UNIVERSITY OF MINNESOTA

CENTER FOR TRANSPORTATION STUDIES

INTELLIGENT TRANSPORTATION SYSTEMS INSTITUTE

Real-Time Traffic Prediction for Advanced Traffic Management Systems: Phase III

Final Report

Prepared by:
Gary A. Davis, Yargos J. Stephanedes, Chang-Jen Lan

Department of Civil Engineering
University of Minnesota

CTS 06-06

Technical Report Documentation Page

1. Report No.	2.	3. Recipients Accession No.	
CTS 06-06			
4. Title and Subtitle Real-Time Traffic Prediction for Advanced Traffic Management Systems: Phase III		5. Report Date May 2006 6.	
7. Author(s) Gary A. Davis, Yorgos J. Stephanedes, Chang-Jen Lan		8. Performing Organization Report No.	
9. Performing Organization Name and Address Department of Civil Engineering University of Minnesota Room 122 CivE 500 Pillsbury Dr S E Minneapolis, MN 55455		10. Project/Task/Work Unit No. CTS Project Number 1995048 11. Contract (C) or Grant (G) No.	
12. Sponsoring Organization Name and Address Intelligent Transportation Systems Institute Center for Transportation Studies University of Minnesota 511 Washington Avenue SE, Suite 200 Minneapolis, MN 55455		13. Type of Report and Period Covered Final Report June 1996 – December 1997 14. Sponsoring Agency Code	
15. Supplementary Notes http://www.cts.umn.edu/pdf/CTS-06-06.pdf			

16. Abstract (Limit: 200 words)

Building on the research performed in Phases I and II of this study, the authors develop in Phase III an algorithm for estimating vehicle turning proportions at signalized intersections using partial detector information. This report demonstrates the efficiency of a maximum likelihood estimator approach to estimate turning proportions over a prediction error minimization principle. This phase enhances the traffic flow models to provide more detail on short-term instability and left-turning behavior at intersections. Finally, the authors develop two types of estimation algorithms using Nonlinear Least Squares and Quasi Maximum Likelihood to estimate the turning proportion parameters. These algorithms are useful tools to generate real-time applications in traffic control.

17. Document Analysis/Descriptors		18.Availability Statement	
signalized intersection, traffic flow, traffic model, Markov processes, detection and identification systems, turning traffic		No restrictions. Document available from: National Technical Information Services, Springfield, Virginia 22161	
19. Security Class (this report) Classified	20. Security Class (this page) Classified	21. No. of Pages 103	22. Price

Real-Time Traffic Prediction for Advanced Traffic Management Systems: Phase III

Final Report

Prepared by:
Gary A. Davis
Yargos J. Stephanedes
Chang-Jen Lan

Department of Civil Engineering
University of Minnesota

May 2006

Intelligent Transportation Systems Institute
Center for Transportation Studies
University of Minnesota

CTS 06-06

ACKNOWLEDGEMENTS

The authors wish to acknowledge those who made this research possible. Funding for the study was provided by the University of Minnesota's Center for Transportation Studies (CTS).

EXECUTIVE SUMMARY

Following Phase II of this project, Phase III continues with development of identification algorithms for estimating vehicle turning proportions on signalized intersections from partial detector information, both in off-line and recursive forms. The identification algorithms are developed based on the Prediction Error Minimization (PEM) principle. To make adequately accurate prediction of the traffic states and flows to support this approach, a markovian compartment traffic model was developed in Phase I for freeway traffic case and calibrated in Phase II for urban traffic case. An effort was also devoted to test the feasibility of the traffic flow models in congestion situations, with an expectation that the flow model is able to capture the highly nonlinear traffic behavior during then, and linked the traffic inputs together to produce accurate outputs. The results seemed promising from both simulation and empirical studies. This phase continues with some further enhancements to the traffic flow models to provide higher level of details on short term instability and left-turning behavior at intersections.

In estimating turning proportions from partial detector information, one major concern is the sufficiency of the detector information to support the estimation, and how the statistical properties such as variability of the estimates are affected by quality of the detector information or types of estimators. One can expect that less information will result in poorer quality of estimates. Phase II addressed this issue by comparing the variabilities of the estimates from full and partial sets of detector configurations using least-squares approaches. This phase extends to demonstrate that, by switching to more efficient maximum likelihood estimator, the quality of estimates is actually improved with a moderate increase on the computation time. A procedure to measure the variabilities

of the estimates for different detector configurations on networks is also proposed to provide traffic practitioners guidelines as to how feasible detector configuration is laid out prior to implementation.

The Markovian compartment model be attainably approximated by the sum of a nonlinear deterministic process and a time-varying linear stochastic process, which in turn leads to the use of the linear system techniques to estimate the unknown traffic states, compute approximate likelihood functions, and develop estimation algorithms to support estimation of model parameters. Two types of estimation algorithms, including Nonlinear Least-Squares (NLS) and Quasi Maximum Likelihood (QML), are developed in this project to off-line and recursively estimate the turning proportion parameters. Both algorithms belong to the class of prediction error minimization methods. The results on off-line estimators demonstrate a benchmark for comparisons of statistical properties, while the recursive estimators are then developed for real-time implementation, particularly for responding to changes in turning proportions over time. Some modifications are also made to adapt the algorithms to track the abrupt changes in actual turning proportions, given the change time is known. The real-data testing results on a two-intersection network demonstrate that these algorithms will be useful tools to generate acceptably accurate turning movement proportions for real-time applications in traffic control.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Real-time traffic controls require that both control actions and prediction/estimation of traffic states and system parameters be operated in real time, in order to accommodate small changes in traffic condition. The effectiveness of control actions and the resulted average traffic performance will mainly depend on the reliability of the detector information and the predictor/estimator (Davis, 1993). The basic traffic flow model used to link the detector information and estimation/prediction has been developed in Phase I for freeway traffic and calibrated in Phase II for urban traffic, while it needs to be enhanced to deal with more complicated situations such as short-term instability and left-turning at signalized intersections. This phase continues with these enhancements to gain some insights into traffic nature.

Next, in estimating turning proportions from partial detector information, one major concern is the sufficiency of the detector information to support the estimation, and how the statistical properties such as variability of the estimates are affected by quality of the detector information or types of estimators. One can expect that less information will result in poorer quality of estimates. Phase II addressed this issue by comparing the variabilities of the estimates from full and partial sets of detector configurations using least-squares approaches. This phase extends to demonstrate that, by switching to more efficient maximum likelihood estimator, the quality of estimates is actually improved with a moderate increase on the computation time. A procedure to measure the variabilities of the estimates for different detector configurations on networks is also proposed to provide traffic practitioners guidelines as to how feasible detector configuration is laid out prior to implementation.

Probably the most critical issue of this project is to develop on-line recursive estimation algorithms to possibly track the time-varying turning movement proportions at intersections using system identification approaches. Cremer and Keller (1981, 1987), Nihan and Davis (1987, 1989), and Bell (1991) have employed least-squares and maximum likelihood approaches to sequentially estimate the turning proportions under full set of detector configuration. It is shown not only that OLS estimations of turning movement proportions are consistent (Nihan and Davis, 1989), but also that a consistent estimator of the general OD matrices can be obtained based on the maximum likelihood principle. Nevertheless, when only partial set of detector counts are available, the above general results on system identification need to be further investigated. Also, due to the nature of the underlying problem, one needs to concern with non-linear extension of the traffic flow model, estimation procedure, and the effect of the information quality on variability of the estimators. Here, in addition to nonlinear least-squares method, a stochastic approach based on the quasi-maximum likelihood criterion is introduced to attack the underlying problem, with a well-defined variance component, i.e., state-dependent covariance matrices in the estimation algorithm, to capture the system uncertainties. It is strongly believed that the variance information contributes substantially to the resulted estimates. These two types of estimation algorithms are developed in this project to off-line and recursively estimate the turning proportion parameters. The results on off-line estimators demonstrate a benchmark for comparisons of statistical properties, while the recursive estimators are then developed for real-time implementation, particularly for responding to changes in turning proportions over time.

The rest of the report is organized in the following fashion. The enhancement to the basic traffic flow model and its calibration will be addressed in Chapter 2. Chapter 3 is devoted to

describing a procedure for measuring variabilities of the turning movement proportion estimator. Estimation algorithms based on Nonlinear Least-Squares (NLS) and Quasi Maximum Likelihood (QML) principles are presented in Chapter 4 in both off-line and recursive forms. Their statistical properties are also investigated. Simulated data sets and a real data set collected in the field will be used to evaluate the effectiveness of each proposed algorithm. Chapter 5 contains the testing results with the real data. Conclusions and directions of future research works are made in Chapter 6.

CHAPTER 2

EXTENSIONS OF TRAFFIC FLOW MODEL

Based on a Markovian compartment idea, a traffic flow model suitable for describing the traffic behavior on signalized intersections has been developed in Phase II of the project. To accommodate the possible transitive flow behavior observed in non-uniform traffic stream and left-turning situations, some extensions to the basic traffic flow model were developed in the phase of the project. Here the basic models are restated to provide the necessary connection to the topic of this chapter.

2.1 Review of the Markovian Traffic Flow Model

Let h, k, and l index the compartments, τ index the basic time intervals (typically 1-5 sec), and let I and j index the input and output nodes. Also define the following notation:

$x_k(\tau)$ = number of vehicles in compartment k at the beginning of time τ,

$y_j(t)$ = number of vehicles crossing output detector j during time interval (t-1,t],

$\mathbf{x}(\tau)$, $\mathbf{y}(t)$ = m-, n-dimensional vectors containing elements $x_k(\tau)$, $y_j(t)$,

$p_k(\mathbf{x}(\tau))$ = the exiting probability from compartment k, as a function of the current states at time τ,

$b_{kl}(t)$ = turning movement proportions from compartment k to l,

$\mathbf{b}(t)$ = p-dimensional vector containing $b_{kl}(t)$,

$q_i(\tau)$ = volume entering at input node I during time interval (τ-1,τ],

$\mathbf{q}(\tau)$ = vector containing $q_i(\tau)$,

δ_{ki} = 1, if input node I is at the upstream boundary of compartment k;

0, otherwise,

$\bar{u}_k$, $\bar{u}_{ek}$ = space mean speed, space mean speed at capacity flow $\bar{u}(x_{ck})$),

x_{ck}, x_{jk} = the numbers of vehicles in compartment k when traffic is at the critical density and jam density, and

L_k, n_l = the length of compartment k and the number of lanes in compartment l.

The distribution of vehicles over compartments is assumed to evolve in time according to the conservation equation

$$x_k(\tau+1) = x_k(\tau) - \sum_l \tilde{y}_{kl}(\tau) + \sum_h \tilde{y}_{hk}(\tau) + \sum_i \delta_{ki} q_i(\tau) \tag{2.1}$$

where $\tilde{y}_{kl}(\tau)$ are random outcomes giving the exit volumes from compartment k to its downstream compartments l, generated as independent multinomial outcomes with parameters $(x_k(\tau), p_k(\mathbf{x}(\tau))b_{kl})$. The general functional form for p_k can be expressed as

$$p_k(\mathbf{x}) = (\textit{Passage probability})\cdot(\textit{Non-blocking probability}) \tag{2.2a}$$

where

$$\textit{Passage probability} = \begin{cases} \dfrac{\Delta}{L_k}\bar{u}_k(x_k), & x_k \le x_{ck} \\ \dfrac{\Delta}{x_k}\dfrac{x_{ck}}{L_k}\bar{u}_{ek}, & x_k \ge x_{ck} \end{cases} \tag{2.2b}$$

$$\textit{Non-blocking probability} = \begin{cases} 1-\left(\dfrac{x_l}{x_{jl}}\right)^{n_l}, & n_l \le 2 \\ 1-\dfrac{2}{n_l}\left(\dfrac{x_l}{x_{jl}}\right)^{2}-\dfrac{n_l-2}{n_l}\left(\dfrac{x_l}{x_{jl}}\right)^{3}, & n_l \ge 3 \end{cases} \tag{2.2c}$$

Finally, given a sequence of input counts $\mathbf{q}(t)$, t=1,...,T, and vector of turning movement proportions $\mathbf{b}(t)$, the aggregated output counts can be generated as

$y_j(t) = \sum_{\tau=1}^{K} \sum_l \tilde{y}_{kl}(\tau)$, if detector j counts exits from compartment k

$= \sum_{\tau=1}^{K} \sum_h \tilde{y}_{hk}(\tau)$, if detector j counts entries into compartment k

where K is the number of basic intervals (τ) in an aggregation interval (indexed by t).

2.2 Accommodation of Instability

Edie and Foote (1961) observed that when encountering denser traffic ahead drivers tend to delay deceleration, leading to shorter spacings and higher flows. Daganzo (1994) tried to capture this effect using dual-mode flow-density curves, which asserted that traffic will be regulated by the higher-capacity curve if such a phenomenon occurs and the lower capacity curve otherwise. Following Daganzo's idea, the proposed flow model can be modified to incorporate this effect. Although the modification can be made in terms of space mean speed and its parameters as in equation (2.2b), for a clearer presentation let us rewrite equation (2.2b) as

$$Passage\ probability = \begin{cases} \dfrac{q_k(x_k)\Delta}{x_k}, & x_k \le x_{ck} \\ \dfrac{q_{ck}\Delta}{x_k}, & x_k \ge x_{ck} \end{cases} \tag{2.3a}$$

where $q_k(x_k)$ is the flow from compartment k, and this relationship can be parameterized by the capacity flow q_{ck} and the critical density. Second, let us define the capacity flow q_{ck} as

$$\begin{aligned} q_{ck} &= q_{ck}^{H}, \quad if \ \frac{x_k}{n_k L_k} \le \frac{x_l}{n_l L_l} \\ q_{ck} &= q_{ck}^{L}, \quad otherwise \end{aligned} \tag{2.3b}$$

where q^H_{ck} and q^L_{ck} are the capacity flows for the higher and lower flow-density curves, respectively. This formulation implies that the traffic operates on the higher flow-density curve when the density in the downstream compartment is greater than in the upstream compartment k.

2.3 Treatments for Signalized Intersections

When there exists a signalized intersection between two adjoining compartments, traffic flow can be interrupted by the traffic lights, and also by conflicting (opposing) flows during a permitted phase. For the interruption by a red light, the non-blocking probability term can simply be set to zero so that no flow exiting from the current compartment is allowed. When the left-turn traffic must yield to an opposing flow in a permitted phase, the modeling aspects become more complicated. The detailed flow modeling for every possible geometric and phasing condition is beyond the scope of this

study, but here a focus will be given to a common case, the exclusive left-turn lane with a permitted left-turn phase.

Since the left-turn lane is exclusive, it can be modeled as a separate compartment. The chance that a left turning vehicle can exit from this compartment and pass through the intersection will depend on the lane-specific distribution of the opposing traffic flows. If it is assumed that the headways h of the opposing traffic in lane number I follows a shifted exponential distributions with the cumulative distribution function

$$\begin{aligned} F_i(h) &= 1 - e^{-\lambda_i(h-T_{\min})}, \quad h \geq T_{\min} \\ &= 0, \qquad\qquad\quad h < T_{\min} \end{aligned} \tag{2.4}$$

then a multi-lane version of a left-turn capacity formulas generalized by Tanner (1967) and Troutbeck (1986) can be written as

$$q_{Lcap} = \frac{\lambda' \left[\prod_i (1 - T_{\min} q_i)\right] e^{-\lambda'(T_c - T_{\min})}}{1 - e^{-\lambda' T_f}} \tag{2.5}$$

where q_i = flow rate in lane I,

$\lambda' = \sum_i \frac{q_i}{1 - T_{\min} q_i}$,

T_c = the critical gap (sec),

T_{min} = The minimum headway in an opposing traffic lane (sec), and

T_f = The follow-up time (sec), defined as the time interval between the departure of two consecutive opposed vehicles that use the same gap under a continuous queue condition.

For a deterministic model, the number of left-turning vehicles which successfully escape from the left-turn compartment during the basic time interval Δ will be at most $q_{Lcap}\Delta$. In the stochastic model, the number of vehicles exiting from the left-turn pocket during each time interval are modeled as binomial outcomes, where the number of trials equals the compartment population x and the success probability is $\min[p(x), q_{Lcap}\Delta/x]$.

2.4 EVALUATION OF THE EXTENDED MODELS

Two empirical studies were conducted to (1) examine the feasibility of extended dual capacity-flow model, and (2) calibrate the traffic flow model in the presence of permitted left turns. Using the data set and testing procedure as described in Sec. 2.2 of the Phase II report, a revised model designed to capture possibly unstable traffic behavior was calibrated in the following. The Bell-shaped function was used to characterize the speed-density relationship. As expected, when the downstream densities are higher, the estimated capacity q_c^H (= 1742.8 vph) is also higher than the capacity q_c^L (= 1476.6 vph) estimated when the downstream densities are lower. The log-likelihood indices before and after incorporating an additional parameter are (-679.68) and (-678.29), resulting in the -2*ln(likelihood ratio) equal to 2.78, and this difference is significant at the $p<0.1$ level ($\chi^2_{1,0.1}=2.71$). This provides some weak evidence supporting Daganzo's 2-mode model.

In the second study, the video camera was installed at a different intersection as depicted in Figure 2.1, covering two opposing approaches at the same time. The left-turn traffic on the northbound approach shared a lane with the through traffic and was only allowed to pass during a permitted phase. The signal was in 2-phase operation. The compartment for the shared left-turn lane

was set to 220 ft, and three lines are drawn on the monitor screen to count the left-turning traffic $y(\tau)$ and the opposing traffic $q_1(\tau)$ and $q_2(\tau)$ during each 5 second interval. At the beginning of each 5-second time interval the left-turning demand $x(\tau)$ appearing in the shared-lane compartment was also counted. During the time period when the data were recorded, most of the through traffic passed through the intersection via the outside lane, and stayed in the outside lane during the red phase. Even when temporarily 'trapped' in the shared lane, almost all through vehicles were able to sneak out without being delayed by the left-turning traffic. Therefore, there were actually very few instances that through traffic was present in the shared lane, and the inner lane functioned primarily as if it were an exclusive left-turn lane. For testing purpose, all of the through vehicles from this lane are eliminated, making the data set suitable for analyzing the effects of opposing traffic on the predicted left-turning flow from an exclusive lane, during permitted phases. The MLE procedure was used to estimate model parameters where in this case the left-turn exit flow was modeled as the binomial outcomes from $x(\tau)$, with probability equal to $\min[p(x(\tau)), q_{Lcap}\Delta/x(\tau)]$. The model parameters include the traffic flow parameters, i.e., free-flow speed and critical density, and the gap acceptance parameters consisting of the critical gap T_c and the follow-up time T_f. These two gap acceptance parameters have been shown to affect the opposed approach capacity (Velan and Van Aerde, 1996). The minimum headway in the opposing traffic lane T_{min} was set to 2 seconds. The results are summarized as follows: u_f = 22.89 mph, k_c = 109.09 vpm, T_c = 7.05 sec, T_f = 2.42 sec, and the corresponding estimate of the saturation flow rate ≈ 1515 vph. Interestingly, the estimated critical gap here is comparable to that estimated by Miller (1972) using a microscopic approach (≈ 6.3 sec), and is also close to the 'first-gap' estimate (≈ 7.2 sec) of the gap sequence by Mahmassani and Sheffi (1981). The follow-up time and saturation flow rate also appear reasonable when compared to values

obtained using the microscopic simulation technique reported in Velan and Van Aerde (1996). Finally, Figure 2.2 shows the time-series plot of predicted and observed exit counts. The adjusted R^2 statistic was 0.58.

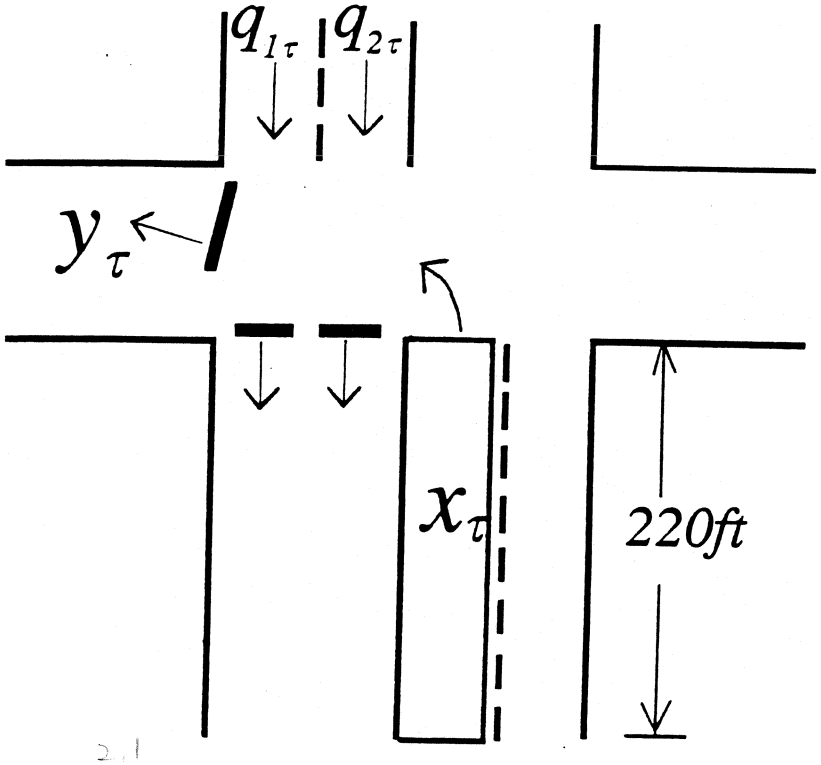
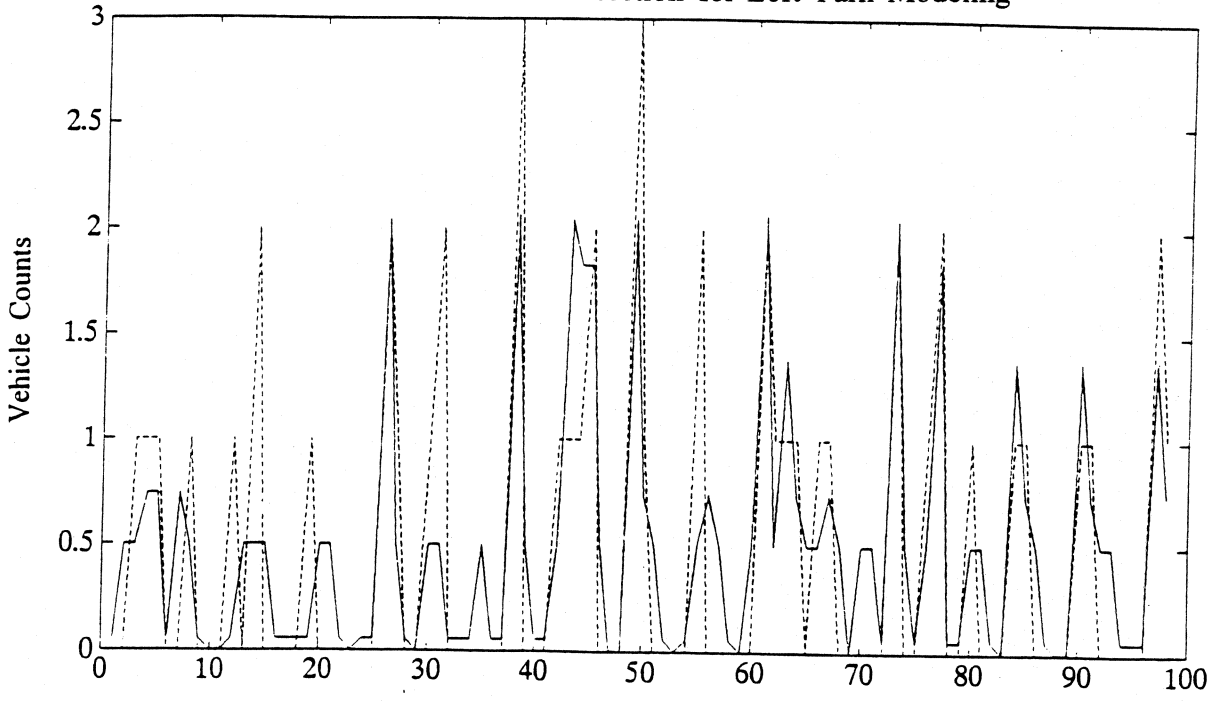


FIGURE 3.7 Data Collection Intersection for Left-Turn Modeling



(Note: solid line: observations, dashed line: 3-parameter)

FIGURE 3.8 Left-Turn Fitted Counts vs Observed Counts

CHAPTER 3

MEASURES OF TURNING MOVEMENT PROPORTION VARIABILITY

Prior to developing algorithms to actually estimate turning movement proportions, it is desirable to consider the issue of parameter identifiability and statistical efficiency of the parameter estimates. As described in Phase II report, if time-series of input and output counts are available from detectors on each intersection leg, it has been established that useful estimates of turning movement proportions can be generated using OLS-based methods. (Nihan and Davis, 1989) What is beneficial to this approach is the information provided by the dense detector configuration. If only a partial set of detectors is available, then the information will be less, producing estimates with less accuracy. Therefore, one of the critical issues is the justification of the sufficiency of detector information to support parameter estimation. This chapter first reviews the definition of identifiability and the numerical test for evaluating the identifiability under a particular detector configuration. A scheme for measuring the variabilities of the parameter estimates under various detector configurations is then introduced.

3.1 Review of Identifiability Problems

To date, many researchers have employed output prediction error minimization methods to estimate traffic model parameters (Cremer and Papageorgiou, 1981; Nihan and Davis, 1987, 1989). The idea is as follows. Let $\mathbf{y}(t)$ be a vector-valued random variable in R^n representing the outcome of the detector outputs. $\mathbf{b}$ is a set of turning movement proportion parameters and an element of a

set **B**, where **B** is a subset of R^p defined by a set of constraint equations. Given a sequence of measured input counts **q**(t), t=1,...,T, trial sets of turning movement proportions, and the Markov traffic flow model, one can compute predicted output flows, $\hat{\boldsymbol{y}}(t, \boldsymbol{b})$, and then the "best" turning movement proportion estimates as those that minimize the predicted errors between $\hat{\boldsymbol{y}}(t, \boldsymbol{b})$ and actual output counts **y**(t), t=1,...,T, based on a quadratic criterion, i.e.,

$$\hat{\boldsymbol{b}} = \underset{\boldsymbol{b}\in\boldsymbol{B}}{\boldsymbol{argmin}} \sum_{t=1}^{T} \left(\boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t, \boldsymbol{b})\right)' \left(\boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t, \boldsymbol{b})\right) \tag{3.1a}$$

The set of constraint equations in the case here can be expressed as

$$\mathbf{Rb} = \mathbf{1},\ \mathbf{b} \geq \mathbf{0} \tag{3.1b}$$

where **1** is a column vector containing all 1's, and **R** is a matrix of ones and zeros guaranteeing that the turning movements proportions out of each approach add to 1.0. The dependence of the predicted outputs on the unobserved state vector makes $\hat{\boldsymbol{y}}(\boldsymbol{b}, t)$ a nonlinear function of the turning movement proportions. This leads to a constrained nonlinear least squares (NLS) problem, which can be solved as long as the problem is well-posed and local minimizing solutions exist. It may turn out, however, that a given configuration of traffic detector placements does not provide enough information to estimate all the desired turning movement proportions. Thus it is important to determine if a proposed detector configuration is adequate for estimating these parameters. A necessary condition for the above NLS problem to be well-posed is that different values for **b** should lead to different values for $\hat{\boldsymbol{y}}(t, \boldsymbol{b})$. This leads to a very relevant property called "structural identifiability" (Bellman and Åström, 1970) which can be defined as follows.

Definition 1 (structurally locally identifiable)

Given inputs **q**(t), the parameter $\boldsymbol{b}$ is structurally locally identifiable if and only if, for almost any $\boldsymbol{b}\in\mathbf{B}$, there exists a neighborhood $N(\boldsymbol{b})$ such that $\hat{\boldsymbol{b}}\in N(\boldsymbol{b})\subset\mathbf{B}$ and $\hat{\boldsymbol{y}}(\hat{\boldsymbol{b}},t)=\hat{\boldsymbol{y}}(\boldsymbol{b},t)$, $\forall t\in\boldsymbol{Z}$ implies that $\hat{\boldsymbol{b}}=\boldsymbol{b}$.

Definition 2 (structurally globally identifiable)

Given inputs **q**(t), the parameter $\boldsymbol{b}$ is structurally globally identifiable if and only if, for almost any $\boldsymbol{b},\hat{\boldsymbol{b}}\in\mathbf{B}$, and $\hat{\boldsymbol{y}}(\hat{\boldsymbol{b}},t)=\hat{\boldsymbol{y}}(\boldsymbol{b},t)$, $\forall t\in\boldsymbol{Z}$ implies that $\hat{\boldsymbol{b}}=\boldsymbol{b}$.

Intuitively, global (local) identifiability states that there exists a one-to-one (locally one-to-one) mapping from the parameter space to the predictor space. Global identifiability implies local identifiability and is hence more appealing, but it is also more difficult to verify (Rothenberg, 1971). For practical use, however, the local identifiability is a good starting point. Grewal and Payne (1976) described an approach to examine the identifiability of the traffic flow parameters in a high-order continuum model. The approach taken here is somewhat similar but is cast in a NLS form.

3.2 An Analytical Procedure for Identifiability Test

Phase II report describes an experimental design to test the parameter identifiability. A procedure is appended here to illustrate its analytical detail. Before proceeding, for simplicity, let us partition the equality constraints in a form of

$$\begin{bmatrix}\boldsymbol{R}_1 & | & \boldsymbol{R}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{b}_1\\ --\\ \boldsymbol{b}_2\end{bmatrix}=\mathbf{1}$$

If $\mathbf{b}_1$ is set of a specific turning parameters with one element from each approach, then $\mathbf{R}_1 = \mathbf{I}$. One can write $\mathbf{R}_1\mathbf{b}_1+\mathbf{R}_2\mathbf{b}_2 = \mathbf{1}$, and $\mathbf{b}_1 = \mathbf{R}_1^{-1}(\mathbf{1}-\mathbf{R}_2\mathbf{b}_2) = \mathbf{1}-\mathbf{R}_2\mathbf{b}_2$. By the chain rule,

$$\begin{aligned}\frac{d\hat{y}(\boldsymbol{b}_2, t)}{d\boldsymbol{b}_2} &= \frac{\partial\hat{y}(\boldsymbol{b}, t)}{\partial\boldsymbol{b}_2} + \frac{\partial\hat{y}(\boldsymbol{b}, t)}{\partial\boldsymbol{b}_1}\frac{\partial\boldsymbol{b}_1}{\partial\boldsymbol{b}_2} \\ &= \frac{\partial\hat{y}(\boldsymbol{b}, t)}{\partial\boldsymbol{b}_2} - \frac{\partial\hat{y}(\boldsymbol{b}, t)}{\partial\boldsymbol{b}_1}\boldsymbol{R}_2\end{aligned} \tag{3.2}$$

Letting

$$\phi(\boldsymbol{b}_2, t) = \frac{d\hat{y}(\boldsymbol{b}_2, t)}{d\boldsymbol{b}_2}$$

one can compute a reduced gradient which gives the derivative of the predicted output counts with respect to the reduced set of turning movement proportion parameters. In the following, for convenience **b** will be referred to the reduced set of parameters unless particularly specified. Then based on system identifiability theory (Seber and Wild, 1989), given a particular detector configuration and actual output counts **y**(t), it is straightforward to verify that the parameters are locally identifiable if the Jacobian matrix

$$\boldsymbol{J}(\boldsymbol{b}) = \begin{bmatrix}\phi(\boldsymbol{b}, 1) \\ \vdots \\ \phi(\boldsymbol{b}, T)\end{bmatrix} \tag{3.3}$$

is nonsingular, or equivalently, the rank of $\boldsymbol{J}(\boldsymbol{b})'\boldsymbol{J}(\boldsymbol{b}) = \Sigma_t \phi(\boldsymbol{b}, t)'\phi(\boldsymbol{b}, t) = p$, in a neighborhood.

In other words, the full rank of the Jacobian matrix implies the local identifiability of parameters. As one can see, the Jacobian matrix $\boldsymbol{J}$ is composed of sub-matrices stacked on each other in order of time, each sub-matrix being the derivative of the predicted output at time t with respect to parameters. Storing this matrix in memory could be awkward if the number of time intervals is large. For convenience, it is often preferable to perform the rank test on $\boldsymbol{J}(\boldsymbol{b})'\boldsymbol{J}(\boldsymbol{b})$.

The Jacobian test can be implemented as follows:

(1) Construct a representation of the system of interest as a compartment model,

(2) Generate a set of simulated input counts $\mathbf{q}(t)$, t=1,...,T, as Poisson outcomes, using reasonable choices for their mean values,

(3) Using a trial set of turning proportions **b**, generate a sequence of predicted outputs flows, compute $\mathbf{J(b)'J(b)}$ and its rank,

(4) Repeat step (3) for other values of **b**, as seems appropriate.

An example is provided here to illustrate how the Jacobian test is performed.

Example

For simplicity, let us begin with an example of a linear estimation problem, which is a special case of the underlying NLS problem. Consider the following T-intersection example (see Figure 3.1), which has six turning movements, indexed from b_1 to b_6. Assume that the travel time between any pair of input-output counts is short and can be ignored. At every time t, the input and output time-series counts are collected and denoted as $q_1(t)$, $q_2(t)$, $q_3(t)$, and $y_1(t)$, $y_2(t)$, $y_3(t)$, t=1,...,T. One can write the linear predictor in a matrix form as

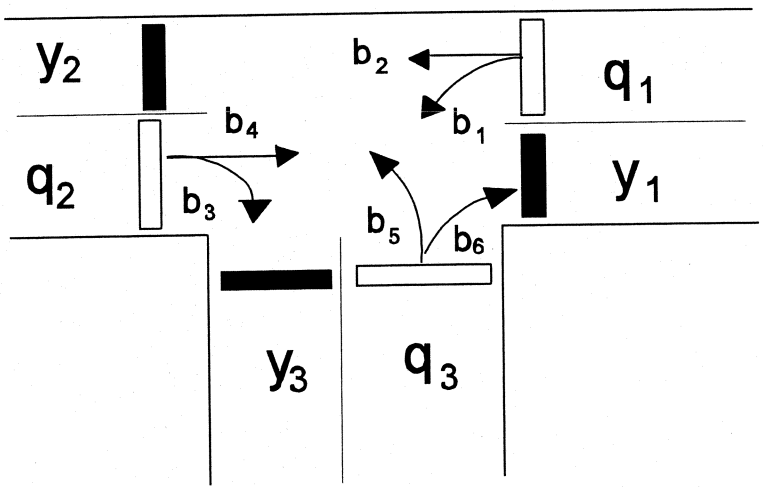


Figure 3.1 A T-Intersection Example for Parameter Identifiability

$$\begin{bmatrix}\hat{y}_1(t)\\ \hat{y}_2(t)\\ \hat{y}_3(t)\end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & q_2(t) & 0 & q_3(t)\\ 0 & q_1(t) & 0 & 0 & q_3(t) & 0\\ q_1(t) & 0 & q_2(t) & 0 & 0 & 0\end{bmatrix} \begin{bmatrix} b_1\\ b_2\\ b_3\\ b_4\\ b_5\\ b_6\end{bmatrix}$$

$$= \boldsymbol{X}(t)\,\boldsymbol{b}$$

Working with a reduced set of parameters, such that $\mathbf{b}_1 = [b_2\ b_4\ b_6]'$ and $\mathbf{b}_2 = [b_1\ b_3\ b_5]'$, the prediction model can be rewritten as

$$\begin{bmatrix} \hat{y}_1(t) - q_2(t) - q_3(t) \\ \hat{y}_2(t) - q_1(t) \\ \hat{y}_3(t) \end{bmatrix} = \begin{bmatrix} 0 & -q_2(t) & -q_3(t) \\ -q_1(t) & 0 & q_3(t) \\ q_1(t) & q_2(t) & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_3 \\ b_5 \end{bmatrix}$$

$$= \boldsymbol{X}(t)\, \boldsymbol{b}_2$$

The rank test is then performed on

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{X}(1) \\ \vdots \\ \boldsymbol{X}(T) \end{bmatrix}$$

With appropriate input counts, the rank test on $\boldsymbol{J}$ will give rank of 3 (full rank), therefore the parameters are identifiable. Now, suppose detector y_3 is malfunctioning and this is detected. Only the time-series counts of $y_1(t)$ and $y_2(t)$ are available. In such a case, the rank test on $\boldsymbol{J}$ could still be of full rank. This example illustrates that, a configuration with one output detector deleted from the full-set of detectors at an intersection can still be identifiable.

3.3 Quality of Detector Information

The success of the rank test does not guarantee that "good" quality of parameter estimates will be obtained from a given identifiable detector configuration. The quality of estimates really depends on the amount of information provided by the detector configuration and the input count data. Even with an identifiable detector configuration, the estimates would not be useful if the quality of the estimates is poor. Ideally, the quality of parameter estimates should be measured by their variability, such as the covariance matrix for the estimates. A simpler alternative is to measure the information based on the condition number of the predicted Jacobian matrix. The condition number of a matrix has been defined in several ways, for example, as the ratio of the largest to the smallest characteristic root (or eigen-value) (Belsley et al., 1980), the ratio of the largest to smallest

singular value (MATLAB, 1994), or the ratio of the norm of the matrix to the norm of the inverse of the matrix (Golub and Van Loan, 1989). The first definition will be undefined when the characteristic root of a matrix is negative or unreal. The third definition is norm-dependent. The singular values are always positive and uniquely determined by the Singular Value Decomposition (SVD). In addition, it has been established (Golub and Van Loan, 1989) that the sensitivity of a least-squares problem can also be analyzed through the SVD, and the smallest singular value is a measure of the closeness of a matrix to singularity. It is therefore recommended that the singular value be used to define the condition number. For convenience, the condition number of the Jacobian matrix is computed via

$$\left(\frac{s_{max}}{s_{min}} \right)^{1/2} \tag{4.4}$$

where s_{max} and s_{min} are the largest and smallest singular values of $\boldsymbol{J}(\boldsymbol{b})'\boldsymbol{J}(\boldsymbol{b})$. Since the singular values or characteristic roots are dependent on the scaling of the data, Greene (1993) suggests normalizing the data by dividing each column in $\boldsymbol{J}$ by the root of its sum of squared of elements (i.e. norm). Higher condition numbers indicate poorer quality of information, suggesting an ill-posed detector configuration. Belsley et al. suggested that values in excess of 20 may indicate a potential problem in linear least-squares problems.

Example

An example is given here to illustrate how the information is quantified by the condition number under various detector configurations. There are four detector configurations being considered for a two-intersection network (see Figures 3.2): a full set of detectors (denoted by

"Full"), a cordon count detectors ("Cordon"), a configuration with a deleted cordon counter at one intersection ("One-less"), and a configuration with a deleted cordon counter at both intersections ("Two-less"). In each simulation run, given a sequence of input counts and a "best" guess of turning movement proportions, one can compute the Jacobian matrices numerically, perform the rank tests, and compute the condition numbers. A total of 50 simulated data sets were generated. The true value of the turning movement proportion used in generating the simulated data were set equal to 0.3, 0.6, and 0.1 for left-turn, through, and right-turn movements. Only the left and through movement proportions, indexed from 1 to 16 (see Figure 3.2), at each approach of intersections were treated as linearly independent parameters. All of these configurations are identifiable based on the rank test of the Jacobian matrix. Figure 3.3 depicts the condition numbers of each configuration resulting from each simulation. As expected, the Full configuration has smallest condition number, and the condition number increases as the number of detectors is reduced from the Full configuration. It is suggested that the Cordon configuration appears to provide adequate information (the average condition number is under 20), while the One-less configuration has a average condition number slightly over 20. Also, the "condition" deteriorates when the cordon counters are eliminated by two. By minimizing the nonlinear sum of squares function (3.1) using the *constr* optimization routine in MATLAB (1994), a sample of turning proportion estimates was generated for each detector configuration. A complete discussion of the off-line NLS estimation is contained in Chapter 4, and will not be duplicated here. Figure 3.4 depicts the parameter estimates across those configurations. One can see that the turning proportion estimates from the Full configuration (marked by '*') match closely with the true values, while some of the estimates from the Cordon configuration (marked by 'o') deviate. The estimates from One-less configuration (marked by '+')

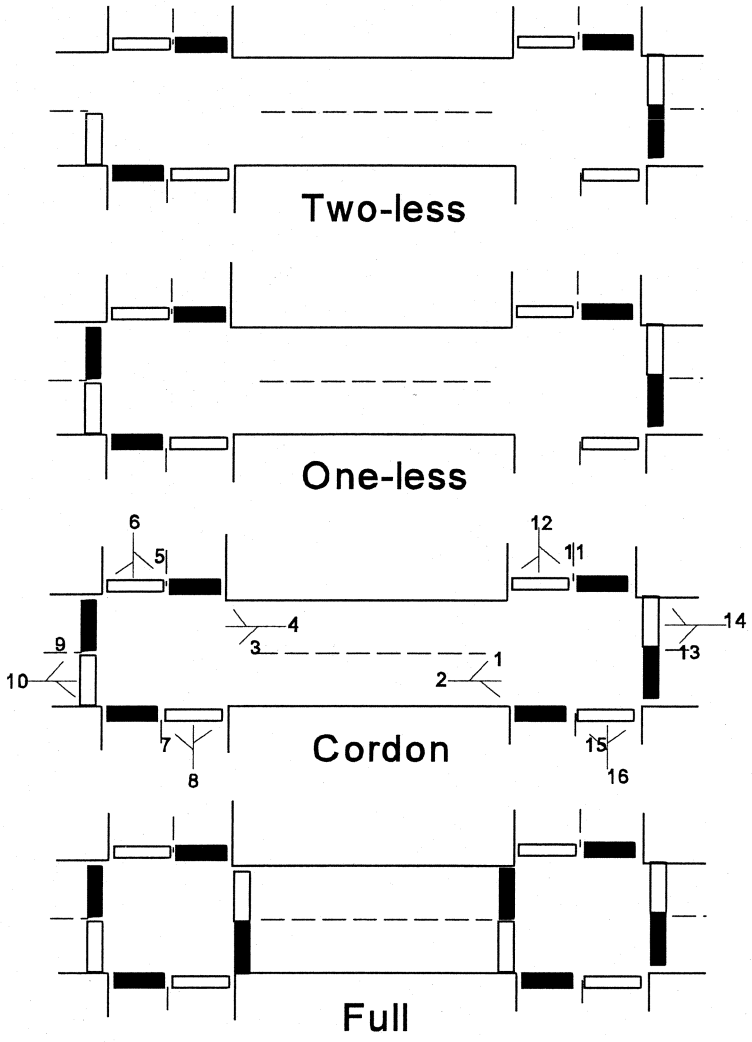


Figure 3.2 Detector Configuration in a Two-Intersection Network

follow closely with the Cordon configuration, indicating that deletion of one detector from the cordon count does not affect quality of estimation. However, when deleting one more output detector from the other intersection, the quality of estimation deteriorates and some of estimates appear to be irregular (b_2, b_{11}, b_{12}). Computing the norm of their deviations from the true values gives a similar indication: 0.1497 ("Full"), 0.2142 ("Cordon"), 0.2286 ("One-less"), and 0.6705 ("Two-less").

The cordon-count configuration is an appealing idea, since detector cost savings would be

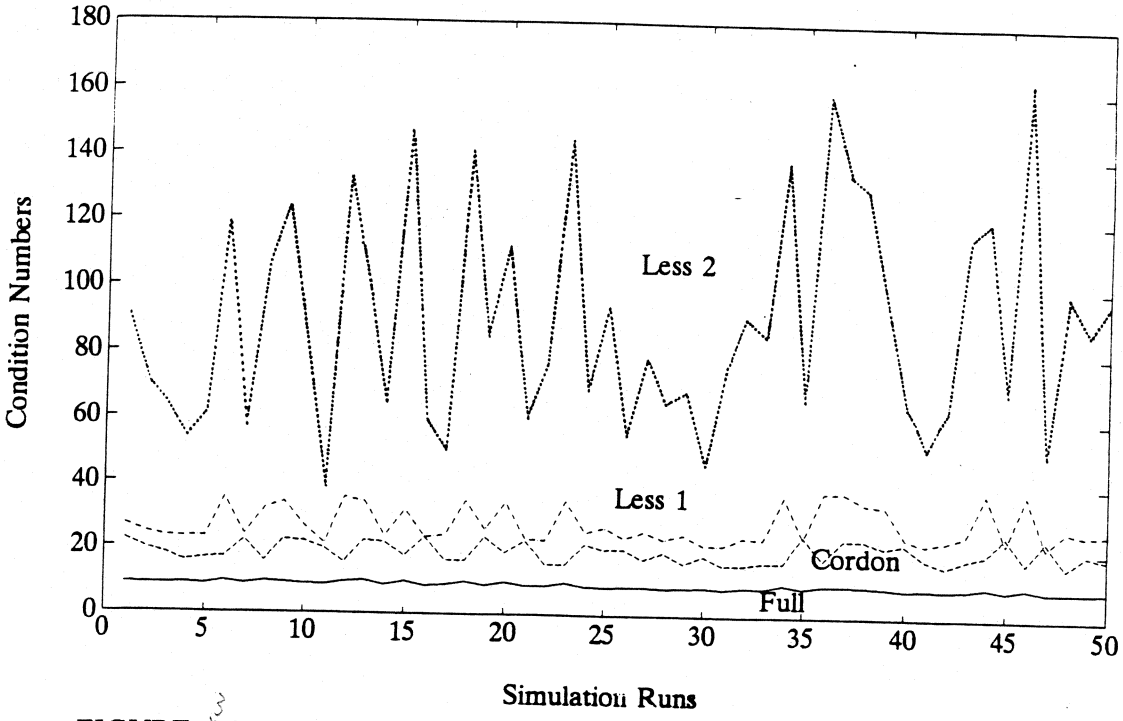


FIGURE 4.3 Condition Numbers Under Various Detector Configurations

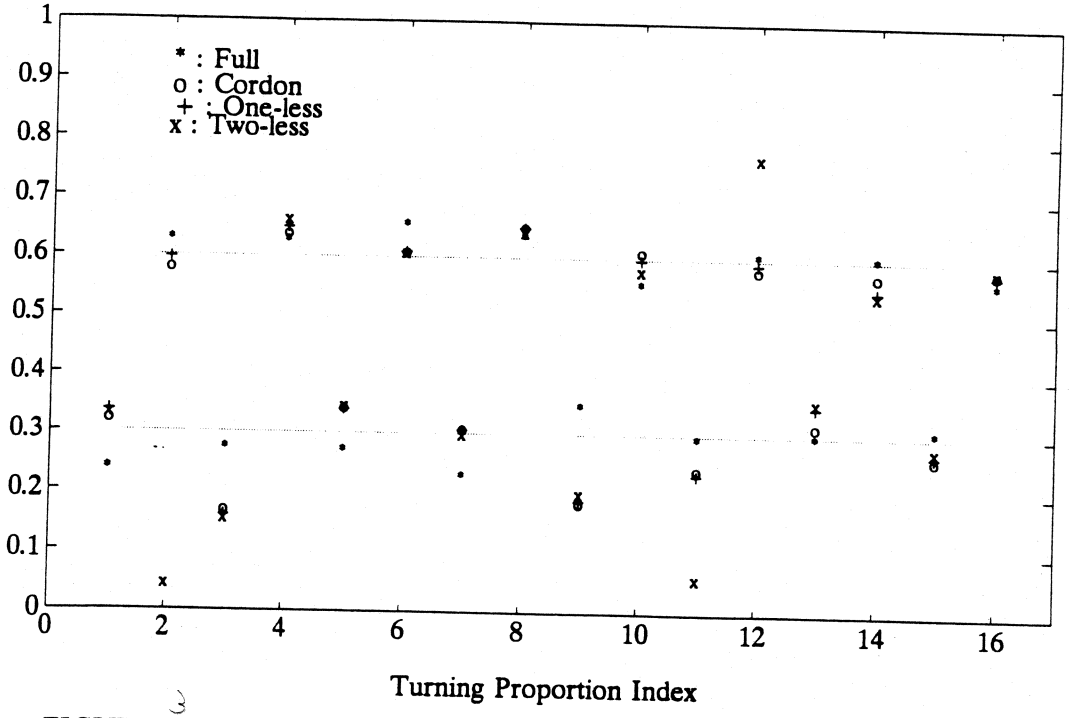


FIGURE 4.4 Parameter Estimates Across Various Detector Configurations

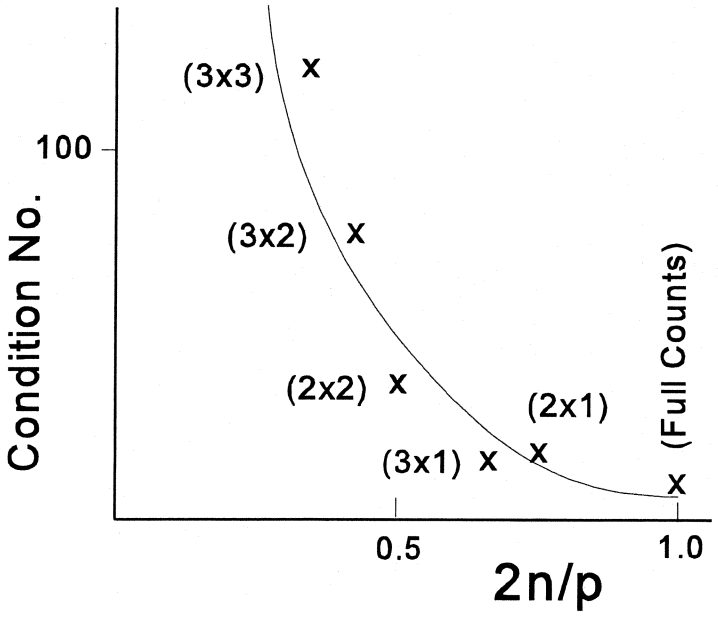


Figure 3.5 Condition No. Against (2n/p) Under Cordon-Count Configurations

CHAPTER 4

ESTIMATION ALGORITHMS

Two types of algorithms are presented for estimating turning proportion parameters in a network of intersections with partial set of detectors. A nonlinear least squares (NLS) algorithm is proposed which estimates the parameters of a deterministic nonlinear system. A Quasi Maximum Likelihood (QML) algorithm is based on the likelihood criterion for estimating parameters of a stochastic time-varying linear system. Both algorithms belong to the class of prediction error minimization (PEM) methods. Section 4.1 gives a brief introduction to the underlying estimation problem. In Section 4.2, the relation between the Markovian traffic flow model and estimation based on the PEM principle is briefly reviewed. A large population approximation of this model for on-line implementations is also presented. Section 4.3 describes how the off-line and recursive versions of these estimators are developed, and sections 4.4 is then devoted to evaluating these estimation algorithms in smulation studies.

4.1 Introduction

The investigation of the problem of automatic identification of the turning movement proportions at intersections using detector counts has resulted in a considerable volume of work. This effort has focussed primarily to seek methodologies for improving the quality of the statistical estimators, under a scenario where a complete set of detectors is available. If counts of inflows from each approach and outflows to each exit are available, and the travel time between input and output

counters is short enough to be ignored, one can image using either the under-determined approach (Jeffreys and Norman, 1977; Mekky, 1979; Van Zuylen, 1979; Hauer et al., 1981) or the over-determined approach (Cremer and Keller, 1981, 1987; Nihan and Davis, 1987, 1989; Kessaci et al., 1989; Bell, 1991) to solve the estimation problem, based on least-square, maximum likelihood, or other criteria. The OLS-based methods have been shown capable of producing unbiased estimates both off-line and recursively (Nihan and Davis, 1987, 1989). However, one of the disadvantages of having such a dense detector configuration is the heavy expenses of installation and maintenance. Besides, evidence shows that the OLS-based methods are unable to produce useful estimates when the travel time between inflow and outflow counters is significant, or when intervening traffic conditions change the travel time (Davis, 1993). Finally, the OLS-based methods may fail locally when detectors malfunction, and the chance of the occurrence of a faulty detector is proportional to number of detectors placed. In view of these concerns, Davis and Lan (1995) proposed an off-line nonlinear least-square (NLS) approach to estimate the turning movement proportions for a network of signalized intersections with a partial set of detectors. This work shed a new light on this topic and indicated that estimation without a complete set of detectors might be possible. However, possible alternative estimators for improving the quality of estimation and a thorough evaluation of the statistical properties of these estimators remain to be addressed. In addition, estimating or predicting the dynamic evolution of turning flows and traffic conditions for supporting real-time traffic control requires algorithms for identifying those parameters recursively.

4.2 The Large Population Approximation

As stated earlier, when only a partial set of detector counts is available on a network of

intersections, the estimation problem is complicated by the fact that the travel time between input and output counters may be varying across time and is not negligible. In this regard, estimation requires a well-supported traffic flow model able to (1) describe the dynamics of traffic conditions on the network, and (2) link the turning proportion parameters embedded in traffic flow model to the predicted output, making estimation based on the prediction error minimization principle possible. The basic and extended traffic flow models had been developed and described in Phase II project and Chapter 2 of Phase III. For real-time implementations, it is desirable to make these nonlinear models tractable and suitable for linear filtering applications. This leads to a need for a plausible approximation to the underlying nonlinear process.

Before proceeding, it needs to show that the traffic flow models are continuously differentiable. It in turn suffices to show that the two-piece passage probability terms are continuously differentiable at $x_k=x_{ck}$, provided that the space mean speed function $\bar{u}(x)$ presumes one of the continuous speed-density relationships. The continuous differentiability is required later to justify an the approximation of the Markov model. Based on the classic notion of macroscopic traffic flow, one can write the flow (suppress k) $q = \frac{\bar{u}(x)x}{L}$. Since $\left(\frac{\partial q}{\partial x}\right)_{x=x_c} = 0$, then

$$\frac{\bar{u}(x_c)}{L} + \frac{\partial \bar{u}(x_c)}{\partial x}\frac{x_c}{L} = 0, \text{ and } \bar{u}_e = \bar{u}(x_c) = -x_c\frac{\partial \bar{u}(x_c)}{\partial x}$$

The derivative of passage probability at $x=x_c$ in the congestion region equals

$$\frac{\Delta}{L}\bar{u}_e\left(-\frac{x_c}{x^2}\right)_{x=x_c} = \frac{\Delta}{L}\left(\frac{-1}{x_c}\right)\left(-x_c\frac{\partial\bar{u}(x_c)}{\partial x}\right) = \frac{\Delta}{L}\frac{\partial\bar{u}(x_c)}{\partial x},$$

which is the derivative at $x=x_c$ in the uncongestion region.

In the underlying models, the output counts are assumed to be aggregated independent multinomial outcomes over the basic time intervals, so their probability distribution is a convolution of functions of the time-delayed state variables $\mathbf{x}(\tau)$. One way to avoid this complexity and set up a state space model easier to manipulate is to include basic output variables in the following augmented state vector, allowing for time aggregation of output counts,

$$\boldsymbol{z}(\tau) \equiv \begin{pmatrix} \boldsymbol{x}(\tau) \\ \boldsymbol{y}(\tau) \end{pmatrix} \tag{4.1}$$

with the state equations written as

$$x_k(\tau+1) = x_k(\tau) - \sum_l \tilde{y}_{kl}(\tau) + \sum_h \tilde{y}_{hk}(\tau) + \sum_i \delta_{ki} q_i(\tau) \tag{4.2}$$

and the observation equation as

$$y_j(\tau+1) = y_j(\tau) + \Sigma_l \tilde{y}_{kl}(\tau), \text{ if detector j counts exits from compartment k} \tag{4.3}$$

$$= y_j(\tau) + \Sigma_h \tilde{y}_{hk}(\tau), \text{ if detector j counts entries into compartment k}$$

Given the continuously differentiable exiting probability functions, as the population of travelers becomes large, one can approximate the above state space model with a first-order auto-regressive model

$$\begin{bmatrix} \boldsymbol{x}(\tau+1) \\ \boldsymbol{y}(\tau+1) \end{bmatrix} - \begin{bmatrix} \bar{\boldsymbol{x}}(\tau+1) \\ \bar{\boldsymbol{y}}(\tau+1) \end{bmatrix} = \boldsymbol{F}(\tau) \left(\begin{bmatrix} \boldsymbol{x}(\tau) \\ \boldsymbol{y}(\tau) \end{bmatrix} - \begin{bmatrix} \bar{\boldsymbol{x}}(\tau) \\ \bar{\boldsymbol{y}}(\tau) \end{bmatrix} \right) + \begin{bmatrix} \boldsymbol{v}_x(\tau) \\ \boldsymbol{v}_y(\tau) \end{bmatrix} \tag{4.4}$$

by applying Large Population Approximation (LPA) (Lehoczky, 1980). The deterministic processes $\bar{\boldsymbol{x}}(\tau)$ and $\bar{\boldsymbol{y}}(\tau)$ obey the nonlinear dynamic equations

$$\begin{aligned} \bar{x}_k(\tau+1) &= f_k(\bar{\boldsymbol{x}}(\tau), \boldsymbol{b}, \bar{\boldsymbol{q}}(\tau)) \\ &= \bar{x}_k(\tau) - \sum_l \bar{x}_k(\tau) p_{kl}(\bar{\boldsymbol{x}}(\tau)) \boldsymbol{b}_{kl} + \sum_h \bar{x}_h(\tau) p_{hk}(\bar{\boldsymbol{x}}(\tau)) \boldsymbol{b}_{hk} + \sum_i \delta_{ki} \bar{q}_i(\tau) \end{aligned} \tag{4.5a}$$

$$\bar{y}_j(\tau+1) = \bar{y}_j(\tau) + g_j(\bar{\boldsymbol{x}}(\tau), \boldsymbol{b}) \tag{4.5b}$$

$= \bar{y}_j(\tau) + \sum_l \bar{x}_k(\tau) p_{kl}(\bar{\boldsymbol{x}}(\tau)) \boldsymbol{b}_{kl}$, if detector j counts exits from compartment k

$= \bar{y}_j(\tau) + \sum_h \bar{x}_h(\tau) p_{hk}(\bar{\boldsymbol{x}}(\tau)) \boldsymbol{b}_{hk}$, if detector j counts entries into compartment k

$\mathbf{F}(\tau)$ is the Jacobian matrix in the form of

$$\boldsymbol{F}(\tau) = \begin{bmatrix} \boldsymbol{F}_{xx}(\tau) & \mathbf{0} \\ \boldsymbol{F}_{yx}(\tau) & \boldsymbol{I} \end{bmatrix} \tag{4.6a}$$

where $\mathbf{F}_{xx} \equiv (\partial \mathbf{f}/\partial \mathbf{x})$, and $\mathbf{F}_{yx} \equiv (\partial \mathbf{g}/\partial \mathbf{x})$.

The covariance matrix of the state variables at time τ,

$$\boldsymbol{P}(\tau) \equiv \begin{bmatrix} \boldsymbol{P}_{xx} & \boldsymbol{P}_{xy} \\ \boldsymbol{P}_{yx} & \boldsymbol{P}_{yy} \end{bmatrix}_{(\tau)}$$

Finally, as a result of the approximation, the mean evolution of the output counts is predicted as

$$\hat{y}_j(t,\mathbf{b}) = \sum_{\tau=1}^{K} g_k(\bar{\boldsymbol{x}}(\tau), \boldsymbol{b}) \tag{4.9}$$

$$= \sum_{\tau}\sum_{l} \bar{x}_k(\tau) p_{kl}(\bar{\boldsymbol{x}}(\tau)) b_{kl}, \text{ if detector j counts exits from compartment k}$$

$$= \sum_{\tau}\sum_{h} \bar{x}_h(\tau) p_{hk}(\bar{\boldsymbol{x}}(\tau)) b_{hk}, \text{ if detector j counts entries into compartment k}$$

The nonlinear state equation and linearized state-space model resulted from LPA will serve the building blocks in constructing and implementing the off-line and recursive estimation algorithms.

4.3 Development of Estimation Algorithms

In the sequel, algorithms for estimating turning movement proportions are developed. First, based on the nonlinear least squares criterion, a NLS algorithm is proposed for estimating parameters of a deterministic nonlinear system. Second, a Quasi Maximum Likelihood (QML) algorithm is based on the likelihood criterion for estimating parameters of a stochastic time-varying linear system. Both algorithms belong to the class of prediction error minimization (PEM) methods (Ljung and Söderström, 1983). These estimators are discussed in both off-line and recursive forms, where off-line estimates serve as a benchmark for comparisons of statistical properties, while the recursive forms are more suitable for real-time implementation, particularly for responding to changes in turning movement proportions over time. The simulation data generated from different example networks and real data collected from the field are used to evaluate the effectiveness of each proposed algorithm.

$$\boldsymbol{e}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t, \boldsymbol{b}) = \boldsymbol{H}(\boldsymbol{z}(t) - \hat{\boldsymbol{z}}(t, \boldsymbol{b})), \tag{4.11}$$

To obtain a necessary condition for the criterion function to be minimized, one takes derivatives of the criterion function with respect to parameters and sets them to zero. Approximating $\hat{\boldsymbol{z}}(t, \boldsymbol{b})$ by using the first-order Taylor's series expansion around **b** yields the following RNLS estimator.

(1) *Time-Update* of the nonlinear state equations in (4.5) describing the mean of the state vector $\mathbf{z}(\tau)$, and of the sensitivity matrix $\Psi \equiv (\partial \hat{\boldsymbol{z}} / \partial \mathbf{b})$ (Mataušek et al., 1980; Chen, 1983) in the following recursion:

$$\psi(\tau+1) = \begin{bmatrix} \boldsymbol{F}_{xx}(\tau) & \mathbf{0} \\ \boldsymbol{F}_{yx}(\tau) & \boldsymbol{I} \end{bmatrix} \psi(\tau) + \begin{bmatrix} \boldsymbol{F}_{xb}(\tau) \\ \boldsymbol{F}_{yb}(\tau) \end{bmatrix} \tag{4.12}$$

where $\mathbf{F}_{xb} \equiv (\partial \mathbf{f} / \partial \mathbf{b})$, $\mathbf{F}_{yb} \equiv (\partial \mathbf{g} / \partial \mathbf{b})$. Note that due to $\hat{\boldsymbol{y}}(\mathbf{b},t)$ being reset to zero at the beginning of each time update for accumulating the output counts, set $\partial \hat{\boldsymbol{y}}(\boldsymbol{b}, t) / \partial \mathbf{b} = \mathbf{0}$ accordingly.

(2) *Measurement-Update*. Once measurements become available at time t, one can update or adjust the parameter estimates according to the following scheme:

$$\begin{aligned} \boldsymbol{K}(t) &= \boldsymbol{P}(t-1)\psi(t)'\boldsymbol{H}'(\boldsymbol{I} + \boldsymbol{H}\psi(t)\boldsymbol{P}(t-1)\psi(t)'\boldsymbol{H}')^{-1} \\ \hat{\boldsymbol{b}}(t) &= \hat{\boldsymbol{b}}(t-1) + \boldsymbol{K}(t)[\boldsymbol{y}(t) - \hat{\boldsymbol{y}}(\boldsymbol{b}, t)] \\ \boldsymbol{P}(t) &= (\boldsymbol{I} - \boldsymbol{K}(t)\boldsymbol{H}\psi(t))\boldsymbol{P}(t-1) \end{aligned} \tag{4.13}$$

This algorithm simply starts with $\mathbf{x}(0)$, $\hat{\boldsymbol{b}}(0)$, and $\Psi(0)=\mathbf{0}$, then calculates $\mathbf{x}(\tau)$ and $\Psi(\tau)$ as time advances. Once measurements $\mathbf{y}(t)$ are available, one updates the parameter estimates $\hat{\boldsymbol{b}}(t)$, resets

$$\boldsymbol{e}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{y}}(t, \boldsymbol{b}) = \boldsymbol{H}(\boldsymbol{z}(t) - \hat{\boldsymbol{z}}(t, \boldsymbol{b})), \tag{4.11}$$

To obtain a necessary condition for the criterion function to be minimized, one takes derivatives of the criterion function with respect to parameters and sets them to zero. Approximating $\hat{\boldsymbol{z}}(t, \boldsymbol{b})$ by using the first-order Taylor's series expansion around **b** yields the following RNLS estimator.

(1) *Time-Update* of the nonlinear state equations in (4.5) describing the mean of the state vector $\mathbf{z}(\tau)$, and of the sensitivity matrix $\Psi \equiv (\partial \hat{\mathbf{z}} / \partial \mathbf{b})$ (Mataušek et al., 1980; Chen, 1983) in the following recursion:

$$\psi(\tau+1) = \begin{bmatrix} \boldsymbol{F}_{xx}(\tau) & \mathbf{0} \\ \boldsymbol{F}_{yx}(\tau) & \boldsymbol{I} \end{bmatrix} \psi(\tau) + \begin{bmatrix} \boldsymbol{F}_{xb}(\tau) \\ \boldsymbol{F}_{yb}(\tau) \end{bmatrix} \tag{4.12}$$

where $\mathbf{F}_{xb} \equiv (\partial \mathbf{f} / \partial \mathbf{b})$, $\mathbf{F}_{yb} \equiv (\partial \mathbf{g} / \partial \mathbf{b})$. Note that due to $\hat{\boldsymbol{y}}(\mathbf{b},t)$ being reset to zero at the beginning of each time update for accumulating the output counts, set $\partial \hat{\boldsymbol{y}}(\boldsymbol{b}, t) / \partial \mathbf{b} = \mathbf{0}$ accordingly.

(2) *Measurement-Update*. Once measurements become available at time t, one can update or adjust the parameter estimates according to the following scheme:

$$\boldsymbol{K}(t) = \boldsymbol{P}(t-1)\psi(t)'\boldsymbol{H}'(\boldsymbol{I} + \boldsymbol{H}\psi(t)\boldsymbol{P}(t-1)\psi(t)'\boldsymbol{H}')^{-1}$$

$$\hat{\boldsymbol{b}}(t) = \hat{\boldsymbol{b}}(t-1) + \boldsymbol{K}(t)[\boldsymbol{y}(t) - \hat{\boldsymbol{y}}(\boldsymbol{b}, t)] \tag{4.13}$$

$$\boldsymbol{P}(t) = (\boldsymbol{I} - \boldsymbol{K}(t)\boldsymbol{H}\psi(t))\,\boldsymbol{P}(t-1)$$

This algorithm simply starts with $\mathbf{x}(0)$, $\hat{\boldsymbol{b}}(0)$, and $\Psi(0)=\mathbf{0}$, then calculates $\mathbf{x}(\tau)$ and $\Psi(\tau)$ as time advances. Once measurements $\mathbf{y}(t)$ are available, one updates the parameter estimates $\hat{\boldsymbol{b}}(t)$, resets

$\partial \hat{y}(\boldsymbol{b}, t) / \partial \mathbf{b}=\mathbf{0}$, and continues the recursion until next measurement becomes available.

4.3.2 *The QML Algorithms*

This algorithm is developed to deal with a stochastic, time-varying linear system. With the approximate model described earlier, given the sequence of measurements $\mathbf{Y}(t) = \{\mathbf{y}(t), t=1,...,T\}$ and an estimate of $\boldsymbol{b}$, one can write the log-likelihood function as

$$L(\boldsymbol{Y}(t), \boldsymbol{b}) = \frac{1}{2}\sum_{t=1}^{T}\left\{\ln|\boldsymbol{P}_{yy}(\boldsymbol{b}, t\,|\,t-1)| + (\boldsymbol{y}(t)-\hat{\boldsymbol{y}}(\boldsymbol{b}, t\,|\,t-1))'\boldsymbol{P}_{yy}(\boldsymbol{b}, t\,|\,t-1)^{-1}(\boldsymbol{y}(t)-\hat{\boldsymbol{y}}(\boldsymbol{b}, t\,|\,t-1))\right\} \tag{4.14}$$

where $||$ denotes the determinant. The QML estimator $\hat{\boldsymbol{b}}_{QML}$ is computed via the following off-line and recursive schemes.

Off-line QML Estimator

Given the $\mathbf{q}(t)$, initial state $\mathbf{x}(0)$, and a trial set of $\hat{\boldsymbol{b}}(0)$ and $\mathbf{P}(0)$, one performs the two-stage recursion as follows:

(1) *Time-update* of the nonlinear state and observation equations (4.5), describing the mean evolution of the state vector $\mathbf{z}(\tau)$, and on the covariance matrices $\mathbf{P}(\tau)$ according to the Lyapunov equation with $\mathbf{P}(\tau) = \mathbf{P}(\tau|\tau\text{-}1)$, and

$$P(\tau|\tau\text{-}1) = F(\tau\text{-}1)P(\tau\text{-}1)F(\tau\text{-}1)'+Q(\tau\text{-}1) \tag{4.15a}$$

(2) *Measurement-update* to obtain the filtered estimates with the above augmented state space model (4.6) using the Kalman filter, once measurements become available at time τ=t.

$$\begin{aligned} \boldsymbol{K}(\tau) &= \boldsymbol{P}(\tau|\tau-1)\boldsymbol{H}'(\boldsymbol{R}(\tau)+\boldsymbol{H}\boldsymbol{P}(\tau|\tau-1)\boldsymbol{H}')^{-1} \\ \hat{\boldsymbol{z}}(\tau) &= \hat{\boldsymbol{z}}(\tau|\tau-1)+\boldsymbol{K}(\tau)\left[\boldsymbol{y}(\tau)-\hat{\boldsymbol{y}}(\boldsymbol{b},\tau)\right] \\ \boldsymbol{P}(\tau) &= (\boldsymbol{I}-\boldsymbol{K}(\tau)\boldsymbol{H})P(\tau|\tau-1) \end{aligned} \tag{4.15b}$$

If no measurement error is assumed, i.e., the measurement error covariance matrix $\mathbf{R}(\tau) = \mathbf{0}$, a simplification of the measurement update procedure yields

$$\begin{aligned} \boldsymbol{P}(\tau) &= \begin{bmatrix} \boldsymbol{P}_{xx}-\boldsymbol{P}_{xy}\boldsymbol{P}_{yy}^{-1}\boldsymbol{P}_{yx} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{(\tau|\tau-1)} \\ \hat{\boldsymbol{x}}(\tau) &= \hat{\boldsymbol{x}}(\tau|\tau-1)+\left(\boldsymbol{P}_{xy}\boldsymbol{P}_{yy}^{-1}\right)_{(\tau|\tau-1)}\left(\boldsymbol{y}(\tau)-\hat{\boldsymbol{y}}(\tau|\tau-1)\right) \end{aligned} \tag{4.16}$$

where $\hat{\boldsymbol{x}}(\tau|\tau\text{-}1)$ and $\mathbf{P}(\tau|\tau\text{-}1)$ are one-step ahead predictions obtained from the time-update step. Similar to the off-line NLS estimation, the recursion can be embedded in the MATLAB routine *constr* to find parameter values minimizing the likelihood function. No derivative computations of the likelihood are required. Note that if measurement error exists and can be quantified properly, the measurement update can be carried out through the regular Kalman filter.

Recursive QML (RQML) Estimator

For the purpose of on-line implementation, the extended Kalman filter (EKF) is perhaps the most well-known among various approaches which have been attempted. The EKF algorithm can

be regarded as the simplified version of the full RQML estimator (Ljung, 1979), where the simplification greatly reduces the computational complexity. In doing so, one needs to further augment the state vector $\mathbf{z}(\tau)$ to include the turning movement proportions by viewing them as constant but unknown states, or a random walk process with the state equation written as $\mathbf{b}(\tau)=\mathbf{b}(\tau-1)+\mathbf{v}_b$, where $\mathbf{v}_b$ is a sequence of independent random vectors with zero mean and covariance matrix $\mathbf{S}(t)$. This algorithm starts with $\hat{\boldsymbol{x}}(0)$, $\hat{\boldsymbol{b}}(0)$, and $\mathbf{P}(0)$, and performs the time-update on the augmented state vector and $\mathbf{P}(\tau)$ as time advances. Once measurements $\mathbf{y}(t)$ are available, it performs a measurement-update on the augmented state vector $\mathbf{z}(t)$ and $\mathbf{P}(t)$ and restarts the time-update with $\hat{\boldsymbol{y}}(t)=\mathbf{0}$.

4.4 Evaluation of Estimation Algorithms

In this section, simulation studies are conducted to examine the quality of linear approximation resulted from LPA, and the statistical properties of the proposed estimation algorithms. As shown before, the linear approximation of the Markov model serves as a building block for constructing the QML algorithm. The quality of such an approximation may directly affect the effectiveness of the estimation algorithm, and therefore it might be necessary to inspect that before proceeding. The second part compares the statistical properties of the NLS and QML estimators from the cordon-count configuration with the OLS estimator computed using a full set of detectors. The third part implements recursive NLS and QML algorithms on the same network and detector configuration using the same simulation data to examine the relative effectiveness of these

algorithms in tracking time-invariant turning movement proportions. Feasibility and possible modifications of these recursive algorithms in tracking stepwise changes of turning parameters across time are also examined. Finally, a real data set collected from a two-signal network is analyzed.

4.4.1 *Simulation Scenarios*

As depicted in Figure 4.1, a simple network with two intersections of two-way streets is used for illustration. Figure 4.2 shows two different detector configurations for this network. Placement scenario I corresponds to the full detectorization, while scenario II corresponds to a cordon count placement where the number of output counters n=6, and the number of linearly independent parameters to be estimated p=16. The simulated traffic counts were generated using the Markovian traffic flow model. Simulated input counts at entry points for each time τ were generated as Poisson outcomes with time-varying $\bar{q}(\tau)$, while the number of vehicles exiting compartment k during time τ were generated as binomial outcomes with $(x_k(\tau), p_k(\mathbf{x}(\tau)))$. The exiting vehicles were then allocated to downstream compartments as multinomial outcomes with probabilities b_{kl}. Traffic signals were given a two-phase fixed timing with 60-sec cycle and 30-sec greens allocated to each phase. A total of 50 simulated data sets were generated, each consisting of 36 5-min traffic counts from each detector.

4.4.2 *Quality of the Approximation*

Using the data generated from scenario II, the means and ±2 standard deviations for six

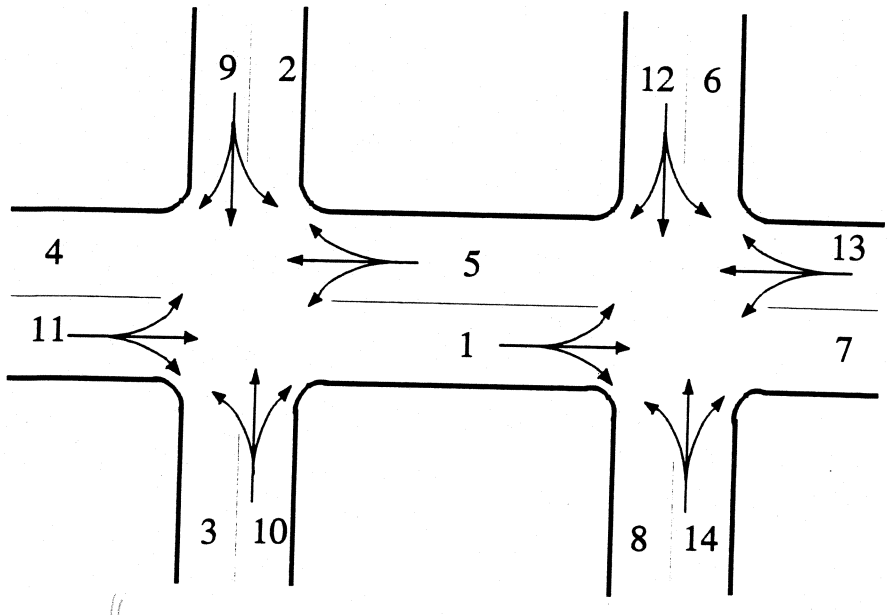


FIGURE 5.1 Two-Intersection Network Example for Simulation Studies

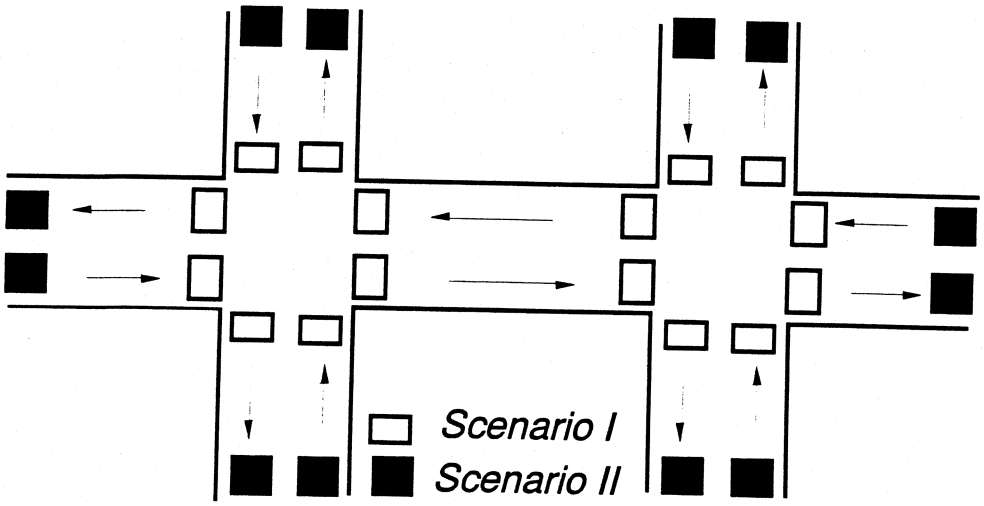
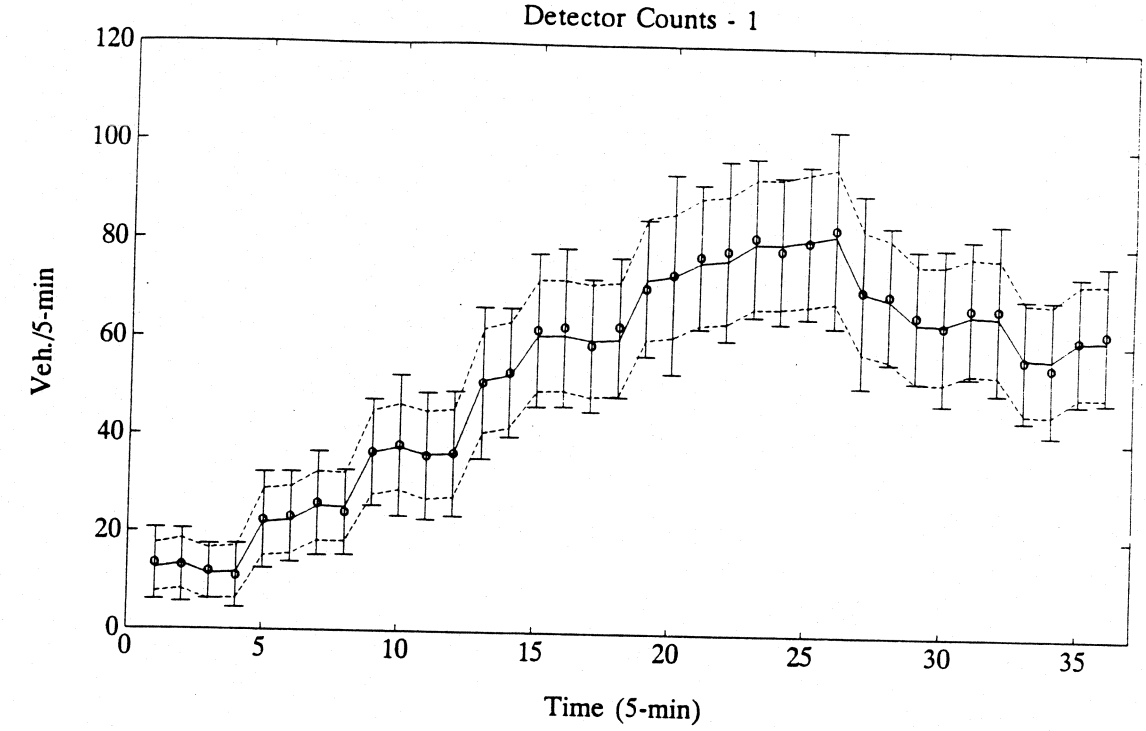


FIGURE 5.2 Two Cordon-Count Detector Configurations for Simulation Studies

simulated output counts were computed and compared with their counterparts computed from the approximated Gaussian process. As shown in Figures 4.3a-f, the deterministic mean trajectories and standard-deviation bounds (in dashed lines) are quite in agreement with the stochastic counterparts (in errorbars) except for a slight under-estimation.

4.4.3 *The Off-line Estimators*

The evaluation of the off-line estimators was performed under both scenarios mentioned above. Under scenario I, the estimates of the turning movement proportions are subject to the equality constraints $\sum_l b_{kl} = 1.0$ and the inequality constraints $\mathbf{0} \le \mathbf{b} \le \mathbf{1}$. Therefore they are computed using the equality constrained least-squares (CLS) solution obtained by orthogonal projection of the OLS estimate onto the constraint set. The inequality constraints were inactive in all cases, so are not dealt with explicitly. Under scenario II, the estimates were generated by the off-line NLS and QML algorithms, accomplished by using a constrained optimization routine in MATLAB. Here, only the left-turn and through proportions (b_{LT} and b_{TH}) at each approach of intersections were treated as independent parameters, with the right-turn proportion then being computed as $b_{RT} = 1-b_{LT}-b_{TH}$. The estimates from these two scenarios were computed for 50 samples. The results regarding the statistical properties of the estimators are compiled in Tables 4.1 and 4.2, where "Bias", "S.D.", "RMSE", "t-stat", and "χ^2_2-stat" columns give the sample bias, standard deviations, root of mean square error, t-statistics, and χ^2-statistics of the estimates. The RMSE is defined as $\sqrt{Bias^2+(S.D.)^2}$, a finite-sample measure of the quality of an estimator. The t-statistic is used to



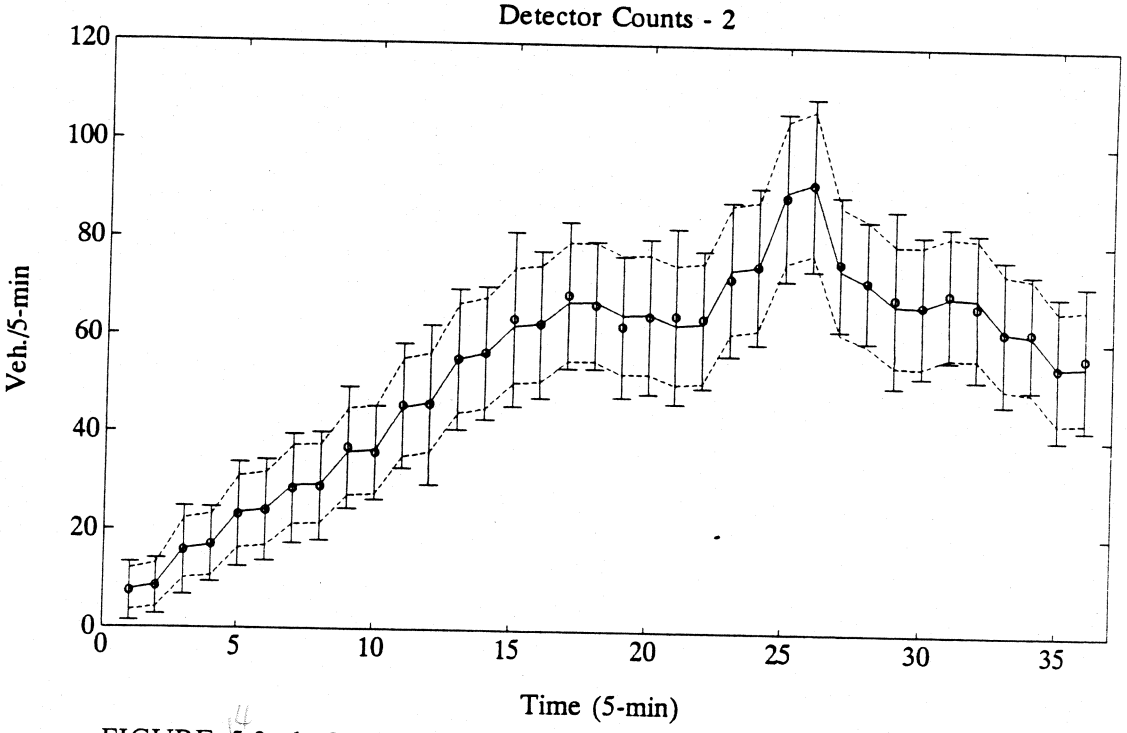
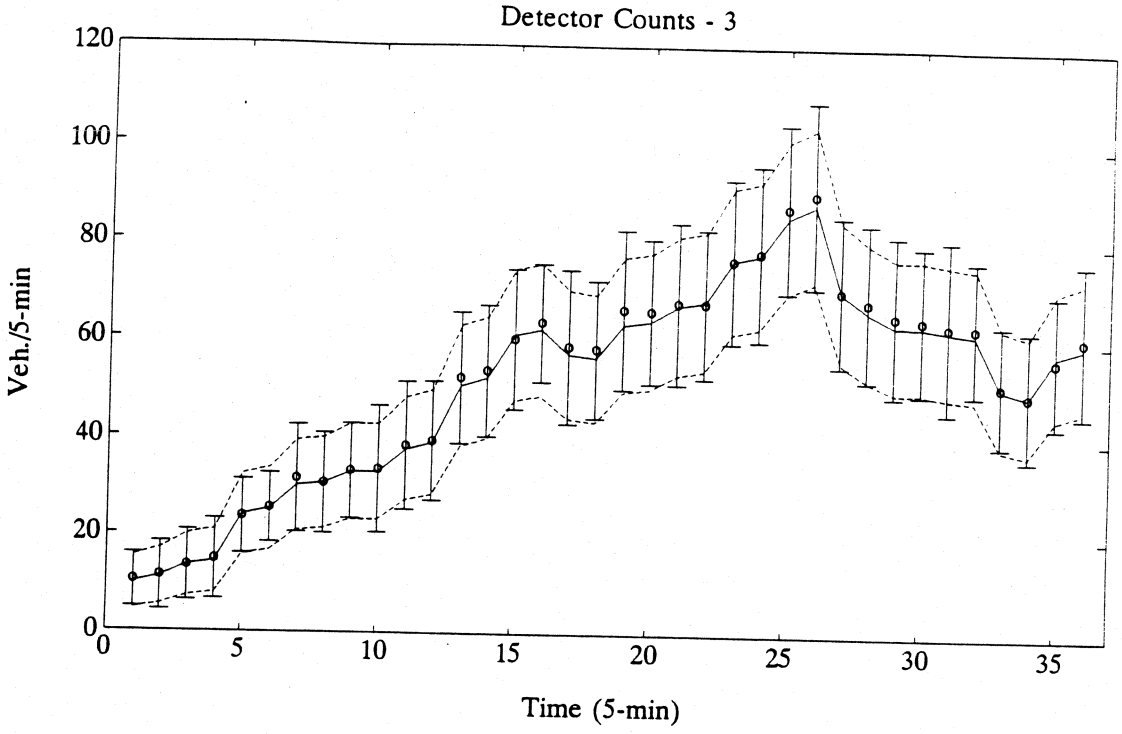


FIGURE 5.3a-b Quality of LP Approximation to Nonlinear Processes



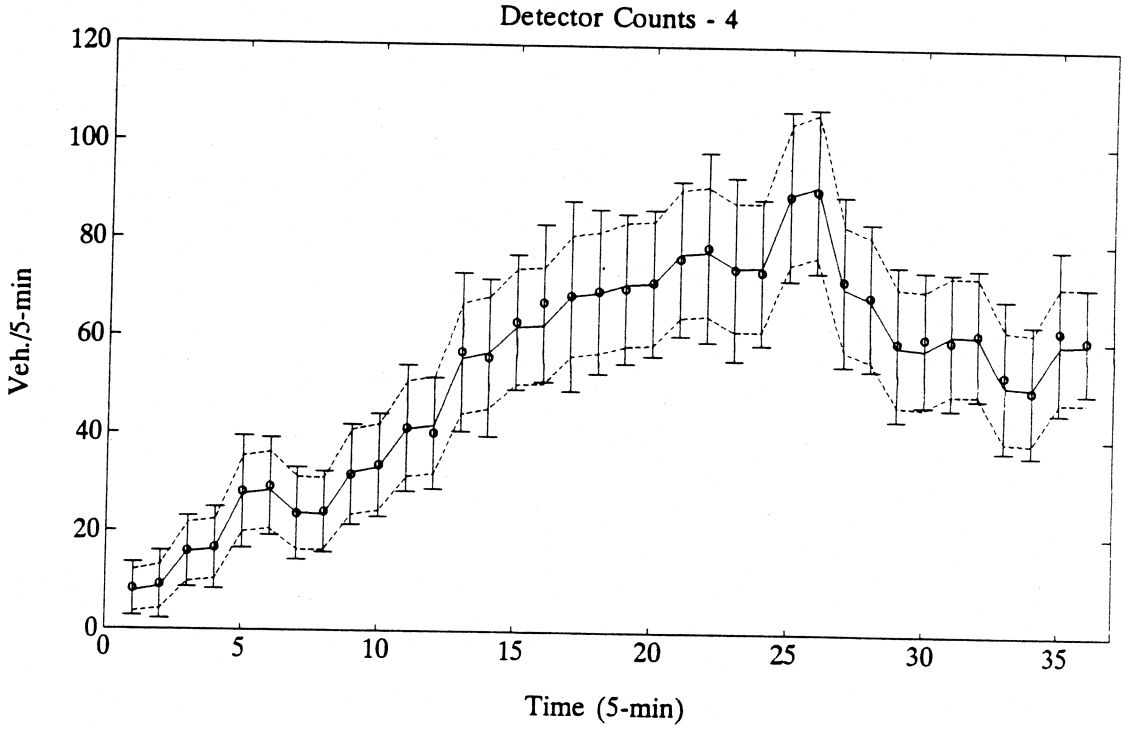
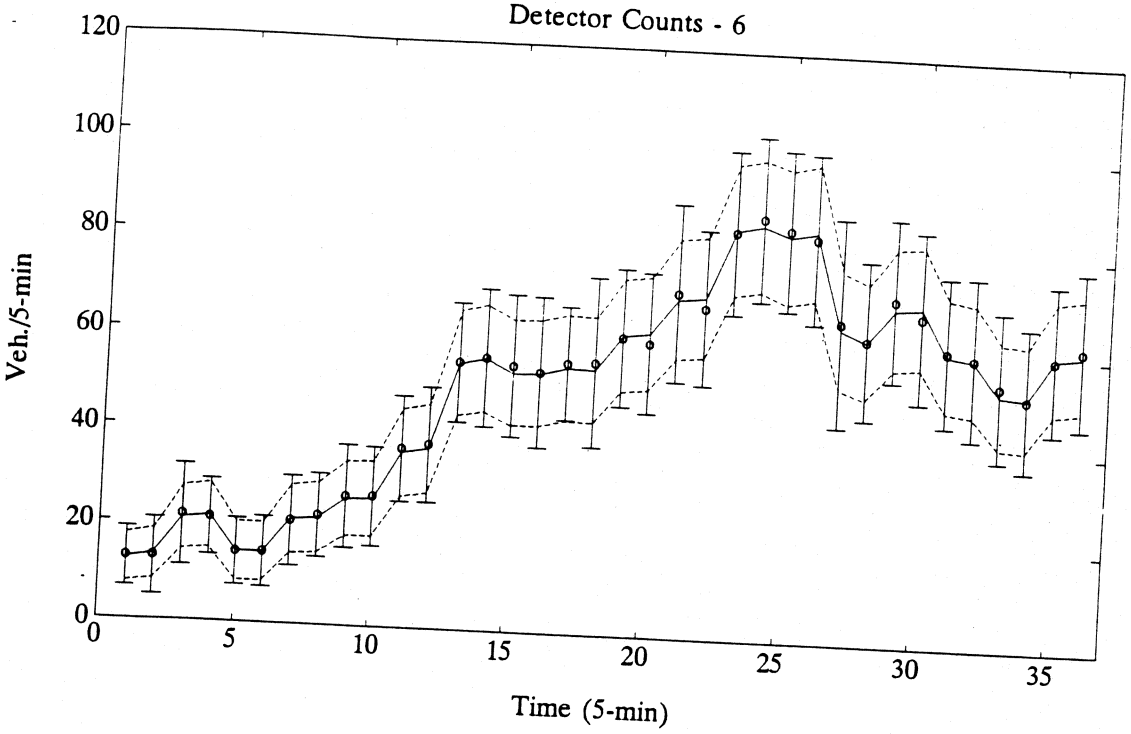


FIGURE 5.3c-d Quality of LP Approximation to Nonlinear Processes



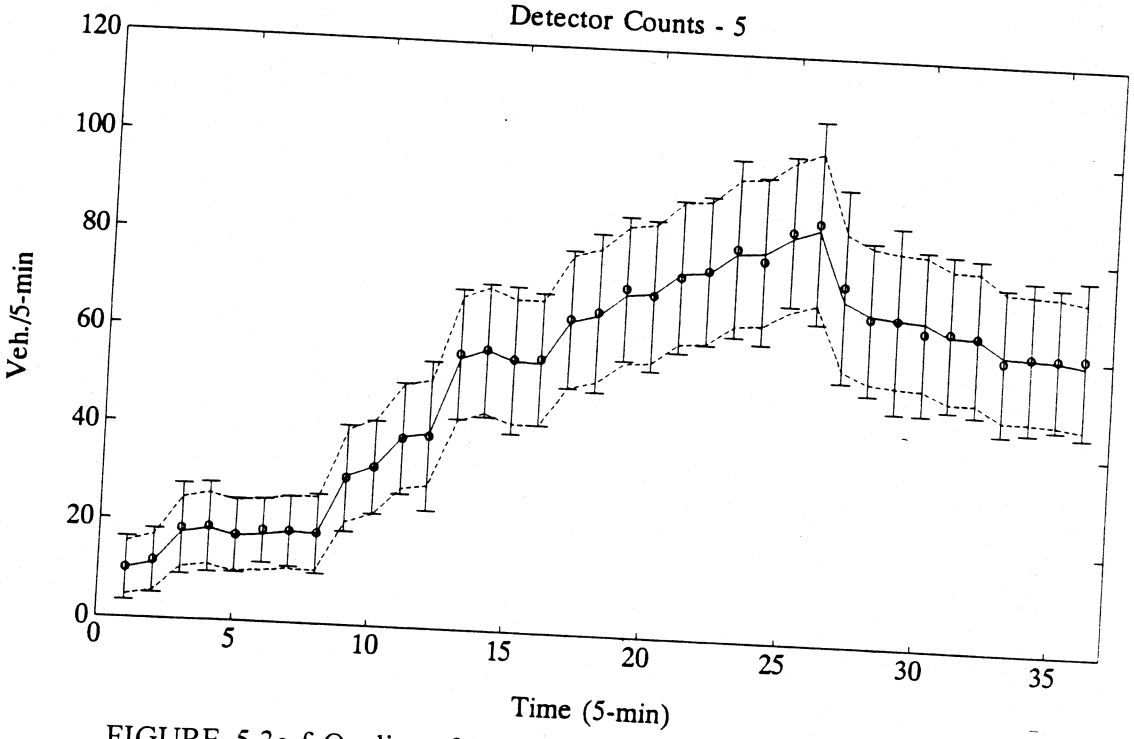


FIGURE 5.3e-f Quality of LP Approximation to Nonlinear Processes

test the hypotheses: $E[\hat{b}_{kl}]=b^0_{kl}$, where b^0_{kl} represents a true value of the turning movement proportion used in generating the simulated data. These were set equal to 0.3, 0.6, and 0.1 for left-turn, through, and right-turn movements. The χ^2_2-statistic is used to test the normality for the estimates. (Bowman and Shenton, 1975) The asterisk "*" in the Tables denotes rejection at the 5% level of significance.

The results for the CLS estimates are consistent with those reported in Nihan and Davis (1989), being unbiased with moderately low standard deviations. As might be expected, the off-line NLS and QML estimates using cordon counts show increases in both bias and variance, probably due to the fact that these estimators work with less information than the CLS estimator. This indicates that the cost savings from using less dense detector placement may be offset by a reduced ability to estimate the turning movement proportions. As a comparison, the NLS estimator is unbiased with higher variance than the QML estimator, while the QML estimator shows some biasedness with lower total RMSE, due to an increase in bias being traded for a decrease in variance. Like the case here, an unbiased estimator with dominating bias does not guarantee that the mean square error is dominated. Often, the ML-based estimator is regarded as a better estimator than LS-based estimator because of its smaller mean square error contributed by the smaller variance (greater efficiency). As shown in Table 4.2, the χ^2 test suggests that almost all of the estimates tend to be normally distributed. ($\chi^2_{2,.95}=5.99$) In terms of computation time, the QML estimator takes more than triple the CPU time to complete the mission, while the average number of function evaluations required is slightly less than the NLS estimator.

TABLE 4.1 Statistical Properties of CLS, NLS, and QML Estimators

Para-meter	True Value	CLS			NLS			QML		
		Bias	S.D.	RMSE	Bias	S.D.	RMSE	Bias	S.D.	RMSE
$b_{1,6}$	0.3	.0018	.0616	.0616	-.0052	.0561	.0564	.0056	.0557	.0560
$b_{1,7}$	0.6	.0028	.0305	.0306	-.0031	.0597	.0597	-.0211	.0553	.0592
$b_{5,3}$	0.3	.0012	.0358	.0358	-.0041	.0648	.0649	.0061	.0524	.0527
$b_{5,4}$	0.6	.0086	.0355	.0366	-.0024	.0452	.0452	-.0302	.0343	.0457
$b_{9,1}$	0.3	.0012	.0346	.0346	-.0057	.0680	.0682	-.0089	.0531	.0538
$b_{9,3}$	0.6	-.0002	.0350	.0350	.0106	.0401	.0415	.0128	.0354	.0376
$b_{10,4}$	0.3	.0025	.0542	.0542	-.0033	.0468	.0469	-.0098	.0436	.0446
$b_{10,2}$	0.6	.0001	.0312	.0312	-.0003	.0338	.0338	-.0020	.0310	.0311
$b_{11,2}$	0.3	-.0034	.0559	.0561	-.0090	.0570	.0577	-.0287	.0502	.0578
$b_{11,1}$	0.6	-.0051	.0547	.0549	.0122	.0647	.0658	.0557	.0440	.0710
$b_{12,7}$	0.3	-.0003	.0309	.0309	-.0051	.0430	.0433	-.0096	.0365	.0377
$b_{12,8}$	0.6	-.0016	.0384	.0384	-.0003	.0386	.0386	-.0009	.0391	.0391
$b_{13,8}$	0.3	-.0041	.0376	.0378	-.0068	.0415	.0421	-.0192	.0337	.0388
$b_{13,5}$	0.6	.0003	.0402	.0402	.0063	.0544	.0547	.0437	.0491	.0658
$b_{14,5}$	0.3	-.0061	.0330	.0336	-.0041	.0525	.0527	.0059	.0425	.0429
$b_{14,6}$	0.6	.0012	.0448	.0449	.0034	.0484	.0485	-.0027	.0458	.0459
Total RMSE	-	.0034	.0420	.0422	.0060	.0519	.0522	.0225	.0446	.0499
No of F.E.		-			762			642		
CPU Time Ratio	-	-			1			3.67		

TABLE 4.2 Biases and Normality Tests of CLS, NLS, and QML Estimators

Parameter	True Value	CLS		NLS		QML	
		t-stat	χ^2-stat	t-stat	χ^2-stat	t-stat	χ^2-stat
$b_{1,6}$	0.3	0.20	2.33	0.64	0.84	0.71	0.61
$b_{1,7}$	0.6	0.63	5.19	0.37	1.66	2.67	0.75
$b_{5,3}$	0.3	0.23	18.85*	0.44	2.48	0.82	2.11
$b_{5,4}$	0.6	1.69	1.10	0.37	0.73	6.17*	0.42
$b_{9,1}$	0.3	0.25	3.36	0.59	0.18	1.17	5.36
$b_{9,3}$	0.6	0.04	0.81	1.85	1.05	2.52*	0.44
$b_{10,4}$	0.3	0.33	3.28	0.50	0.07	1.57	0.46
$b_{10,2}$	0.6	0.01	0.31	0.06	4.41	0.45	26.94*
$b_{11,2}$	0.3	0.43	0.13	1.10	1.28	4.01*	1.38
$b_{11,1}$	0.6	0.66	3.79	1.32	1.27	8.86*	1.08
$b_{12,7}$	0.3	0.06	0.55	0.84	2.41	1.85	2.25
$b_{12,8}$	0.6	0.29	0.56	0.05	0.13	0.17	0.64
$b_{13,8}$	0.3	0.76	1.32	1.15	0.34	3.97*	1.40
$b_{13,5}$	0.6	0.06	1.73	0.81	0.12	6.23*	1.80
$b_{14,5}$	0.3	1.30	0.63	0.55	1.83	0.97	1.73
$b_{14,6}$	0.6	0.19	0.22	0.49	2.96	0.41	1.96

4.4.4 *The Recursive Estimators*

A more interesting issue in practice, particularly for real-time control, is how these algorithms can be applied to track the changes of the parameters over time, and how fast they can adapt to these changes. To answer the second question, the same simulation data set as in the off-line estimation is used to execute the recursive algorithms RNLS and RQML with different initial values: for $\mathbf{b}_{LT}(0) = 0.5$ and $\mathbf{b}_{TH}(0) = 0.4$. Note that since the parameter is time invariant in this experiment, one may assume a constant model for parameter transition, i.e., $\mathbf{b}_{t+1}=\mathbf{b}_t$, in RQML. In terms of mean values, both algorithms take about 20 iterations to finally converge to the true values. Their performances, however, are distinguished by the evolution of the variance bounds. In the entire course of recursion, RQML outperforms RNLS with tighter bounds, indicating that RQML is a more efficient algorithm and so may be more useful in practice. Figures 4.4a-p contain all 16 parameters' recursive estimates over 72 time intervals, where solid lines are 50-sample means and ±2 standard deviations for RQML, and dashed lines are for RNLS.

Both algorithms described above have one common characteristic, that the algorithm's gain $\mathbf{K}(t)$ decreases significantly after a certain number of iterations. Thereafter, the adaptation of estimates is insensitive to changes in the parameter values. To be able to track the changes in parameters over time, such as stepwise changes, the recursive algorithms need to be modified. There are many possible alternatives. For RNLS, one efficient way is to incorporate a discount factor in the criterion function as follows

$$\min \sum_{t=1}^{T} \left(\boldsymbol{y}(t)-\hat{\boldsymbol{y}}(\boldsymbol{b},t)\right)'\left(\boldsymbol{y}(t)-\hat{\boldsymbol{y}}(\boldsymbol{b},t)\right)\boldsymbol{d}^{T-t}, \tag{4.17}$$

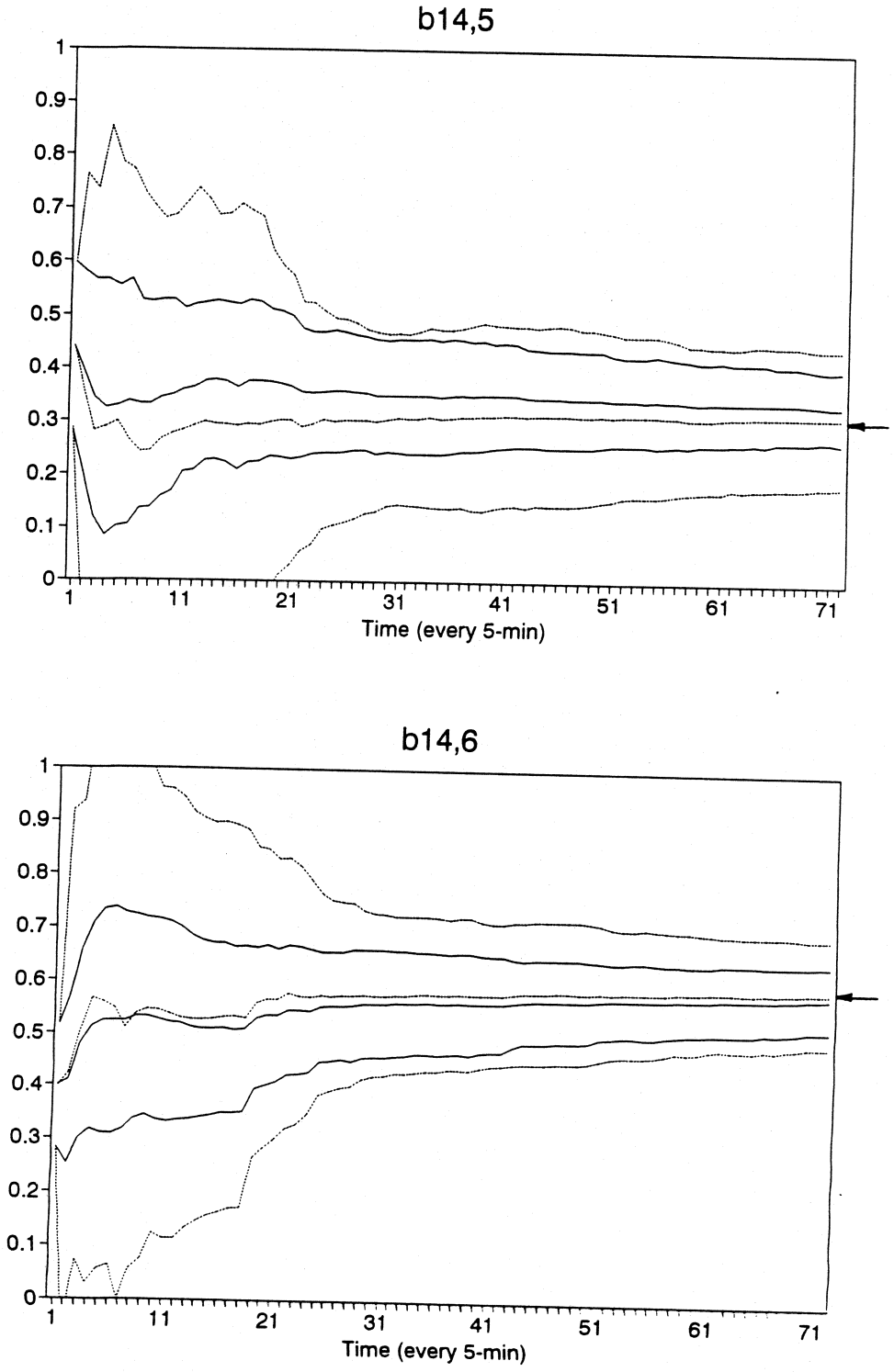
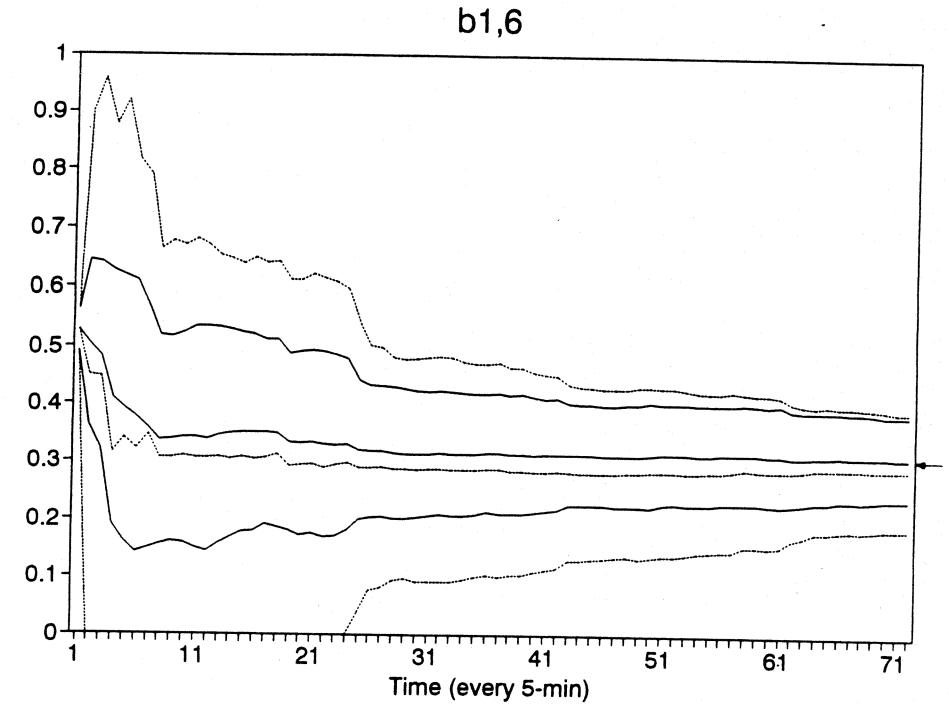


FIGURE 5.4o-p Recursive Estimates of Constant Turning Proportion Parameters



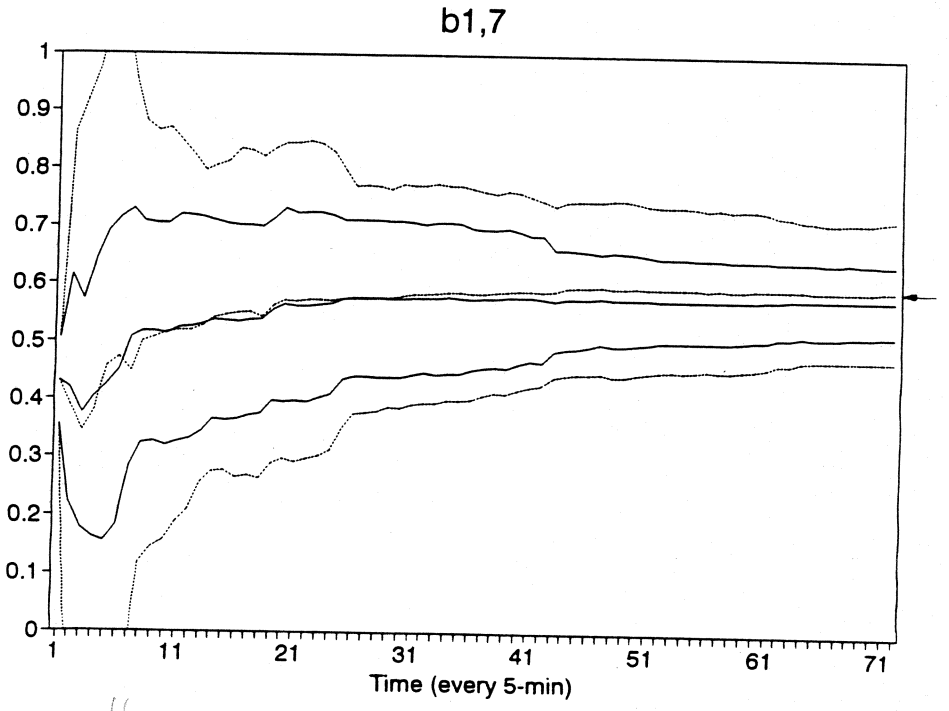
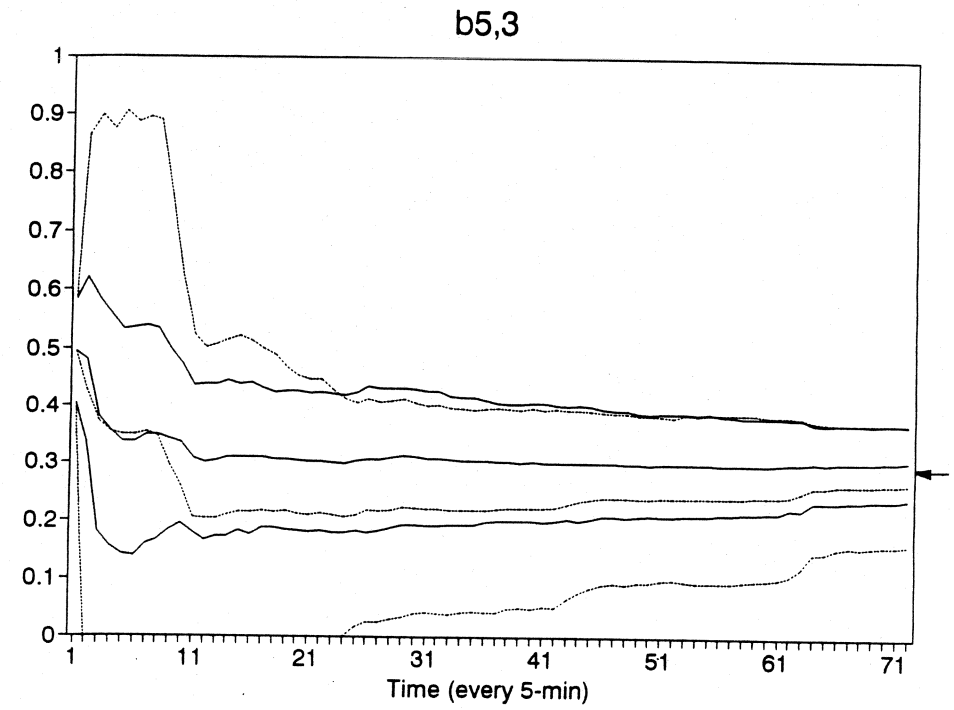


FIGURE 5.4a-b Recursive Estimates of Constant Turning Proportion Parameters



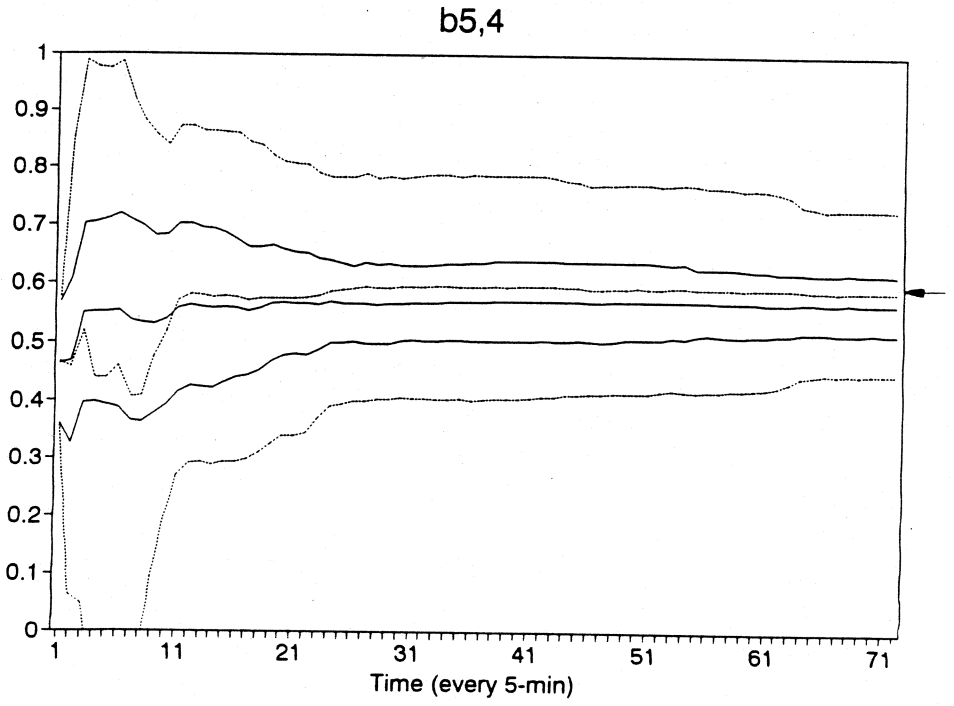
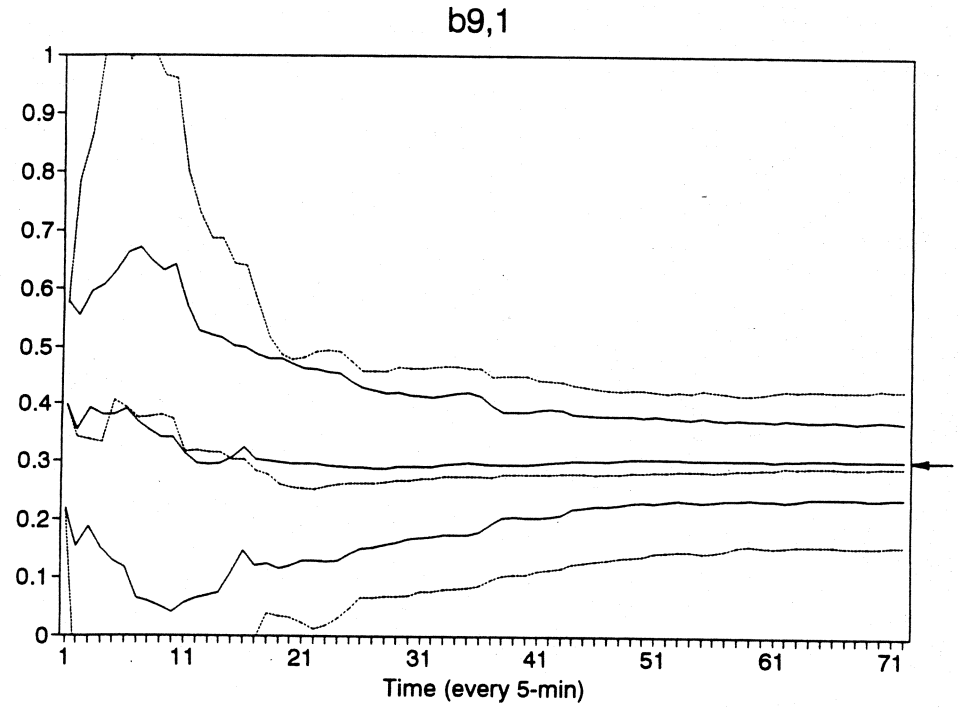


FIGURE 5.4c-d Recursive Estimates of Constant Turning Proportion Parameters



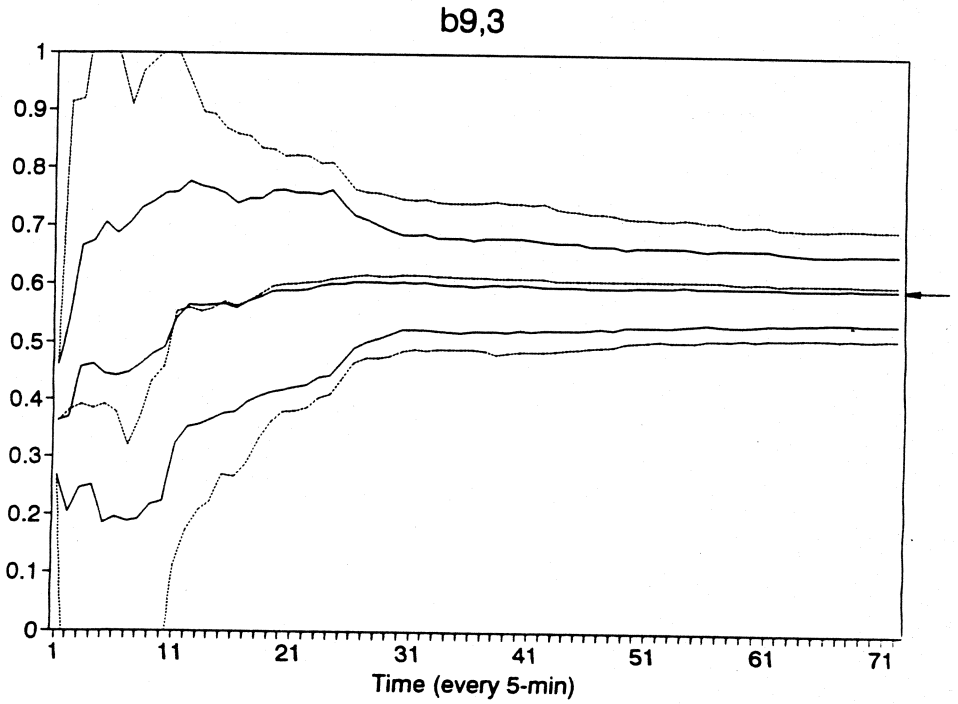
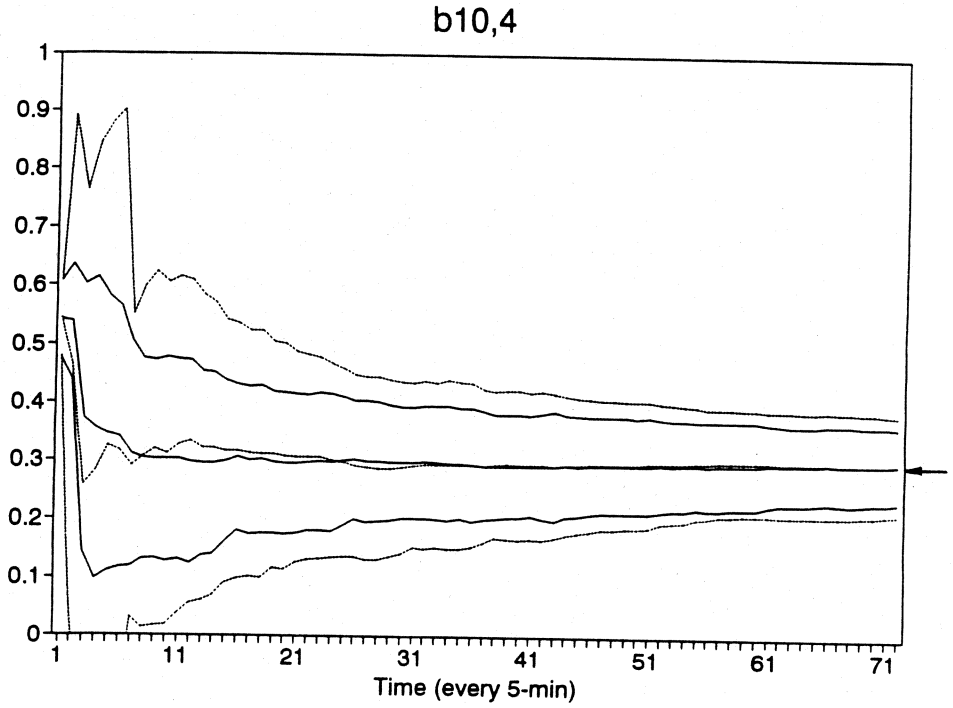


FIGURE 5.4e-f Recursive Estimates of Constant Turning Proportion Parameters



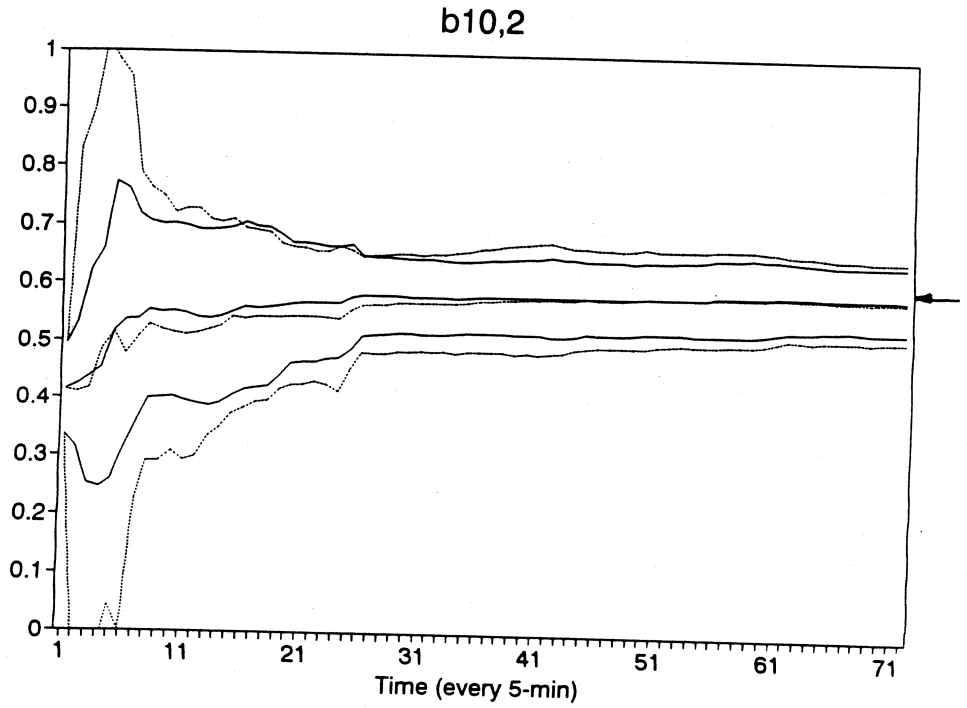
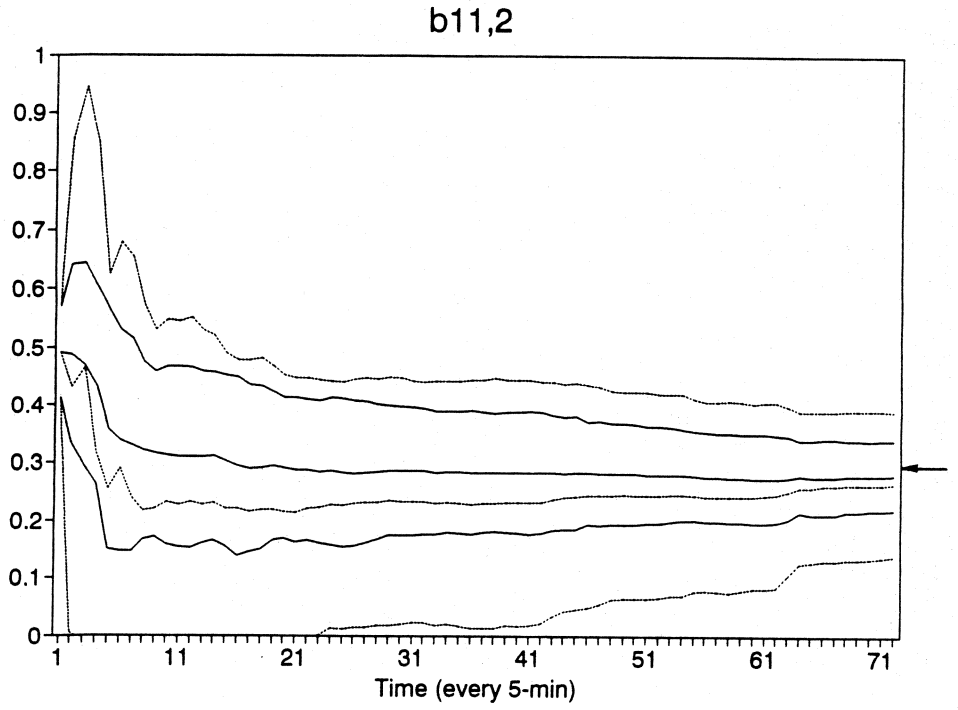


FIGURE 5.4g-h Recursive Estimates of Constant Turning Proportion Parameters



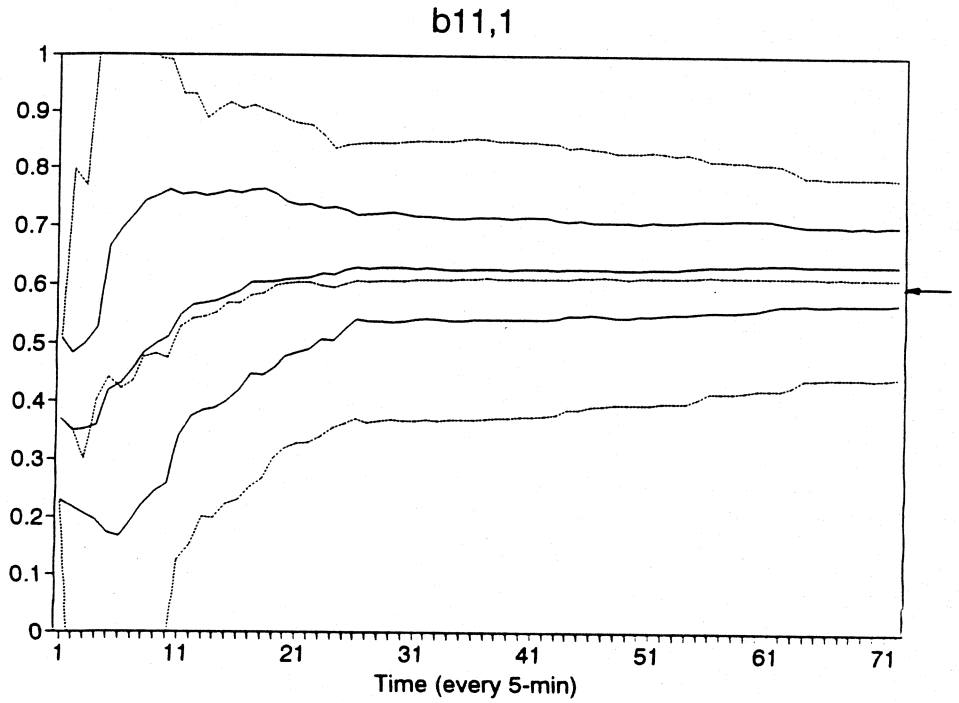


FIGURE 5.4i-j Recursive Estimates of Constant Turning Proportion Parameters

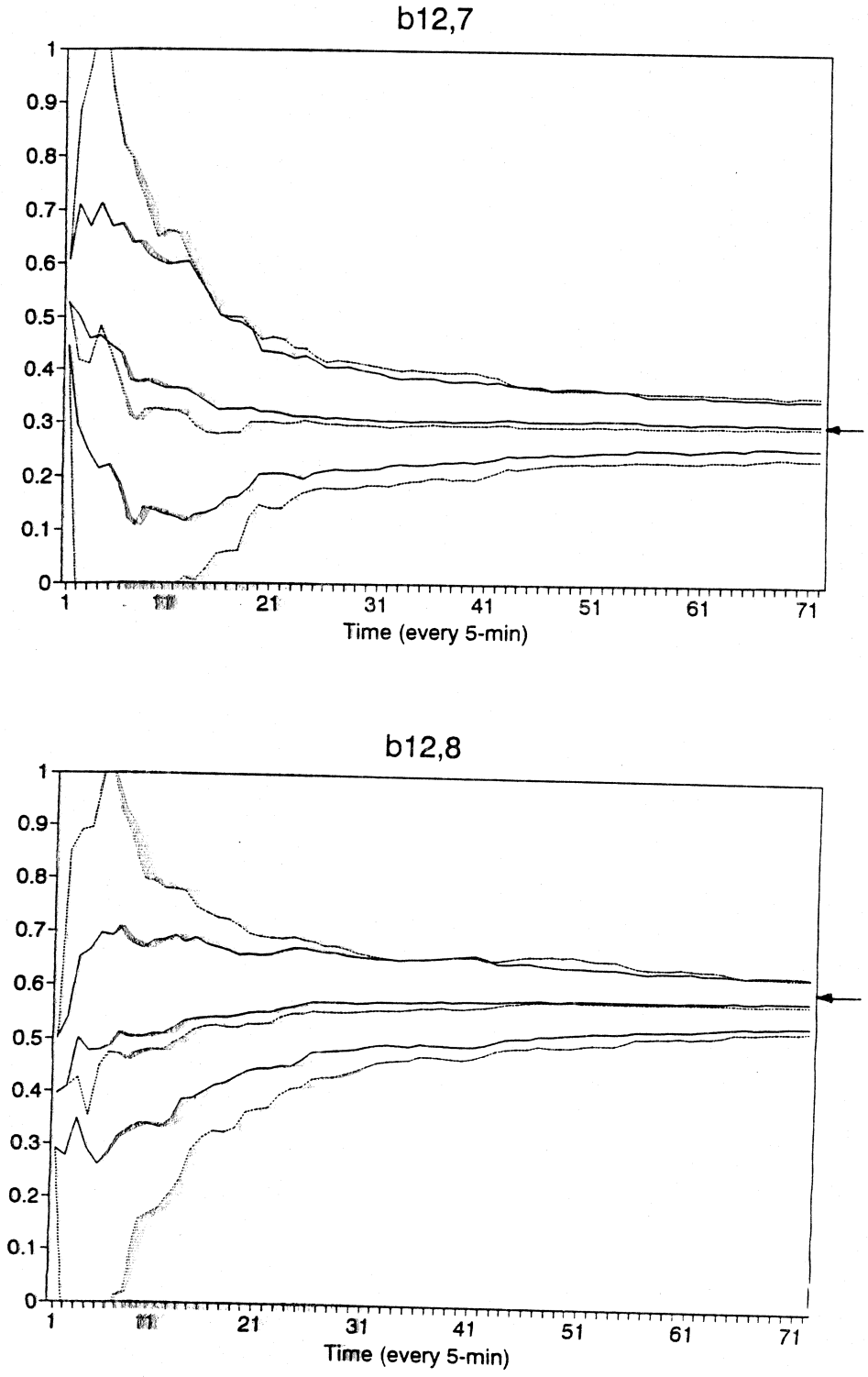
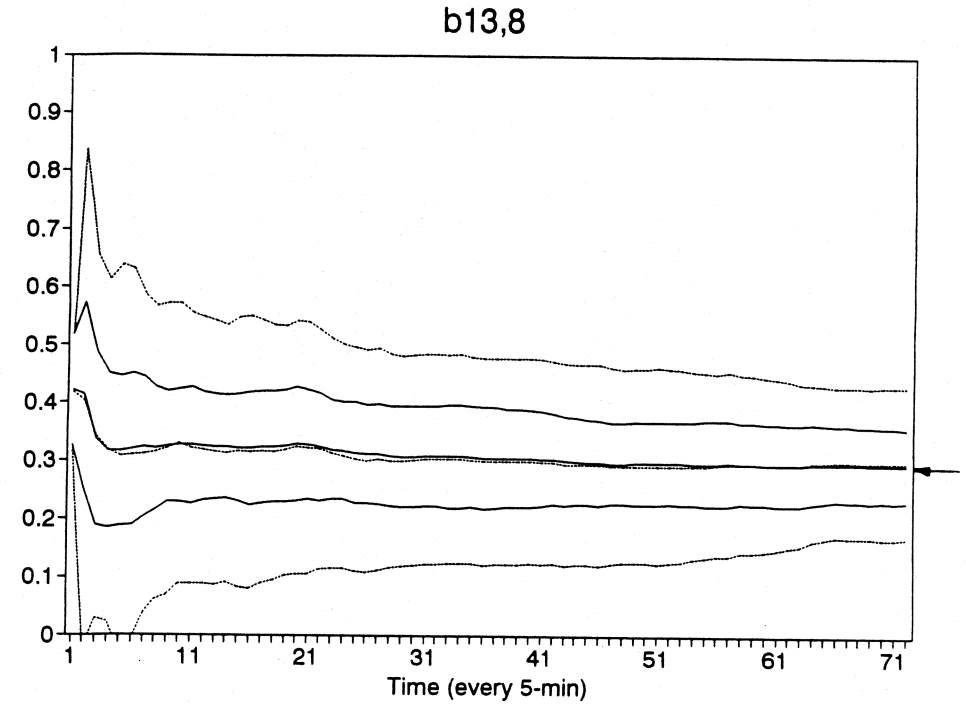


FIGURE 5.4-1 Recursive Estimates of Constant Turning Proportion Parameters



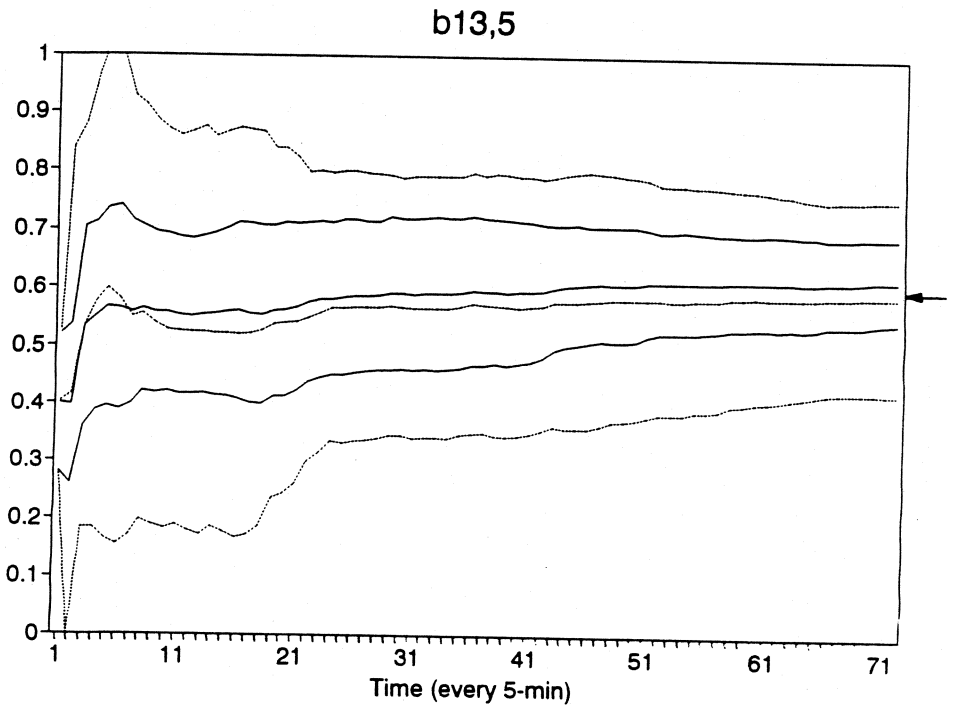


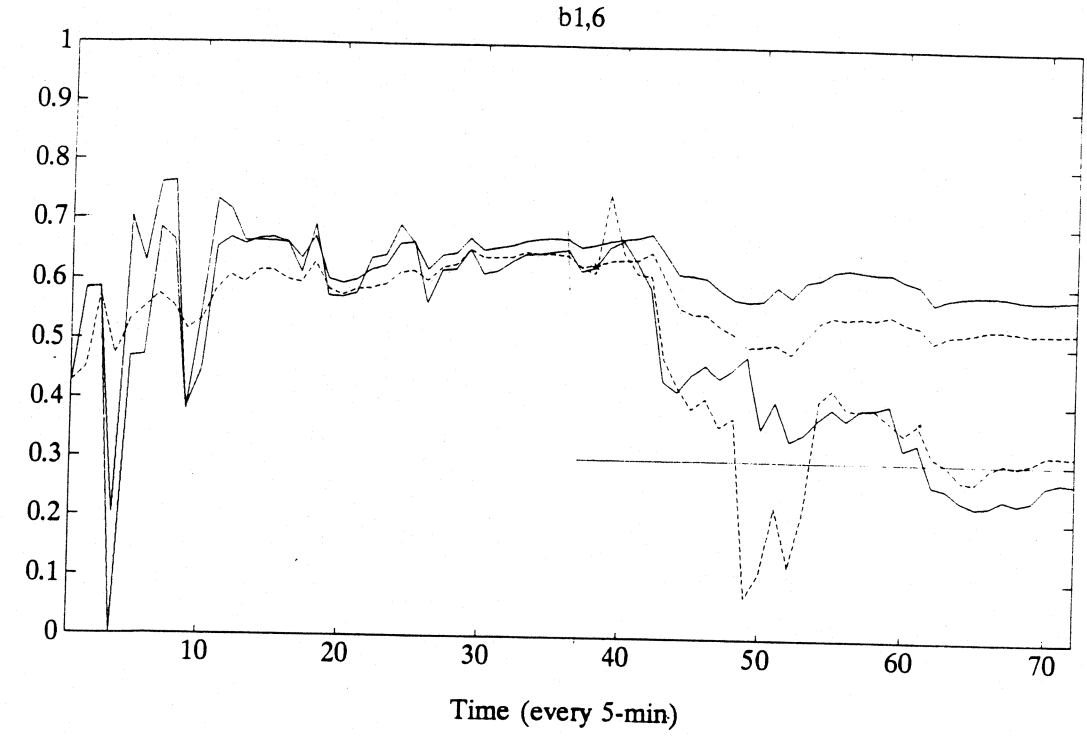
FIGURE 5.4m-n Recursive Estimates of Constant Turning Proportion Parameters

allowing heavier weights to be placed on more recent measurements, with the effects of previous measurements on the criterion being discounted at an exponential rate. The d is a design parameter remains between 0 and 1. The resulting modified RNLS algorithm produces the following measurement-update step

$$\begin{aligned} \boldsymbol{K}(t) &= \boldsymbol{P}(t-1)\phi(t)'\boldsymbol{H}'(d\boldsymbol{I}+\boldsymbol{H}\phi(t)\boldsymbol{P}(t-1)\phi(t)'\boldsymbol{H}')^{-1} \\ \hat{\boldsymbol{b}}(t) &= \hat{\boldsymbol{b}}(t-1)+\boldsymbol{K}(t)[y(t)-\hat{y}(\boldsymbol{b},t)] \\ \boldsymbol{P}(t) &= (\boldsymbol{I}-\boldsymbol{K}(t)\boldsymbol{H}\phi(t))\boldsymbol{P}(t-1)/d \end{aligned} \tag{4.18}$$

For the QML estimator, two common schemes, as discussed in Goodwin and Sin (1984), are (1) covariance resetting, and (2) covariance modification. The first scheme is simply to reset **P**(t) when one suspects that a change in parameters has occurred. This will require detection of parameter changes, for timely resetting or for the resetting to not be triggered at the wrong time due to a "false alarm." Based on a similar concept, the second scheme is to add positive terms to **P**(t) and revitalize the algorithm. One natural way is to assume the transition of the parameters follows a random walk with an appropriate covariance term. This covariance can be determined off-line by using the EM algorithm (Fahrmeir and Tutz, 1994), or be treated as a design parameter to be specified prior to implementation. The author experimented with both schemes using several simulation runs. For simplicity, it was assumed that the time when the parameters change is known. The simulation data are generated from a scenario where $\mathbf{b}_{LT}(t)=0.7$ and $\mathbf{b}_{TH}(t)=0.2$ for first 36 time intervals and $\mathbf{b}_{LT}(t)=0.3$ and $\mathbf{b}_{TH}(t)=0.6$ for next 36 time intervals. The recursions start with $\hat{\boldsymbol{b}}_{LT}(0)=0.5$ and $\hat{\boldsymbol{b}}_{TH}(0)=0.4$.

After several trials with simulation data, it was found difficult to assign appropriate covariance terms to the random walk model for RQML with covariance modification to track the parameter changes. The covariance resetting scheme seemed more effective, as long as the time when parameters change is appropriately identified. As depicted in Figures 4.5a-p, solid lines represent the RQML estimates with and without covariance resetting, and dashed lines represent the RNLS estimates with and without factor discounting. The discount factor was chosen as 0.87 after several trials. It can be seen that both modified algorithms are able to adapt to stepwise changes in parameters within a reasonable number of iterations, while the original algorithms respond much slowly. The discounted RNLS algorithm seems to exhibit smoother transition behavior than the RQML algorithm with covariance resetting after the parameters being changed.



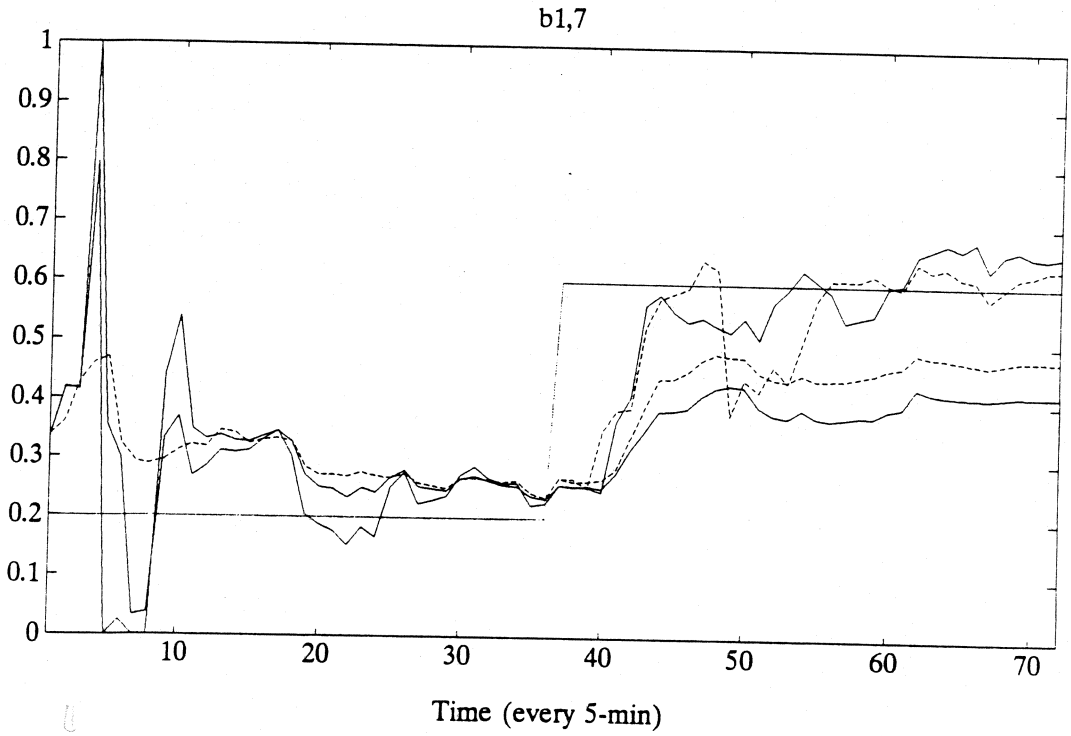
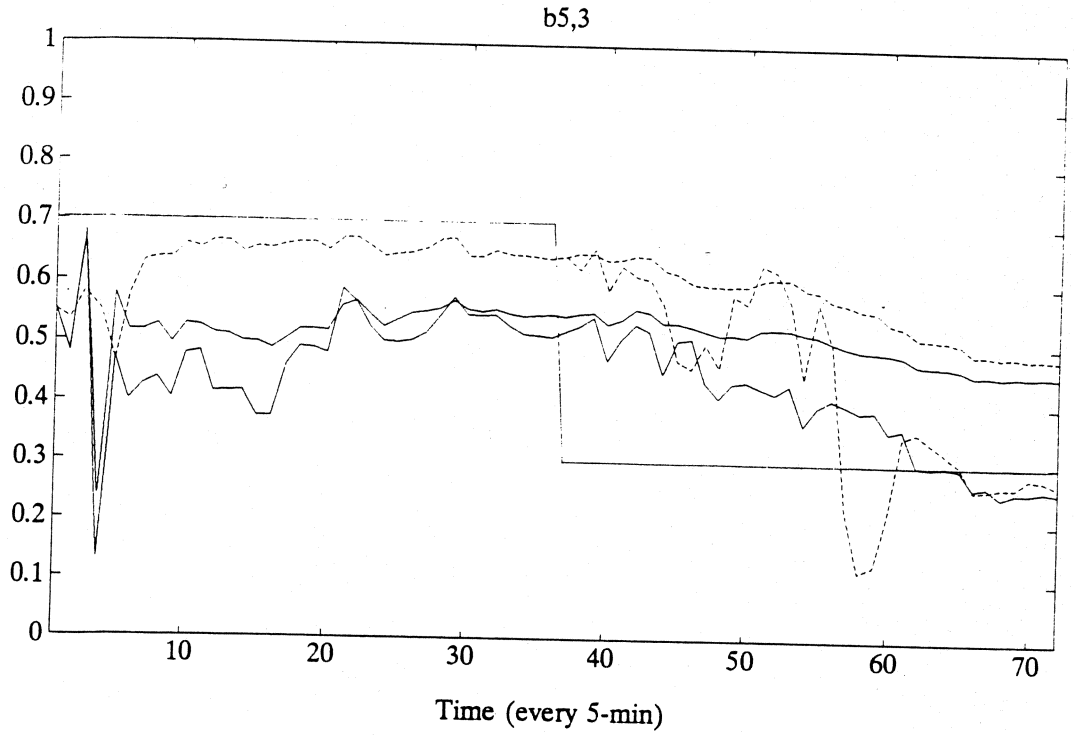


FIGURE 5.5a-b Recursive Estimates of Step-changed Turning Proportion Parameters



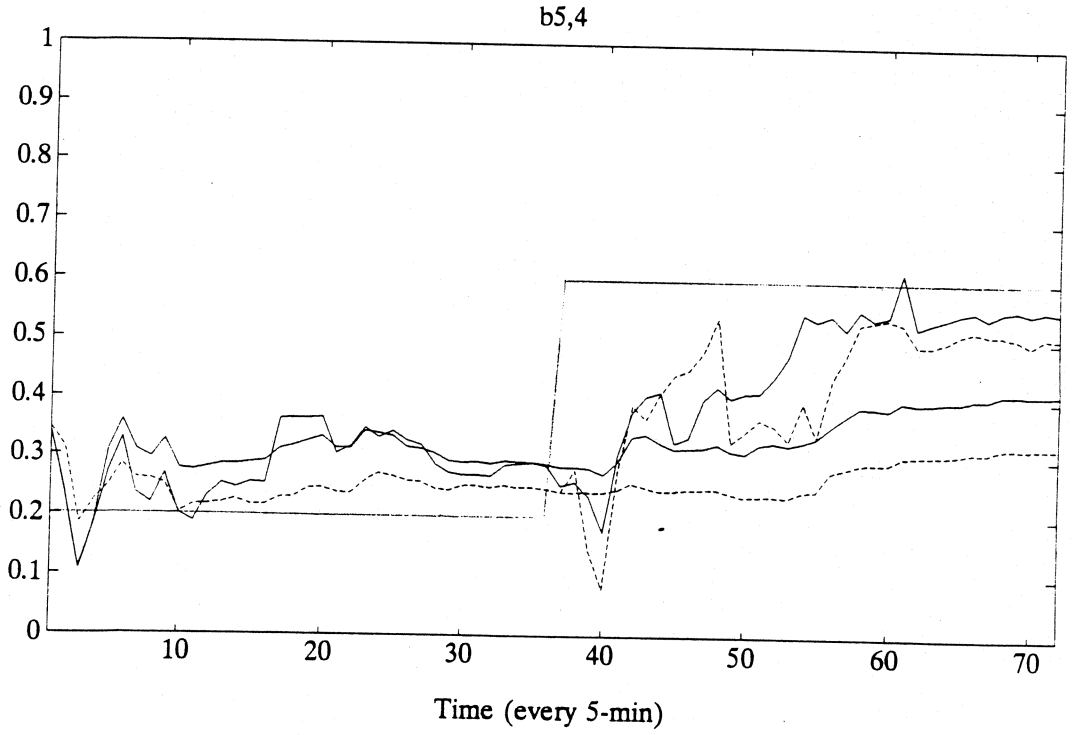
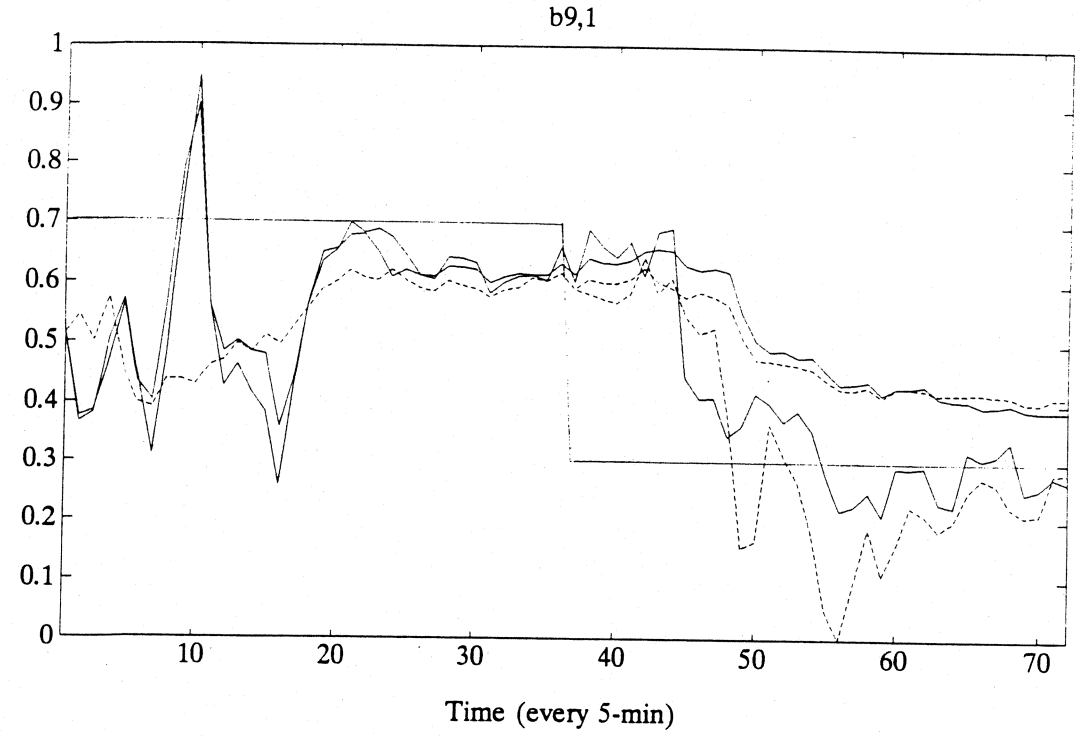


FIGURE 5.5c-d Recursive Estimates of Step-changed Turning Proportion Parameters



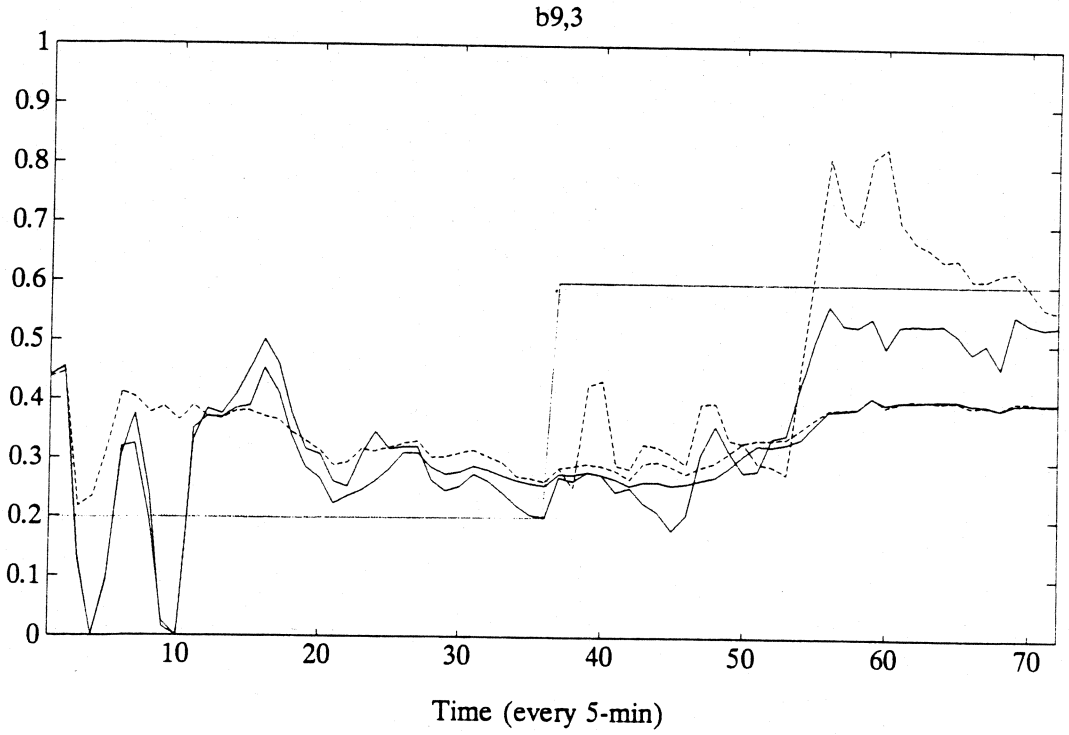
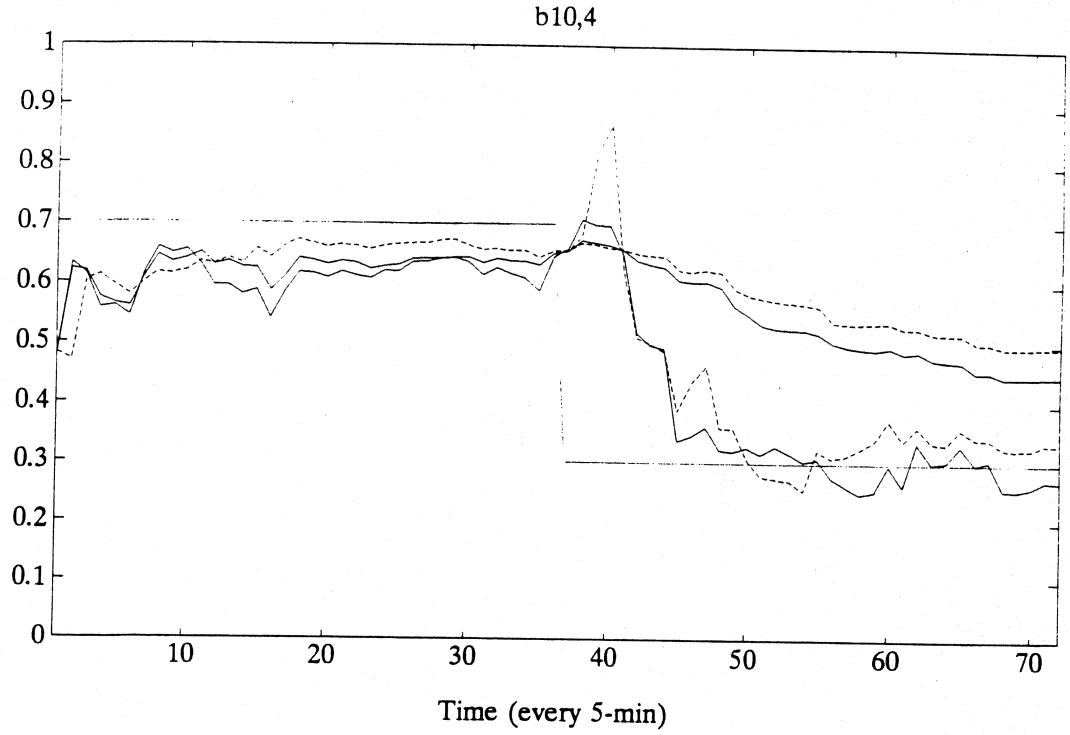


FIGURE 5.5e-f Recursive Estimates of Step-changed Turning Proportion Parameters



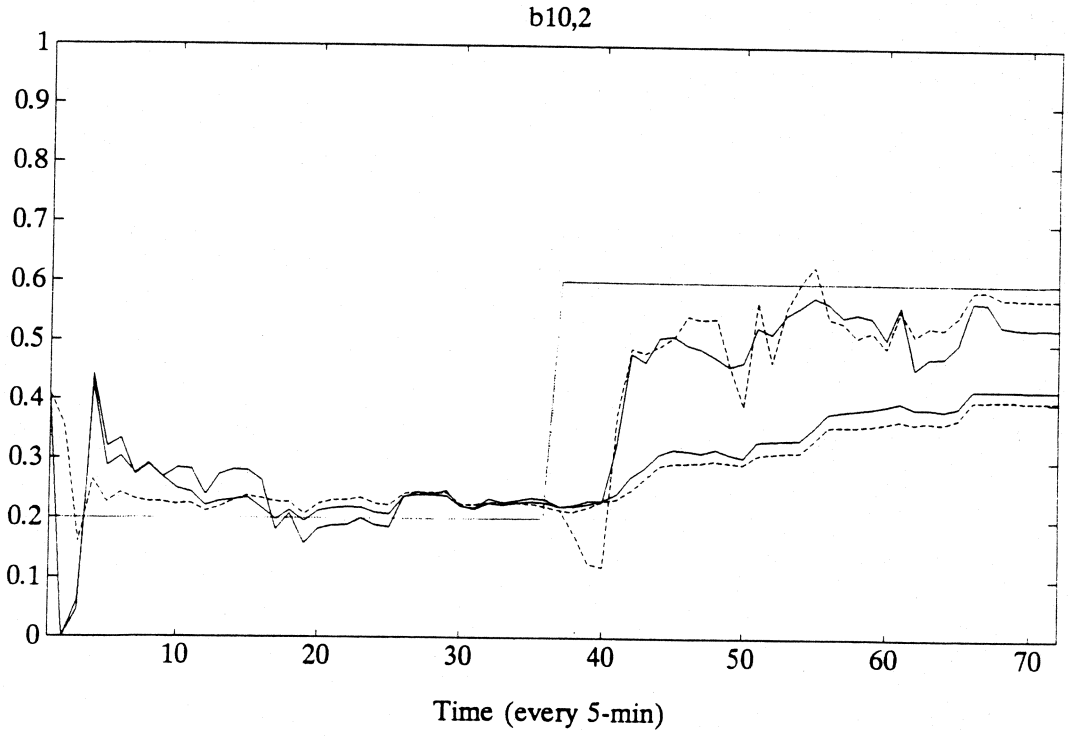
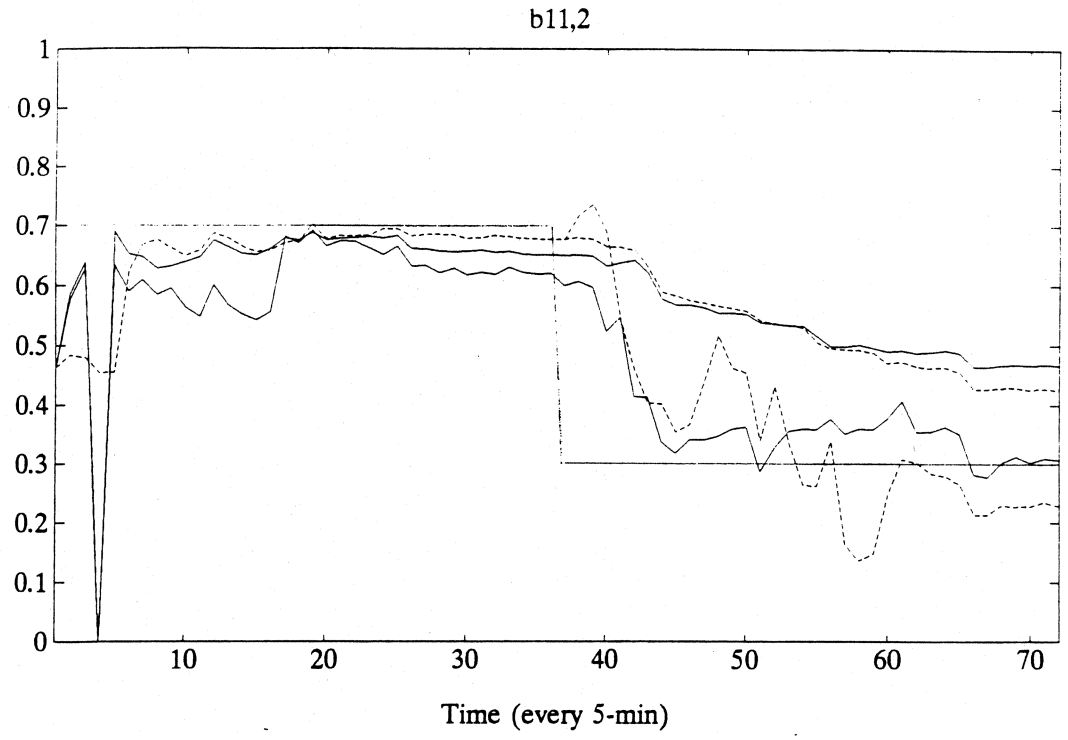


FIGURE 5.5g-h Recursive Estimates of Step-changed Turning Proportion Parameters



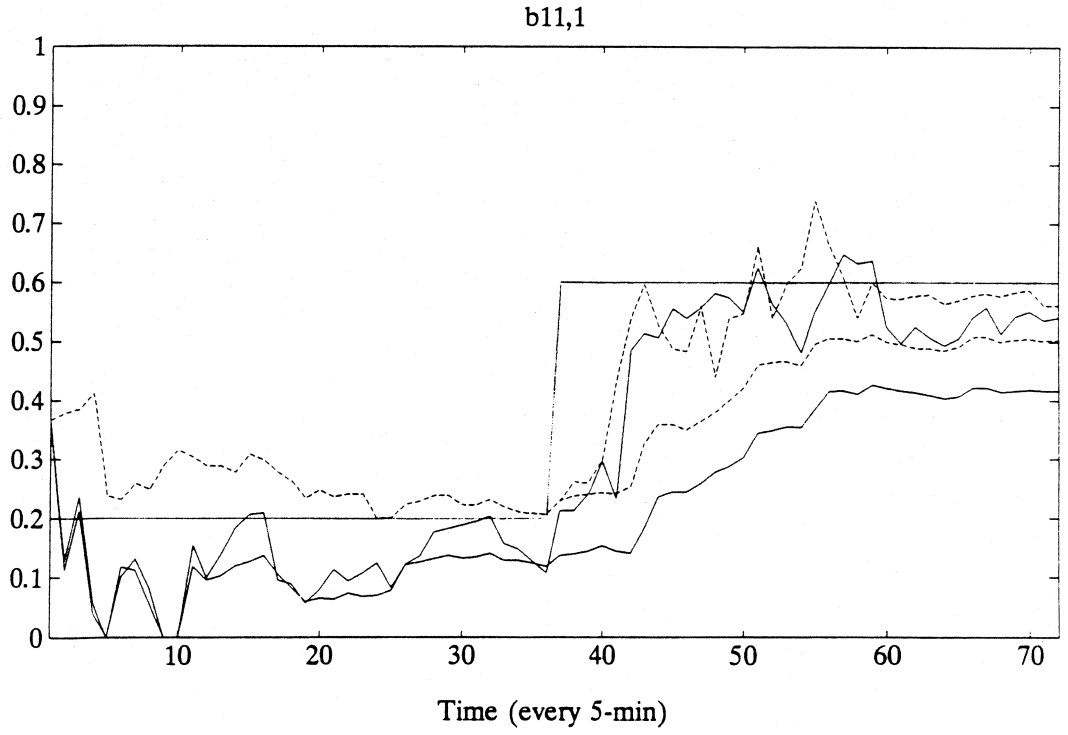
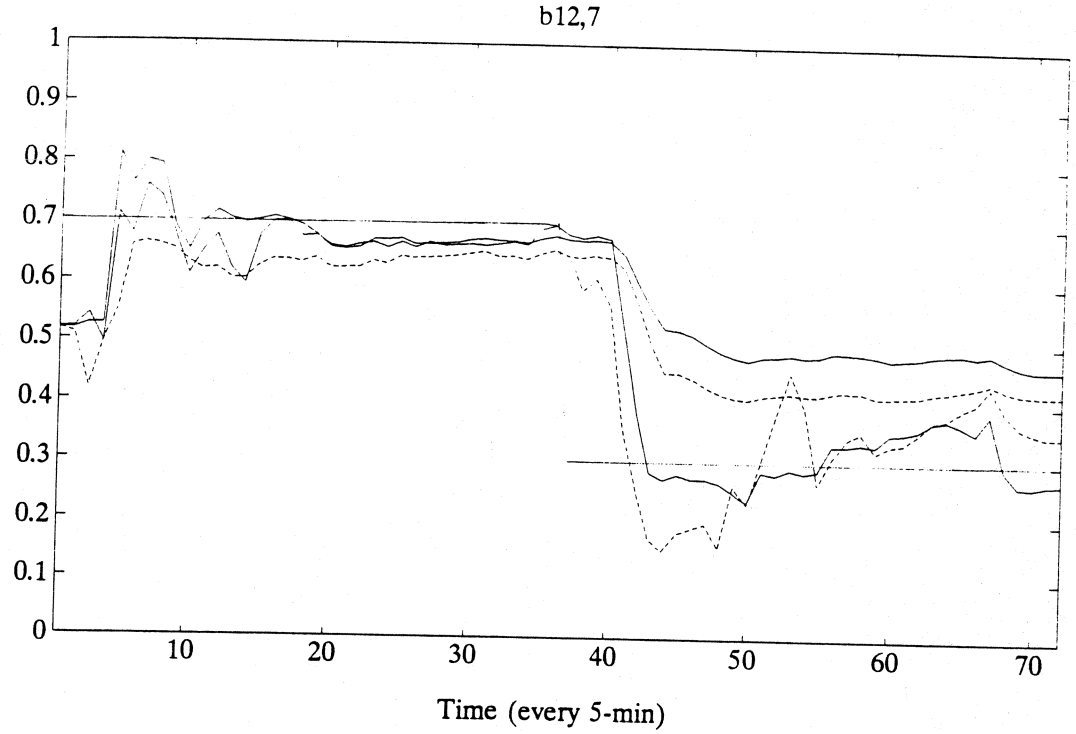


FIGURE 5.5i-j Recursive Estimates of Step-changed Turning Proportion Parameters



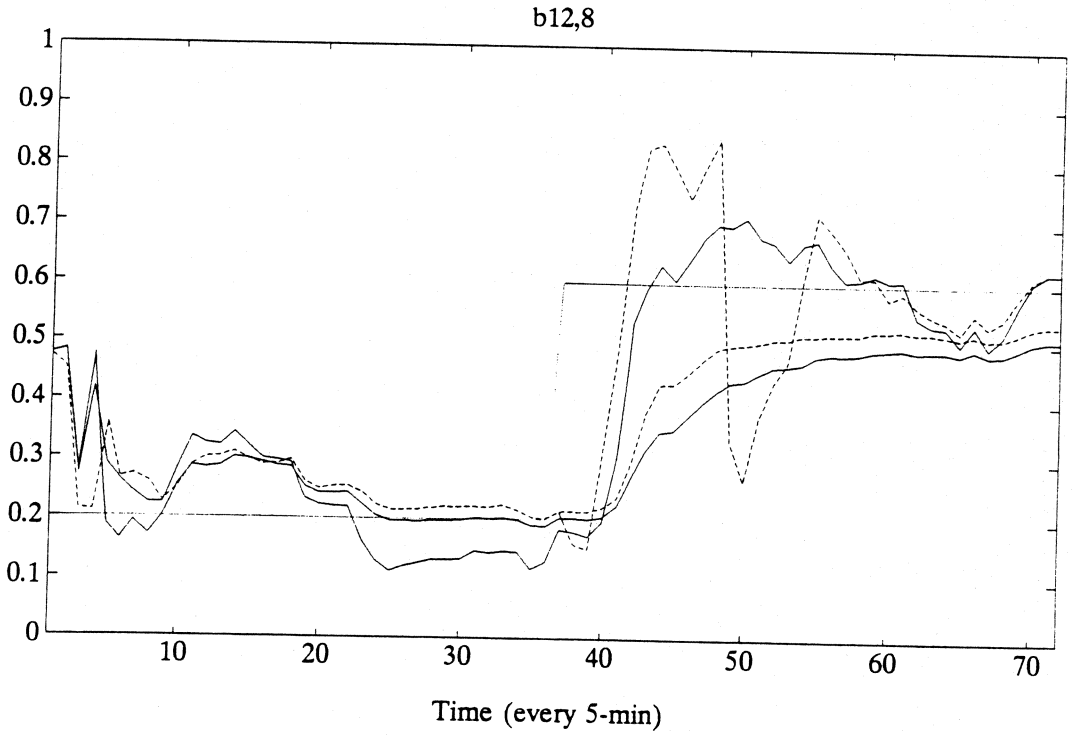
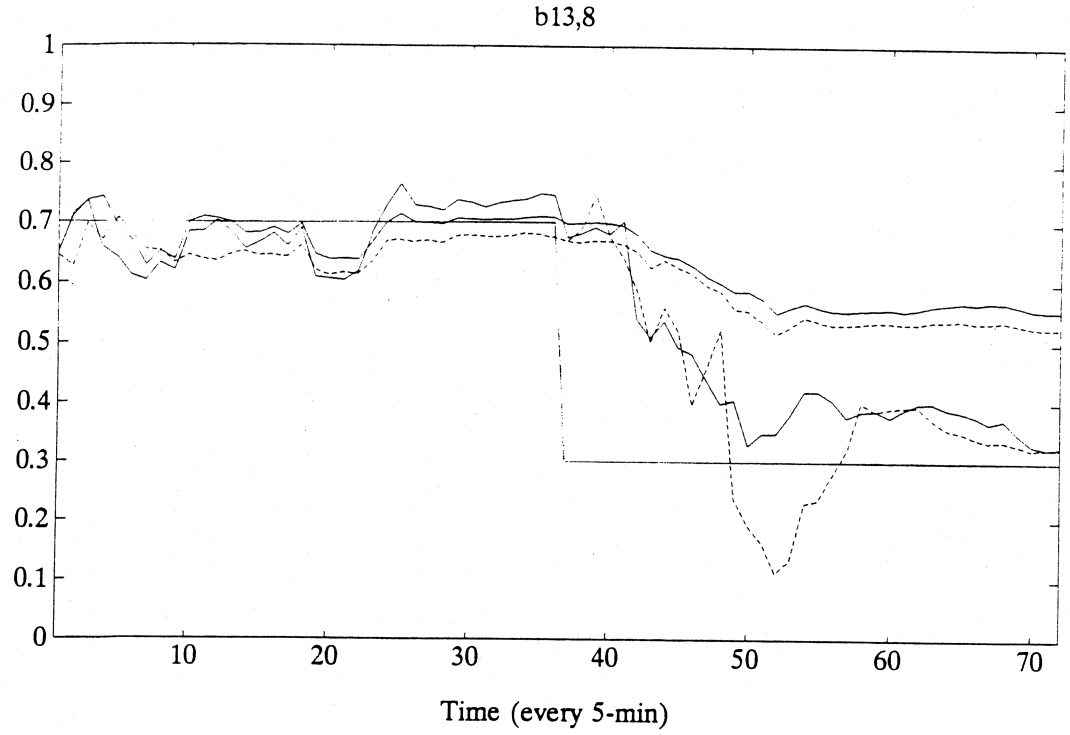


FIGURE 5.5k-l Recursive Estimates of Step-changed Turning Proportion Parameters



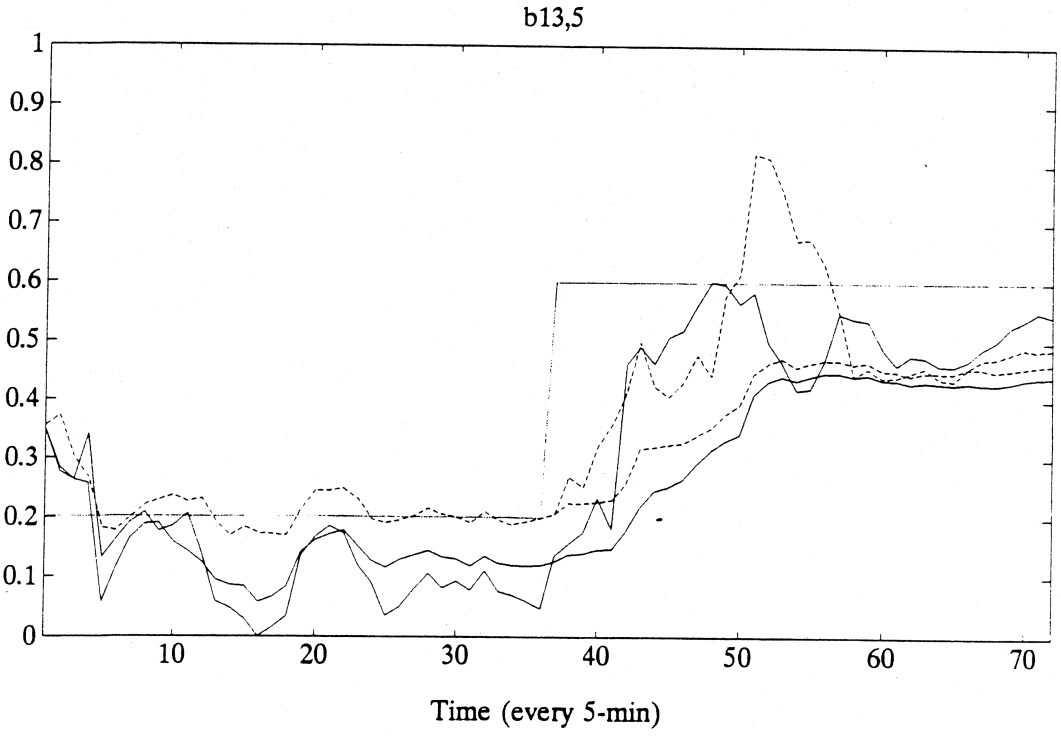
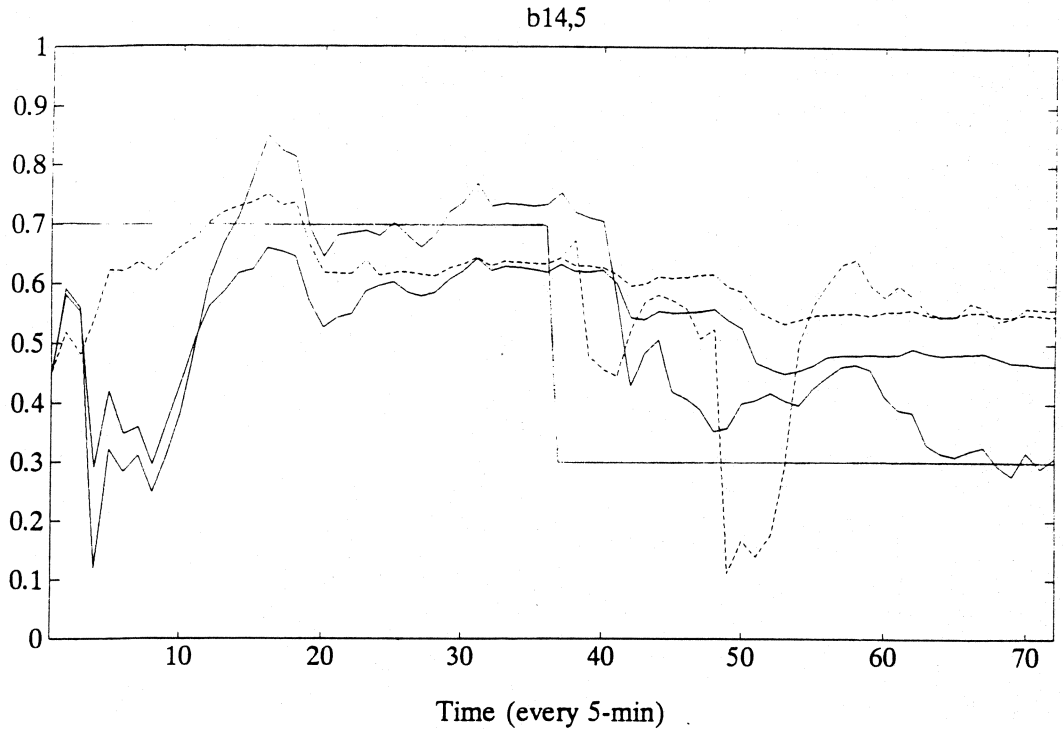


FIGURE 5.5m-n Recursive Estimates of Step-changed Turning Proportion Parameters



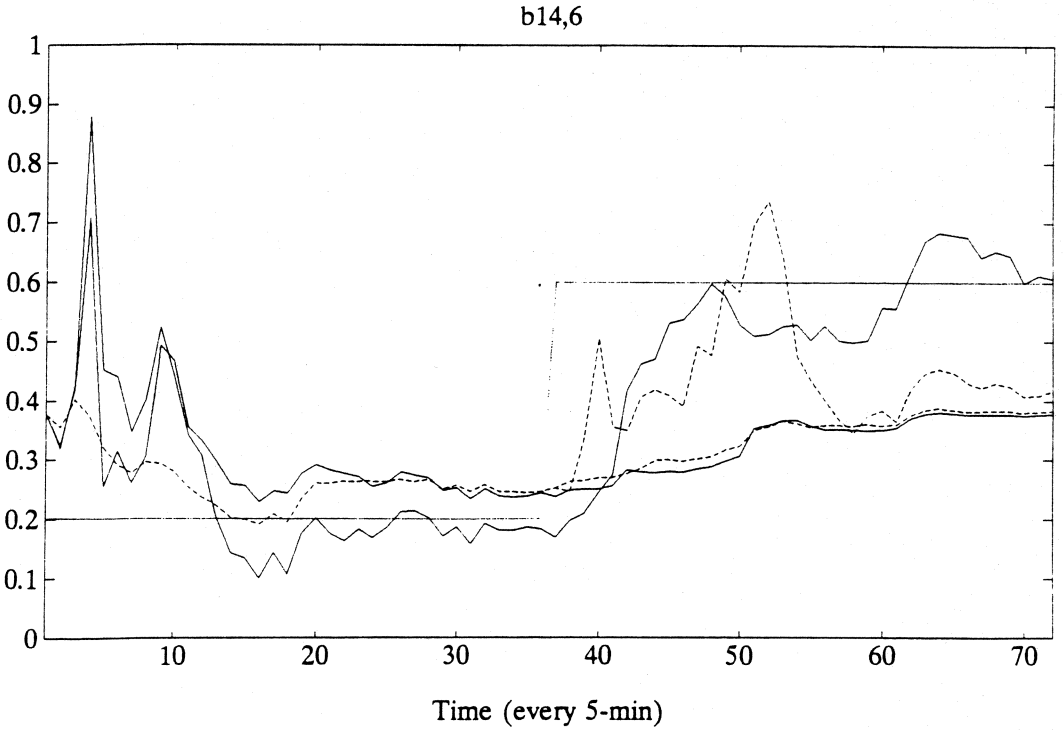


FIGURE 5.5o-p Recursive Estimates of Step-changed Turning Proportion Parameters

CHAPTER 5

TESTS WITH REAL DATA

5.1 Data Collection Procedure

Data collection of the turning movement counts on a system of intersection is a fairly challenging task. Manual counting requires a great deal of man power coordinated at different sites. A problem which immediately arises is the difficulty of checking the accuracy of data after the survey is complete. A solution to that is to record the traffic using a video camera and extract data from the tape. Even better yet, one could apply video sensing technology such as AUTOSCOPE (Michalopoulos et al., 1991). Unfortunately, direct access to this advanced technology was not available for this project. It was decided to select two closely-located intersections separated by only 220 ft, on four-lane, two-way Washington Ave. near downtown Minneapolis (as shown in Figure 5.1), so that an individual camera installed on top of a nearby five-story building could cover two intersections completely. The recording took place from 4:15 pm to 5:45 pm, during the afternoon peak period. The traffic sometimes spilled back to upstream intersections and blocked the entries in both directions. Especially on the inbound direction, a high frequency of bus boarding often completely blocked one lane traffic. The pedestrian traffic was also heavy, and therefore the left-turning or right-turning traffic from side streets often needed to yield and was forced to stop. The signal on the T-intersection operated in two phases, while the other signal operated in three phases, with one protected phase for mainline outbound traffic and left turns, followed by a phase with permitted left turns for both directions, and finally the third phase for the side street.

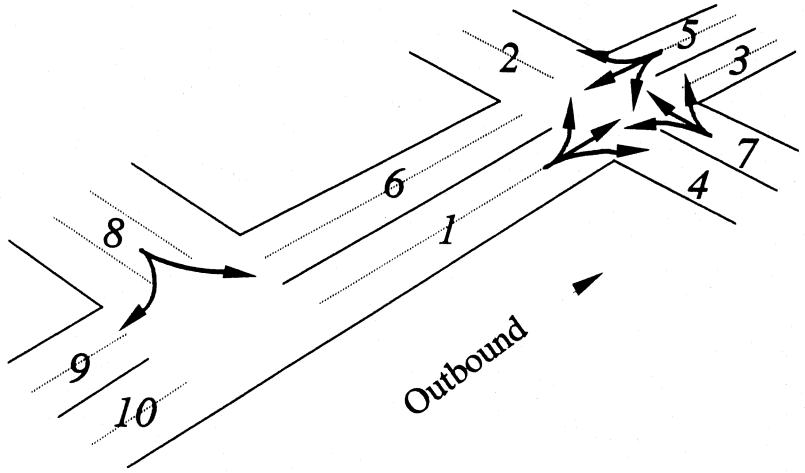


FIGURE 5.6 Data Collection Site on Washington Ave.

Therefore, the permitted left-turn model described in Chapter 2 can be applied here. Both signals have a common cycle of 100 seconds. Data were extracted from the tape every 5 seconds, so counts can be aggregated into any large interval. The count aggregation interval was set at two-minutes, and that gave us 45 observations in total.

5.2 The Results

In the underlying process, it was assumed that the turning counts from an approach, $y_{kl}(t)$, were independent multinomial outcomes with population equal to the total exiting counts $n_k(t)$ and turning probabilities $b_{kl}(t)$. If $b_{kl}(t)$ is constant over time, Nihan and Davis (1989) showed that the maximum likelihood estimates of the turning proportions is $\hat{b}_{kl} = \Sigma_t y_{kl}(t) / \Sigma_t n_k(t)$. A χ^2 test was used to formally justify the assumptions of (a) the multinomial distribution of the turning volumes, and (b) the constancy of the turning movement probabilities. The same procedure is applied here to test this data. Figures 5.2a-k depict those turning proportions as functions of time along with the ML estimates (in dashed lines). Based on the expression described in Nihan and Davis (1989), one can compute the expected frequencies of turning counts and plot against observed counterparts. As depicted in Figures 5.3a-k, although the expected and observed frequencies do not meet closely in some of the plots, the χ^2 test (summarized in Table 5.1) gives no evidence to reject these assumptions at 5% significance level, except for y_{81}. The degrees of freedom were calculated as the number of bins minus two.

A likelihood ratio test was used to investigate the constancy assumption, by dividing the whole data set into three 30-min subsets and testing the equality hypothesis of the ML estimates of

turning proportions in each 30-min interval. The likelihood function for the turning counts is

$$\prod_t \frac{n_k(t)!}{\prod_l y_{kl}(t)!} \prod_l b_{kl}^{y_{kl}(t)} \tag{4.19}$$

Similarly, the MLE of turning proportions in each 30-min interval $\bar{b}_{kl}(k)$, k=1 to 3, can be determined as the ratio of the sum of total turning counts over the sum of total exiting counts over the 30-min interval. These MLEs are depicted as solid lines in Figures 5.2a-k. Define the null hypothesis

$$H_0: \bar{b}_{kl}(1)=\bar{b}_{kl}(2)=\bar{b}_{kl}(3)$$

where number of independent turning parameters considered in the hypothesis is equal to 2 for approach 1, 5, 7, and 1 for approach 8 (T-intersection). Then one can compute the test statistic - $2*(\ln L_0 - \ln L_1)$ for each approach, where L_0, L_1 are the maximum likelihood values for null and alternative hypothesis. The degrees of freedom for χ^2-statistic are equal to 4 for approaches 1, 5, 7, and 2 for approach 8. The results are listed in the following Table.

Approach No.	1	5	7	8
χ^2-stat (df)	0.77 (4)	3.74 (4)	6.24 (4)	2.54 (2)

Again, the likelihood ratio test suggests no evidence to reject the constancy assumption at a 5% significance level.

Next, to test the independence assumption, first the autocorrelation of the standardized

residuals is examined. The standardization accounts for the heteroscedasticity that may exist in the time-series measurements when they have different variances due to different exiting volumes. The standardized residuals are defined as

$$d_{kl} = \frac{y_{kl} - n_k \hat{b}_{kl}}{\sqrt{n_k \hat{b}_{kl}(1-\hat{b}_{kl})}} \tag{4.20}$$

From the computed autocorrelations depicted in Figures 5.4a-k, one can safely claim that all of the turning counts are uncorrelated across time. Second, a Lagrange multiplier test due to Breusch and Pagan (1980) is used to test the hypothesis that the within-group correlations are zeros, where the "group" is referred to the intersection approach from which traffic turns. The statistic is computed as the multiplication of number of samples (45) with the summation of the off-diagonal, within-group squared residual correlation coefficients in either lower or upper triangle of the correlation matrix. Compared to the critical value (61.4) at 5% significance level for 45 degrees of freedom, the computed statistic, 41.9, suggests no evidence to reject the hypothesis. These establish evidences that turning counts are uncorrelated both across time and group (approach). However, the non-correlation does not imply independence unless the variables are normally distributed. Based on the correlation coefficients (listed in TABLE 5.2) computed from normal probability plots of individual residuals (Filliben, 1975), it was found that some of the residuals do not appear normal, but deviate not far from normality. Note that the normality of the marginal distribution does not imply the multinomial normality. In summary, it is concluded that there exit evidences in which all of the turning counts are uncorrelated and most of them might be generated from

independent multinomial processes.

Since the turning proportions are constant over time, the tracking algorithms RNLS without factor discounting and RQML without covariance resetting are expected to perform reasonably well. As mentioned earlier, heavy pedestrian traffic and one-lane blockage due to frequent bus boarding inbound may make traffic flow prediction a little bit difficult. It turns out, as shown in Figures 5.5a-g that, except for b_{76} and b_{72}, both NLS algorithm (short dashed line) and QML algorithm (long dashed line) track the mean of the actual turning proportions fairly well. Similar to the simulation analysis, the recursive estimates fluctuate during the first few iterations. This suggests that, for practical use, one must let the algorithms run for a while, until the estimates settle down to the right places.

TABLE 5.1 Chi-square Test for Constancy of Turning Counts on Washington Ave.

Turning Elements	χ^2-stat (df[1])
y_{12}	5.28 (4)
y_{13}	1.17 (5)
y_{14}	5.22 (4)
y_{54}	0.001 (1)
y_{56}	5.47 (5)
y_{52}	4.85 (3)
y_{76}	3.69 (4)
y_{72}	0.24 (1)
y_{73}	0.49 (2)
y_{81}	9.77* (4)
y_{89}	5.17 (3)

Note: [1] degree of freedom
* Reject the normality hypothesis at the 5% significance level

TABLE 5.2 Normality Test of Standardized Residuals of Turning Counts on Washington Ave.

Standardized Residuals	Correlation Coefficients
y_{12}	0.9833
y_{13}	0.9845
y_{14}	0.9693*
y_{54}	0.9489*
y_{56}	0.9860
y_{52}	0.9720*
y_{76}	0.9859
y_{72}	0.9652*
y_{73}	0.9817
y_{81}	0.9902
y_{89}	0.9902

* Reject the normality hypothesis at the 5% significance level (critical value = 0.974)

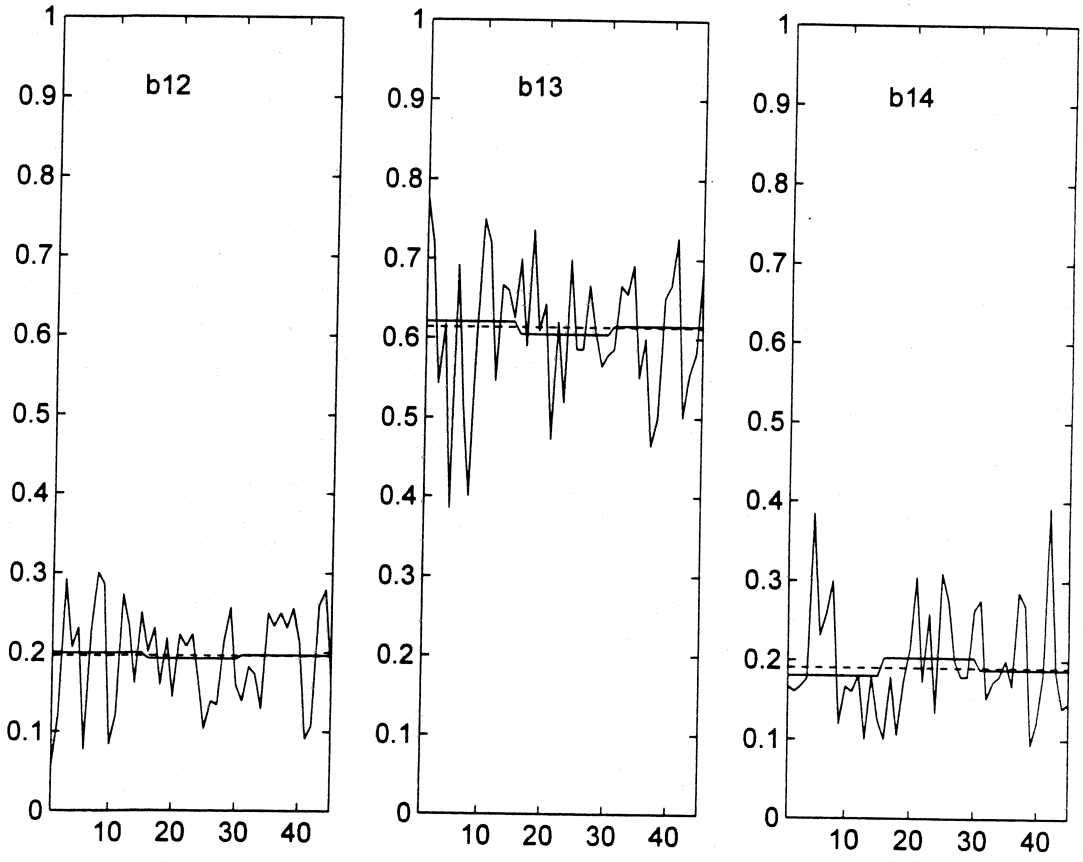


FIGURE 5.7a-c Time-Series Plots of Turning Proportions on Washington Ave.

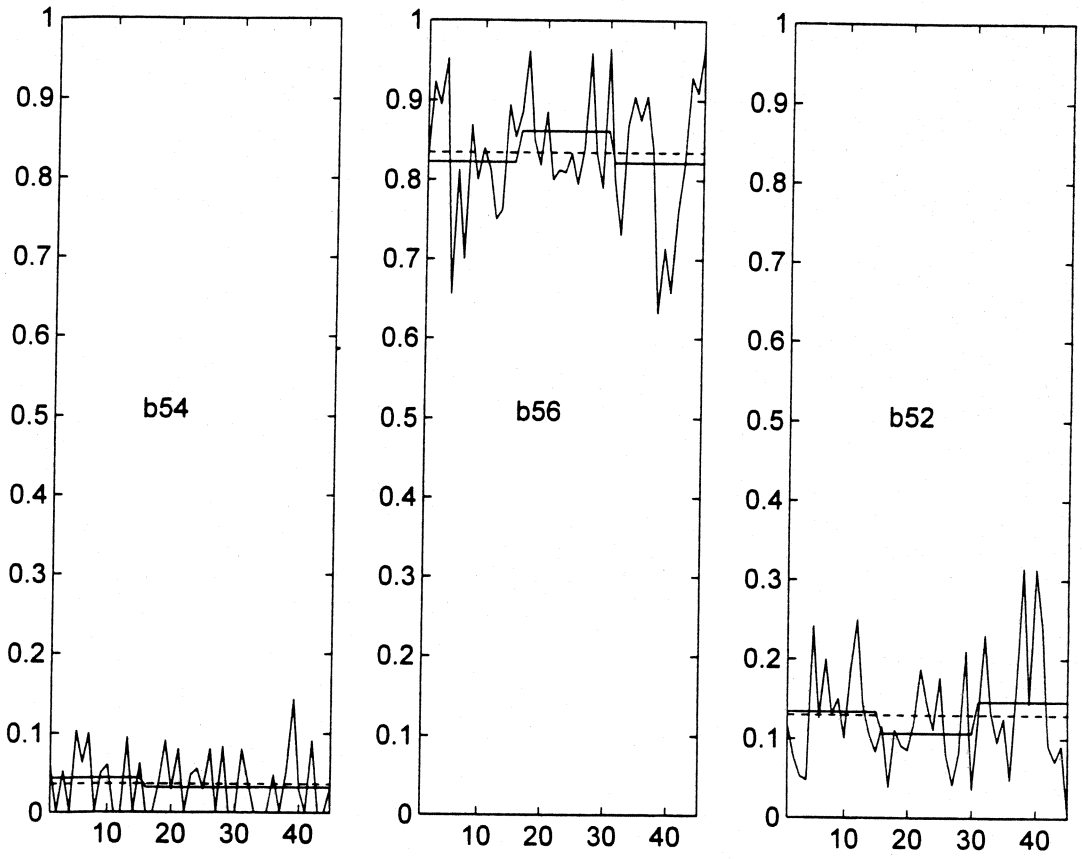


FIGURE 5.7d-f Time-Series Plots of Turning Proportions on Washington Ave.

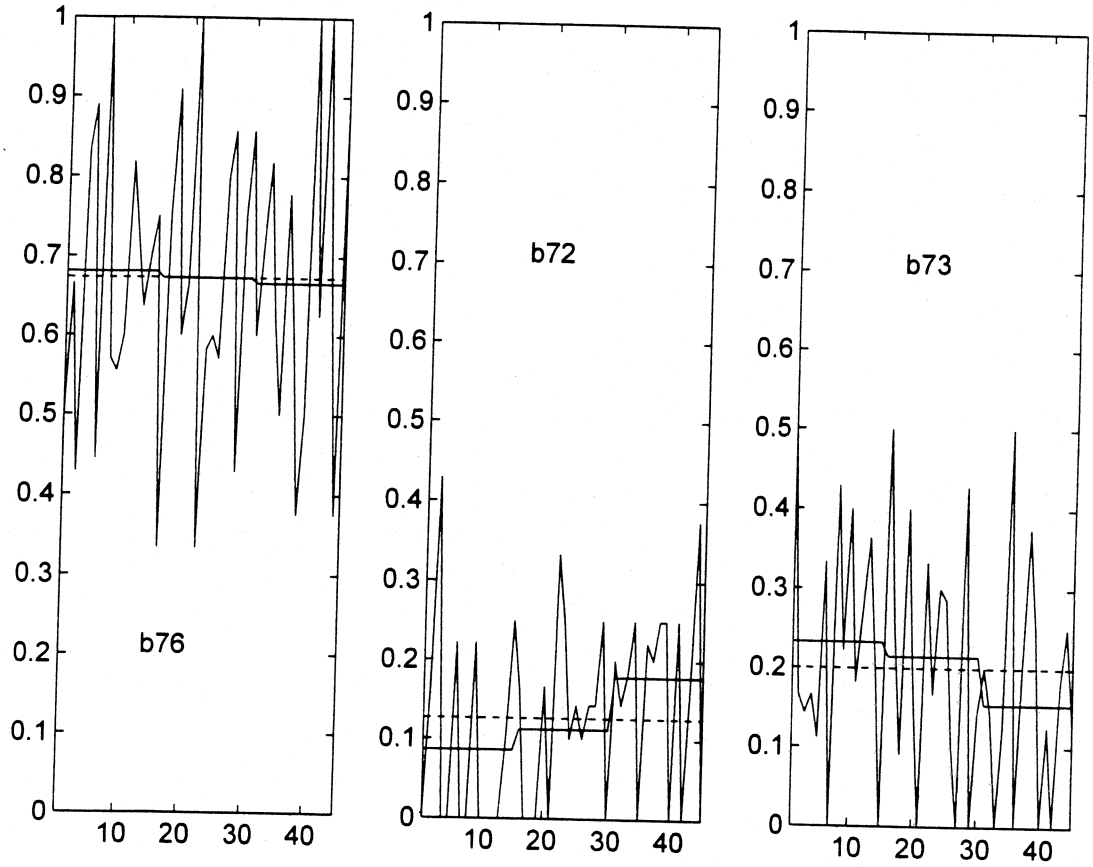


FIGURE 5.7g-i Time-Series Plots of Turning Proportions on Washington Ave.

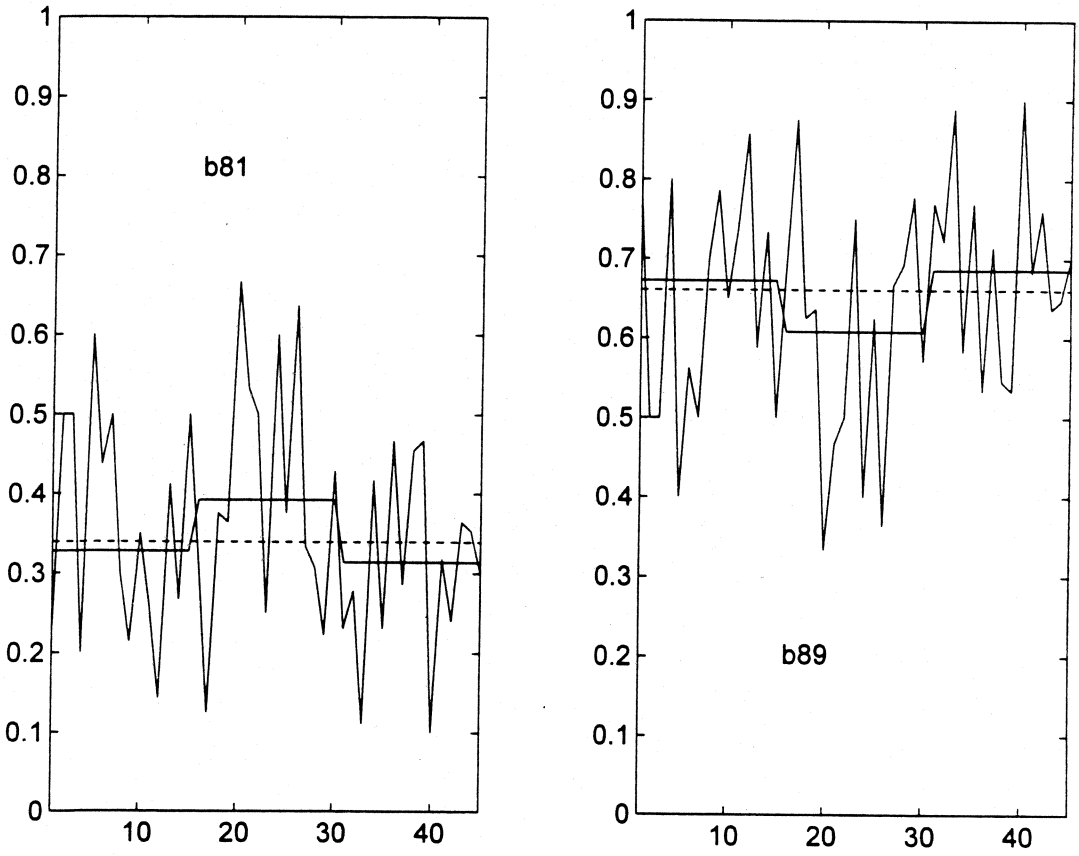


FIGURE 5.7j-k Time-Series Plots of Turning Proportions on Washington Ave.

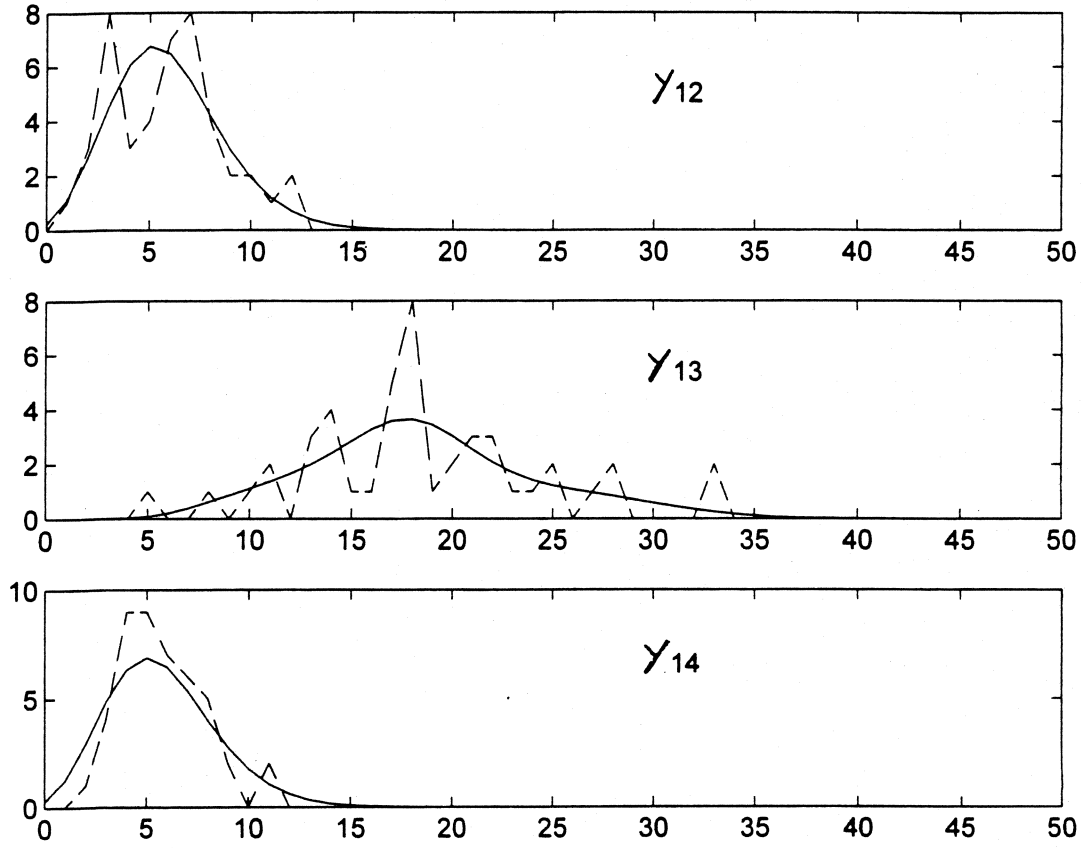


FIGURE 5.8a-c Expected vs Observed Frequencies of Turning Counts on Washington Ave.

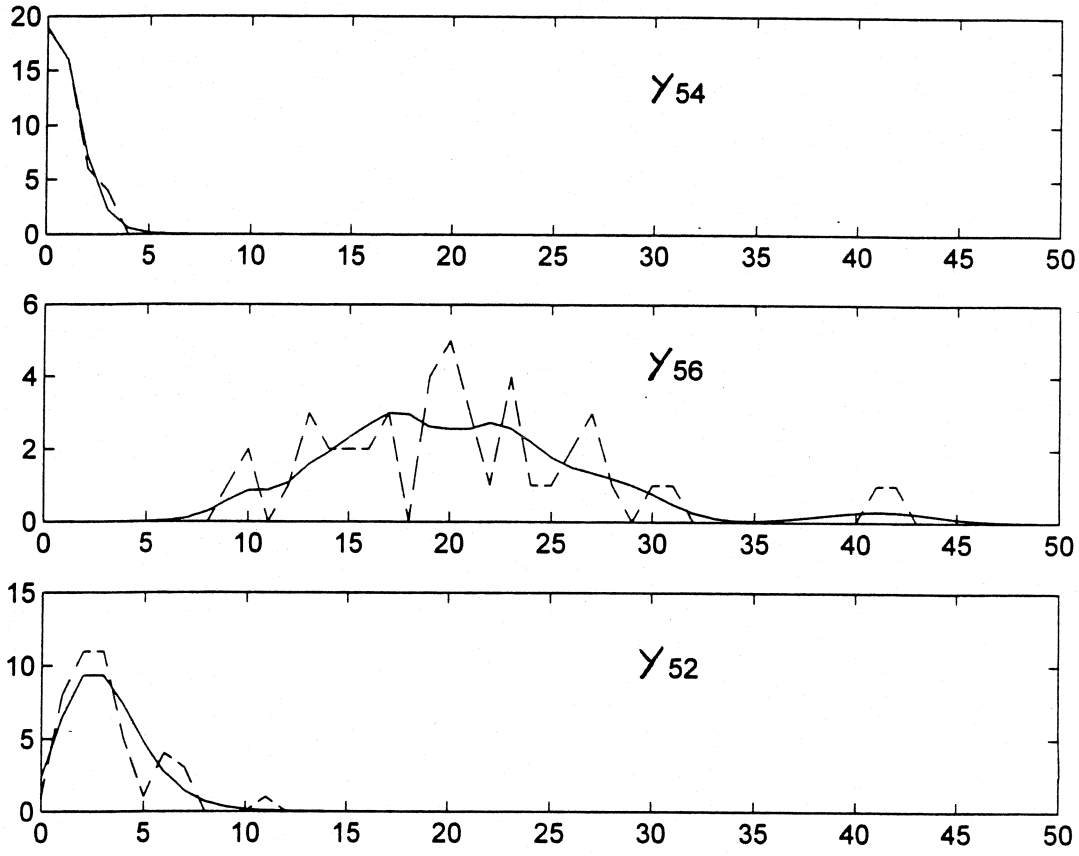


FIGURE 5.8d-f Expected vs Observed Frequencies of Turning Counts on Washington Ave.

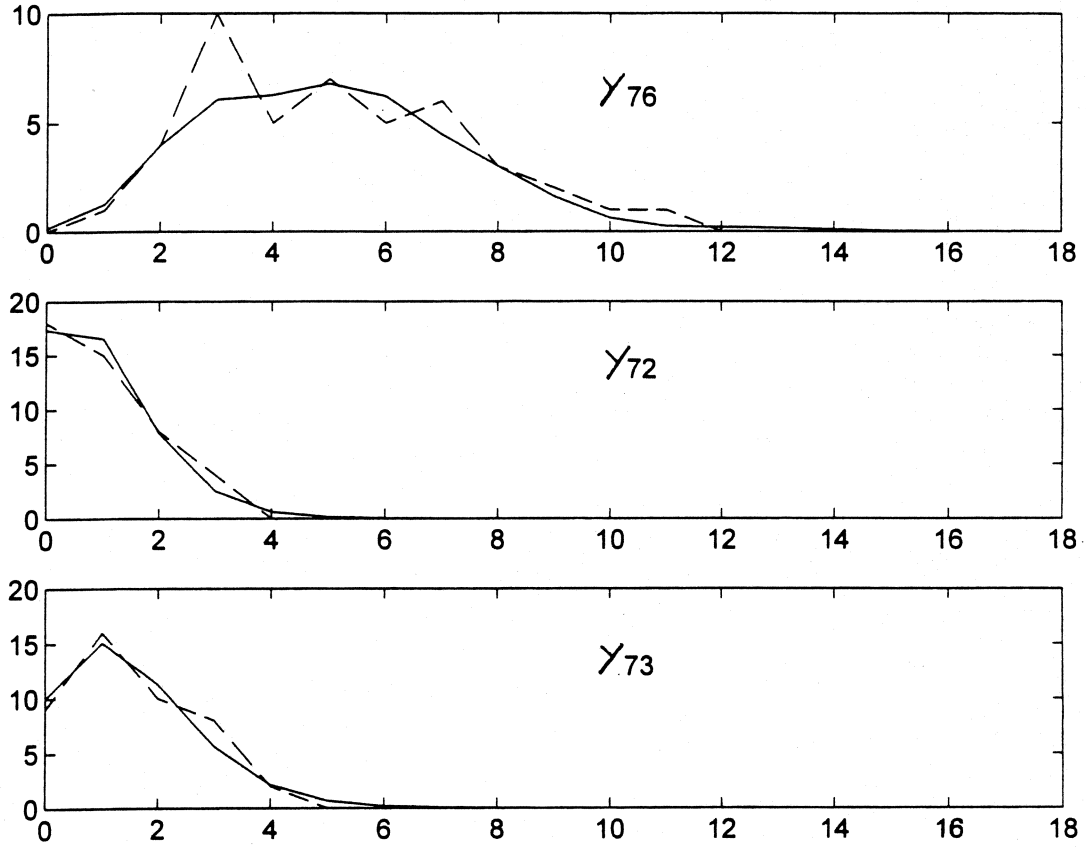


FIGURE 5.8g-i Expected vs Observed Frequencies of Turning Counts on Washington Ave.

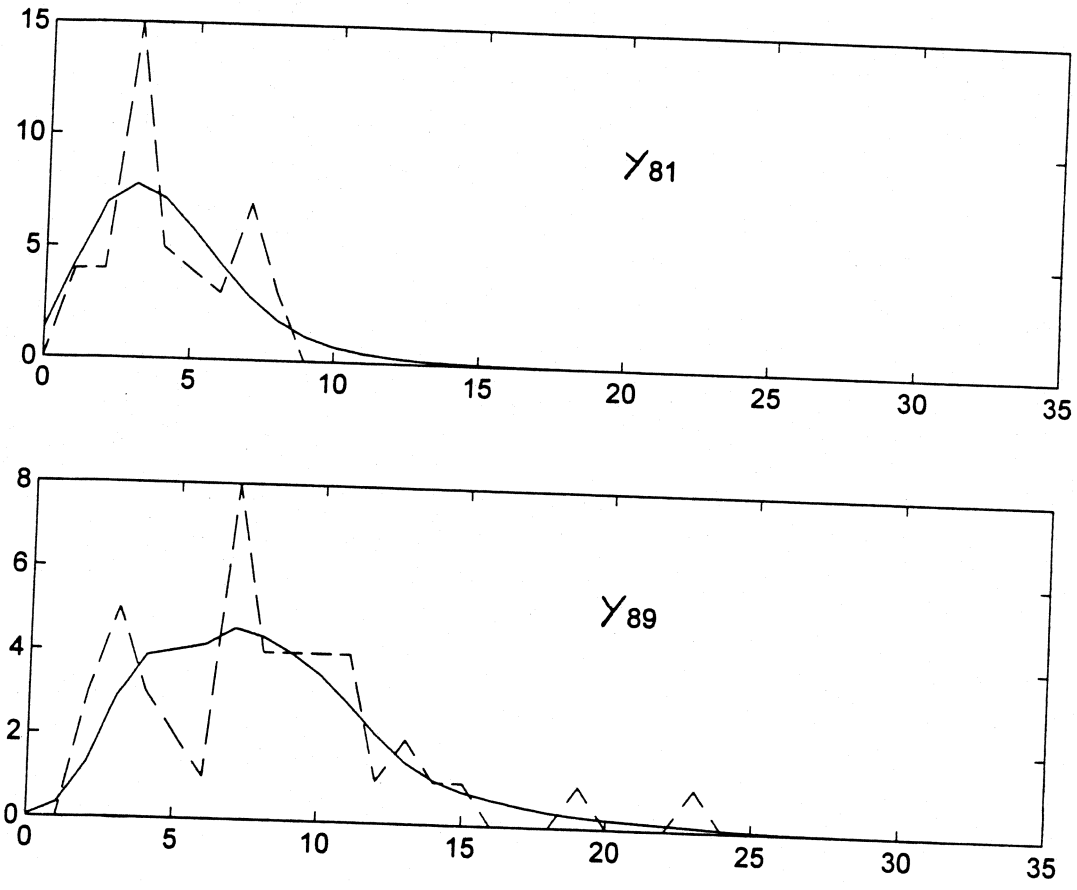


FIGURE 5.8j-k Expected vs Observed Frequencies of Turning Counts on Washington Ave.

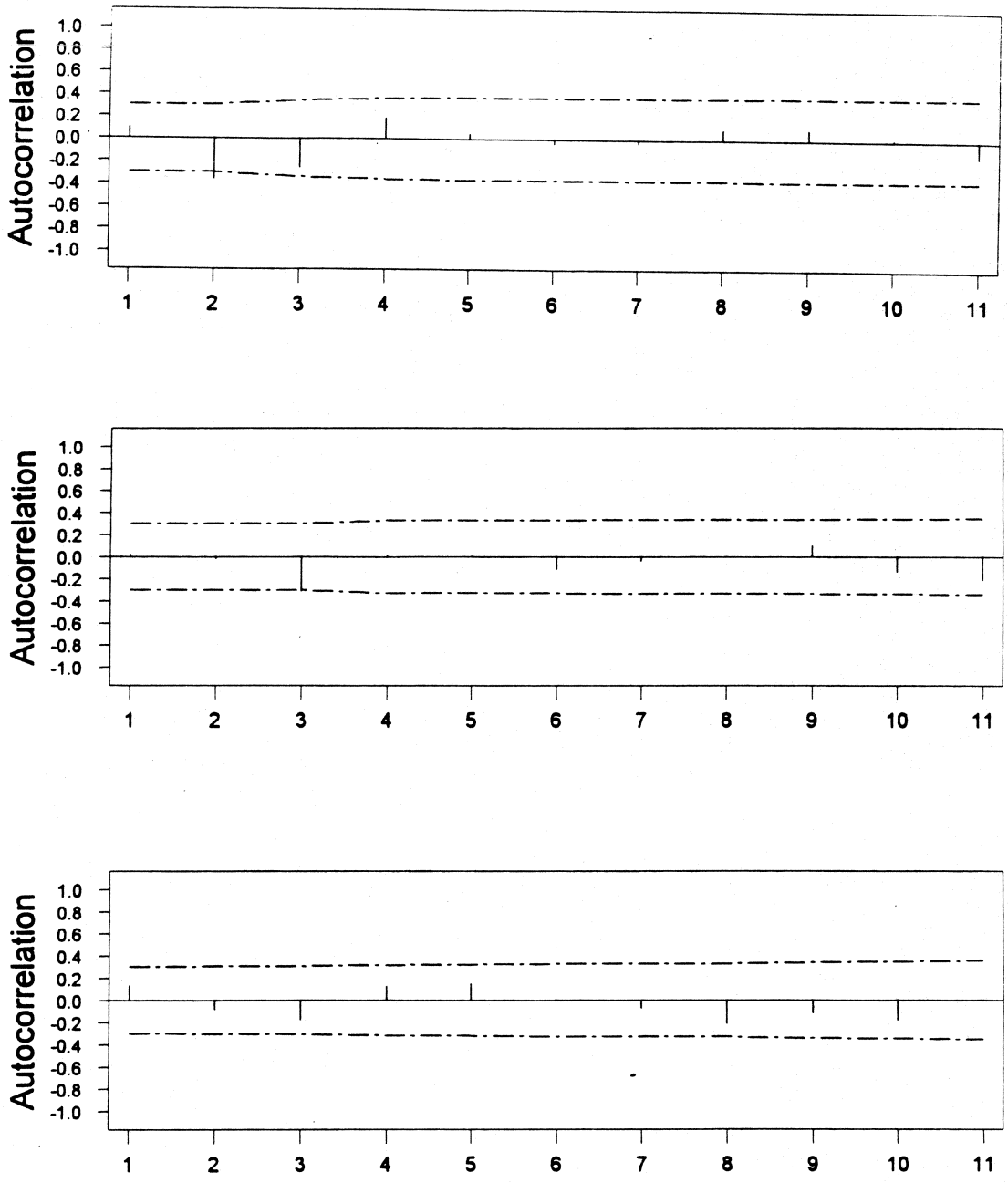


FIGURE 5.9a-c Autocorrelations of the Standardized Residuals of Turning Counts on Washington Ave.

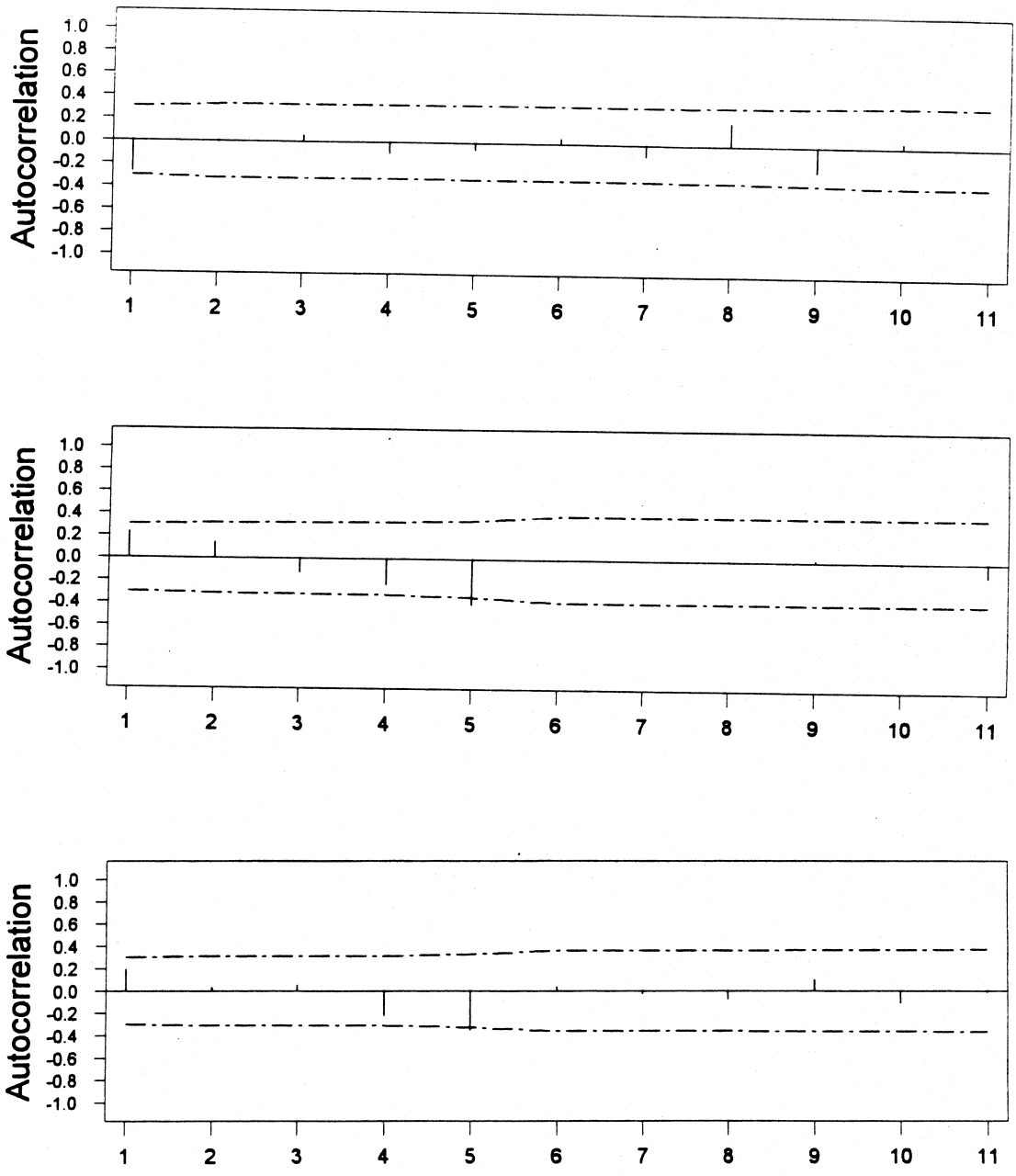


FIGURE 5.9d-f Autocorrelations of the Standardized Residuals of Turning Counts on Washington Ave.

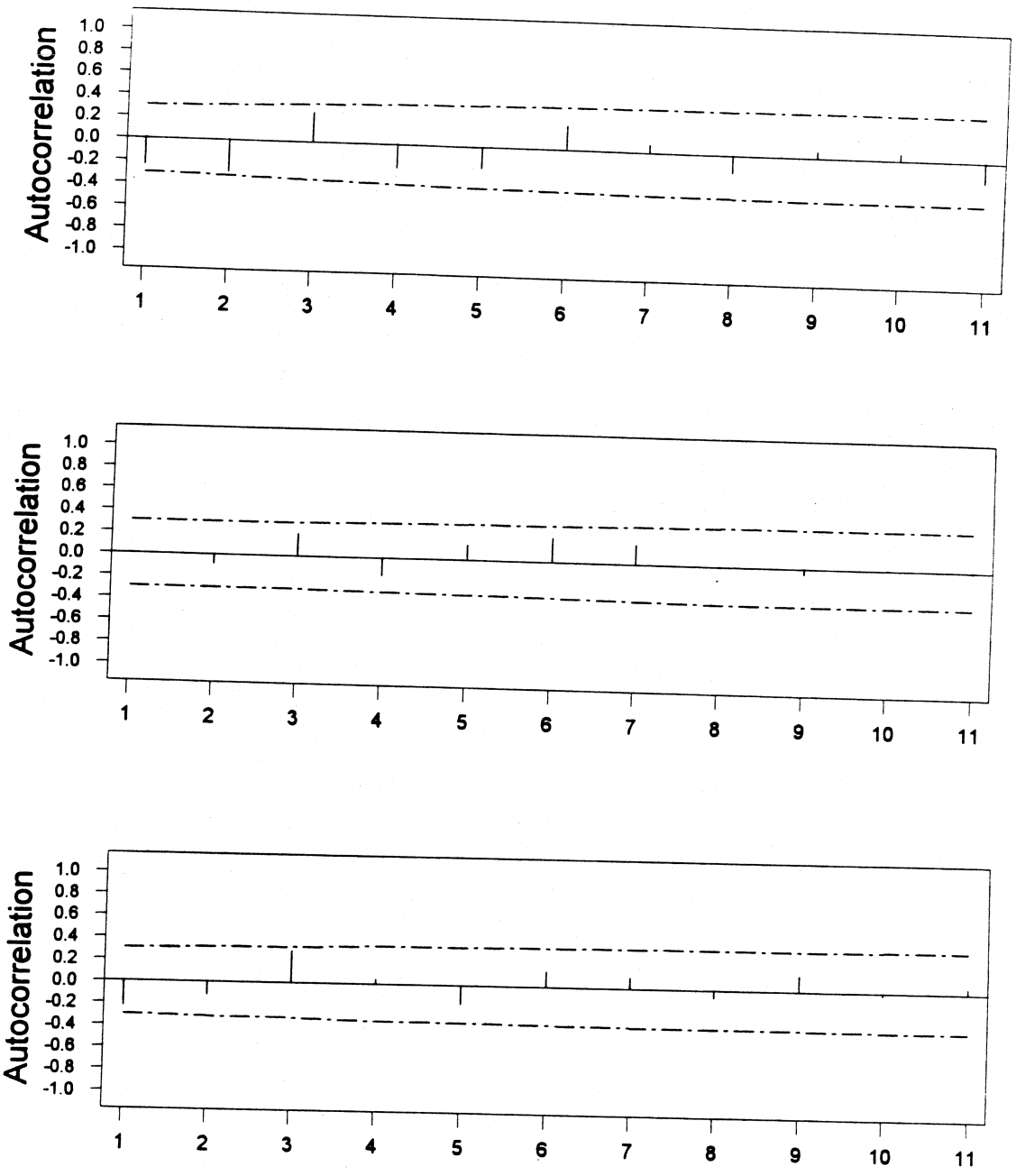


FIGURE 5.9g-i Autocorrelations of the Standardized Residuals of Turning Counts on Washington Ave.

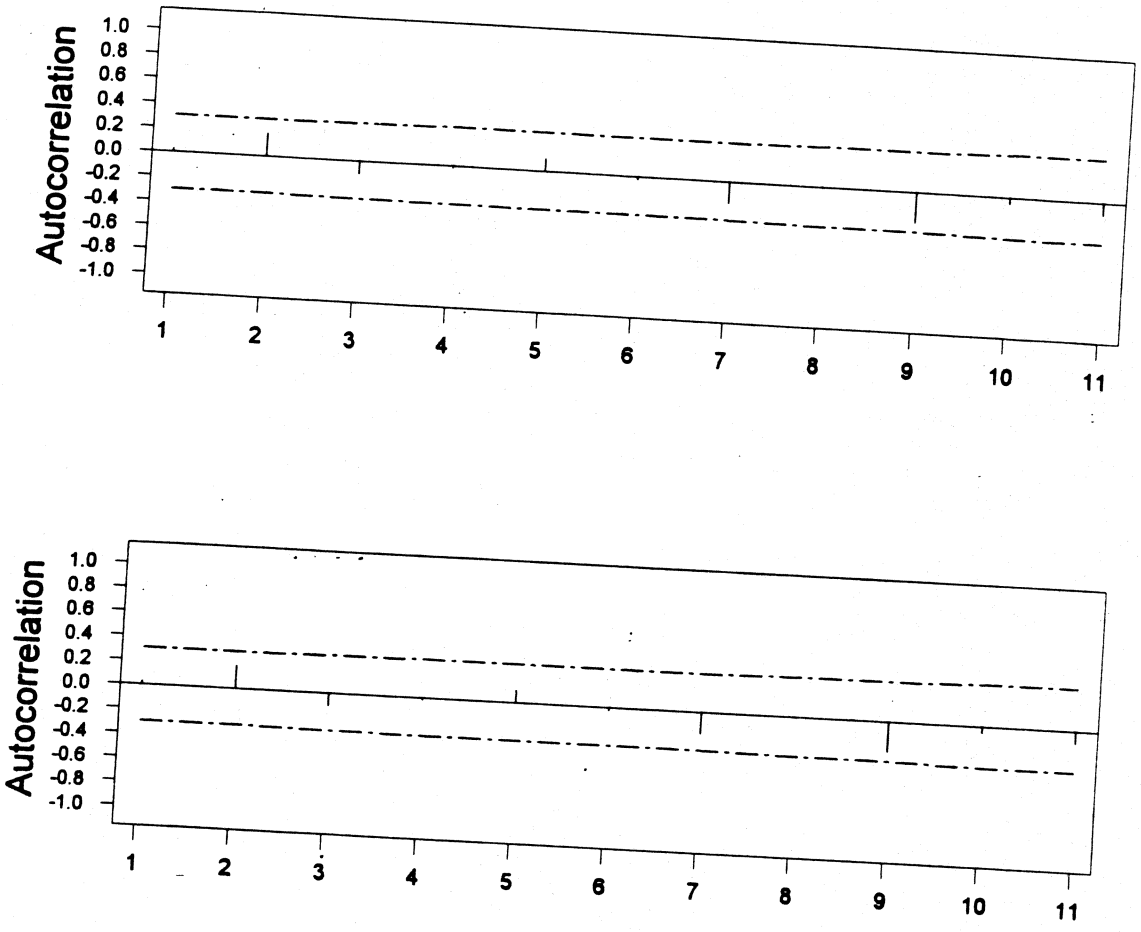


FIGURE 5.9j-k Autocorrelations of the Standardized Residuals of Turning Counts on Washington Ave.

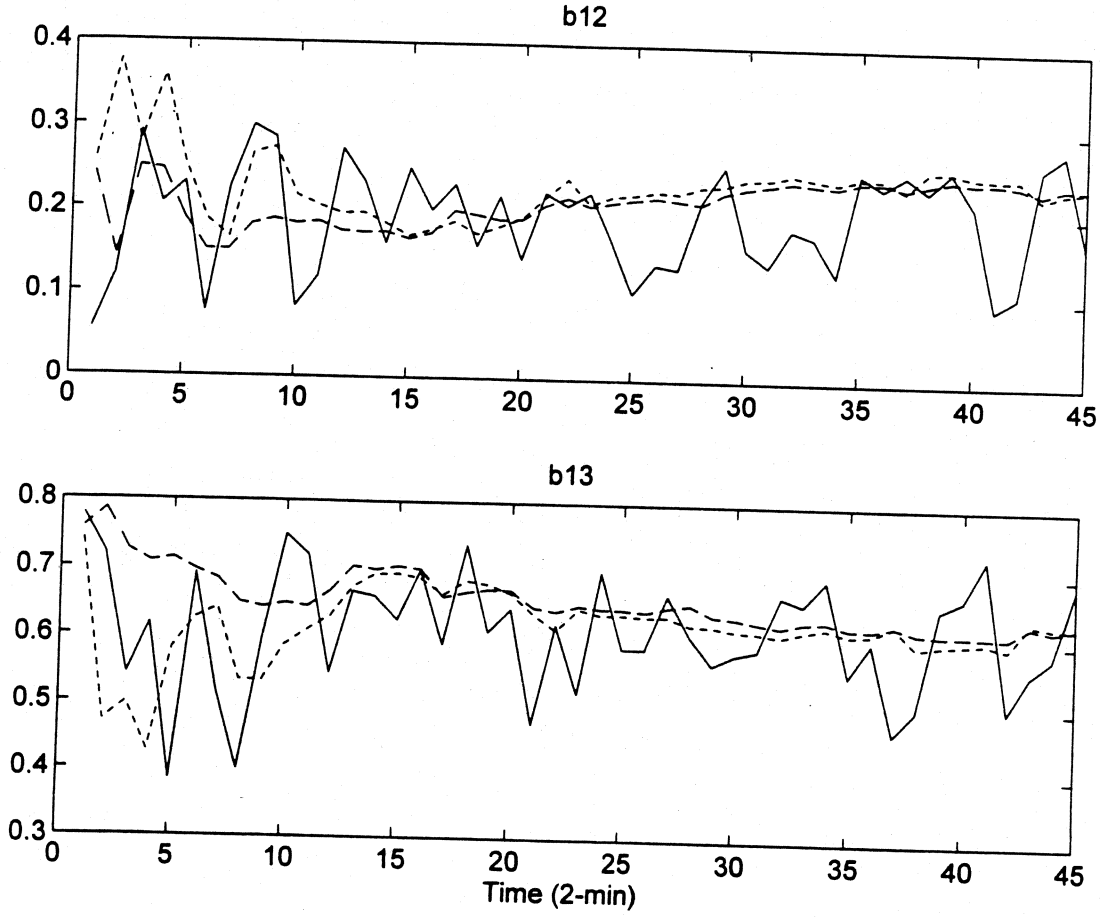


FIGURE 5.10a-b Recursive Estimates of Turning Proportions on Washington Ave.

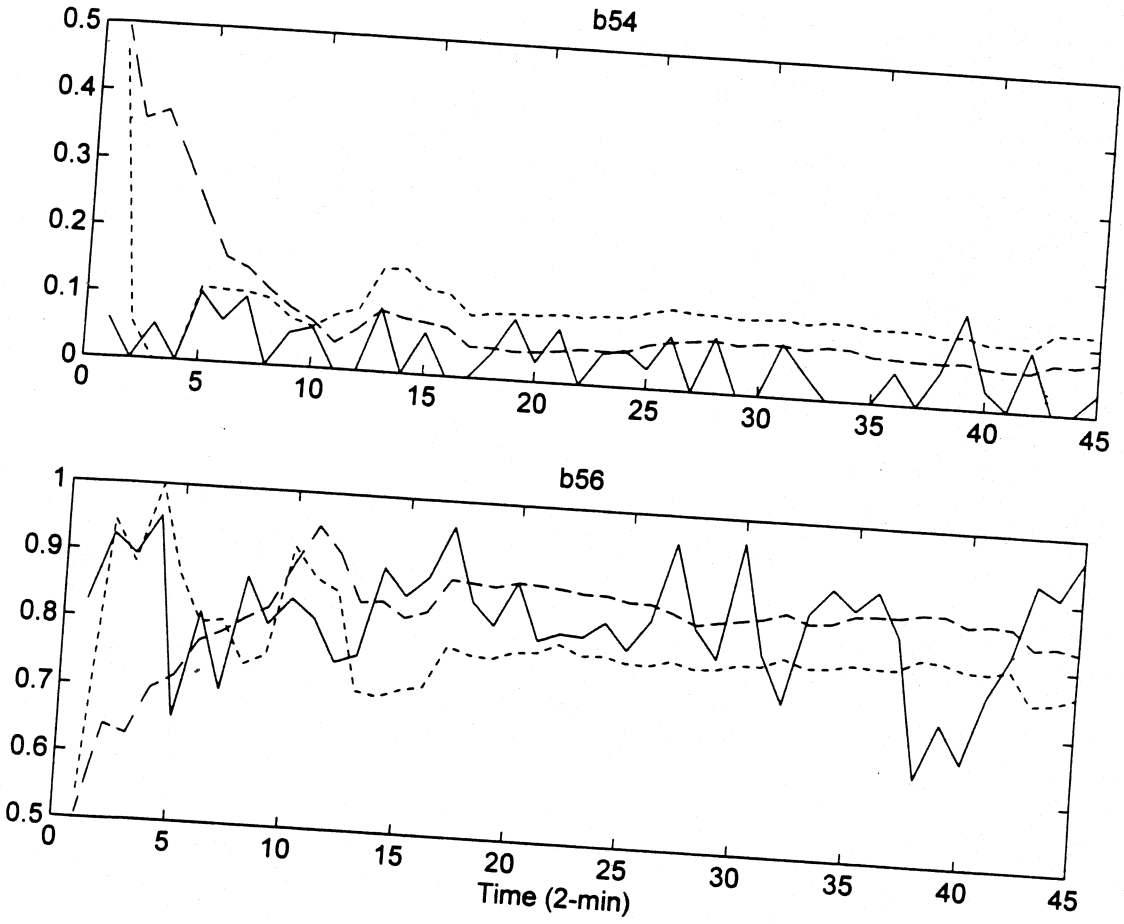


FIGURE 5.10c-d Recursive Estimates of Turning Proportions on Washington Ave.

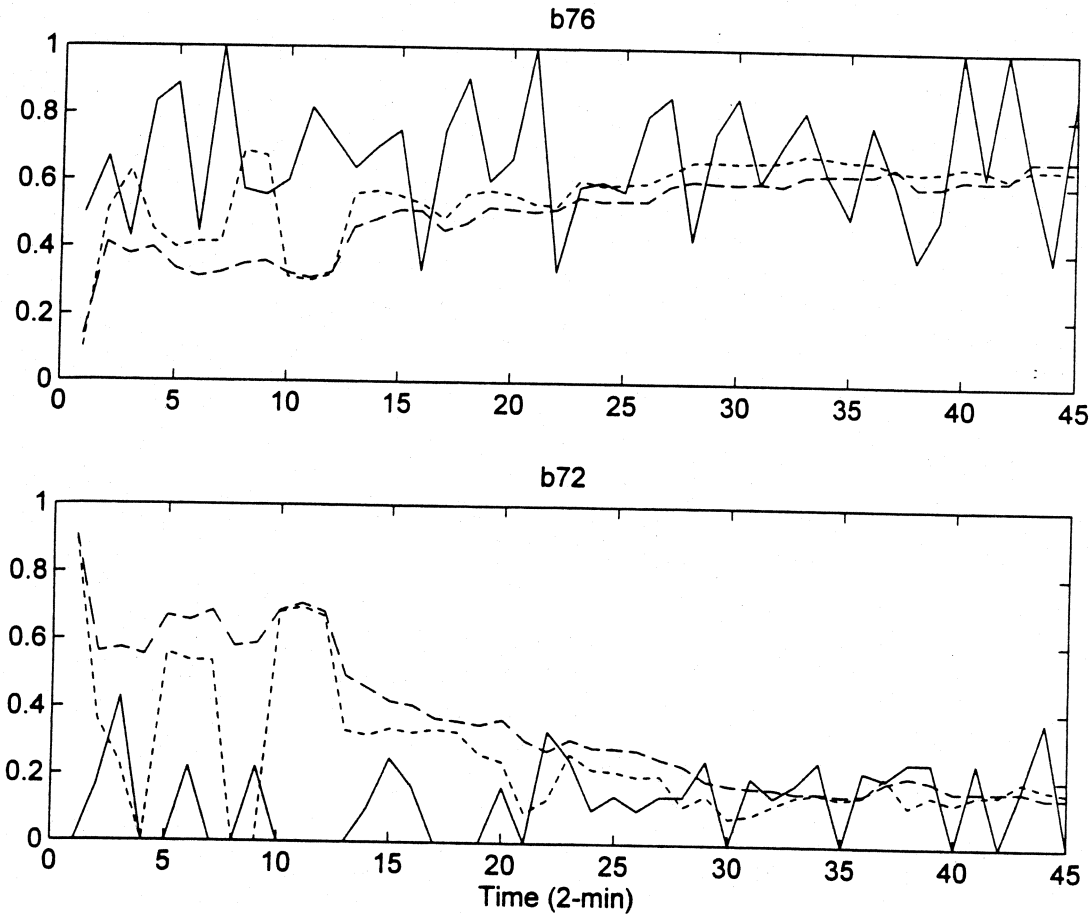


FIGURE 5.10e-f Recursive Estimates of Turning Proportions on Washington Ave.

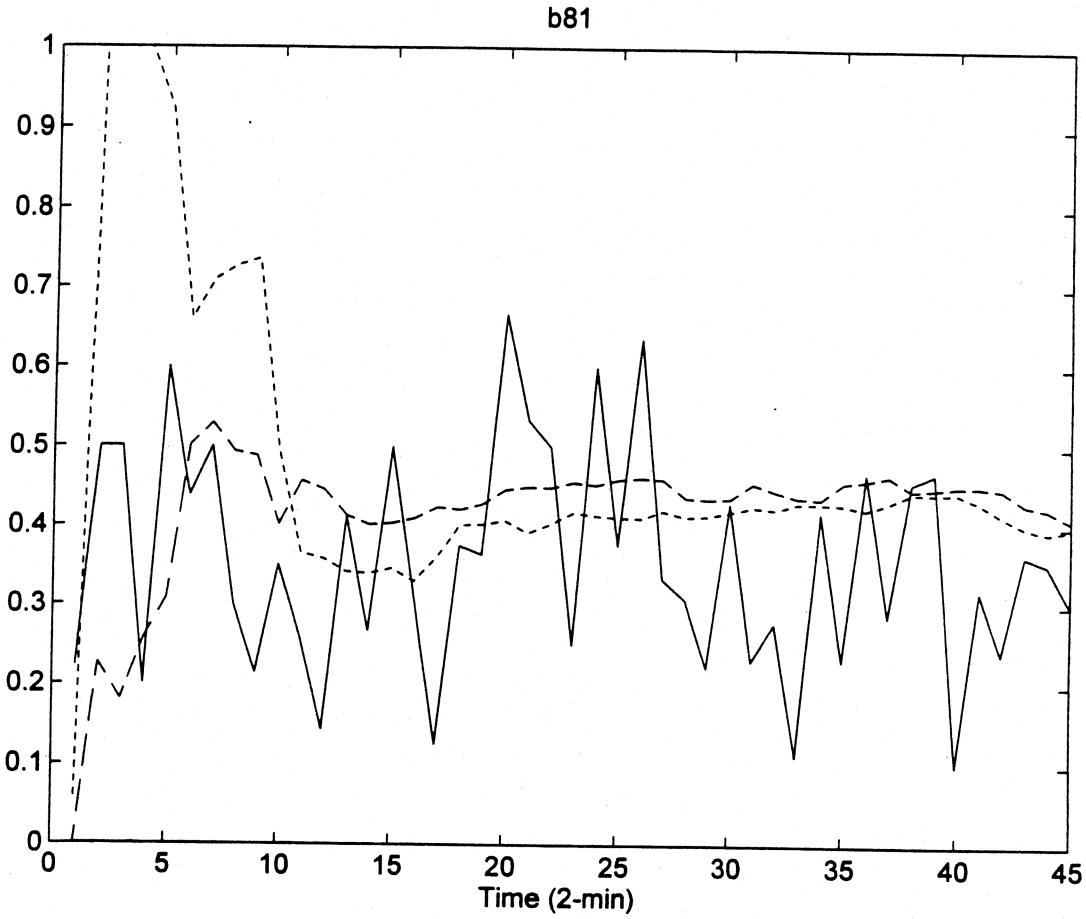


FIGURE 5.10g Recursive Estimates of Turning Proportions on Washington Ave.

CHAPTER 6

CONCLUSIONS

6.1 Conclusions

The first objective of this phase of project was to enhance the basic traffic flow model for better predicting the detector output counts and distribution of traffic states as inputs in the estimation algorithms. Base on a set of macroscopic, discrete-time traffic flow models based on the Markovian compartment concept developed and calibrated in Phase I&II, the proposed model was modified to capture an interesting traffic instability observed by Edie and Foote. Tested with a real data set, the modified model performed reasonably and produced a slightly better data fit than did the original. This indicates that the proposed model indeed has a potential to duplicate complex traffic phenomena, a long-term objective of high-order models. Although it has effects analogous to those of the high-order models, the proposed model is more tractable and has fewer parameters to be calibrated. Finally, by leading naturally to state-space models, the proposed model provides a basis for on-line tracking of turning movement proportions as well as the dynamic evolution of traffic states on the general network.

The second objective was to develop a procedure to possibly measure the variability of the turning movement parameter estimates under different degree of information embedded in various detector configurations. The condition number of the Jacobian matrix computed from each simulation run was used to quantify the amount of detector information which might indicate a potential problem in estimation if its value is in excess of 20.

Regarding the third objective, off-line and recursive algorithms were presented for estimating turning movement proportions when only a partial set of detector counts is available, using Nonlinear Least Square (NLS) and Quasi Maximum Likelihood (QML) approaches. The quality of Large Population Approximation (LPA) was first examined. It turned out that the deterministic mean trajectories and confidence bounds were in agreement with their stochastic counterparts, with the confidence bounds for the deterministic process being slightly under-estimated. Compared to the Ordinary Least Square (OLS) estimator under full detector configuration, the off-line NLS and QML using cordon counts showed increased bias and variances, due to the fact that these estimators work with less detector information. As a comparison, the NLS estimator is unbiased while the QML estimator shows biases in some parameter estimates. In terms of total Mean Square Error (MSE), however, the QML estimator is better than NLS due to an increase in bias being offset by a decrease in variance. As far as the computation time is concerned, the QML estimator takes more than triple the CPU time, while the average number of function evaluations required is less than the NLS estimator.

The efficiency of the off-line QML estimator is also demonstrated in the recursive estimates. The confidence intervals computed from 50 samples indicated that the recursive QML estimator is in fact a more efficient estimator than the recursive NLS estimator. This suggests that some of the loss of accuracy due to less information provided by a looser detector configuration might be offset by switching to a more efficient estimation approach. To track the changes in parameters across time, the discounted RNLS and the RQML with covariance resetting schemes were introduced. The results showed that, compared to the original algorithms, both modified algorithms adapt more quickly to the stepwise changes in parameters within a reasonable number of iterations, provided

that the time when parameters change is appropriately identified. Finally, in the real-data study both NLS and QML algorithms tracked the means of the actual turning proportions reasonably well. For practical use, it is suggested to let the algorithms run for a while until the estimates settle down to the right places, because the recursive estimates may fluctuate during the first few iterations.

6.2 Future Studies

This avenue of research never seems to end. Several relevant work must be classified as future studies that could be continued in the near future. Areas for future studies are itemized below.

(1) One can always improve the traffic flow models to a further level of detail. Regarding the Markovian traffic flow model proposed here, a certain amount of detail might be necessary to give a more precise description of real traffic conditions for other purpose. For example, the ability to handle the lane discontinuity situation due to a bottleneck, road construction, or non-recurrent incident are of interest.

(2) Sufficient and necessary conditions for locally or globally identifying the parameters in the approximated system is appealing, although it is very difficult to obtain, especially for the nonlinear systems.

(3) The recursive algorithms performed well in the small-scale network under the cordon-count configuration, namely, an arterial network containing two inter-sections. The condition numbers for grid networks with cordon counts indicate that the information provided by the cordon-count configuration declines as network gets bigger. One can remedy this problem by decomposing the network or adding internal detectors in the network, and this leads to

an optimal design problem for the detector configuration.

Although not directly relevant to the main theme of this thesis, the detection of change points in parameters deserves serious attention. Not only would the detection algorithm be useful in the estimation of time-varying parameters, but it also could be used as a tool to detect abnormal events, such as incidents, for better management of traffic flow.

REFERENCES

Bell, M. G. H. (1991) The real time estimation of origin-destination flows in the presence of platoon dispersion, *Transportation Research*, 25B, pp. 115-125.

Bellman, R. and Åström, K. J. (1970) On structural identifiability. *Math. Biosci.*, 7, pp. 329-339.

Belsley, D., E. Kuh, and R. Welsch. (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Bowman, K. O. and L. R. Shenton. (1975) Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. *Biometrika* 62: pp. 243-250.

Breusch, T. and A. Pagan. (1980) The LM Test and Its Applications to Model Specification in Econometrics. *Review of Economics Studies*, 47, pp. 239-254.

Chen, Y. C. (1983) Convergence study of two real-time parameter estimation schemes for nonlinear system. *Nonlinear Stochastic Problems*, pp. 79-86.

Cremer, M. and Keller, H. (1981) Dynamic identification of O-D flows from traffic counts at complex intersections. *Proc. 8th Int. Symp. Transportation and Traffic Theory*, Toronto University.

Cremer, M. and Keller, H. (1987) A new class of dynamic methods for identification of origin-destination flows, *Transportation Research*, 21B, pp. 117-132.

Cremer, M. and Papageorgiou, M. (1981) Parameter identification for a traffic flow model. *Automatica*, Vol. 17, No. 6, pp. 837-843.

Daganzo, C. F. (1994) The Cell Transmission Model - A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research -B*, Vol. 28B, No. 4, pp. 269-287.

Davis, G. A. (1993) Estimating freeway demand patterns and impact of uncertainty on ramp controls. *Journal of Transportation Engineering*, Vol 119, No. 4, pp. 489-503.

Davis, G. A. and C. J. Lan. (1995) Estimating intersection turning movement proportions from less-than-complete sets of traffic counts. *Transportation Research Record* 1510, National Academy Press, Washington, D.C., pp. 53-59.

Davis, G. A. and N. Nihan. (1991) A Stochastic Process Approach to Estimating OD Parameters from Time-Series of Traffic Counts. *Transportation Research Record* 1328, TRB, Washington, D.C., pp. 36-42.

Edie, L. C. and R. S. Foote. (1961) Experiments on Single-Lane Flow in Tunnels. *Proceedings of the 1st International Symposium on the Theory of Traffic Flow*, pp. 175-192.

Fahrmeir, L. and G. Tutz (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag, New York.

Filliben, J. J. (1975) *The Probability Plot Correlation Coefficient Test for Normality*. Technometrics, 17, no. 1, pp. 111-117.

Golub, G. H. and C. F. Van Loan. (1989) *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland.

Goodwin, G. C. and K. S. Sin (1984) *Adaptive Filtering Prediction and Control*. Prentice-Hall, Inc., New Jersey.

Greene, W. H. (1993) *Econometric Analysis*. Macmillan Publishing Company, New York.

Grewal, M. S. and H. J. Payne (1976) Identification of Parameters in a Freeway Traffic Model, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-6, No. 3, pp. 176-185.

Hauer, E., Pagitsas, E., and Shin, B. T. (1981) Estimation of turning flows from automatic counts. *Transportation Research Record* 795, TRB, pp. 1-7.

Jeffreys, M. and M. Norman. (1977) On finding realistic turning flows at road junctions. *Traffic Engineering Control*, 18(1), pp. 19-21, 25.

Kessaci, A., J. L. Farges, and J. J. Henry (1989) On-line estimation of turning movements and saturation flows in PRODYN. *Proceeding of the 6th IFAC/IFIP/IFORS Symposium* on Transportation, Paris.

Lan, C-J and G. A. Davis (1997) Empirical assessment of a markovian traffic flow model. Paper Accepted for Publication in *Transportation Research Record*.

Lehoczky, J. P. (1980) Approximations for interactive Markov chains in discrete and continuous time. *Journal of Mathematical Sociology*, Vol. 7, pp. 139-157.

Ljung, L. (1979) Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control*, Vol. AC-24, No. 1, pp. 36-50.

Ljung, L and Söderström, T. (1983) *Theory and Practice of Recursive Identification*. The MIT Press, Cambridge, MA.

Mahmassani, H and Y. Sheffi. (1981) Using Gap Sequences to Estimate Gap Acceptance Functions. *Transportation Research -B*, V. 15B, No. 3, pp. 143-148.

Mataušek, M. R. and Stanković, S. S. (1980) Robust real-time algorithm for identification of non-linear time-varying systems. *Int. J. Control*, Vol. 31, No. 1, pp. 79-94.

MATLAB, Version 4.2c. (1994) The MathWorks, Inc.

Mekky, A. (1979) On estimating turning flows at road junctions. *Traffic Engineering Control*, 20(10), pp. 486-487.

Michalopoulos, P. G., B. Wolf, and R. Benke. (1991) Testing and Field Implementation of the Minnesota Video Detection System (AUTOSCOPE). *Transportation Research Record* 1287, TRR, Washington, D.C., pp. 176-184.

Miller, A. J. (1972) Nine Estimators of Gap Acceptance Parameters. *Proceeding of 5th International Symposium on the Theory of Traffic Flow and Transportation* (Edited by G. F. Newell), pp. 215-235, Elsevier, NY.

Nihan, N. and Davis, G. A. (1987) Recursive estimation of origin-destination matrices from input/output counts, *Transportation Research*, 21B, pp. 149-163.

Nihan, N. and Davis, G. A. (1989) Application of prediction-error minimization and maximum likelihood to estimate intersection O-D matrices from traffic counts, *Transportation Science*, 23, pp. 77-90.

Rothenberg, T. (1971) Identification in Parametric Models. *Econometrica*, 39, pp. 577-591.

Seber, G. and Wild, C. (1989) *Nonlinear regression*, Wiley and Sons, New York.

Tanner, J. C. (1967) The Capacity of an Uncontrolled Intersection. *Biometrika*, 54, 3 and 4, pp. 657-658.

Transportation Research Board. (1985) *Highway Capacity Manual*, TRB, National Research Council, Washington, D.C.

Troutbeck, R. J. (1986) Average Delay at an Unsignalized Intersection with Two Major Streams Each Having a Dichotomized Headway Distribution. *Transportation Science*, Vol. 20, No. 4, pp. 272-286.

Van Zuylen, H. J. (1979) The estimation of turning flows on a junction. *Traffic Engineering Control*, 20(11), pp. 539-541.

Velan, S. M. and M. Van Aerde. (1996) Gap Acceptance and Approach Capacity at Unsignalized Intersections. *ITE Journal*, Vol. 66, No. 3, pp. 40-45.