ESTIMATING EXPECTED SOJOURN TIMES FOR A MARKOV RENEWAL PROCESS[1]

Manuel D. Rossetti† and Gordon M. Clark‡

† Department of Systems Engineering
University of Virginia
Thornton Hall
Charlottesville, VA 22903

‡ Department of Industrial, Welding and Systems Engineering
The Ohio State University
210 Baker Systems
1971 Neil Avenue
Columbus, OH 43210

Key Words and Phrases: Markov renewal processes; sojourn times; least squares regression; method of moments

ABSTRACT

We consider the estimation of the expected sojourn time in a Markov renewal process under the data condition that only the counts of the exits from the states are available for fixed intervals of time. For analytical and illustrative purposes we concentrate on the two-state process case. We present least squares and method of moments estimators and compare their statistical properties both analytically and empirically. We also present modified estimators with improved properties based upon an overlapping interval sampling strategy. The major results indicate that the least squares estimator is biased in general with the bias depending on the size of the sampling interval and the first two moments of the sojourn time distribution function. The bias becomes negligible as the size of the sampling interval increases. Analytical and empirical results indicate that the method of moments estimator is less sensitive to the size of the sampling interval and has slightly better mean squared error properties than the least squares estimator.

[1]Published in Communications in Statistics: Theory and Methods, Volume 24, Issue 2, pp. 553-579, 1995

1 INTRODUCTION

Let $(X, T) = \{X_n, T_n, n = 0, 1, \ldots\}$ be a Markov renewal process, see for example Çinlar(1975), where we define X_n to be the state of the process just after the n^{th} event and T_n to be the time of the n^{th} event. The process $\{X_n\}$ is a finite state Markov chain and the inter-event time or sojourn time, $Y_n = T_{n+1} - T_n$, has a distribution dependent on X_n. The purpose of this paper is to present estimators for the means of the sojourn time distribution functions under a strict data availability condition.

To motivate the data availability condition, consider a server processing heterogeneous customers. The objective is to estimate the mean service times for each customer type under a data availability condition commonly encountered in shop floor systems. Specifically, we can observe the total service time and production quantities by product type during fixed-length time intervals, such as shifts. More generally, we assume that we can only observe the state exiting counting processes associated with the MRP over fixed intervals of time. We can not observe the actual sojourn times nor can we observe state transitional count data, i.e. specific data concerning the exact times and types of state transitions is unavailable.

Moore and Pyke(1968) examined estimators of the transition distributions of a MRP based on maximum likelihood estimation(MLE). In Moore and Pyke(1968), sojourn times and transitional count data are observable; however, in our case these data are unavailable so that a straight forward MLE approach is not possible. Markov renewal processes and their statistical estimation are also discussed in Karr(1986). The nonparametric maximum likelihood estimators discussed in Karr(1986) do not apply here because of our data availability constraint. Instead, we derive and evaluate estimators based on least squares(LS) and on the method of matching moments(MOM).

The rest of this paper is structured as follows. First, we give the application which motivated our research. In Section 2, we present the definition and moment solutions for the exiting counting processes associated with the Markov renewal process. In Section 3, we present our basic least squares estimator, and in Section 4 we present a competing estimator based upon the method of matching moments. In Section 5, we evaluate empirically the performance of the estimators via simulation, and in Section 6 we present a modified version of the estimators with improved performance. Finally, we summarize our results and discuss future research.

2 DEFINITION AND MOMENT SOLUTIONS

We begin this section with an application which motivates the estimation problem examined in this paper. We then define the state-exiting counting processes and present solutions for the moments of the counting processes.

2.1 Motivating Application

This research evolved from a capacity planning project where there was a need to accurately estimate service times in multi-server, multi-class customer queueing systems. The major constraint associated with the project was that there was to be minimal effort expended in data collection. Bar code scanning stations between work centers having multiple servers allow for the automatic collection of customer arrival and departure information, but not the amount of time spent in service for each customer. These data do allow for calculation of total service time exclusive of idle time and for production counts. We refer the reader to Rossetti(1992) for more details of the system configuration. In order to develop and analyze estimators for the more general multi-server, multi-class customer queueing system, we decided to approximate the system behavior with a single-server system which has a underlying Markov renewal process structure where time is service time exclusive of idle time.

Assume that we have a server which processes heterogeneous customers, and that the selection of the next customer occurs according to a Markov chain. The server spends a random amount of time servicing the customer where the service time distribution depends only upon the type of customer selected. If we define the state of the process as the type of customer the server is currently serving, then the sojourn time distributions for the states of the MRP correspond to the service time distribution functions for the customers. In the next section, we give a precise definition for the exiting counting processes described in this example.

2.2 Exiting Counting Process

We begin with a definition and some notation to precisely define the counting process for the number of times a state is exited for the MRP examined in this paper.

Definition 1 *Let* $\{X_n, n \geq 0\}$ *be a Markov Chain with a finite state space,* $S = \{1, 2, \ldots, K\}$. *Let* $\{Y_n, n \geq 0\}$ *be a sequence of non-negative random variables where* $n \geq 1$ *and* $Y_0 = 0$. *Let* $T_n = Y_0 + Y_1 + \cdots + Y_n$, $n \geq 0$. *Let* $(X, T) = \{X_n, T_n; n \geq 0\}$ *be a Markov renewal process with the property*

$$\Pr\{X_{n+1} = j, T_{n+1} - T_n \leq t \mid X_0, \ldots, X_n; T_0, \ldots, T_n\} = \Pr\{T_{n+1} - T_n \leq t \mid X_n\}$$

Let $\{X(t), t \geq 0\}$ *be the semi-Markov process engendered by* (X, T), *where* $\{X(t), t \geq 0\}$, *is defined as* $X(t) = X_n$ *if and only if* $T_n \leq t \leq T_{n+1}$. *The random variables* T_n *and* Y_n *are called the* n^{th} *transition epoch and the* n^{th} *sojourn time of the process and* $F_i(t) = \Pr\{Y_n \leq t \mid X_{n-1} = i\}$ *is called the sojourn time distribution function for state* i. *Define* θ_i *and* δ_i^2 *to be the mean and the variance of the nth sojourn time,* Y_n, *whose distribution function is* $F_i(t)$. *Let* $p_{ij} = \Pr\{X_n = j \mid X_{n-1} = i\}$ *be the Markov chain single step transition probabilities. Let* $N(t) = \max\{n : T_n \leq t\}$, *so that* $N(t)$ *is the number of transition epochs in* $(0, t]$. *Let* $N_j(t)$ = *number of times* $X_n = j$ *for* $0 \leq n \leq N(t) - 1$ *with* $N_j(t) = 0$ *for all* j *if* $N(t) = 0$. *Thus* $N_j(t)$ *represents the number of times state* j *has been exited in the interval* $(0, t]$. *We call the* $N_j(t)$ *the Exiting Counting Processes(ECP's) associated with the MRP.*

In terms of notation, our goal is to estimate the means, θ_i, of the distribution functions, $F_i(t)$, from only the count of the number of times each state has been exited, $N_i(t)$. The transition probabilities are also unknown, but we will show that they are not needed in estimating the mean sojourn times. We develop a least squares estimator and an estimator based on the method of moments. In order to analyze the statistical properties of these estimators, we need expressions for the moments of the ECP which we present in the following section.

2.3 ECP Moment Solutions

In this section, we present expressions for the unconditional equilibrium moments of the ECP associated with the MRP. In general, the moment expressions can be quite complicated. The following Theorem establishes the conditional moments of the ECP and shows the relationship between the ECP results and the state entering counting process. We denote the Laplace-Stieltjes transform of a real-valued function $G(\cdot), t \geq 0$, as $\tilde{G}(s) = \int_0^\infty e^{-st} dG(t)$ and $\mathbf{G}(\cdot) = [G_{ij}(\cdot)]$ as a matrix of functions. Also, we let $\text{diag}(d_1, \ldots, d_n)$ represent a matrix with $d_i, i = 1, \ldots, n$ on the diagonal and zeros elsewhere, $\mathbf{I} = diag(1, \ldots, 1)$, and $\mathbf{G}^d = \text{diag}(G_{11}, \ldots, G_{nn})$.

Theorem 1 *Let $L_j(t)$ = number of times state j is entered or the number of times $X_n = j$ for $0 < n \leq N(t)$. Define $Q_{ij}(t) = p_{ij}F_i(t)$. Let $H_{ij}(t) = \mathrm{E}[L_j(t) \mid X(0) = i]$, $M_{ij}(t) = \mathrm{E}[N_j(t) \mid X(0) = i]$, $W_{ij}(t) = \mathrm{E}\left[N_j^2(t) \mid X(0) = i\right]$, and $h_{i(jk)}(t) = \mathrm{E}[N_j(t)N_k(t) \mid X(0) = i]$, where $h_{i(jk)}(t) = h_{i(kj)}(t)$ for $j \neq k$, and $B(t) = \text{diag}(F_1(t), \ldots, F_K(t))$. Let $\vec{\mathbf{h}}^{(jk)}(t) = \left[h_{1(jk)}(t), h_{2(jk)}(t), \ldots, h_{K(jk)}(t)\right]$ and let $\vec{\mathbf{m}}^{(jk)}(t)$ be a $n \times 1$ vector with elements*

$$m_i = \begin{cases} 0 & \text{if } i \neq j \text{ and } i \neq k \\ M_{jk}(t) & \text{if } i = j \\ M_{kj}(t) & \text{if } i = k \end{cases}$$

Then

$$\tilde{\mathbf{M}}(s) = \left[\mathbf{I} - \tilde{\mathbf{Q}}(s)\right]^{-1} \tilde{\mathbf{B}}(s) \tag{1}$$

$$\tilde{\mathbf{M}}(s) = \left[\tilde{\mathbf{H}}(s) + \mathbf{I}\right] \tilde{\mathbf{B}}(s) \tag{2}$$

$$\tilde{\mathbf{W}}(s) = \tilde{\mathbf{M}}(s)\left\{2\left[\tilde{\mathbf{M}}(s)\tilde{\mathbf{B}}^{-1}(s)\right]^d - \mathbf{I}\right\} \tag{3}$$

$$\tilde{\vec{\mathbf{h}}}^{(jk)}(s) = \tilde{\mathbf{M}}(s)\tilde{\mathbf{B}}^{-1}(s)\tilde{\vec{\mathbf{m}}}^{(jk)}(s) \tag{4}$$

The proof is outlined in the appendix. For a more detailed discussion of results relating to the moment solutions of the ECP, we refer the reader to Rossetti and Clark(1994). In order to be able to utilize moment expressions in the analysis of estimators for θ_i, we develop asymptotic expansions for the moments of the ECP in terms of the moments of $F_i(t)$ and the transition probabilities of the embedded Markov chain. Throughout the remainder of the paper, we assume that the embedded Markov chain has two-states and is ergodic, i.e. $0 < p_{ii} < 1$, for i= 1,2. We do this for two reasons. First, the reduction to two states allows for easier analytical investigation which highlights the essential properties of the estimators. Second, the estimation method is not limited to two states, so that the methodology can be applied in the larger state space case. We discuss how the larger state space case might be handled in our conclusions. The following result presents the equilibrium moment solutions necessary for analysis in this paper.

Result 1 *Let $(X,T) = \{X_n, T_n; n \geq 0\}$ be a Markov renewal process as defined in Definition 1 with state space, $S = \{1,2\}$ and $0 < p_{ii} < 1, i = 1,2$. Assume that the processes, $N_j(t)$, have existed in the distant past and that we start observing the process at a randomly selected point in time which we define to be $t = 0$. We denote this process as the equilibrium process. Furthermore, suppose that the Markov chain $\{X_n, n \geq 0\}$, associated with $\{X(t), t \geq 0\}$ is ergodic. Let α_i, σ_i^2 be the mean and variance of the recurrence time for states $i = 1,2$ and $\pi_j = lim_{n\to\infty} Pr\{X_n = j\}$. Then $\mathrm{E}[N_1(t)] = t/\alpha_1$, $\mathrm{E}[N_2(t)] = t/\alpha_2$ and*

$$\mathrm{E}\left[N_1^2(t)\right] = B_1 + B_2 t + B_3 t^2/2 + o(1) \quad (5)$$

$$\mathrm{E}\left[N_2^2(t)\right] = D_1 + D_2 t + D_3 t^2/2 + o(1) \quad (6)$$

$$\mathrm{E}\left[N_1(t)N_2(t)\right] = A_1 + A_2 t + A_3 t^2/2 + o(1) \quad (7)$$

where $B_2 = \sigma_1^2/\alpha_1^3$, $B_3 = 2/\alpha_1^2$, $D_2 = \sigma_2^2/\alpha_2^3$, $D_3 = 2/\alpha_2^2$, $A_3 = 2/\alpha_1\alpha_2$ and

$$A_2 = \frac{\pi_1\pi_2}{\theta^3}\left[\delta^2 - \theta_1\theta_2\left(\frac{2}{\beta} - 1\right)\right] \quad (8)$$

with $\theta = \pi_1\theta_1 + \pi_2\theta_2$, $\beta = 2 - p_{11} - p_{22}$, and $\delta^2 = \pi_1\delta_1^2 + \pi_2\delta_2^2$.

The proof utilizes the results of Theorem 1 to develop the equilibrium exiting processes and then asymptotically expands the moment's Laplace-Stieltjes transforms. Since this process involves considerable algebraic manipulations, we refer the interested reader to Rossetti and Clark(1994) for the details. The coefficients B_1, D_1, and A_1 are complicated functions of the p_{ij} and the first three moments of $F_i(t)$ and are given in Rossetti and Clark(1994). We will utilize the following approximation in our analysis which involves dropping the coefficients B_1, D_1, and A_1 since they would be small in comparison to the other terms for large t.

Definition 2 *The moments* $\mathrm{E}[N_1^2(t)]$, $\mathrm{E}[N_2^2(t)]$, *and* $\mathrm{E}[N_1(t)N_2(t)]$ *have the form* $k_0 + k_1 t + k_2 t^2/2 + o(1)$. *We define the approximation by dropping the* k_0 *and* $o(1)$ *terms so that*

$$\mathrm{E}\left[N_1^2(t)\right] \doteq B_2 t + B_3 t^2/2 \tag{9}$$

$$\mathrm{E}\left[N_2^2(t)\right] \doteq D_2 t + D_3 t^2/2 \tag{10}$$

$$\mathrm{E}\left[N_1(t)N_2(t)\right] \doteq A_2 t + A_3 t^2/2 \tag{11}$$

In Rossetti and Clark(1994), we analytically and empirically evaluate the above approximation and one other approximation in order to assess their accuracy. For the case of $F_i(t) = 1 - e^{-t/\theta_i}, i = 1, 2$, we have the following result for the approximation given in Definition 2.

Result 2 *The moments* $\mathrm{E}[N_1^2(t)]$, $\mathrm{E}[N_2^2(t)]$, *and* $\mathrm{E}[N_1(t)N_2(t)]$ *have the form* $k_0(1 - e^{-\gamma t}) + k_1 t + k_2 t^2/2$. *Let*

$$\mathrm{AE}(\cdot) = |true - approximation| \tag{12}$$

$$\mathrm{ARE}(\cdot) = \frac{|true - approximation|}{|true|} \tag{13}$$

For fixed parameters $\mathrm{AE}(\cdot)$ *has the following form*

$$\mathrm{AE}(\cdot) = \left|k_0(1 - e^{-\gamma t})\right|$$

so that

	$\mathrm{E}[N_1^2(t)]$	$\mathrm{E}[N_2^2(t)]$	$\mathrm{E}[N_1(t)N_2(t)]$
$\lim_{t\to+\infty} \mathrm{AE}(\cdot)$	$\lvert k_0\rvert$	$\lvert k_0\rvert$	$\lvert k_0\rvert$
$\lim_{t\to 0^+} \mathrm{AE}(\cdot)$	0	0	0
$\lim_{t\to+\infty} \mathrm{ARE}(\cdot)$	0	0	0
$\lim_{t\to 0^+} \mathrm{ARE}(\cdot)$	$\left\lvert 1 - \frac{\sigma_1^2}{\alpha_1^2}\right\rvert$	$\left\lvert 1 - \frac{\sigma_2^2}{\alpha_2^2}\right\rvert$	$+\infty$

and for fixed t, $\mathrm{AE}(\cdot)$ *and* $\mathrm{ARE}(\cdot)$ *may be arbitrarily large depending on the choice of parameters.*

The proofs of $\mathrm{AE}(\cdot)$ for ($t \to 0^+$, and $t \to +\infty$) follow immediately from the functional form of $\mathrm{AE}(\cdot)$. The proofs of $\mathrm{ARE}(\cdot)$ for ($t \to +\infty$) and ($t \to 0^+$)

follow from an application of L'Hospital's rule. The result for fixed t follows by noting that $(k_0 \to \infty)$ as $(p_{11} \to 1$ and $p_{22} \to 1)$. Plots of the error functions indicated that the error drops off quickly as t gets large.

To examine the average behavior of the approximation, we randomly selected $n = 10000$ parameter values according to a uniform distribution such that the range of parameters was $0 < p_{11}, p_{22} < 1$ and $1 < \theta_1, \theta_2 < 100$ and then computed ARE(·). Table 1 reports the statistics for sampled absolute relative errors for $t = 120, 480$. The results indicate that in terms of absolute relative error the approximation is quite good over a reasonable range of parameters when t is large compared to the parameters. The form of the approximation also allows for easier analysis. The next sections will utilize the approximation in analyzing the statistical properties of the proposed estimators.

Table 1: Sample Results for ARE(·).

t		ARE(E $[N_1^2(t)]$)	ARE(E $[N_2^2(t)]$)	ARE(E $[N_1(t)N_2(t)]$)
120	$\bar{x}$	0.0065	0.0081	0.1457
	s^2/n	1.61 $\times 10^{-6}$	3.91 $\times 10^{-5}$	0.0105
	n	10000	10000	10000
	min	1.09 $\times 10^{-8}$	7.4 $\times 10^{-9}$	1.03 $\times 10^{-8}$
	max	11.44	14.15	972.19
480	$\bar{x}$	0.001	0.001	0.01
	s^2/n	9.0 $\times 10^{-8}$	2.3 $\times 10^{-7}$	5.03 $\times 10^{-5}$
	n	10000	10000	10000
	min	7.0 $\times 10^{-10}$	5.0 $\times 10^{-10}$	6.0 $\times 10^{-10}$
	max	2.69	3.59	65.92

where $\bar{x}$ is the sample average and s^2 is the sample variance.

3 LEAST SQUARES ESTIMATOR

In this section, we present the least squares estimator for the means of the sojourn time distribution functions. Again, we limit our analysis to the equilibrium version of a two state MRP with $0 < p_{ii} < 1$, for i= 1,2. We assume that we can observe the MRP for a predetermined length of time which can be divided into intervals of fixed length, τ, in which only the count of the number of times each state was exited is given. Consider the following linear model

$$Y_i = \sum_{k=1}^{K} X_{ik}\theta_k + \epsilon_i, \ (i = 1, \ldots, n) \tag{14}$$

where $K = 2$ is the number of states, n is the number of fixed length intervals, $Y_i = \tau$ is the amount of time in interval i, X_{ik} is the number of times state k is exited during interval i, θ_k is the mean of the sojourn time distribution function, and ϵ_i is the model error term. From least squares theory, we have the following.

Estimator 1 *Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be the least squares estimators of θ_1 and θ_2 for the model given by Equation (14). Denote the equilibrium process for the ECP by $\{N_k^e(t)\}$. Let*

$$X_{ik} = N_k^e((i)\tau) - N_k^e((i-1)\tau), \ i = 1, 2, \ldots, n \tag{15}$$

where τ is a fixed interval size. Let $\overline{X}_1$, $\overline{X}_2$, $\overline{X}_1^{(2)}$, $\overline{X}_2^{(2)}$, and $\overline{X_1X_2}$ be the sample moments. Then,

$$\hat{\theta}_1 = \frac{\tau\left((\overline{X}_1)(\overline{X}_2^{(2)}) - (\overline{X_1X_2})(\overline{X}_2)\right)}{(\overline{X}_1^{(2)})(\overline{X}_2^{(2)}) - (\overline{X_1X_2})^2} \tag{16}$$

$$\hat{\theta}_2 = \frac{\tau\left((\overline{X}_2)(\overline{X}_1^{(2)}) - (\overline{X_1X_2})(\overline{X}_1)\right)}{(\overline{X}_1^{(2)})(\overline{X}_2^{(2)}) - (\overline{X_1X_2})^2} \tag{17}$$

The estimator follows immediately from least squares theory and substitution of the sample moments. Note that since $\mathrm{E}[X_{ik}] = \tau/\alpha_k$ and $\alpha_k = \left(\sum_{j=1}^{K} \pi_j\theta_j\right)/\pi_k$ the expectation of Equation (14) yields $\mathrm{E}[\epsilon_i] = 0$. We note that in practice, $\hat{\theta}_1$ and $\hat{\theta}_2$ could be negative and rejected as viable estimates; however, we do not address that problem in this paper. The following theorem states the convergence property of the least squares estimators.

Theorem 2 *Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be the least squares estimators of θ_1 and θ_2. Assume that $\{X_{ik}, i = 1, 2, \ldots,\}$ is ergodic with $\mathrm{E}[X_{ik}] < \infty$ and $\mathrm{E}[X_{ik}^2] < \infty$ then as $n \to \infty$, and provided $\mathrm{E}[N_1^2(\tau)]\,\mathrm{E}[N_2^2(\tau)] - (\mathrm{E}[N_1(\tau)N_2(\tau)])^2 \neq 0$, we have*

$$\hat{\theta}_1 \overset{wp1}{\to} \hat{\theta}_1^a$$

$$= \frac{\tau\left(\mathrm{E}\left[N_1(\tau)\right]\mathrm{E}\left[N_2^2(\tau)\right]-\mathrm{E}\left[N_1(\tau)N_2(\tau)\right]\mathrm{E}\left[N_2(\tau)\right]\right)}{\mathrm{E}\left[N_1^2(\tau)\right]\mathrm{E}\left[N_2^2(\tau)\right]-\left(\mathrm{E}\left[N_1(\tau)N_2(\tau)\right]\right)^2} \tag{18}$$

$$\hat{\theta}_2 \overset{wp1}{\to} \hat{\theta}_1^a$$

$$= \frac{\tau\left(\mathrm{E}\left[N_2(\tau)\right]\mathrm{E}\left[N_1^2(\tau)\right]-\mathrm{E}\left[N_1(\tau)N_2(\tau)\right]\mathrm{E}\left[N_1(\tau)\right]\right)}{\mathrm{E}\left[N_1^2(\tau)\right]\mathrm{E}\left[N_2^2(\tau)\right]-\left(\mathrm{E}\left[N_1(\tau)N_2(\tau)\right]\right)^2} \tag{19}$$

The proof appears in the appendix. By ergodic, we mean that there exists a finite constant to which the sample mean converges as the sample size increases to infinity; see for example Karlin and Taylor(1975, pg. 487–488). We note that the $\hat{\theta}_i^a$ are functions of the first, second, and covariance moments for the ECP of the MRP.

3.1 Approximations for Bias, Variance, and MSE

In this section, we give approximations for the bias, variance, and mean squared error of the least squares estimator in terms of the parameters of the MRP using the approximations given in Definition 2.

Corollary 1 *Let* $\mathrm{Bias}\left[\hat{\theta}_i\right] = \mathrm{E}\left[\hat{\theta}_i\right] - \theta_i$*. Given Result 1, Theorem 2, and Definition 2 we have,*

$$\mathrm{Bias}\left[\hat{\theta}_i\right] \doteq \theta_i\left(\frac{1-B_c}{B_c}\right) \tag{20}$$

where $B_c = 1 + b_c/\tau$ *and* $b_c = \delta^2/\theta$*.*

The proof appears in the appendix. This result indicates that the least squares estimator is biased in general with the approximate bias dependent on τ, δ^2, and θ. We note that as τ increases the bias should improve since the biasing constant, B_c, approaches one.

Regression in the standard linear model assumes that the regressor variables are non-random, the error term ϵ_i has mean zero and constant variance and that ϵ_i and ϵ_j are uncorrelated, see for example Graybill(1976, pg. 171; 216). In our case the regressor variables, X_{ik}, are random variables and the linear relationship is postulated between the expected values of the random variables involved in the model. The Y_i are constants and the errors will be correlated in general. The following results are offered as approximations for the behavior of the estimators. We shall also utilize the relationships to motivate a sampling method which improves the quality of the estimates.

Result 3 *Let MSE_R be the mean squared error of the residuals about the regression based on standard linear model regression assumptions. Given Result 1 and Definition 2, we have*

$$\widehat{\mathrm{MSE}}_R \doteq \frac{n\tau^2}{n-2}\left(\frac{b_c}{b_c+\tau}\right) \tag{21}$$

The proof appears in the appendix.

Result 4 *Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be the least squares estimators. Given Result 3, Definition 2 and assuming standard linear model regression assumptions, we have*

$$\widehat{\mathrm{Var}}\left[\hat{\theta}_1\right] \doteq \left(\frac{\tau^2}{T-2\tau}\right)\left(\frac{b_c}{b_c+\tau}\right)\left(\left(\frac{\pi_2\theta}{\pi_1}\right)\left(\frac{\beta}{2-\beta}\right)+\frac{\theta_1^2}{b_c+\tau}\right) \tag{22}$$

$$\widehat{\mathrm{Var}}\left[\hat{\theta}_2\right] \doteq \left(\frac{\tau^2}{T-2\tau}\right)\left(\frac{b_c}{b_c+\tau}\right)\left(\left(\frac{\pi_1\theta}{\pi_2}\right)\left(\frac{\beta}{2-\beta}\right)+\frac{\theta_2^2}{b_c+\tau}\right) \tag{23}$$

where $T = n\tau$ is the total observation time.

The proof appears in the appendix.

Remark 1 If ($\delta_1^2 \to 0$ and $\delta_2^2 \to 0$) then $b_c \to 0$ and $B_c \to 1$ so that $\hat{\theta}_1^a \overset{\text{approx}}{\to} \theta_1$ and $\hat{\theta}_2^a \overset{\text{approx}}{\to} \theta_2$. Also, if $b_c \to 0$ then $\widehat{\mathrm{Var}}\left[\hat{\theta}_1\right] \overset{\text{approx}}{\to} 0$ and $\widehat{\mathrm{Var}}\left[\hat{\theta}_2\right] \overset{\text{approx}}{\to} 0$. Thus, the bias and variance properties of the least squares estimators should improve as the sojourn distribution functions become more deterministic.

Remark 2 Let $c_i = \delta_i/\theta_i$ be the coefficient of variation for the sojourn time distribution function for state i. Suppose, $0 \le c_i^2 \le k_i$, $i = 1, 2$ where k_i is a constant, then

$$1 \le B_c \le 1 + \left(\frac{\nu_1\theta_1 + \nu_2\theta_2}{\tau}\right)\max(k_1, k_2)$$

where $\nu_j = \pi_j\theta_j/\theta$. So that for large τ, B_c can be made close to one, and for small coefficients of variation B_c is close to one. Thus, the bias should improve as τ increases and as c_i decreases.

Remark 3 The approximation for the estimate of the variance of $\hat{\theta}_i$ increases as π_i decreases due to the π_i term in the denominator of Equations 22 and 23. This makes intuitive sense since if π_i decreases there is less chance that the state will be exited and thus less data available for estimating that state's mean sojourn time. Also, the approximation for the estimate of the variance of $\hat{\theta}_i$ increases as $2-\beta = p_{11}+p_{22}$ approaches zero. We note that as $2-\beta$ approaches zero the two state Markov renewal process behaves like an alternating renewal process.

4 METHOD OF MOMENTS ESTIMATOR

In this section, we present an estimator for the mean of the sojourn time distribution function based upon the method of matching moments. The difficulty with applying the method of moments to the two state problem is that the number of parameters to estimate is six, (θ_1, θ_2, δ_1^2, δ_2^2, p_{11}, and p_{22}), while we only have first moment, second moment, and covariance moment information, (four independent equations). In order to reduce the complexity of the equations, we will utilize the approximations given in Definition 2 and combine unneeded terms to form nuisance parameters. The next result presents the method of moments estimator.

Estimator 2 *Let $\tilde{\theta}_1$ and $\tilde{\theta}_2$ be the method of moments estimators. Given Result 1 and Definition 2, we have*

$$\tilde{\theta}_1 = \frac{\tau\left((\overline{X}_1)(\overline{X_2^{(2)}}) - (\overline{X_1X_2})(\overline{X}_2)\right)}{(\overline{X}_1)^2(\overline{X_2^{(2)}}) - 2(\overline{X}_1)(\overline{X}_2)(\overline{X_1X_2}) + (\overline{X}_2)^2(\overline{X_1^{(2)}})} \tag{24}$$

$$\tilde{\theta}_2 = \frac{\tau\left((\overline{X}_2)(\overline{X_1^{(2)}}) - (\overline{X_1X_2})(\overline{X}_1)\right)}{(\overline{X}_1)^2(\overline{X_2^{(2)}}) - 2(\overline{X}_1)(\overline{X}_2)(\overline{X_1X_2}) + (\overline{X}_2)^2(\overline{X_1^{(2)}})} \tag{25}$$

provided $(\overline{X}_1)^2(\overline{X_2^{(2)}}) - 2(\overline{X}_1)(\overline{X}_2)(\overline{X_1X_2}) + (\overline{X}_2)^2(\overline{X_1^{(2)}}) \neq 0$

The proof appears in the appendix.

Corollary 2 *Let $\tilde{\theta}_1$ and $\tilde{\theta}_2$ be the method of moments estimators. Assume that $\{X_{ik}, i = 1, 2, \ldots, \}$, as defined in Equation (15), is ergodic with $\mathrm{E}\,[X_{ik}] < \infty$ and $\mathrm{E}\,[X_{ik}^2] < \infty$ then as $n \to \infty$, we have*

$$\tilde{\theta}_1 \overset{wp1}{\to} \tilde{\theta}_1^a \doteq \theta_1 \tag{26}$$

$$\tilde{\theta}_2 \overset{wp1}{\to} \tilde{\theta}_1^a \doteq \theta_2 \tag{27}$$

The proof appears in the appendix.

Remark 4 Corollary 2 states that the asymptotic values of the method of moments estimators are *approximately* equal to the true values of the parameters, but note that we need τ large enough to make the error in the approximations negligible. Note also that the p_{ij} became nuisance parameters and are not needed in the estimation of the sojourn times.

We note that the $N_i(\tau)$ are asymptotically normal, see for example Heyman and Sobel(1982, pg. 114) and Taga(1963, pg. 8). Suppose that $N_1(\tau)$ and $N_2(\tau)$ are bivariate normal, then the sample moments, $\overline{X}_1$, $\overline{X}_2$, $\overline{X}_1^{(2)}$, $\overline{X}_2^{(2)}$, and $\overline{X_1X_2}$ are maximum likelihood estimators(MLE's) of the corresponding moment parameters of the bivariate normal distribution; see for example Rao(1973), pg. 447). Thus, since the method of moments estimator is a function of the moments for $N_i(\tau)$, the MOM estimator is an asymptotic function of MLE's and thus the MOM is asymptotically a MLE.

5 EMPIRICAL EVALUATION OF ESTIMATORS

In this section, we present an empirical comparison of the least squares estimator and the method of moments estimator. The cases we examined are given in Table 2. The comparison is given in terms of estimates for the bias, variance, and mean squared error of the estimators. In all cases, we simulated $T = 50,000$ minutes of the MRP. The observation period, T, was divided into fixed intervals of size τ and the number of times each state was exited during each interval was collected. The estimates were then computed using the equations given in Estimators 1 and 2. Each simulation was replicated r times to produce r estimates of the estimators. We call these r replications, micro-replications. Each micro-replication provides an estimate of the bias, variance, and mean squared error of the estimators. Each of the r replications were replicated R times to provide a sampling distribution for the estimates of bias, variance, and mean squared error. In the tables $\bar{x}$ represents the average and s^2 represents the unbiased sample variance of the estimates of $\widehat{\text{Bias}}$, $\widehat{\text{Var}}$, and $\widehat{\text{MSE}}$ from the R macro-replications. The simulations were performed using Simscript II.5. The transition to the initial state j was selected according to the steady state probability of transition π_j.

Table 2: Experimental Test Cases

$$F_i(t) = 1 - e^{-t/\theta_i}$$
$$\theta_1 = 20,\ \theta_2 = 10$$

Case	p_{11}	p_{22}
1	0.1	0.1
2	0.9	0.9
3	0.9	0.1
4	0.1	0.9

The results of the experimental test cases are given in Tables 3 and 4. The results indicate that the bias associated with the least squares estimator can be quite severe for small values of τ. The method of moments estimator suffers a smaller increase in bias as τ decreases. The variance of both estimators can be significant enough to make estimation doubtful as in Case 1 Table 3. Increasing the number of intervals while decreasing their length reduces the variance of the estimators; however, the bias is significantly increased for the least squares estimator, especially for $\hat{\theta}_1$. In terms of $\widehat{\text{MSE}}$, in Table 3 the estimators are competitive while for Table 4 the method of moments estimator appears to have better mean squared error properties. Overall, the method of moments estimator appears more robust to the selection of the parameter τ.

Table 3: Empirical Results $n = 100$, $\tau = 500$

R = 51, r = 70, $\theta_1 = 20$, $\theta_2 = 10$

			LS		MOM	
Case			$\hat{\theta}_1$	$\hat{\theta}_2$	$\tilde{\theta}_1$	$\tilde{\theta}_2$
1	$\bar{\hat{\theta}}$	$\bar{x}$	18.9198	10.1367	19.539	10.4686
		s^2	0.306	0.301	0.325	0.321
	$\widehat{\text{Bias}}$	$\bar{x}$	-1.0802	0.1367	-0.4610	0.4686
		s^2	0.306	0.301	0.325	0.321
	$\widehat{\text{Var}}$	$\bar{x}$	19.631	19.4703	20.9381	20.7725
		s^2	10.710	10.121	12.135	11.482
	$\widehat{\text{MSE}}$	$\bar{x}$	21.0982	19.7837	21.4697	21.3069
		s^2	14.589	10.793	13.805	13.266
2	$\bar{\hat{\theta}}$	$\bar{x}$	19.2993	9.7572	19.9303	10.0762
		s^2	0.009	0.005	0.010	0.005
	$\widehat{\text{Bias}}$	$\bar{x}$	-0.7007	-0.2428	-0.0697	0.0762
		s^2	0.009	0.005	0.010	0.005
	$\widehat{\text{Var}}$	$\bar{x}$	0.4466	0.2966	0.4801	0.3170
		s^2	0.006	0.002	0.007	0.002
	$\widehat{\text{MSE}}$	$\bar{x}$	0.9464	0.3601	0.4946	0.3274
		s^2	0.02	0.004	0.007	0.002
3	$\bar{\hat{\theta}}$	$\bar{x}$	19.2521	9.8673	19.986	10.2426
		s^2	0.009	0.529	0.009	0.572
	$\widehat{\text{Bias}}$	$\bar{x}$	-0.7479	-0.1327	-0.014	0.2426
		s^2	0.009	0.529	0.009	0.572
	$\widehat{\text{Var}}$	$\bar{x}$	0.5416	31.479	0.5887	33.9096
		s^2	0.001	35.122	0.011	40.645
	$\widehat{\text{MSE}}$	$\bar{x}$	1.1093	32.015	0.5976	34.529
		s^2	0.031	35.948	0.011	41.011
4	$\bar{\hat{\theta}}$	$\bar{x}$	19.3078	9.8056	19.7551	10.0327
		s^2	0.198	0.003	0.207	0.003
	$\widehat{\text{Bias}}$	$\bar{x}$	-0.6922	-0.1944	-0.2449	0.0327
		s^2	0.198	0.003	0.207	0.003
	$\widehat{\text{Var}}$	$\bar{x}$	10.9474	0.1505	11.471	0.1577
		s^2	3.273	0.001	3.609	0.001
	$\widehat{\text{MSE}}$	$\bar{x}$	11.6203	0.1910	11.7338	0.1616
		s^2	4.975	0.001	4.462	0.001

Table 4: Empirical Results $n = 500$, $\tau = 100$

R = 51, r = 70, $\theta_1 = 20$, $\theta_2 = 10$

			LS		MOM	
Case			$\hat{\theta}_1$	$\hat{\theta}_2$	$\tilde{\theta}_1$	$\tilde{\theta}_2$
1	$\bar{\hat{\theta}}$	$\bar{x}$	15.3695	10.377	17.913	12.0946
		s^2	0.030	0.033	0.043	0.043
	$\widehat{\text{Bias}}$	$\bar{x}$	-4.6305	0.377	-2.087	2.0946
		s^2	0.030	0.033	0.043	0.043
	$\widehat{\text{Var}}$	$\bar{x}$	2.1088	2.0875	2.8652	2.8443
		s^2	0.097	0.114	0.187	0.207
	$\widehat{\text{MSE}}$	$\bar{x}$	23.5803	2.2617	7.2629	7.2738
		s^2	2.541	0.127	0.818	0.906
2	$\bar{\hat{\theta}}$	$\bar{x}$	16.7161	9.0369	19.4780	10.5301
		s^2	0.004	0.001	0.005	0.001
	$\widehat{\text{Bias}}$	$\bar{x}$	-3.2839	-0.9631	-0.5220	0.5301
		s^2	0.004	0.001	0.005	0.001
	$\widehat{\text{Var}}$	$\bar{x}$	0.1880	0.0802	0.2640	0.1137
		s^2	0.001	0.00018	0.002	0.00028
	$\widehat{\text{MSE}}$	$\bar{x}$	10.9761	1.0088	0.5418	0.3961
		s^2	0.160	0.004	0.008	0.002
3	$\bar{\hat{\theta}}$	$\bar{x}$	16.6726	9.1783	19.9061	10.9586
		s^2	0.003	0.079	0.003	0.111
	$\widehat{\text{Bias}}$	$\bar{x}$	-3.3274	-0.8217	-0.0939	0.9586
		s^2	0.003	0.079	0.003	0.111
	$\widehat{\text{Var}}$	$\bar{x}$	0.1707	4.4597	0.2492	6.3686
		s^2	0.001	0.493	0.002	1.052
	$\widehat{\text{MSE}}$	$\bar{x}$	11.2449	5.2121	0.2614	7.3965
		s^2	0.118	0.869	0.002	1.305
4	$\bar{\hat{\theta}}$	$\bar{x}$	16.4738	9.1135	18.4071	10.1829
		s^2	0.029	0.001	0.037	0.001
	$\widehat{\text{Bias}}$	$\bar{x}$	-3.5261	-0.8865	-1.5929	0.1829
		s^2	0.029	0.001	0.037	0.001
	$\widehat{\text{Var}}$	$\bar{x}$	1.971	0.0396	2.4729	0.0501
		s^2	0.149	0.00004	0.241	0.00006
	$\widehat{\text{MSE}}$	$\bar{x}$	14.4335	0.8261	5.0461	0.0843
		s^2	1.575	0.002	0.596	0.00014

6 IMPROVED ESTIMATORS

In this section, we present a modification of the fixed interval sampling method which allows for improved estimator performance. Again, we assume that we have a fixed total observation period, T, that can be divided into n intervals of length, τ. For each interval, only the number of times a state is exited is observable, but we have the ability to vary the length of the intervals. The motivation behind varying the length of the intervals comes from an examination of the mean squared error for the least squares estimator. For fixed T, the length of τ affects the bias and variance of the least squares estimator. As τ increases the bias decreases and the variance increases. As τ decreases the bias increases and the variance decreases. Figure 1 illustrates how the approximate mean square error of the least squares estimator, derived from Corollary 1 and Result 4, varies according to fixed interval size for Case 1 of Table 2. In the figure, the total observation period, T, and the number of intervals, n, is fixed so that only τ varies. Note how the mean square error decreases and then begins to increase as τ increases. This behavior suggests that the mixing of smaller size intervals with larger size intervals may reduce the effect of the bias increase while gaining a variance reduction. We call the types of estimators presented in this section mixing estimators.

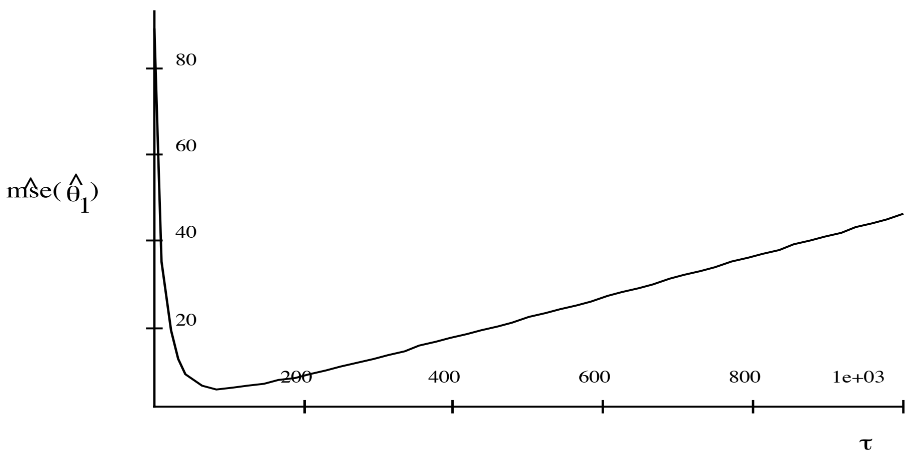


Figure 1: Approximate $\widehat{\mathrm{MSE}}(\hat{\theta}_1)$ vs Fixed Interval Size τ for Case 1

Figure 2 shows examples of two types of mixing strategies, non-overlapping intervals (NOLI) and overlapping intervals (OLI). The sample path, T, is divided into n_m intervals of size τ_m, such that $T \approx n_m \tau_m$, for $m = 1, \ldots, M$. We note that many other mixing strategies are possible. For the least squares estimator, we examined two basic strategies for performing the estimation, namely

1. perform the regression on each interval size separately and average the estimates of the parameters, i.e. let $\hat{\theta}_{im}$ be the least squares estimate for interval size m and state i. If we denote the averaged estimator as $\hat{\theta}_i^A$ then $\hat{\theta}_i^A = (1/M) \sum_{m=1}^{M} \hat{\theta}_{im}$ would be the estimator for θ_i.

2. combine the different sized intervals into one regression so that Y_i in Equation 14 would vary with the interval size. The estimates would simply be the standard least squares solution to linear model given in Equation 14. We denote this estimator as $\hat{\theta}_i^R$.

With any of the mixing strategies and estimation methods there will be a complex tradeoff between the degrees of freedom in the regression, dependence within the intervals, and dependence across intervals which can effect the bias and variance properties of the estimators. In Table 5, we present the results of applying OLI to the the test cases given in Table 2.

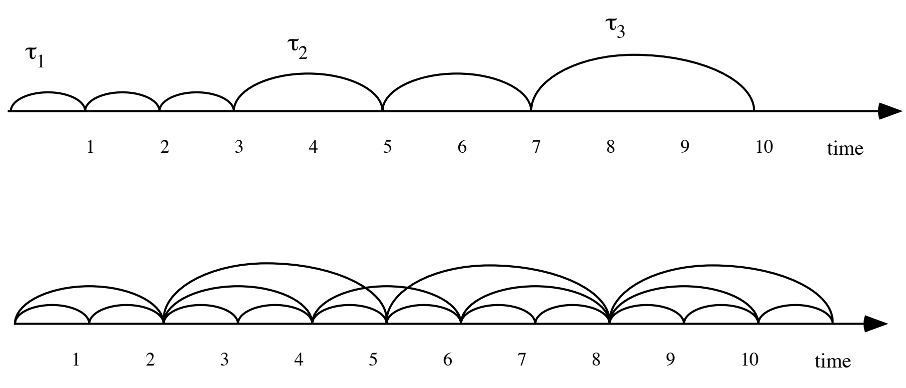


Figure 2: Example Mixing Strategies

Table 5: OLI Empirical Results

R = 51, r = 70, $\theta_1 = 20$, $\theta_2 = 10$
$n_1 = 200$, $n_2 = 100$, $n_3 = 66$, $T = 50000$
$\tau_1 = 250$, $\tau_2 = 500$, $\tau_3 = 750$

			LS				MOM	
Case			$\hat{\theta}_1^R$	$\hat{\theta}_2^R$	$\hat{\theta}_1^A$	$\hat{\theta}_2^A$	$\tilde{\theta}_1^A$	$\tilde{\theta}_2^A$
1	$\bar{\hat{\theta}}$	$\bar{x}$	18.722	10.3319	18.6223	10.2385	19.3572	10.6507
		s^2	0.196	0.194	0.206	0.203	0.220	0.217
	$\widehat{\text{Bias}}$	$\bar{x}$	-1.278	0.3319	-1.3777	0.2385	-0.6248	0.6507
		s^2	0.196	0.194	0.206	0.203	0.220	0.217
	$\widehat{\text{Var}}$	$\bar{x}$	13.225	13.0909	13.8356	13.7046	14.8066	14.6667
		s^2	4.285	3.588	4.712	3.994	5.39	4.536
	$\widehat{\text{MSE}}$	$\bar{x}$	15.0504	13.3909	15.9357	13.9609	15.4359	15.3032
		s^2	5.926	3.922	6.775	4.282	6.038	5.439
2	$\bar{\hat{\theta}}$	$\bar{x}$	19.2752	9.7778	19.1446	9.715	19.9043	10.1021
		s^2	0.007	0.003	0.007	0.003	0.007	0.003
	$\widehat{\text{Bias}}$	$\bar{x}$	-0.7248	-0.2222	-0.8554	-0.2850	-0.957	0.1021
		s^2	0.007	0.003	0.007	0.003	0.007	0.003
	$\widehat{\text{Var}}$	$\bar{x}$	0.3816	0.2360	0.3727	0.2294	0.4069	0.2479
		s^2	0.004	0.002	0.004	0.001	0.005	0.002
	$\widehat{\text{MSE}}$	$\bar{x}$	0.9136	0.2886	1.1109	0.3139	0.4234	0.2614
		s^2	0.017	0.003	0.021	0.003	0.005	0.002
3	$\bar{\hat{\theta}}$	$\bar{x}$	19.2376	9.9532	19.0904	9.8751	19.9756	10.3339
		s^2	0.006	0.34	0.006	0.336	0.007	0.365
	$\widehat{\text{Bias}}$	$\bar{x}$	-0.7624	-0.0468	-0.9096	-0.1249	-0.0244	0.3339
		s^2	0.006	0.34	0.006	0.336	0.007	0.365
	$\widehat{\text{Var}}$	$\bar{x}$	0.4342	22.5902	0.4325	22.7526	0.4762	24.646
		s^2	0.007	18.982	0.007	19.008	0.008	22.219
	$\widehat{\text{MSE}}$	$\bar{x}$	1.0217	22.926	1.266	23.098	0.4834	25.1153
		s^2	0.024	18.959	0.031	19.029	0.008	22.311
4	$\bar{\hat{\theta}}$	$\bar{x}$	19.2209	9.8143	19.1318	9.7663	19.6693	10.0422
		s^2	0.1615	0.002	0.1633	0.0021	0.1712	0.0022
	$\widehat{\text{Bias}}$	$\bar{x}$	-0.7791	-0.1857	-0.868	-0.2337	-0.3307	0.0422
		s^2	0.1615	0.002	0.1633	0.0021	0.1712	0.0022
	$\widehat{\text{Var}}$	$\bar{x}$	8.3114	0.1175	8.3919	0.1182	8.8287	0.1247
		s^2	1.514	0.00036	1.547	0.00038	1.7184	0.00044
	$\widehat{\text{MSE}}$	$\bar{x}$	0.0767	0.1541	9.3056	0.1749	9.1059	0.1287
		s^2	2.9194	0.00043	3.1977	0.00055	2.499	0.00056

For the OLI experiments, we selected three interval sizes, (τ_1 = 250, τ_2 = 500, τ_3 = 750), to represent small, medium, and large size intervals with the observation of the τ_3 interval starting after 500 minutes in order to give $n_3 = 66$. If we compare the OLI least squares results to the least squares results of Table 3, we see a slight increase in bias and a significant decrease in variance for those cases in which least squares had high variance. There does not appear to be a significant difference between the $\hat{\theta}_i^A$ and $\hat{\theta}_i^R$ estimators although $\hat{\theta}_i^R$ may have slightly better performance. The method of moments has similar results. Figures 3 and 4 illustrate the mean squared error of the estimators as compared to the base interval size of $\tau = 500$. The basic conclusion is that OLI mixing strategy can significantly improve the mean squared error properties of the least squares and method of moments estimators with the improvement coming at the cost of a small increase in bias. Empirical results also indicated that NOLI gives similar improvements. In essence, the mixing strategies allow the overall estimation process to be less sensitive to a bad choice of interval size.

LS = $\hat{\theta}_1$ LS R = $\hat{\theta}_1^R$ LS A = $\hat{\theta}_1^A$ MOM A = $\tilde{\theta}_1^A$

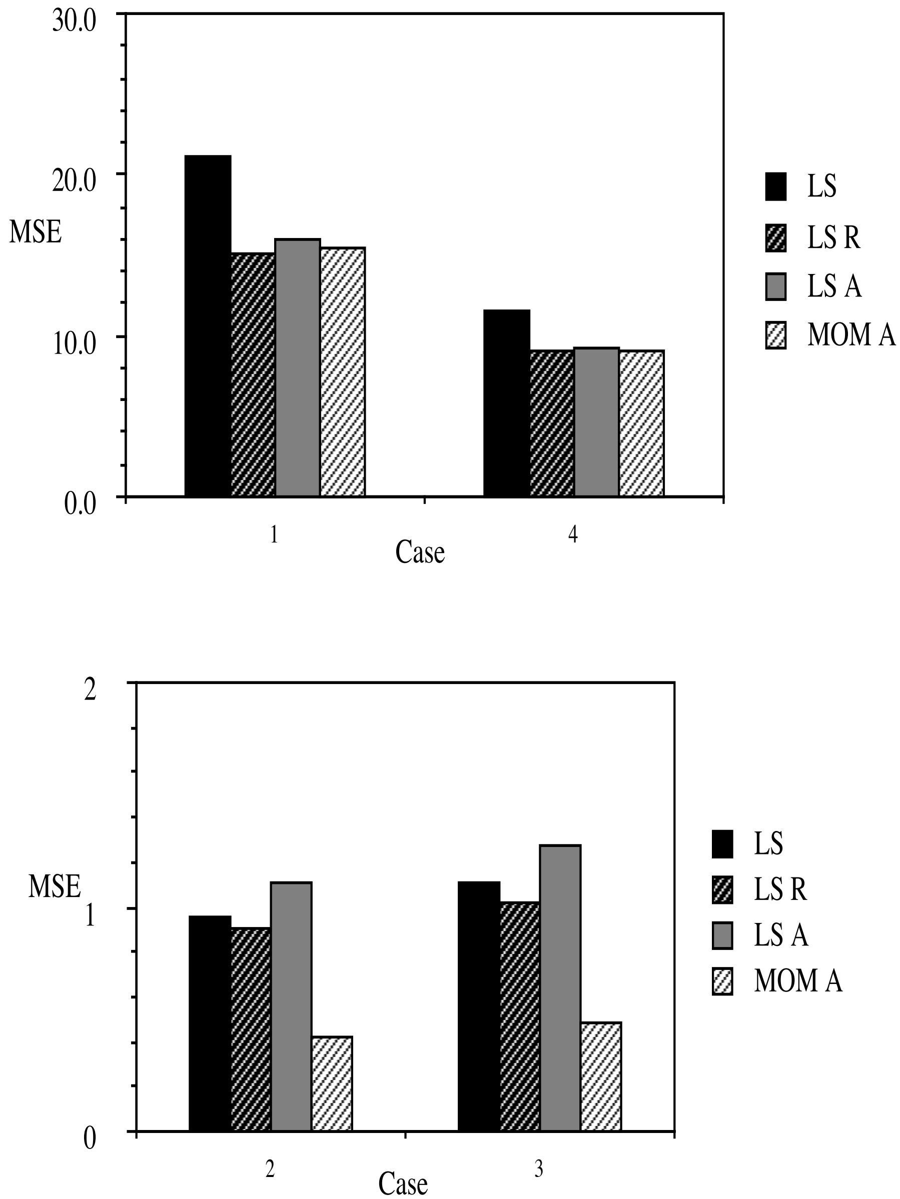
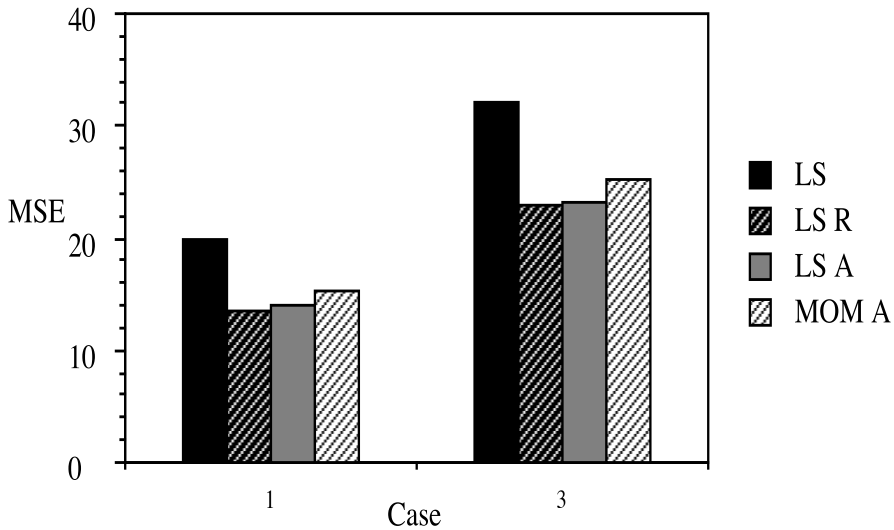


Figure 3: Mean Squared Error Comparison θ_1

$\text{LS} = \hat{\theta}_2$ $\text{LS R} = \hat{\theta}_2^R$ $\text{LS A} = \hat{\theta}_2^A$ $\text{MOM A} = \tilde{\theta}_2^A$



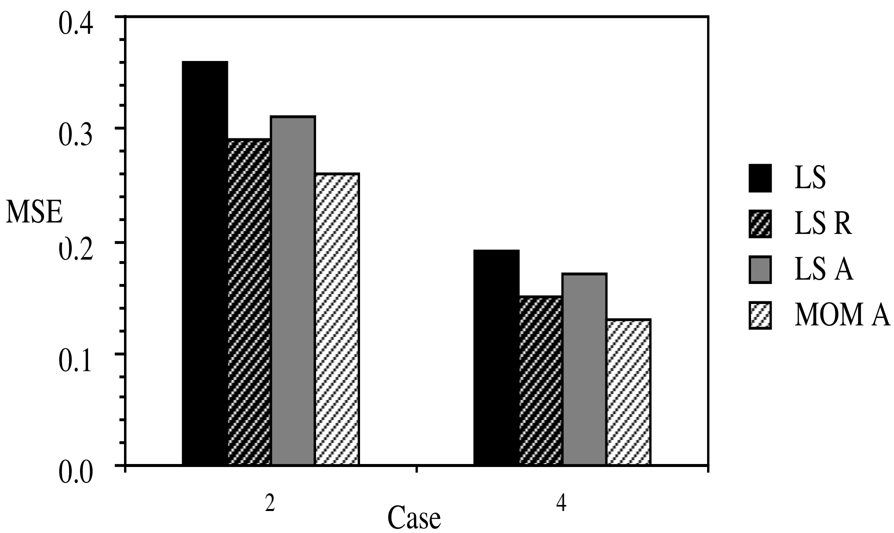


Figure 4: Mean Squared Error Comparison θ_2

7 SUMMARY AND CONCLUSIONS

This paper examined the estimation of the mean of the sojourn time distribution function for a Markov renewal process under a strict but realistic data availability condition. The contributions include

1. The definition of the state exiting counting process and analytical formulation of the solutions for the moments of the ECP's.
2. Analytical evaluation of the least squares estimator in terms of approximations for the bias, variance, and mean squared error properties.
3. Development of a MOM estimator and analytical evaluation in terms of approximate bias.
4. Empirical comparison of the least squares and MOM estimators.
5. Presentation of a data collection strategy which improves the overall performance of the estimators.

We found that the least squares estimator is biased in general. The bias increases as the fixed interval length decreases, and decreases as the fixed interval length increases. The variance of the least squares estimator may be quite large depending upon the parameters of the MRP. The variance increases as the fixed interval length increases, and decreases as the fixed interval length decreases for a fixed total observation period. The variance of the estimator corresponding to state j increases(decreases) as the steady state probability associated with the embedded Markov chain for state j decreases(increases). The variance also increases as the two state Markov renewal process becomes more like an alternating renewal process. We showed that for deterministic sojourn times the bias and the variance of the least squares estimators should be small (approximately zero). Finally, we showed that the asymptotic values of the MOM estimators should be approximately equal to the true parameters dependent upon the selection of a large enough fixed interval size.

The empirical results confirm the analytical results and also indicate that the MOM estimator appears more robust to the size of the fixed interval and may have slightly better MSE properties. The results for the OLI mixing strategy indicate that the strategy may increase bias slightly while having the potential to significantly decrease the variance in those cases where the variance of the estimator was originally large. As such, the mixing strategies offer protection against a bad choice for the size of the fixed interval.

We note that while our analysis was limited to the two state MRP the results should be applicable to larger state space situations by collapsing the state space down to two states. If we wanted to estimate θ_i, the first step would be to aggregate all states other than state i into a single state labeled as state two. We would then perform the analysis on the two states. The process would then be repeated for all the other states to achieve estimates for all of the θ_i. Finally, the least squares mean sojourn time estimation methodology could also be directly applied to the larger state space situation by expanding the linear model to include the counting processes for the larger number of states. Further analytical and empirical research needs to be done to examine the affect of the aggregation process on the estimator's performance.

APPENDIX

Proofs of the lemmas, theorems, corollaries, and other results are given in this appendix.

Proof for Theorem 1:
By renewal arguments, one can show that,

$$\tilde{M}_{ij}(s) = \begin{cases} \tilde{F}_i(s) \sum_{k=1}^{K} p_{ik} \tilde{M}_{kj}(s) & \text{if } i \neq j \\ \tilde{F}_i(s) \left[1 + \sum_{k=1}^{K} p_{ik} \tilde{M}_{kj}(s)\right] & \text{if } i = j \end{cases} \tag{28}$$

$$\tilde{W}_{ij}(s) = \begin{cases} \tilde{F}_i(s) \sum_{k=1}^{K} p_{ik} \tilde{W}_{kj}(s) & \text{if } i \neq j \\ \tilde{F}_i(s) \left[1 + \sum_{k=1}^{K} p_{ik} \left(\tilde{W}_{kj}(s) + 2\tilde{M}_{kj}(s)\right)\right] & \text{if } i = j \end{cases} \tag{29}$$

$$\tilde{h}_{ijk}(s) = \begin{cases} \tilde{F}_i(s) \sum_{\ell=1}^{K} p_{i\ell} \tilde{h}_{\ell jk}(s) & \text{if } i \neq j, i \neq k, j \neq k \\ \tilde{F}_i(s) \left[\sum_{\ell=1}^{K} p_{i\ell} \left(\tilde{h}_{\ell jk}(s) + \tilde{M}_{\ell k}(s)\right)\right] & \text{if } i = j \text{ and } j \neq k \end{cases} \tag{30}$$

See Theorem 1 in Rossetti and Clark(1994) for more details. From Çinlar(1975, pp. 11-12) and Pyke(1961b), we have

$$\tilde{\mathbf{H}}(s) = \left[\mathbf{I} - \tilde{\mathbf{Q}}(s)\right]^{-1} - \mathbf{I}$$

Thus, the results follow by placing the equations for $\tilde{M}_{ij}(s)$, $\tilde{W}_{ij}(s)$, and $\tilde{h}_{ijk}(s)$ in matrix form, noting the definition of matrix multiplication and solving the matrix equations. □

Proof for Theorem 2:

By construction, $\{N_k^e(t)\}$ has stationary increments, see Wolff(1989, pg 109–110). Let

$$X_{ik} = N_k^e((i)\tau) - N_k^e((i-1)\tau),\ i = 1, 2, \ldots, \tag{31}$$

where τ is a fixed interval size. Thus $\{X_{ik}, i = 1, 2, \ldots,\}$ is a stationary process. Assume that $\{X_{ik}, i = 1, 2, \ldots,\}$ is ergodic with $\mathrm{E}[X_{ik}] < \infty$ and $\mathrm{E}[X_{ik}^2] < \infty$ then by Theorem 5.6 of Karlin and Talyor(1975, pg. 487–488), we have that as $n \to \infty$,

$$\overline{X}_1 = (1/n)\sum_{i=1}^{n} X_{i1} \overset{wp1}{\to} \mathrm{E}[X_{11}] = \mathrm{E}[N_1(\tau)] \tag{32}$$

$$\overline{X}_2 = (1/n)\sum_{i=1}^{n} X_{i2} \overset{wp1}{\to} \mathrm{E}[X_{12}] = \mathrm{E}[N_2(\tau)] \tag{33}$$

$$\overline{X}_1^{(2)} = (1/n)\sum_{i=1}^{n} X_{i1}^2 \overset{wp1}{\to} \mathrm{E}\left[X_{11}^2\right] = \mathrm{E}\left[N_1^2(\tau)\right] \tag{34}$$

$$\overline{X}_2^{(2)} = (1/n)\sum_{i=1}^{n} X_{i2}^2 \overset{wp1}{\to} \mathrm{E}\left[X_{12}^2\right] = \mathrm{E}\left[N_2^2(\tau)\right] \tag{35}$$

From Serfling(1980, pg. 24–26), we have that

$$\overline{X_1 X_2} = (1/n)\sum_{i=1}^{n} X_{i1}X_{i2} \overset{wp1}{\to} \mathrm{E}[X_{11}X_{12}] = \mathrm{E}[N_1(\tau)N_2(\tau)] \tag{36}$$

The results given in equations (18) and (19) follow from an application of the theorems given in Serfling(1980, pg. 24–26) provided that the denominator $\neq 0$. □

Proof for Corollary 1:

Fix the interval size at τ. Let $\mathrm{E}[N_1(\tau)] = F_1\tau$, $\mathrm{E}[N_2(\tau)] = G_1\tau$. For $\hat{\theta}_1^a$, substitute into Equation 18 to yield

$$\hat{\theta}_1^a = \frac{\tau^3 (F_1 D_2 - G_1 A_2)}{(B_2 D_2 - A_2^2)\,\tau^2 - \left(\frac{1}{2}B_2 D_3 + \frac{1}{2}B_3 D_2 - A_2 A_3\right)\tau^3} \tag{37}$$

Using Result 1 yields the desired result. The result for $\hat{\theta}_2^a$ is similar. □

Proof for Result 3:

From Graybill [7, pg. 171,216], assume the following standard linear model assumptions. Let $\mathbf{Y}$ be an $n \times 1$ observable vector of random variables; let $\mathbf{X}$ be an $n \times p$ matrix $(n > p)$ of known fixed numbers; let $\boldsymbol{\theta}$ be a $p \times 1$ vector of unknown parameters; let $\boldsymbol{\epsilon}$ be an $n \times 1$ unobservable vector of random variables with $\mathrm{E}[\boldsymbol{\epsilon}] = \vec{0}$ and $\mathrm{Cov}[\boldsymbol{\epsilon}] = \mathrm{MSE}_R\, \mathbf{I}$ where $\mathrm{MSE}_R > 0$ is unknown and $\mathbf{I}$ is the $(n \times n)$ identity matrix. Thus, we are assuming that $\boldsymbol{\epsilon}$ has an unknown distribution with each ϵ_i having mean zero and constant variance and that ϵ_i and ϵ_j are uncorrelated. According to standard linear model regression theory, see Graybill(1976, pg. 217) an estimate of the MSE for the residuals about the regression is given by

$$\widehat{\mathrm{MSE}}_R = \left(\frac{1}{n-p}\right)\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}}\right) \tag{38}$$

where $\hat{\boldsymbol{\theta}}$ is the least squares estimator of $\boldsymbol{\theta}$. Application to the linear model

$$Y_i = \sum_{k=1}^{2} X_{ik}\theta_k + \epsilon_i,\ (i = 1, \ldots, n) \tag{39}$$

yields

$$\widehat{\mathrm{MSE}}_R = \frac{1}{n-2}\left(\sum_{i=1}^{n} Y_i^2 - \hat{\theta}_1 \sum_{i=1}^{n} X_{i1}Y_i - \hat{\theta}_2 \sum_{i=1}^{n} X_{i2}Y_i\right) \tag{40}$$

For our sampling method, we note that $Y_i = \tau$, substitution yields

$$\widehat{\mathrm{MSE}}_R = \frac{n\tau}{n-2}\left(\tau - \left(\hat{\theta}_1\overline{X}_1 + \hat{\theta}_2\overline{X}_2\right)\right) \tag{41}$$

Substitution of the results of Theorem 2 and Corollary 1 yields the approximation. □

Proof for Result 4:

This result follows the same argument as the proof of Lemma 3 after noting that by standard linear model regression theory estimates of

$$\widehat{\mathrm{Var}}\left[\hat{\theta}_1\right] = \frac{\widehat{MSE}_R}{n}\frac{\overline{X_2^2}}{(\overline{X_1^2})(\overline{X_2^2}) - (\overline{X_1X_2})^2} \tag{42}$$

$$\widehat{\mathrm{Var}}\left[\hat{\theta}_2\right] = \frac{\widehat{MSE}_R}{n}\frac{\overline{X_1^2}}{(\overline{X_1^2})(\overline{X_2^2}) - (\overline{X_1X_2})^2} \tag{43}$$

and that MSE = $\text{Bias}^2 + \text{Var}$. Substitution of the results of Theorem 2 and Corollary 1 yields the approximation. □

Proof for Estimator 2:

We have that $\mathrm{E}[N(\tau)] = \tau/\theta$, $\mathrm{Var}[N_1(\tau)] \doteq \frac{\pi_1^3}{\theta^3}\left(\frac{\delta^2}{\pi_1} + \frac{\pi_2\theta_2^2\lambda}{\pi_1^2}\right)\tau$, $\mathrm{Var}[N_2(\tau)] \doteq \frac{\pi_2^3}{\theta^3}\left(\frac{\delta^2}{\pi_2} + \frac{\pi_1\theta_1^2\lambda}{\pi_2^2}\right)\tau$, and $\mathrm{Cov}[N_1(\tau), N_2(\tau)] \doteq \frac{\pi_1\pi_2}{\theta^3}(\delta^2 - \theta_1\theta_2\lambda)\,\tau$ where $\lambda = (2/\beta) - 1$. Noting that π_1, π_2, and θ are directly estimable from the data indicates that we have four equations with four unknowns θ_1, θ_2, λ, and δ^2. Solve $\mathrm{Cov}[N_1(\tau), N_2(\tau)]$ for δ^2 and substitute into the equations for $\mathrm{Var}[N_1(\tau)]$ and $\mathrm{Var}[N_2(\tau)]$ and after some algebraic manipulation we have

$$\frac{\theta_1}{\theta_2} = \frac{\pi_1\mathrm{Var}[N_2(\tau)] - \pi_2\mathrm{Cov}[N_1(\tau), N_2(\tau)]}{\pi_2\mathrm{Var}[N_1(\tau)] - \pi_1\mathrm{Cov}[N_1(\tau), N_2(\tau)]} \tag{44}$$

Substituting in the sample moments into Equation 44 and using $\tau = \overline{X}_1\theta_1 + \overline{X}_2\theta_2$ yields the desired result. □

Proof for Corollary 2:

Let $\hat{L}_1$ and $\hat{D}_{mom}$ be the numerator and denominator in Equation (24). By the same convergence arguments given in Theorem 2, we have that

$$\begin{aligned}
\hat{D}_{mom} &\stackrel{wp1}{\to} \mathrm{E}[N_1(\tau)]^2\,\mathrm{E}\left[N_2^2(\tau)\right] - 2\mathrm{E}[N_1(\tau)]\,\mathrm{E}[N_2(\tau)]\,\mathrm{E}[N_1(\tau)N_2(\tau)] + \\
&\qquad \mathrm{E}[N_2(\tau)]^2\,\mathrm{E}\left[N_1^2(\tau)\right] = \hat{D}^a_{mom} \\
\hat{L}_1 &\stackrel{wp1}{\to} \tau\left(\mathrm{E}[N_1(\tau)]\,\mathrm{E}\left[N_2^2(\tau)\right] - \mathrm{E}[N_1(\tau)N_2(\tau)]\,\mathrm{E}[N_2(\tau)]\right) = \hat{L}^a_1
\end{aligned}$$

Thus, $\tilde{\theta}_1 \stackrel{wp1}{\to} \hat{L}^a_1/\hat{D}^a_{mom}$, provided $\hat{D}^a_{mom} \neq 0$. Substitution of the moment equations given in Definition 2 yields

$$\tilde{\theta}^a_1 \doteq \frac{\tau^3(F_1D_2 - G_1A_2)}{(F_1^2D_2 + G_1^2B_2 - 2F_1G_1A_2)\,\tau^3} \tag{45}$$

Substitution of the equations given in Result 1 yields the desired results. The derivation for $\tilde{\theta}^a_2$ is similar. □

ACKNOWLEDGEMENTS

The authors would like to thank AT&T Network Systems in Columbus, Ohio for their support of this research especially, Martin Gussman, our project liaison.

BIBLIOGRAPHY

Çinlar, E. (1975) *Introduction to Stochastic Processes.* Prentice-Hall, Englewood Cliffs, N. J.

Çinlar, E. (1968) "Some Joint Distributions for Markov Renewal Processes." *J. Austrl. Statist.* 10 pp 8–20.

Graybill, F. A. (1976) *Theory and Application of the Linear Model*, Wadsworth Publishing Company, Inc.

Heyman, D. P. and Sobel, M. J. (1982) *Stochastic Models in Operations Research, Volume 1*, McGraw– Hill , Inc.

Hunter, J. J. (1969) "On the Moments of Markov Renewal Processes." *Adv. Appl. Prob.* 1 pp 188–210.

Karlin, S., and Taylor, H. M. (1975) *A First Course in Stochastic Processes*, 2nd ed. Academic Press, Inc.

Karr, A. F. (1986) *Point Processes and Their Statistical Inference*, Marcel Dekker, Inc.

Kshirsagar, A. M. and Gupta, Y. P. (1967) "Asymptotic Values of the First Two Moments in Markov Renewal Processes." *Biometrika* 54, 3 and 4, pp 597– 603.

Moore, E. H. and Pyke, R. (1968) "Estimation of the Transition Distributions of a Markov Renewal Process." *Ann. Inst. Stat. Math.* 20 pp. 411– 424.

Pyke, Ronald, (1961b) "Markov Renewal Processes with Finitely Many States." *Ann. Math. Stat.* 33 pp. 1243–1259.

Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*, 2nd ed. John Wiley and Sons, Inc., New York.

Rossetti, M. D. (1992) "Queueing Service Parameter Estimation From Event Oriented Production Control Data," PhD Dissertation, The Ohio State University, Columbus, Ohio.

Rossetti, M. D. and Clark, G. M. (1994) "Moment Solutions for State Exiting Counting Processes of a Markov Renewal Process." *Working Paper SYS-MDR-2-1994*, Department of Systems Engineering, University of Virginia, Thornton Hall, Charlottesville, VA 22903.

Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, Inc.

Taga, Y. (1963) "On the Limiting Distributions in Markov Renewal Processes with Finitely Many States," *Ann. Inst. Statist. Math.*, 15, pg. 1–10.

Wolff, R. W. (1989) *Introduction to Stochastic Modeling and the Theory of Queues*, Prentice– Hall.