T) < 0$  and  $g(0) > 2$ , then there is such a root. Hence, we compute these quantities:

$$g(0) = \frac{T}{S} \sum_{k=1}^N \left( \frac{Y_k}{x_k} \right) \quad (49)$$

and

$$g'(S/T) = -\frac{4}{S^2} \sum_{k=1}^N Y_k (\bar{x} - x_k). \quad (50)$$

If indeed  $g'(S/T) < 0$  and  $g(0) > 2$ , then we find the unique root of (48) in  $(0, S/T)$  by a convenient method. Since derivatives are readily available, it is natural to use a Newton method (starting at 0 since we do not want the root at  $S/T$ ).

We always allow for a root to (48) with  $a \leq 0$ . However, using the constraint (9), we consider  $\hat{a} = 0$  and  $\hat{b}_N^M$  in (44) as a candidate optimum.

Next, we consider the possibility of  $a > S/T$ . However, by (47),  $a > S/T$  if



and only if  $b < 0$ . We treat the case  $b < 0$  by repeating the procedure above on the reverse-time problem, where  $\bar{Y}_k^* = Y_{N-k+1}$ ,  $b = -b^*$ ,  $a + bT = a^*$ . Let  $g^*$  be  $g$  in (48) for the reverse-time problem. Note that for the reverse-time problem

$$g^{*'}(S/T) = -g'(S/T) \tag{51}$$

for  $g'(S/T)$  in (50). Hence, there is at most one root in  $(0, S/T)$  for the two problems.

Finally, we evaluate the likelihood for each of the candidate solutions:

- (i) root of  $g(a) = 2$  in  $(0, S/T)$ ,  $b = 2(S - aT)/T^2$ ,
- (ii)  $a = a^* + b^*T$ ,  $b = -b^*$  for  $a^*$  root of  $g^*(a) = 2$  in  $(0, S/T)$  and  $b^* = 2(S - a^*T)/T^2$ ,
- (iii)  $a = S/T$ ,  $b = 0$ ,
- (iv)  $a = 0$ ,  $b = 2S/T^2$ ,
- (v)  $a = -bT$ ,  $b = -2S/T^2$ ,

with the understanding that we consider (ii) only if there is no solution to (i) and (iii)–(v) only if there is no solution to (i) and (ii). By (41) and (47), to evaluate (iii)–(v) it suffices to calculate

$$\sum_{k=1}^N Y_k \ln(a + bx_k) \tag{53}$$

for each possibility and choose the one yielding the maximum. (Note that  $a + bx_k > 0$  for all  $k$  in each case.) In this way we obtain our estimates  $(\hat{a}_N^M, \hat{b}_N^M)$ .

### 6. Simulation results

We wrote a simulation program (in C) to randomly generate the mutually independent Poisson random variables  $Y_k$  with means  $\lambda_k$  in (2) and carry out the estimation procedures. In table 1 we give the sample standard deviations of the estimators  $\hat{a}$  and  $\hat{b}$  for each of the methods (OLS, ML and TWLS) in a representative set of nine cases in which the constraints in (9) are very rarely violated, each obtained from 1000 replications. The cases are specified by the parameter four-tuple  $(n, T, a, b)$ . The constraints in (9) were in fact never violated in this sample. Hence, in this sample the associated sample means of  $\hat{a}$  and  $\hat{b}$  were consistently close to the true values  $a$  and  $b$ .

From table 1 we see that the standard deviations for ML and TWLS are almost identical, and are consistently less than for OLS. (The only minor exception is the last case in which the actual slope is zero). However, all three sample standard deviations are quite close. Moreover, the observed OLS and TWLS standard

Table 1  
A comparison of the sample standard deviations of the estimators  $\hat{a}$  and  $\hat{b}$  in a representative set of cases based on 1000 replications of the simulation.

Case				OLS		ML		TWLS	
$N$	$T$	$a$	$b$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$
12	12	100	25	7.41	1.32	6.96	1.25	6.96	1.25
120	12	100	25	7.80	1.34	7.42	1.27	7.42	1.27
12	6	100	25	9.03	2.94	8.80	2.86	8.80	2.87
12	12	100	4.167	5.84	0.913	5.81	0.908	5.81	0.908
12	12	10	2.5	2.43	0.417	2.31	0.403	2.30	0.400
120	12	10	2.5	2.44	0.418	2.30	0.398	2.30	0.397
12	24	10	2.5	1.99	0.187	1.79	0.163	1.78	0.163
12	12	10	4.167	2.72	0.488	2.50	0.455	2.49	0.453
12	12	10	0.0	1.94	0.295	1.95	0.296	1.94	0.295

deviations, are quite accurately predicted by the formulas in sections 3 and 4 (as is easily verified).

In table 2 we give the  $l_2$  distances between the estimates  $\hat{a}$  and  $\hat{b}$  for different pairs of methods, based on the same 1000 replications. For example, the  $l_2$  distance for  $a$  by OLS and ML is

$$\|\hat{a}^0 - \hat{a}^M\|_2 = \left( \frac{1}{1000} \sum_{i=1}^{1000} (\hat{a}_i^0 - \hat{a}_i^M)^2 \right)^{1/2}. \quad (54)$$

Table 2 shows that ML tends to produce nearly the same estimates as TWLS.

The information in table 2 can also be seen from plots of the estimator pairs, e.g.,  $\{(\hat{a}_i^0, \hat{a}_i^M) : 1 \leq i \leq 1000\}$ . We illustrate by contrasting OLS-ML and ML-TWLS for the case (12, 12, 100, 25) in fig. 2. This figure confirms that in this case

$$|\text{TWLS} - \text{ML}| < |\text{OLS} - \text{ML}|.$$

In tables 3 and 4 we present the sample means and standard deviations of the estimators  $\hat{a}$  and  $\hat{b}$  when  $T = 12$ ,  $a = 0$  and  $b = 10$  for several different values of  $N$ , again based on 1000 replications. In tables 5 and 6 we present the same results for the case  $T = 12$ ,  $a = 0$  and  $b = 100$ . Since  $a = 0$ , the nonnegativity conditions (9) are violated about half the time in these examples. Fig. 3 plots estimator pairs for the case in tables 5 and 6.

From tables 3 and 5, we see that there is a significant bias in the intercept estimate due to the adjustment to satisfy (9). From tables 3 and 5, we see that the

Table 2  
A comparison of the  $l_2$ -distances between estimators  $\hat{a}$  and  $\hat{b}$  for pairs of methods, in a representative set of cases based on 1000 replications of the simulation.

Case		$\lambda$		OLS-ML		TWLS-ML	
$N$	$T$	$a$	$b$	$\hat{a}$	$\hat{b}$	$\hat{a}$	$\hat{b}$
12	12	100	25	2.61	0.435	0.16	0.026
120	12	100	25	2.56	0.426	0.17	0.029
12	6	100	25	2.14	0.71	0.26	0.087
12	12	100	4.167	0.60	0.10	0.13	0.020
12	12	10	2.5	0.82	0.137	0.16	0.027
120	12	10	2.5	0.85	0.142	0.17	0.029
12	24	10	2.5	0.95	0.080	0.09	0.008
12	12	10	4.167	1.14	0.190	0.18	0.029
12	12	10	0.0	0.14	0.024	0.14	0.023

OLS mean and standard deviation agree with formulas (23) and (24). We also see that ML and TWLS are significantly better than OLS in this case; there is both a smaller bias and a smaller sample standard deviation.

From tables 4 and 6, we see that there is relatively little bias in the slope. Moreover, the three methods do not differ greatly for the slope, although TWLS and ML actually have significantly less bias than OLS. However, this bias differential is small compared to the sample standard deviation.

Fig. 3 tells a similar story pictorially. From fig. 3, we see that the ML and TWLS estimators are close for the intercept, but the ML and OLS estimators are not. The same is true for the slope, but the difference between ML and OLS is not nearly so great.

Fig. 4 shows the estimator pairs for the same case as fig. 3, except the number of measurement intervals is increased from  $N = 12$  to  $N = 120$ . From fig. 4 we see that the agreement between ML and TWLS decreases. Moreover, increasing  $N$  helps ML and TWLS, but not OLS. However, the ML bias does not steadily decline as  $N$  increases the way the TWLS bias does.

In conclusion, the major advantage of ML over OLS is in estimating the intercept when  $a \approx 0$  (or  $a + bT$  when  $a + bT \approx 0$ ).

## 7. Model tests

In this section we discuss how we can test whether the nonhomogeneous Poisson process with a linear rate  $\lambda(t) = a + bt$  is a reasonable model for data over the interval  $[0, T]$ . The possible alternative hypotheses include: (i) a Poisson process with nonlinear rate, (ii) a non-Poisson process with linear rate, and (iii) a non-Poisson process with nonlinear rate. We do not directly examine any of these

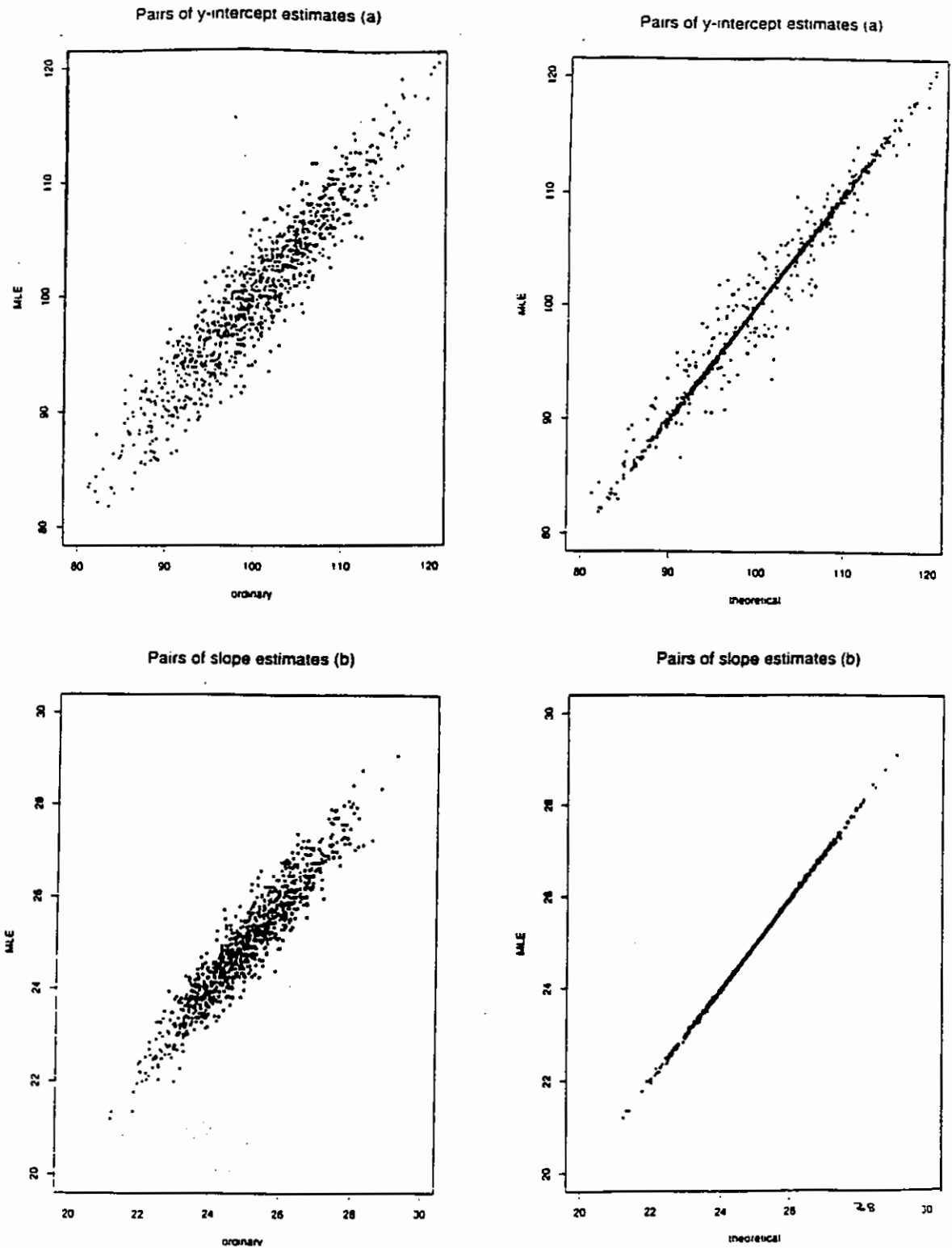


Fig. 2. Estimator pairs for the case  $T = 12$ ,  $N = 12$ ,  $a = 100$  and  $b = 25$ .

Table 3

The sample mean and standard deviation of the  $y$ -intercept estimator  $\hat{a}$  as a function of the method and  $N$  for the case  $T = 12$ ,  $a = 0$  and  $b = 10$ , based on 1000 replications.

$N$	Sample mean			Sample standard deviation		
	OLS	TWLS	ML	OLS	TWLS	ML
4	1.32	1.05	1.06	1.95	1.62	1.62
8	1.32	0.98	0.98	1.91	1.43	1.43
12	1.32	0.80	0.81	1.91	1.23	1.25
24	1.22	0.68	0.70	1.80	1.14	1.16
120	1.21	0.53	0.64	1.88	1.04	1.11
240	1.28	0.45	0.60	1.87	1.02	1.01
1200	1.27	0.37	0.62	1.94	1.19	1.04
2400	1.43	0.29	0.63	1.96	0.75	1.05

Table 4

The sample mean and standard deviation of the slope estimator  $\hat{b}$  as a function of the method and  $N$  for the case  $T = 12$ ,  $a = 0$  and  $b = 10$ , based on 1000 replications.

$N$	Sample mean			Sample standard deviation		
	OLS	TWLS	ML	OLS	TWLS	ML
4	9.839	9.832	9.831	0.52	0.45	0.45
8	9.840	9.845	9.845	0.53	0.45	0.45
12	9.840	9.875	9.872	0.52	0.43	0.43
24	9.851	9.892	9.890	0.52	0.43	0.43
120	9.850	9.911	9.891	0.65	0.42	0.42
240	9.832	9.919	9.893	0.50	0.39	0.40
1200	9.855	9.949	9.907	0.52	0.41	0.40
2400	9.821	9.962	9.906	0.54	0.42	0.43

alternatives here. However, following standard statistical practice, we can detect a poor fit of our linear Poisson model from statistics that fall outside the main region of their distributions under the linear Poisson model.

A simple direct procedure is to compare the observed sample standard deviations of the estimators  $\hat{a}$  and  $\hat{b}$  based on actual data with the distributions of those same estimators from the linear Poisson model. The linear Poisson case can be described by the histograms from 1000 replications in the examples here.

We can also exploit more structure. We consider the two cases: multiple replications and a single replication. For  $n$  independent replications, we note that the distributions of the estimators  $\hat{a}$  and  $\hat{b}$  by any of our methods should be approximately normal by the central limit theorem. Indeed, the normality was demonstrated by looking at histograms of the 1000 observations in the examples of table 1.

Table 5

The sample mean and standard deviation of the  $y$ -intercept estimator  $\hat{a}$  as a function of  $N$  and the method for the case of  $T = 12$ ,  $a = 0$  and  $b = 100$ , based on 1000 replications.

$N$	Sample mean			Sample standard deviation		
	OLS	TWLS	ML	OLS	TWLS	ML
4	4.05	3.30	3.30	6.04	5.05	5.05
8	3.94	2.91	2.90	5.80	4.18	4.16
12	4.05	2.49	2.51	5.83	3.83	3.85
24	3.93	2.43	2.44	5.83	3.49	3.51
120	3.89	1.89	1.94	5.82	3.30	3.12
240	4.24	1.52	1.95	6.39	2.95	3.20
1200	3.86	1.29	1.88	5.76	3.08	3.10
2400	3.91	1.06	1.95	5.78	2.10	3.12

Table 6

The sample mean and standard deviation of the slope estimator  $\hat{b}$  as a function of  $N$  and the method for the case  $T = 12$ ,  $a = 0$  and  $b = 100$ , based on 1000 replications.

$N$	Sample mean			Sample standard deviation		
	OLS	TWLS	ML	OLS	TWLS	ML
4	99.04	98.95	98.95	3.29	2.90	2.90
8	99.07	99.09	99.09	3.18	2.76	2.75
12	99.02	99.22	99.21	3.28	2.73	2.73
24	99.07	99.23	99.23	3.29	2.66	2.69
120	99.08	99.41	99.40	3.27	2.66	2.63
240	99.00	99.58	99.43	3.29	2.51	2.56
1200	99.18	99.68	99.49	3.19	2.59	2.53
2400	99.08	99.71	99.41	3.27	2.48	2.64

However, when the nonnegativity conditions (9) are frequently violated, as in the examples in table 2, the  $y$ -intercept  $\hat{a}$  tends to have approximately a truncated normal distribution instead of a normal distribution, with parameters as in (23) and (24).

Henceforth, suppose that there are  $n$  replications and that the normal approximation is appropriate. Let  $\bar{a} = n^{-1} \sum_{i=1}^n \hat{a}_i$  and  $\bar{b} = n^{-1} \sum_{i=1}^n \hat{b}_i$ ; let  $\sigma_a^2$  and  $\sigma_b^2$  be the actual variance of  $\hat{a}$  and  $\hat{b}$ . The normality implies that

$$V_a \equiv \sum_{i=1}^n (\hat{a}_i - \bar{a})^2 / \sigma_a^2 \quad \text{and} \quad V_b \equiv \sum_{i=1}^n (\hat{b}_i - \bar{b})^2 / \sigma_b^2, \quad (55)$$

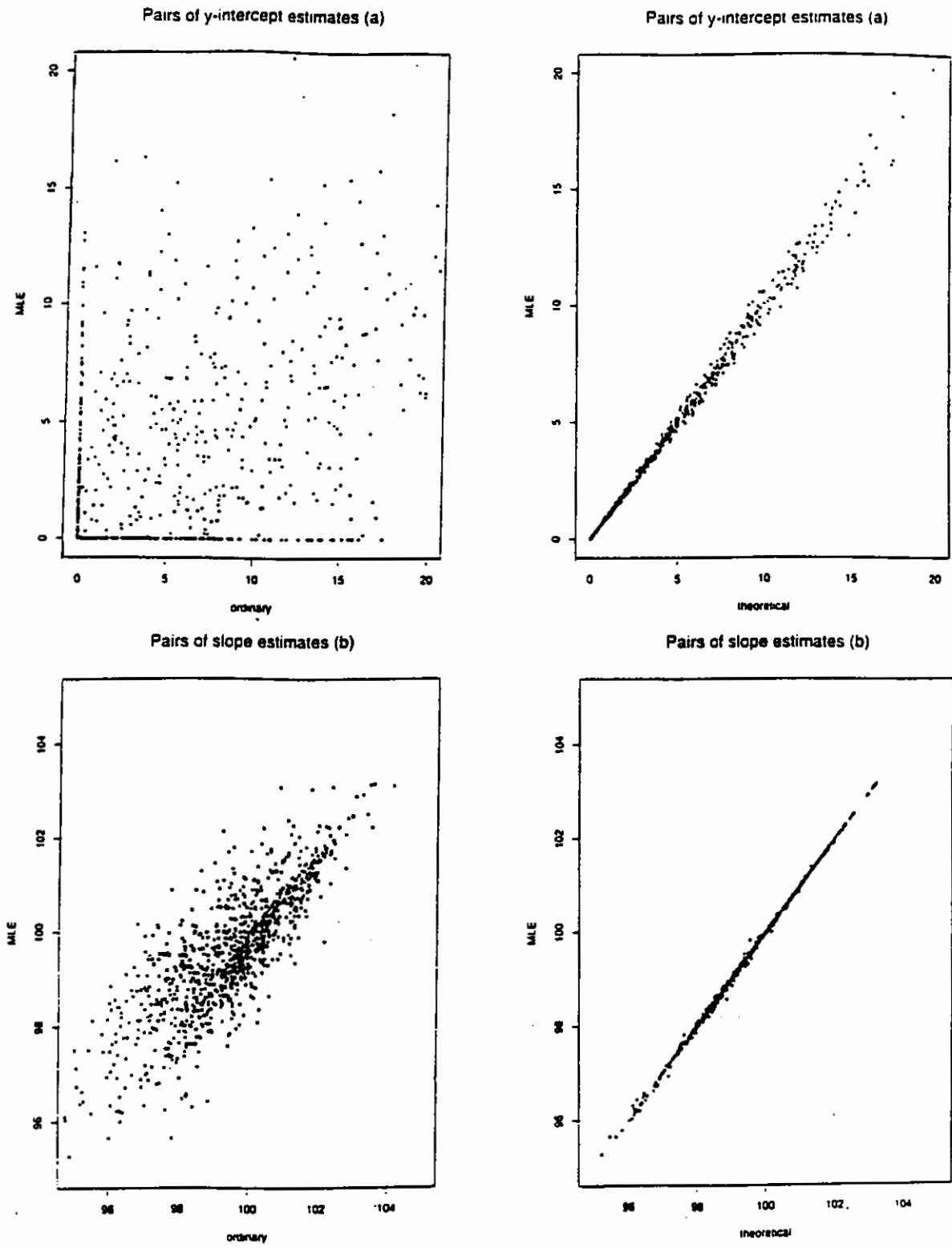


Fig. 3. Estimator pairs for the case  $T = 12, N = 12, a = 0$  and  $b = 100$ .

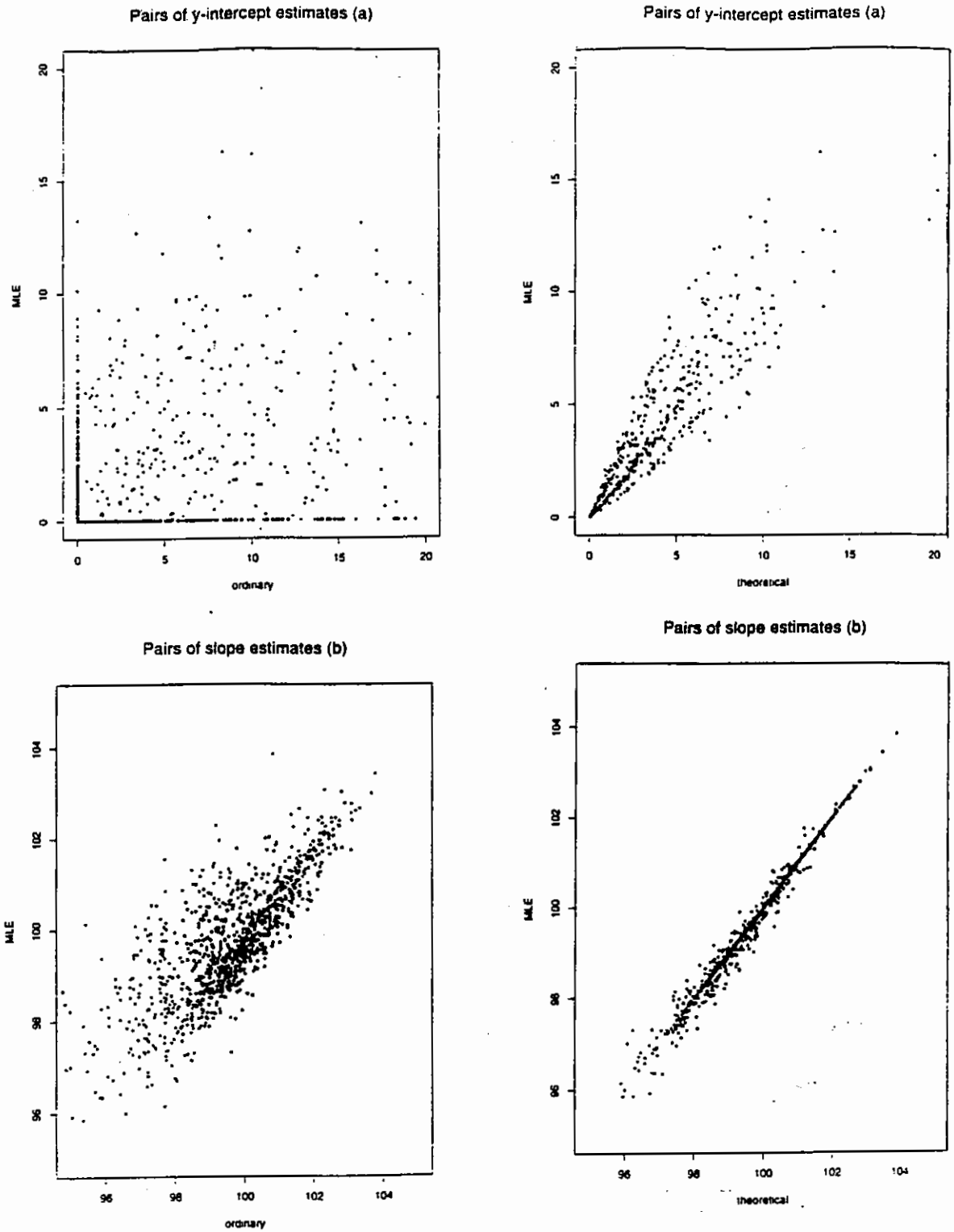


Fig. 4. Estimator pairs for the case  $T = 12$ ,  $N = 120$ ,  $a = 0$  and  $b = 100$ .



should both have chi-square distributions with  $n - 1$  degrees of freedom. For OLS, we approximate  $\sigma_a^2$  and  $\sigma_b^2$  by the exact OLS variances associated with the estimated parameter pair  $(\bar{a}, \bar{b})$  determined in section 3. For IWLS and ML, we approximate  $\sigma_a^2$  and  $\sigma_b^2$  by the exact TWLS asymptotic variances associated with the estimated parameter pair  $(\bar{a}, \bar{b})$ . Both of these exact variances are relevant only under the assumption that the constraints in (9) are not violated too frequently, which we check in our calculations of the estimate.

For a single replication, we assume that the means  $EY_k$  are sufficiently large that we can regard the Poisson variables  $Y_k$  as being approximately normally distributed. Then we use the transformation to convert the unequal variance problem to the equal variance problem, which was used to justify the optimality of TWLS. Since we do not know the real parameters  $a$  and  $b$ , we use our estimates. The normalized sum of squares

$$U \equiv \sum_{k=1}^N \hat{\sigma}_k^{-2} \left( Y_k - (\hat{a} + \hat{b}x_k) \frac{T}{N} \right)^2, \tag{56}$$

where

$$\hat{\sigma}_k^2 = (\hat{a} + \hat{b}x_k) \frac{T}{N},$$

should thus be approximately chi-square with  $N - 2$  degrees of freedom.

These chi-square distributional properties should enable us to roughly gauge model consistency. The test in (56) is able to capture significant departures from the linear Poisson model as occur in the telephone call data in hours immediately after a rate change.

### 8. Proof of theorem 1

First note that  $b = 0, a = S/T$  solves (42) and (43) if and only if

$$\sum_{k=1}^N x_k Y_k = \frac{T}{2} \sum_{k=1}^N Y_k$$

and note the same is true for (5) and (6); then the weights in (15) are constant, so that  $\bar{x} = T/2$ . Henceforth, we consider solutions with  $b \neq 0$ .

Let

$$A_j = \sum_{k=1}^N \frac{x_k^j}{a + bx_k} \tag{57}$$

for  $j = 0, 1, 2$  and note that

$$aA_0 + bA_1 = N \quad \text{and} \quad aA_1 + bA_2 = \frac{NT}{2}, \quad (58)$$

so that  $A_1 = (N - aA_0)/b$ ,

$$A_2 = \frac{NT}{2b} - \frac{Na}{b^2} + \frac{a^2}{b^2}A_0, \quad (59)$$

$$\bar{x} = \frac{N}{bA_0} - \frac{a}{b}. \quad (60)$$

The rest of the proof is contained in the following two lemmas.

#### LEMMA 1

Eq. (6) holds, i.e.,  $\hat{a} = (N\bar{Y}/T) - \hat{b}\bar{x}$  for weights (15) with  $\lambda_k = \hat{a} + \hat{b}x_k$  and  $\hat{b} \neq 0$  if and only if  $(\hat{a}, \hat{b})$  satisfies (42).

#### *Proof*

By (57) and (60), we can write (6) as

$$a = \frac{N}{TA_0} \sum_{k=1}^N \frac{Y_k}{a + bx_k} - b \left( \frac{N}{bA_0} - \frac{a}{b} \right),$$

which we see is equivalent to (42). □

#### LEMMA 2

Assuming that (42) holds, eq. (5) is equivalent to (43).

#### *Proof*

First note that, by (57)–(60),

$$\begin{aligned} \sum_{k=1}^N \frac{(x_k - \bar{x})^2}{a + bx_k} &= \sum_{k=1}^N \frac{x_k^2 - 2\bar{x}x_k + \bar{x}^2}{a + bx_k} \\ &= \frac{N}{b} \left( \frac{T}{2} + \frac{a}{b} - \frac{N}{bA_0} \right). \end{aligned}$$

Hence, using (42), we can write (5) as

$$b = \frac{\frac{N}{T} \sum_{k=1}^N \frac{x_k Y_k}{a + bx_k} - \bar{x}T}{\frac{N}{b} \left( \frac{T}{2} + \frac{a}{b} - \frac{N}{bA_0} \right)}$$

or, equivalently,

$$\frac{NT}{2} + \frac{Na}{b} - \frac{N^2}{bA_0} - \frac{N}{T} \left( \sum_{k=1}^N \frac{x_k Y_k}{a + bx_k} + \frac{NT}{bA_0} - \frac{Ta}{b} \right),$$

which we see is equivalent to (43). □

### Acknowledgements

We thank our colleague Stephen G. Eick for assistance in the early phase of this research, Kingshuk Choudhury for pointing out that theorem 1 can be regarded as a consequence of the theory of generalized linear models as in McCullagh and Nelder [14], and the reviewers for their helpful comments.

### References

- [1] I.V. Basawa and B.L.S. Prakasa Rao, *Statistical Inference for Stochastic Processes* (Academic Press, New York, 1980).
- [2] M. Brown, Statistical analysis of non-homogeneous Poisson processes, in: *Stochastic Point Processes: Statistical Analysis, Theory and Applications*, ed. P.A.W. Lewis (Wiley, New York, 1972) pp. 67–89.
- [3] R.J. Carroll, Adapting for heteroscedasticity in linear models, *Ann. Statist.* 4(1982)1224–1233.
- [4] G.L. Choudhury, D.M. Lucantoni and W. Whitt, Numerical solution of piecewise-stationary  $M_t/G_t/1$  queues, *Oper. Res.*, to appear.
- [5] A.C. Cole, Progress in forecasting and traffic measurement techniques as seen at recent teletraffic congresses, in: *Teletraffic Science for New Cost-Effective Systems, Networks and Services, ITC-12*, ed. M. Bonatti (Elsevier, Amsterdam, 1989) pp. 198–206.
- [6] J.L. Davis, W.A. Massey and W. Whitt, Sensitivity to the service-time distribution in the non-stationary Erlang loss model, *Manag. Sci.*, 14(1995)1107–1116.
- [7] N.R. Draper and H. Smith, *Applied Regression Analysis* (Wiley, New York, 1966).
- [8] V. Frost and B. Melamed, Traffic modeling for telecommunication networks, *IEEE Commun. Magazine* 32(1994)70–81.
- [9] R.W. Hall, *Queueing Methods for Services and Manufacturing* (Prentice Hall, Englewood Cliffs, NJ, 1991).
- [10] J.M. Holtzman and D.L. Jagerman, Estimating peakedness from arrival counts, *Proc. 9th Int. Teletraffic Congress*, Torremolinos, Spain.
- [11] W. Leland, M. Taqqu, W. Willinger and D. Wilson, On the self-similar nature of ethernet traffic, *IEEE/ACM Trans. Networking* 2(1994).

- [12] W.A. Massey and W. Whitt, Networks of infinite-server queues with nonstationary Poisson input, *Queueing Systems* 13(1993)183–250.
- [13] W.A. Massey and W. Whitt, Stationary-process approximations for the nonstationary Erlang loss model, *Oper. Res.*, to appear.
- [14] P. McCullagh and J.A. Nelder, *Generalized Linear Models* (Chapman and Hall, London, 1983).
- [15] K.S. Meier-Hellstern, P.E. Wirth, Y. Yan and D.A. Hoeflin, Traffic models for ISDN data users: office automation application, in: *Teletraffic and Datatraffic in a Period of Change, ITC-13*, eds. A. Jensen and V.B. Iversen (Elsevier, Amsterdam, 1991) pp. 167–172.
- [16] A.M. Mood and F.A. Graybill, *Introduction to the Theory of Statistics*, 2nd ed. (McGraw-Hill, New York, 1963).
- [17] K. Rahko, Measurements for control and modelling of teletraffic, in: *Teletraffic and Datatraffic in a Period of Change, ITC-13*, eds. A. Jensen and V.B. Iversen (Elsevier, Amsterdam, 1991) pp. 609–614.
- [18] J.W. Roberts, *Performance Evaluation and Design of Multiservice Networks* (Commission of the European Communities, Luxembourg, 1992).
- [19] D.L. Snyder and M.I. Miller, *Random Point Processes in Time and Space*, 2nd ed. (Springer, New York, 1991).
- [20] M.R. Taaffe and K.L. Ong, Approximating  $Ph(t)/M(t)/S/C$  queueing systems, *Ann. Oper. Res.* 8(1987)103–116.
- [21] S. Weisberg, *Applied Linear Regression* (Wiley, New York, 1985).