# A Small Community Through Trip Rate Methodology

By

Michael Anderson, Ph.D., Mr. Yasir Abdullah
Civil & Environmental Engineering Department
University of Alabama in Huntsville
Huntsville, Alabama

Prepared by

**UTCA**

University Transportation Center for Alabama
The University of Alabama, The University of Alabama in Birmingham, and
The University of Alabama at Huntsville

UTCA Report Number 03312
February 28, 2005

Technical Report Documentation Page

| 1.Report No FHWA/CA/OR- | 2.Government Accession No. | 3.Recipient Catalog No. |
|---|---|---|

**4.Title and Subtitle**

A Small Community Through Trip Rate Methodology

**5.Report Date**

February 2005

**7. Authors**

Michael Anderson and Yasir Abdullah

**8. Performing Organization Report No.**

UTCA Report 03312

**9. Performing Organization Name and Address**

Civil and Environmental Engineering Department
The University of Alabama in Huntsville
Huntsville, AL 35899

**10.Work Unit No.**

**11.Contract or Grant No.**

DTSR0023424

**12. Sponsoring Agency Name and Address**

University Transportation Center for Alabama
The University of Alabama
Box 870205, 271 H M Comer Mineral Industries Building
Tuscaloosa, AL 35487-0205

**13.Type of Report and Period Covered**

Final Report: January 1, 2003 – January 1, 2004

**14. Sponsoring Agency Code**

**15. Supplementary Notes**

**16. Abstract**

This Report examines a new methodology to predict external trip exchanges in small urban communities. The paper documents a study performed on several small communities in Alabama and includes the methodology, data analysis, model development and validation. The new model presented in this paper contains a community location factor and is shown to provide improved results when compared to existing model.

**17.Key Words**

Student Seminar, TRB Annual Meeting

**18.Distribution Statement**

| 19.Security Class (of this report) | 20.Security Class (of this page) | 21.No. of pages | 22.Price |
|---|---|---|---|
| | | 57 | |

# Contents

# List of Figures

# List of Tables

# Executive Summary

This report documents the development of a model to predict the percent through trips in small urban areas, using limited and readily available data, so the process conforms to the needs of small communities. A non-intrusive data collection methodology, utilizing video cameras, was performed to collect external-external and external-internal traffic movements for several small communities in Alabama. The trip percentages obtained from two sample cities in the data collection effort were used to develop equations to predict traffic flows, which were applied to the other cities in the study. One model developed, utilizing a single equation for traffic prediction, proved superior for traffic prediction when compared to previously developed and published equations. The new model provides the state of Alabama a tool to forecast trip movements in small communities where traditional travel demand models are not available.

# CHAPTER 1
# INTRODUCTION

The transportation planning process is used to provide information for decision makers regarding future transportation activities and requirements. Most of the conventional planning processes that are quite data intensive and costly were designed for use in large urban areas and metropolitan cities. This study focused on developing a model to find the percentage of through trips in small urban areas with limited data, so the decision support process for such communities will become affordable.

Due to a variety of factors, the popularity of small urban areas has increased over the past several decades; in the 1980s, almost 40 percent of the United States population lived in communities with a population between 2,500 and 50,000 (Census, 1983). This generated a greater awareness of transportation problems in small urban areas (Khisty and Rahi, 1990). In Alabama, of the current 4,040,857 residents, only 940,754 (or 23 percent) reside in one of the six cities with more than 50,000 residents (Population, 1998). This implies that the majority of the state's residents live in small urban areas. Unfortunately, the studies that have been conducted to determine the transportation needs of small areas are limited. Moreover, studies that were completed previously to facilitate the synthesis of a through-trip table for urban areas of less than 50,00 population have become outdated and provide unreasonable figures (Anderson, 1999).

The external – external trips and external – internal trips are of great importance for the development of infrastructure and highway bypasses in these small communities. Trips passing from one part of the state through the community to another part of the state are known as external-external trips. The trips entering the community and returning on the same route are known as external-internal trips. The principle methodology for determining through-trip rates cited in the literature is based on the effects of road functional classification, average daily traffic characteristics, vehicle type, route continuity and population of the community (Modlin, 1982). One alternative methodology that has been presented recently focused on the economic impact of neighboring communities to the study city; however, this method has not seen widespread use (Anderson, 1999).

This study intends to develop a methodology to assist transportation professionals in calculating the through-trip rates for a community based on a combination of community characteristics, roadway facility type, and economic characteristics of the surrounding communities.

## 1.1 Background Information

The current transportation planning process relies on urban transportation planning (UTP) models that are typically oriented to provide regional, 24-hour modeling capabilities. In general, these models provide adequate estimates of current and future average daily traffic on a region-wide basis or for major flows such as through freeway interchanges or along principal corridors (Khisty and Rahi, 1990).

Traffic volume is the basic factor in determining the deficiency of transportation corridors. Growth in traffic volume and the corresponding need for new or improved facilities may be anticipated by combining a future land use plan along with an understanding of the relationship between land use and transportation. In major planning studies, these relations are usually established from data available from external, internal, truck, and origin-destination (O-D) surveys. This cannot be done in all small communities because the cost of conducting such surveys to facilitate the development of unique simulation models for individual areas is prohibitive (Modlin, 1982). This escalating cost mandates a synthetic procedure for determining travel patterns in small urban areas.

## 1.2 Objective of The Study

The objective of this study is to examine and develop a methodology to assist transportation professionals in calculating the through-trip rates for a community, based on community characteristics, roadway facility type, and regional economic characteristics. The proposed methodology is developed specifically to benefit communities with populations below 50,000 and is intended to support decisions related to highway construction.

Since communities with populations below 50,000 are not required to maintain a metropolitan planning organization, few of these communities perform any conventional transportation planning activities. Unfortunately, many of these communities are facing transportation questions regarding the current infrastructure available and the merit of proposed highway improvements, most notably highway bypasses (Anderson, 1999). This research effort attempts to develop a methodology to assist these small urban areas in determining the likelihood that travelers on the roadways are destined for the community or beyond.

## 1.3 Report Tasks

For the successful completion of the study, the work was partitioned into four tasks. The tasks were defined to be a literature and community review, collection of data, design of the model, and evaluation of the model. Each of these steps is addressed in more detail in the following sections.

### *1.3.1 Literature, Community and Market Review*

Appropriate literature regarding external through-trip rates was reviewed to better understand existing methods. This review was necessary to evaluate the need of the new through-trip rate methodology for small urban areas. A detailed market study was conducted to find the best possible methods for traffic data collection and the most feasible and affordable devices to be used. In addition, individual communities were reviewed to identify existing roadway traffic counts and available data, as well as potential sources for these data.

### *1.3.2 Data Collection and Analysis*

The data collection effort for the through-trip rate estimating methodology involved video surveillance of vehicles entering and exiting the communities of interest on major roads, and manually recording vehicle license numbers for a few single lane roads. The data were collected for all major entrances and exits for each study community, for approximately ten hours. The vehicle information was examined to determine which vehicles entered and exited along the same roadway, which would be classified as external-internal trips. Vehicles entering and exiting using two different routes where defined as through trips.

### *1.3.3 Model Design*

After collecting and analyzing the data, specific community characteristics and economic considerations were used to develop a model that best reflects community travel patterns. This model is based on a statistical analysis of the data related to community, roadway and economic variables. The through-trip rates were determined to be a function of selected community variables, roadway approach to the community, and surrounding communities.

### *1.3.4 Evaluation of the New Methodology*

The new methodology had to be validated with data collected during the data collection period of this project and evaluated against existing methods used to determine the through-trip rates. This evaluation would determine if the new methodology provides more accurate traffic forecasts, which is the primary data in supporting infrastructure decisions. After this analysis, a conclusion could be reached about the applicability of the new through-trip rate methodology to forecast traffic flow in small urban communities.

## 1.4 Document Organization

The report contains seven chapters. The first chapter discusses the need for developing a methodology to predict through trips for small communities, the objective of the study and the tasks involved. The second chapter provides a review of previous studies conducted in this area and the methods used for data collection and analysis. The third chapter explains the methodology adopted and the practical tasks that were executed for the completion of the study. In the fourth chapter, a statistical regression analysis develops three models to predict the percent trips. The fifth chapter was dedicated to validating the models for a similar community, comparing the results of the new model with existing models, and discussing achieved results. The sixth chapter draws conclusions obtained from this study, and discusses areas for future improvements. The final chapter includes the references for the report.

## CHAPTER TWO
## LITERATURE REVIEW

This report was conducted to assist traffic planners in small urban areas in predicting through trips for the study city. As traffic patterns, mode of transport, economic and social characteristics have undergone a drastic change during recent years, most existing models to forecast through trips have become outdated (Modlin, 1982). Previous studies conducted, and the methods used to develop past models are included in this chapter. A review of transportation planning, statistical analysis methods, and data collection technology will be discussed in the following sections.

### 2.1 Travel Models for Small Urban Communities

Before examining travel models, a brief review of small communities is provided in the following section. The review examines the role of small communities and specific travel model developments directed towards these communities.

#### *2.1.1 Small Communities*

The concept of a small community, also known as a rural community, cannot be narrowly defined. The concept of a rural community depends largely on an individual's experience. Many residents of large cities consider anything outside a large metropolitan area as rural; residents of sparsely populated agricultural areas think of even small cities as urban.

In practice, the U.S. Department of Transportation defines rural in two ways: first, for functional classification and outdoor advertising regulations, rural is anything outside of an area with a population of 5,000. Second, for planning purposes, rural is considered to be anything outside of metropolitan areas with a population of 50,000 or greater (FHWA, 2002). This definition leaves a lot of room for significant differences within these categories. Therefore, it is prudent to describe rural based upon what we see across the country. For the purpose of this report, "rural" or "small community" was defined to be non-metropolitan cities with populations less than 50,000.

#### *2.1.2 Travel Models*

The development of travel models for small urban communities is not a new idea. Several methods have been developed to provide forecasts based on specific input assumptions. Khisty and Rahi, 1987 identified thirteen modeling methodologies applicable for small urban areas, and a recently published report by Martin and McGuckin (1998) provide the basis for this study. They are discussed briefly in this section.

The principle methodology developed by Modlin (1982) to facilitate the synthesis of a through-trip table for small urban areas was a formidable achievement at that time. The effects of functional classification, average daily traffic (ADT), percentage of trucks, route continuity and urban area population were significantly correlated with through-trip patterns. A set of simple multiple regression expressions was developed to the percentage of through-trip ends at each station and distribute these trips among stations. The independent significant variables were the ADT at the destination station, percentage of truck excluding panels and pickups at the destination, percentage of through trips at the destination, and route continuity as a dummy variable (Modlin, 1982).

In a recently conducted study by Anderson (1999), it was found that these models were useful when they met certain specifications, but they could not be used as a reliable model to find the through-trip rates for small areas. From this observation, it was concluded that there was a need to develop a methodology, based on a combination of the community characteristics, facility type, and economic characteristics of the surrounding communities.

Low's model suggests that traffic volumes are determined one link at a time, primarily as a function of the relative probability that one link will be used in preference to another (Low, 1972). Interzonal trip probabilities were assigned to the network, using traffic assignment procedures, to produce estimated trip probabilities on a link-by-link basis. These individual link probabilities were used, along with actual traffic counts, to develop regression equations correlating interzonal trip probabilities and traffic counts. These regression equations are used to estimate link volumes, using the interzonal trip probabilities obtained with horizon year trip generation results (Khisty and Rahi, 1990).

In 1983, Neumann et al. proposed an alternate modeling approach directed towards smaller urban areas. The model directly estimates area wide, all-purpose trip production rates (Nuemann et al., 1983). The model distributes and assigns zonal socioeconomic variables directly to the study area network. To use the model, external trips were deducted from the total ground counts to obtain internal trips of the area. Entering the ground counts as the dependent variable and socio-economic variables as independent variables develops linear regression models. The resulting regression coefficients were the estimates of the area wide, all-purpose trip production rates (Khisty and Rahi, 1990). As with the model developed by Low (1972), this regression model is again used to forecast horizon year traffic.

Khisty and Al-Zahrani proposed another small area model, referred to as the internal volume-forecasting (IVF) model based on Low's model. This model incorporates changes to reduce the errors in Low's model. The model recommends replacing the zonal production and attraction characteristics with direct estimates of trip productions and attractions, in other words it uses employees per zone and job positions per zone as the basis for trip generation. In addition, the model recommends the use of total employment available in the study area as a factor in developing the trip interchange index assigned to the network (Khisty and Rahi, 1990).

It can be seen that all the three models require high quality, system wide traffic counts and assume stability of mathematical relationships between development and horizon years. For many rural communities, traffic count data are not readily available. Furthermore, trip generation equations and distribution parameters will likely change significantly in small areas with infrastructure changes, negating the previously developed regression equations. Therefore, a new model that is not dependent on existing traffic counts and is responsive to significant traffic changes is desired (Anderson and Souleyrette, 2000).

The latest model developed for external travel estimation was reprinted in the NCHRP Report 365 (Martin and McGuckin, 1998). A model developed in North Carolina estimated through-trip ends at a station on the cordon of a study area. The model used functional classification, the ADT at the external station, the percentage of trucks (excluding vans and pick ups), the percentage of vans and pickups, and the population of the study area (Martin and McGuckin, 1998). Equation 2-1 gives the percent through trips at an external station i.

$$Y_i = 76.76 + 11.22 \times I - 25.74 \times PA - 042.18 \times MA + 0.00012 \times ADT_i + 0.59 \times PTKS_i - 0.48 \times PPS_i - 0.000417 \times POP \quad \text{(Eq. 2-1)}$$

Where:
$Y_i$ = percentage of the ADT at external station I, that are through trips,
I = interstate (0 or 1),
PA = principal arterial (0 or 1),
MA = minor arterial (0 or 1),
$ADT_i$ = average daily traffic at external station i,
$PTKS_i$ = percentage of trucks excluding vans and pickups at external station i,
$PPS_i$ = percentage of vans and pickups at external station i, and
POP = population inside the cordon area.

In Equation 2-1, an external station can be only one of the three functional classifications. For each functional class, the value of the variable is 1; for the other two, the value will be 0. The distribution of estimated through-trip ends from an external station to each one of other external stations is required to obtain a through trip matrix. Modlin (1982) developed three equations based on functional classes to estimate the distribution of through trips that enter the analysis area at an origin external station (i) and pass through a destination station (j). For each interchange, the functional class of the destination station dictates which equation is to be used (Martin and McGuckin, 1998).

Interstate:

$$Y_{ij} = -2.70 + 0.21 \times PTTDES_j + 67.86 \times RTECON_{ij} \quad \text{(Eq. 2-2)}$$

Principal Arterial:

$$Y_{ij} = -7.40 + 0.55 \times PTTDES_j + 24.68 \times RTECON_{ij} + 45.62 \times ADT_j / \sum ADT_j \quad \text{(Eq. 2-3)}$$

Minor Arterial:

$$Y_{ij} = -0.63 + 86.68 \times ADT_j / \sum ADT_j + 30.04 \times RTECON_{ij} \quad \text{(Eq. 2-4)}$$

Where:
$Y_{ij}$ = percentage distribution of through-trip ends from origin station i to destination station j,
$PTTDES_j$ = percentage through-trip ends at destination station j,
$RTECON_{ij}$ = route continuity between stations i and j: 1 = Yes, 0 = No, and
$ADT_j$ = average daily traffic at destination station j.

Station to station trip movements can be estimated using a simple factoring procedure, which uses an external station's portion of the total trips. Some efforts have to be made to ascertain the existing through movement patterns either by reference to earlier studies of the area or by general observations (Martin and McGuckin, 1998). The key for using this model is knowing the ADT by direction for trucks and autos at each external station.

## 2.2 Regression Analysis

Regression analysis is the most widely used statistical technique for investigating and modeling the relationship between variables (Montgomery and Peck, 1992). In a statistical problem, if we let the dependent variable (response) be *Y* and the independent variable (regressor) is *X*, the reasonable form of linear relationship between the two variables is shown in Equation 2-5.

$$Y = \beta_0 + \beta_1 X \qquad \text{(Eq. 2-5)}$$

The concept of regression analysis deals with finding the best relationship between *X* and *Y*, quantifying the strength of that relationship, and the use of methods that allow for prediction of the response values given values of regressor *X*. In the equation, $\beta_0$ is known as the intercept and $\beta_1$ as the slope. The statistical error between the observed value of *Y* and the straight line ($\beta_0 + \beta_1 X$) is represented by $\in$. The more plausible regression model is represented as in Equation 2-6

$$Yi = \beta_0 + \beta_1 Xi + \in i \qquad \text{(Eq. 2-6)}$$

In most applications, as the case with this study, there will be more than one regressor that helps to explain *Y*. The multiple regression equation is used in such situations. It is shown in Equation 2-7.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 +. \ldots+ \beta_n X_n + \in \qquad \text{(Eq. 2-7)}$$

where *Y* is the natural dependent variable, and $X_1, X_2, \ldots, X_n$ are the independent variables. The resulting analysis is termed as multiple linear regression analysis. The term linear is used to indicate that the model is linear in the parameters $\beta_0$, $\beta_1$, .., $\beta$n, not because *Y* is a linear function of *X*'s. The important objective of regression analysis is to estimate the unknown parameters in the regression model (Montgomery and peck, 1992 and Walpole et al., 2002).

The second phase of regression analysis is called model adequacy checking, in which the appropriateness of the model is studied and the quality of the fit is ascertained. The outcome of the adequacy checking may indicate either that the model is reasonable or that the original fit must be modified. Thus, the regression analysis is an iterative procedure, in which data lead to model selection, and a fit of the model to the data is produced. The quality of the fit is then investigated, leading either to modification of the model or adoption of the model (Walpole et al., 2002).

## 2.3 Methods and Equipment Used for Data Collection

Currently, many techniques are used for collecting transportation data in the field. For manual forms of data collection, methods range from using pencil and paper to, more recently, laptop computers running specialized software programs. In recent times, travel time has become an emerging factor in transportation studies. A common method for determining travel time is to place people in the field to observe license plate numbers at different points, and then key them into portable computers. Software that matches license plates from two different source files is then used to find matches and calculate travel times (Washburn, 2002 and Washburn and Nihan, 1997)

### *2.3.1 License Plate Readers (LPR)*

During the duration of this study, it was found that there are specialized license plate readers (LPR) available in the market. These LPRs use a unique, patented approach to capture license plate images. By using external trigger signals, the system can read both retro-reflective and non-retro reflective license plates, meaning that it will capture images of 100 percent of the vehicles entering and exiting the area (Army, 2002). The LPR automatically generates and archives lane and data information and a time stamp for each image. In addition to the alphanumeric read, the LPR also provides state and country information. Its menu-driven operator interfaces are customized for specific applications and make it extremely user-friendly. These systems are usually used for long-term installations and are highly expensive.

### *2.3.2 Speech Recognition of License Plate Data*

Speech recognition for on-site license plate data collection is another method that has been studied recently. It has emerged as a potentially viable tool for data collection with very distinct advantages over some existing manual methods (Washburn, 2002). The speech recognition application developed by Washburn (2002) was actually a combination of the two existing methods for performing license plate data collection: voice recorder entry and portable computer keyboard entry.

Voice recorders offer a convenient way to collect data, as speaking is the only continuous task that is required. For the license plate data collection, they offer the advantage of allowing the user to keep his/her eyes and attention focused on the license plates at all times. This leads to larger sample sizes and greater accuracy rates. This especially becomes more prevalent in situations where the observer may be required to use binoculars to see the license plates clearly. The main difficulty with this method arises in associating an accurate time-stamp with each plate entry on a consistent basis (Washburn, 2002 and Washburn and Nihan, 1997).

Portable computer keyboard entry methods are also suited to the same purpose. The user can type license plate characters into a special program that automatically time-stamps the entries, based upon the internal computer clock. The use of specialized software programs also eliminates the need for any further data transcribing after the data collection effort (Washburn, 2002). A quality software program will allow the data collectors to immediately post process (e.g., plate match) the license plate data without any further modifications. The main drawback of this method is that the accuracy of license plate entries is often directly correlated with the typing ability of the data collector (Washburn, 2002).

The developed speech recognition application combines the advantages of both the voice recorder and keyboard entry techniques and eliminates each one's disadvantages, while also providing a hands-free form of data entry (Washburn, 2002). Development of this type of application did prove to be very challenging. The essential components of this application are the license plate data collection program, speech recognition engine, grammar and vocabulary file, and the user interface that represents the “front-end” for the main data collection program (Washburn, 2002).

When we look at the average accuracy rate for each method, we can see that the keyboard entry method has a 97 percent, the voice recognition has 95 percent, and speech recognition method has 93% accuracy (Washburn, 2002). This means the accuracy rates obtained with the keyboard entry method are still very good, on an average about four percent better than those obtained with speech recognition. On the other hand, the execution speed of the data collection program still remains a constraint for the voice recognition method, which means that during periods of high traffic flow on the arterial, the program is unable to process the speech input at the same rate the data collector is able to speak the license plates (Washburn, 2002 and Washburn and Nihan, 1997).

Interestingly it was observed that there is a larger pool of correct license plates yielded by the speech recognition method than the keyboard entry method (Washburn, 2002). The speech recognition method averaged about an 11 percent higher sample rate than the keyboard entry method. As a result, for many of the data collection exercises a larger correct sample was collected with speech recognition than with the keyboard, despite a slightly lower accuracy rate with the speech recognition method. Although speech recognition has its own advantages, this technology is still relatively immature. As technology matures and computing power increases, it may see increased utility over keyboard entry for field data collection. At that time, the need for using a keyboard for input could become a thing of the past (Washburn, 2002 and Washburn and Nihan, 1997).

### *2.3.3 Video License Plate Data*

Video license plate surveys have been in use for more than a decade. In this method, video cameras of adequate quality are properly set up, and then the license plate images are captured for post collection data reduction (Shuldiner, 1994). Problems arising for setting up video cameras have been resolved in various studies. The data reduction activity, however, is quite tedious. In most cases, the letters and numbers on a license plate are manually transcribed to a data file. This manual process is time consuming, and expensive (Shuldiner, 1994). The transcription step contains ample opportunities for mistakes. Although automated license plate readers are being implemented with success elsewhere, their dependence on high-end equipment makes them too expensive for most applications (Shuldiner, 1994 and Shuldiner and Hudson, 1996).

Video license plate surveys involve four basic steps: survey planning, data collection, and data reduction and analysis, and results reporting. Approximately 50 percent of the total cost of a video license plate survey is spent on only the data reduction step (Gupta et al., 2002). Data reduction translates the tapes made during data collection into files that are suitable for input into analysis. This process usually requires a person to watch each videotape, then manually enter time and plate information into a computer file. In their work, Gupta and his team focused on stream lining the data reduction step. Their product is supposed to be a process that is less labor intensive, more accurate, and more cost-effective. Image processing and optical character recognition technologies are being employed in this field. By using specialized hardware and good-quality video (e.g., Super VHS), it is possible to use computers to process video data for license plate recognition (Gupta et al., 2002).

Image processing is a tool that helps immensely in training computers to act like humans. One example is reading license plate data (Gupta et al., 2002). Optical character recognition technology involves reading text from number plates and translating the images into a form that the computer can manipulate. Most OCR systems use a combination of hardware and software to recognize characters, although some inexpensive systems do it entirely through software (Webopedia, 2002). A systematic approach to OCR includes segmentation of the image into individual characters, feature extraction, and character matching. By using these techniques the video license plate data reduction (VLPDR) software was developed, which can identify video frames that contain vehicles and discard the remaining frames.

VLPDR can locate and read the time stamps in most of these frames. Although VLPDR cannot read the license plate numbers into a data file, this final step is made easier by a user-friendly graphical user interface (GUI). VLPDR saves a significant amount of manual data reduction; it can reduce the time involved by an estimated 60 percent (Gupta et al., 2002). However, the amount of labor savings can vary, depending on the parameters chosen by the user (Gupta et al., 2002).

### 2.4 Summary of Literature Review

Taking into consideration all the available methods for data collection and looking into the feasibility of obtaining them within the limits of this study, both advanced technological and manual methods were used during this project. High-resolution video cameras were used to capture the images of number plates of vehicles moving on high-density multi-lane roads, and paper and pencil were used on minor single lane roads. The obtained information was entered into a database using desktop computers. The data was matched and analyzed using specialized software.

# CHAPTER 3
# METHODOLOGY

This chapter presents the methodology adopted and data collection activity required to develop the through-trip model for small urban communities. The first section describes the study's technical approach. The second section discusses the selected communities and their respective demographics. The third section documents the data collection effort, which was the most challenging part of the study. The fourth section describes the data processing effort. The chapter concludes with a summary of the methodology.

## 3.1 Technical Approach

The development of the methodology to predict the through-trip rates for small urban communities used a non-intrusive data collection method for several communities within Alabama. The communities were selected depending on a variety of considerations, specifically proximity to other major cities, population, economic considerations, and geographical location.

Each community was considered as a single zone, with various external stations connecting the community to the state, removing internal-internal trips from our study. The external-external trips were trips passing from one part of the state through the community to another part of the state. The internal-external trips were trips entering the community and returning to their origin.

Video surveillance was used to collect vehicle travel information for the study communities. Video surveillance allowed the data collection team to automate the process of recording vehicles as they entered or departed the community being studied. This process reduced the inconvenience to passing motorists as the vehicles were allowed to pass at highway speeds and guaranteed accurate recording of vehicles, as the data collectors were not busy with manual recording of data. The collected data was then manually transcribed into a computer data base file. After the collection of data, roadway, community and economic characteristics were used to develop a methodology to predict external trip rates.

## 3.2 Community Review

Eight communities in Alabama were selected as case study cities, and for model development and validation. For the selected cities, a site visit was performed to determine the roads in the community that would be considered as major entrance/exit points for the community. The selected cities and their demographics are shown in Table 3-1.

**Table 3-1 Community demographics**

| Community | Population |
| --- | --- |
| Alexander City | 15,008 |
| Arab | 7,174 |
| Hartselle | 12,019 |
| Roanoke | 6,563 |
| Russellville | 8,971 |
| Sylacauga | 12,616 |
| Talladega | 15,143 |
| Troy | 13,935 |

The communities selected for the study are located throughout Alabama. The population of each community is between 5,000 and 15,000. Available social and economic data help make the new model more practical, because the existing models were developed for larger communities, or they are outdated due to changes in social and economical conditions.

The ADT for the major through routes was obtained from the 1996 Alabama Traffic Flow Map, prepared by the Alabama Department of Transportation. The ADT values for the selected roadways are shown in Table 3-2. All the roads used in this study were either U.S. or State highways (Traffic, 1997). Additionally, the number of lanes (NL) on the road at each collection point was recorded, as it will influence the traffic characteristics of the respective roadway.

**Table 3.2 ADT and number of lanes**

| Urban Area | Cordon Station | ADT | NL |
|---|---|---|---|
| **ALEXANDER CITY** | AL 22 W | 7130 | 2 |
| | AL 63 N | 8730 | 2 |
| | US 280 W | 12,510 | 4 |
| | AL 22 E | 6,100 | 2 |
| | AL 63 S | 5,940 | 2 |
| | US 280 E | 12,070 | 4 |
| **ARAB** | US 231 N | 13060 | 4 |
| | AL 69 W | 6620 | 2 |
| | US 231 S | 6720 | 4 |
| | AL 69 E | 8950 | 4 |
| **HARTESELLE** | US 31 N | 19790 | 4 |
| | AL 36 W | 7220 | 2 |
| | US 31 S | 14410 | 4 |
| | AL 36 E | 10580 | 2 |
| **ROANOKE** | US 431 N | 5960 | 4 |
| | AL 22 W | 3120 | 2 |
| | US 431 S | 3750 | 2 |
| | AL 22 E | 3670 | 2 |
| **RUSSELLVILLE** | US 43 N | 18,130 | 4 |
| | AL 24 W | 4,200 | 2 |
| | US 43 S | 11,630 | 4 |
| | AL 24 E | 8,080 | 4 |
| **SYLACAUGA** | US 280 W | 17,300 | 4 |
| | US 231 S | 5,420 | 2 |
| | US 280 E | 9,410 | 4 |
| | AL 148 E | 3,700 | 2 |
| | AL 21 N | 9,570 | 2 |
| **TALLADEGA** | AL 77 N | 10,610 | 2 |
| | AL 21 S | 7,280 | 2 |
| | AL 77 S | 3,700 | 2 |
| | AL 22 N | 9,990 | 2 |
| **TROY** | US 231 N | 17,930 | 4 |
| | AL 29 S | 3,600 | 2 |
| | AL 87 S | 6,910 | 2 |
| | US 231 S | 17,630 | 4 |
| | AL 29 N | 4,480 | 2 |

The observation points were selected in such a way that internal-internal trips were minimized. In addition, great care was given to position the data collector in an area where the vehicle's speed are minimized, as this helps in viewing the license plates with more clarity through the video screen.

### 3.3 Data Collection and Project Management

The first practical step in the development of the model was the data collection. Twenty-one students of the University of Alabama in Huntsville were hired for this purpose. Eight video cameras (VHS-C) and video cassettes, batteries, and chargers were obtained. After conducting an initial check and experimenting with the functionality of the camera systems, the data collection team was trained to use the system.

Scheduling the trips, organizing the data collection team and executing the job in the field were challenging as well as fulfilling tasks. All precautions were taken to avoid inconvenience that may occur during the process. Since the filming of license plates of vehicles may create confusion among the public, concerned agencies like the Federal Highway Administration, Alabama Department of Transportation and city police officials were informed about the event ahead of time. Some local newspapers published articles about the survey to inform the public.

In each city, the filming of cars started in early morning and lasted until evening, approximately ten hours of daytime traffic was recorded by the camcorders. Each data collector was assigned to a pre-determined roadway station. He or she was in charge of one side of the assigned roadway. The effect of sun on the image was very evident, especially in the east-west direction; the inclined position of the sun at certain times caused blur images. The camera was moved and adjusted to avoid this effect as much as possible. Other difficulties faced during the data collection effort were rain and fog.

One person within each study community was appointed as team coordinator. He or she was in charge of meeting with police officials, inspecting all data collection stations on the roadway, providing onsite help for personnel, and providing temporary relief for breaks. Recharging batteries and replacement of faulty equipment were some of the problems that needed immediate attention. Wireless communication was useful in keeping team members connected to the coordinator. The effort of some team members who volunteered to write down the number plates manually was of great help, especially when equipment failed.

The survey trips were conducted on working days Monday – Thursday, so the regular community traffic flow was obtained. Holidays and Fridays were avoided because people were assumed to be engaged in special events on those days. Overall, the data collection process was smooth and the local communities were generally cooperative.

### 3.4 Data Reduction and Processing

The video license plate images captured at the entry/exit locations were reduced and processed at a lab in the Civil and Environmental Engineering Department of the University of Alabama in Huntsville. Two Pentium III 500 MGHZ with 256MB memory were used for this purpose. A popular computer hardware device for capturing video images and video analysis software were used to view license plate data.

The flexibility of the hardware and software allowed data entry personnel to stop the recorded image when a vehicle number plate was visible. With the vehicle image stopped on the screen, the data entry personnel were able to step forward and backward viewing every frame (taken at a rate o 30/second) making it possible to obtain the best image quality. The letters and numbers on a license plate were entered into a database. This manual process, as indicated in (Gupta et al., 1996) was time consuming, and expensive, however, much less expensive than using automated LPR technology.

An average of 6-10 hours was required to view each ninety minute videocassette and record the data in the data file. The time depended heavily on the clarity of the recorded images, the number of lanes at each station, and the density of the traffic. With approximately 10 hours of data collected for all external stations in each of the eight cities, the total time to reduce the data was almost 4,000 hours.

Within the database, six fields were created in the design table view. The first field was an internal identification number. It provided the total number of vehicles passing the station according to their sequence of arrival. The second field was the license plate numbers of the vehicles, which was the most crucial and essential factor of the entire data collection process. This factor was used to match the trips made by vehicles at each station, and calculate the percent of internal-external and external-external trips. The third field was a description of the vehicle including the make, color and model. The second and third fields were filled manually as text form. The fourth, fifth and sixth were used to identify vehicle type (Car, SUV/ Van/ Pick up, or Truck). All these fields were filled manually through a check box.

Once all the videocassettes were viewed and the license plate data were recorded in the database, the next step was to edit and remove obvious errors and combine all data at each station into a single table. The percentages of cars, SUVs/ vans/ pick ups, and commercial trucks were determined based on the check-box data. As the ADT obtained was the average from both incoming and outgoing directions at each station, the average percentage of trucks, SUVs and cars from both the entering and exiting stations was used in this study. Table 3-3 shows the distribution of vehicles at each location for the three cities.

**Table 3-3 Percent distribution of vehicle classes**

| Urban Area | Cordon Station | Car (%) | SUV/Van/Pick-up (%) | Truck (%) |
|---|---|---|---|---|
| ALEXANDER CITY | US 280 W | 54.04 | 33.61 | 12.35 |
| | AL 22 W | 47.95 | 42.20 | 9.81 |
| | AL 63 S | 48.42 | 41.87 | 9.70 |
| | US 280 E | 42.63 | 48.13 | 9.19 |
| | AL 22 E | 46.53 | 45.54 | 7.91 |
| | AL 63 N | 50.86 | 39.73 | 9.41 |
| ARAB | US 231 N | 49.50 | 45.90 | 4.60 |
| | AL 69 W | 47.10 | 43.40 | 9.50 |
| | US 231 S | 48.50 | 45.50 | 6.00 |
| | AL 69 E | 47.30 | 45.30 | 7.40 |
| HARTESELLE | US 31 N | 51.10 | 44.40 | 4.50 |
| | AL 36 W | 41.80 | 52.60 | 5.60 |
| | US 31 S | 46.90 | 49.00 | 4.10 |
| | AL 36 E | 45.00 | 50.20 | 4.80 |
| ROANOKE | US 431 N | 46.00 | 46.50 | 7.50 |
| | AL 22 W | 51.50 | 40.30 | 8.20 |
| | US 431 S | 50.60 | 36.60 | 12.80 |
| | AL 22 E | 42.70 | 49.80 | 7.50 |
| RUSSELLVILLE | US 43 N | 47.77 | 46.89 | 5.33 |
| | AL 24 W | 46.69 | 47.40 | 5.91 |
| | US 43 S | 46.18 | 44.28 | 9.54 |
| | AL 24 E | 46.31 | 44.57 | 9.12 |
| SYLACAUGA | US 280 W | 42.51 | 44.84 | 12.65 |
| | US 231 S | 50.99 | 38.91 | 10.10 |
| | US 280 E | 47.47 | 43.07 | 9.46 |
| | AL 148 E | 47.47 | 42.51 | 10.02 |
| | AL 21 N | 49.10 | 43.03 | 7.87 |
| TALLADEGA | Due to technical problems in the video taping, the data for Talladega, Alabama was rendered unusable for the purposes of this study. Therefore, this city was removed from further consideration in the project. | | | |
| TROY | US 231 N | 42.39 | 40.02 | 17.59 |
| | AL 29 S | 45.59 | 43.35 | 11.05 |
| | AL 87 S | 54.85 | 38.47 | 6.68 |
| | US 231 S | 44.51 | 39.76 | 15.72 |
| | AL 29 N | 46.92 | 40.36 | 12.72 |

The next step in the data processing was to remove all the vehicles with unclear number plates from the database, as these vehicles created unwanted errors during the matching process. For example, if the license plate was not visible due to a blurred image, the data entry person might enter ‘BLUR’ in the number plate field. When the software matched tables, this entry was repeated several times, and consequently matched a large number of vehicles, which in reality were completely different vehicles. For this reason, these unwanted entries were removed from the data table. This was done by running a query to list the license plates in ascending order, and then manually selecting all the unwanted entries and deleting them from the table. The new edited table was saved with a different table name.

All the edited tables were brought into a single Access database. Then a query was run to match two tables; one of these tables contained incoming vehicles and the other table outgoing vehicles. Each incoming station was matched with all the outgoing stations from the community. The trips using the same roadway were identified as internal-external or external-internal trips; and the trips passing to other outgoing stations were identified as external-external trips. Since the internal-internal trips were out of scope of this study, no effort was made to collect these trips. The percent of trips from each incoming route to each outgoing route was calculated by dividing the matching number of vehicles on the incoming route and each outgoing route with the total number of matching vehicles from that incoming road to all the outgoing roads. The obtained percentages of trips for the three communities are shown in Tables 3-4 through 3-10.

**Table 3-4. Percent trips between stations (Alexander City)**

| Urban Area | Origin Station | Destination Station | % Of Vehicles ($Y_{ij}$) | Type of Trip |
|---|---|---|---|---|
| Alexander City | US 280 W | US 280 W | 40 | E-I, I-E |
| | | AL 22 W | 5 | E-E |
| | | AL 63 S | 12 | E-E |
| | | US 280 E | 37 | E-E |
| | | AL 22 E | 2 | E-E |
| | | AL 63 N | 4 | E-E |
| | AL 22 W | AL 22 W | 65 | E-I, I-E |
| | | AL 63 S | 5 | E-E |
| | | US 280 E | 10 | E-E |
| | | AL 22 E | 3 | E-E |
| | | AL 63 N | 3 | E-E |
| | | US 280 W | 13 | E-E |
| | AL 63 S | AL 63 S | 72 | E-I, I-E |
| | | US 280 E | 11 | E-E |
| | | AL 22 E | 2 | E-E |
| | | AL 63 N | 4 | E-E |
| | | US 280 W | 9 | E-E |
| | | AL 22 W | 1 | E-E |
| | US 280 E | US 280 E | 55 | E-I, I-E |
| | | AL 22 E | 1 | E-E |
| | | AL 63 N | 3 | E-E |
| | | US 280 W | 25 | E-E |
| | | AL 22 W | 3 | E-E |
| | | AL 63 S | 14 | E-E |
| | AL 22 E | AL 22 E | 33 | E-I, I-E |
| | | AL 63 N | 13 | E-E |
| | | US 280 W | 14 | E-E |
| | | AL 22 W | 3 | E-E |
| | | AL 63 S | 14 | E-E |
| | | US 280 E | 23 | E-E |
| | AL 63 N | AL 63 N | 6 | E-I, I-E |
| | | US 280 W | 11 | E-E |
| | | AL 22 W | 0 | E-E |
| | | AL 63 S | 28 | E-E |
| | | US 280 E | 28 | E-E |
| | | AL 22 E | 28 | E-E |

**Table 3-5. Percent trips between stations (Arab)**

| Urban Area | Origin Station | Destination Station | % Of Vehicles($Y_{ij}$) | Type of Trip |
|---|---|---|---|---|
| Arab | US 231 N | US 231 N | 62 | E-I, I-E |
| | | AL 69 W | 9 | E-E |
| | | US 231 S | 16 | E-E |
| | | AL 69 E | 13 | E-E |
| | AL 69 W | AL 69 W | 48 | E-I, I-E |
| | | US 231 S | 11 | E-E |
| | | AL 69 E | 25 | E-E |
| | | US 231 N | 16 | E-E |
| | US 231 S | US 231 S | 47 | E-I, I-E |
| | | AL 69 E | 15 | E-E |
| | | US 231 N | 30 | E-E |
| | | AL 69 W | 8 | E-E |
| | AL 69 E | AL 69 E | 63 | E-I, I-E |
| | | US 231 N | 17 | E-E |
| | | AL 69 W | 11 | E-E |
| | | US 231 S | 9 | E-E |

**Table 3-6 Percent trips between stations (Hartselle)**

| Urban Area | Origin Station | Destination Station | % Of Vehicles($Y_{ij}$) | Type of Trip |
|---|---|---|---|---|
| Hartselle | US 31 N | US 31 N | 27 | E-I, I-E |
| | | AL 36 W | 12 | E-E |
| | | US 31 S | 38 | E-E |
| | | AL 36 E | 23 | E-E |
| | AL 36 W | AL 36 W | 42 | E-I, I-E |
| | | US 31 S | 20 | E-E |
| | | AL 36 E | 30 | E-E |
| | | US 31 N | 8 | E-E |
| | US 31 S | US 31 S | 65 | E-I, I-E |
| | | AL 36 E | 13 | E-E |
| | | US 31 N | 15 | E-E |
| | | AL 36 W | 7 | E-E |
| | AL 36 E | AL 36 E | 75 | E-I, I-E |
| | | US 31 N | 5 | E-E |
| | | AL 36 W | 9 | E-E |
| | | US 31 S | 11 | E-E |

**Table 3-7 Percent trips between stations (Roanoke)**

| Urban Area | Origin Station | Destination Station | % Of Vehicles ($Y_{ij}$) | Type of Trip |
|---|---|---|---|---|
| Roanoke | US 431 N | US 431 N | 77 | E-I, I-E |
| | | AL 22 W | 7 | E-E |
| | | US 431 S | 11 | E-E |
| | | AL 22 E | 5 | E-E |
| | AL 22 W | AL 22 W | 70 | E-I, I-E |
| | | US 431 S | 5 | E-E |
| | | AL 22 E | 5 | E-E |
| | | US 431 N | 20 | E-E |
| | US 431 S | US 431 S | 43 | E-I, I-E |
| | | AL 22 E | 5 | E-E |
| | | US 431 N | 46 | E-E |
| | | AL 22 W | 7 | E-E |
| | AL 22 E | AL 22 E | 62 | E-I, I-E |
| | | US 431 N | 24 | E-E |
| | | AL 22 W | 8 | E-E |
| | | US 431 S | 6 | E-E |

**Table 3-8 Percent trips between stations (Russellville)**

| Urban Area | Origin Station | Destination Station | % Of Vehicles($Y_{ij}$) | Type of Trip |
|---|---|---|---|---|
| Russellville | US 43 N | US 43 N | 67 | E-I, I-E |
| | | AL 24 W | 14 | E-E |
| | | US 43 S | 4 | E-E |
| | | AL 24 E | 14 | E-E |
| | AL 24 W | AL 24 W | 65 | E-I, I-E |
| | | US 43 S | 1 | E-E |
| | | AL 24 E | 22 | E-E |
| | | US 43 N | 12 | E-E |
| | US 43 S | US 43 S | 57 | E-I, I-E |
| | | AL 24 E | 2 | E-E |
| | | US 43 N | 7 | E-E |
| | | AL 24 W | 35 | E-E |
| | AL 24 E | AL 24 E | 64 | E-I, I-E |
| | | US 43 N | 17 | E-E |
| | | AL 24 W | 14 | E-E |
| | | US 43 S | 5 | E-E |

**Table 3-9 Percent trips between stations (Sylacauga)**

| Urban Area | Origin Station | Destination Station | % Of Vehicles ($Y_{ij}$) | Type of Trip |
|---|---|---|---|---|
| Sylacauga | US 280 W | US 280 W | 49 | E-I, I-E |
| | | US 231 S | 25 | E-E |
| | | US 280 E | 8 | E-E |
| | | AL 148 E | 4 | E-E |
| | | AL 21 N | 14 | E-E |
| | US 231 S | US 231 S | 71 | E-I, I-E |
| | | US 280 E | 2 | E-E |
| | | AL 148 E | 1 | E-E |
| | | AL 21 N | 15 | E-E |
| | | US 280 W | 11 | E-E |
| | US 280 E | US 280 E | Insufficient Data | E-I, I-E |
| | | AL 148 E | Insufficient Data | E-E |
| | | AL 21 N | Insufficient Data | E-E |
| | | US 280 W | Insufficient Data | E-E |
| | | US 231 S | Insufficient Data | E-E |
| | AL 148 E | AL 148 E | 69 | E-I, I-E |
| | | AL 21 N | 15 | E-E |
| | | US 280 W | 8 | E-E |
| | | US 231 S | 7 | E-E |
| | | US 280 E | 1 | E-E |
| | AL 21 N | AL 21 N | 72 | E-I, I-E |
| | | US 280 W | 5 | E-E |
| | | US 231 S | 16 | E-E |
| | | US 280 E | 1 | E-E |
| | | AL 148 E | 6 | E-E |

**Table 3-10 Percent trips between stations (Troy)**

| Urban Area | Origin Station | Destination Station | % Of Vehicles ($Y_{ij}$) | Type of Trip |
|---|---|---|---|---|
| Troy | US 231 N | US 231 N | 36 | E-I, I-E |
| | | AL 29 S | 14 | E-E |
| | | AL 87 S | 29 | E-E |
| | | US 231 S | 21 | E-E |
| | | AL 29 N | 0 | E-E |
| | AL 29 S | AL 29 S | 83 | E-I, I-E |
| | | AL 87 S | 9 | E-E |
| | | US 231 S | 4 | E-E |
| | | AL 29 N | 2 | E-E |
| | | US 231 N | 2 | E-E |
| | AL 87 S | AL 87 S | 84 | E-I, I-E |
| | | US 231 S | 3 | E-E |
| | | AL 29 N | 2 | E-E |
| | | US 231 N | 6 | E-E |
| | | AL 29 S | 5 | E-E |
| | US 231 S | US 231 S | 54 | E-I, I-E |
| | | AL 29 N | 3 | E-E |
| | | US 231 N | 3 | E-E |
| | | AL 29 S | 17 | E-E |
| | | AL 87 S | 24 | E-E |
| | AL 29 N | AL 29 N | 75 | E-I, I-E |
| | | US 231 N | 2 | E-E |
| | | AL 29 S | 8 | E-E |
| | | AL 87 S | 11 | E-E |
| | | US 231 S | 4 | E-E |

The problems faced during data processing were many; the most relevant was human error. Especially when reading and recording data, the accuracy depended heavily on how each individual perceived the vehicle. Even though data entry personnel were given specific instructions, the diversity of human perception had an impact on the obtained data. For example, a vehicle might be recorded as a maroon car, with tag number 52B478C. Another person may have seen it as a red car with tag number 528478C or 52B4780. Because of this, only the most legible data were used.

In addition, it was decided not to consider the description field for the matching process, because it had high variation. The situation was completely different when it came to the few places where the license plate numbers were written manually. In those cases, the output was comparatively better, but it could only be applied on routes with lower speeds and volumes.

## 3.5 Summary of the Methodology

The methodology to collect and develop data for the case study cities was presented in this chapter. The next chapter uses the data to develop a new prediction model. The validation of the model is examined in a later chapter. For this study, two communities located at different regions of Alabama, were selected for the model development; the other communities were used for model validation.

# CHAPTER FOUR
# STATISTICAL ANALYSIS AND MODEL DEVELOPMENT

The development of a model to predict through trips for small urban communities could be dependent on several variables. The external survey conducted using camcorders provided information on trip interchanges. Apart from the trip origin/destination, a number of other variables were used to model the external travel. The model development was conducted in three stages: first, a model was developed to estimate through trips at external stations; second, a model was developed to distribute the through trips between stations; and third, a combined model was developed to predict the percentage of trips from one station to another, whether it was internal-external or external-external.

Since this study focused on small communities, and the roadways passing through such communities were either U.S. or state highways, functional class was not considered an important variable to include in the study, deviating from previous studies where functional class was of high importance. However, all of the routes were either two-lane or four-lane highways, which was considered an important variable to consider. As the number of lanes increased, the traffic movement increased on that road and it usually acted as a corridor for the community to the outside world. There was also an observed increase in the flow of traffic on highways which connected the community to neighboring major attraction points. These attraction points were identified as large businesses, major cities or transportation facilities.

Before beginning the statistical analysis section, a brief description is given of the potential predictors or model variables. Variables were identified by reviewing previous studies, and incorporating existing roadway and socio-economic characteristics. In addition, the location of concerned communities and its accessibility to nearby major cities or expressways was also taken into consideration.

## 4.1 Initial Predictors

The initial predictors for three models are mentioned in the following three sections.

### *4.1.1 Estimation of Through Trips at External Stations*

Previous research efforts have shown that the following variables influence through-trip percent.

- Average Daily Traffic ($ADT_i$) at the external station i.
- Population (POP) of the study community.
- Percentage of trucks ($PT_i$) at the external station i.
- Percentage of SUVs, Vans and Pickups ($PSVP_i$) at the external station i.

During the community review, two additional predictors were added that were assumed to influence the through-trip pattern.

- A nearby major city or expressway ($NMC_i$) acted as a major point of attraction. It was observed that the roads connecting these attraction points acted as the corridor from the community to the outside world. The previously developed models were not able to address this factor as they generally focused on trips in a single larger community. This could be one of the reasons why those models could not be implemented effectively for

- small communities. For this classification, the value of the functional variable will be either 0 or 1, depending on whether a nearby attraction exists.
- Number of lanes ($NL_i$) of the cordon of the external station i.

### *4.1.2 Distribution of Through Trips Between Stations*

From the previous research, the following predictors were identified to have an influence on the distribution of through trips between stations.

- Percentage of through trips ($Y_j$) ends at destination station j.
- Population (POP) of the study community.
- Route continuity ($RTECON_{ij}$) between stations i and j: 1 = Yes, 0 = No.
- The Average Daily Traffic ($ADT_j$) at the destination station.
- The ratios of the ADT at the destination station to the sum of ADT at all destination stations combined ($ADT_j / \sum ADT_j$).

From the community review, three more potential predictors were identified that potentially influenced the distribution of through trips between stations.

- Nearby major city or expressway ($NMC_j$) towards the destination station j.
- Number of lines ($NL_j$) at the destination station cordon.
- Percentage of trucks ($PT_j$) at the destination station j.

### *4.1.3 Single Model for Percent Trips Between Two Stations*

The predictors consist of the major predictors from the other two models and one additional factor that differentiate the external-internal trips from the through trips.

- Average daily traffic at the origin station ($ADT_i$) and destination station ($ADT_j$).
- Number of lanes at the origin station ($NL_i$) and destination station ($NL_j$).
- Nearby major city at origin station (NMCi) or destination station ($NMC_j$).
- Percent of trucks at origin ($PT_i$) and destination ($PT_j$).
- Route continuity from the origin station to the destination station ($RTECON_{ij}$).
- Internal-External or External-Internal factor (IEF), if the vehicles were coming back on the same road. 1 = Yes, 0 = No.

## 4.2 Model to Estimate Through Trips at External Stations

Two cities, Arab and Roanoke, were used for model development. The values of all variables were collected (See Table 4-1) and statistical software was used to run the regression analysis. A stepwise regression analysis was conducted and the insignificant predictors were eliminated (See Figure 4-1).

**Table 4-1 Predictors for Arab and Roanoke**

| Arab | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Station** | **$Y_i$** | **$NL_i$** | **$ADT_i$** | **$PT_i$** | **$PSVP_i$** | **POP** | **$NMC_i$** |
| US 231 N | 38 | 4 | 13060 | 4.6 | 45.9 | 7174 | 1 |
| AL 69 W | 52 | 2 | 6620 | 9.5 | 43.5 | 7174 | 0 |
| US 231 S | 53 | 4 | 6720 | 6 | 45.5 | 7174 | 0 |
| AL 69 E | 37 | 4 | 8950 | 7.4 | 45.3 | 7174 | 1 |
| Roanoke | | | | | | | |
| US 431N | 23 | 4 | 5960 | 7.5 | 46.5 | 6563 | 1 |
| AL 22W | 30 | 2 | 3120 | 8.2 | 40.3 | 6563 | 0 |
| US 431S | 58 | 2 | 3750 | 12.8 | 36.7 | 6563 | 0 |
| AL 22 E | 38 | 2 | 3670 | 7.5 | 49.8 | 6563 | 0 |

```
Alpha-to-Enter: 0.15  Alpha-to-Remove: 0.15

Response is    Yi    on  6 predictors, with N =    8

Step          1        2        3
Constant     -55.63   -77.78   -29.05

NLi             8.7      8.0      7.4
T-Value        2.71     3.73     2.72
P-Value       0.225    0.065    0.072

ADTi        0.00471  0.00416  0.00382
T-Value        3.08     6.72     5.08
P-Value       0.200    0.021    0.015

PTi            5.94     5.72     4.46
T-Value        4.13     5.55     4.84
P-Value       0.151    0.031    0.017

PSVPi          0.87     0.80
T-Value        1.38     1.71
P-Value       0.400    0.229

POP          -0.005
T-Value       -0.42
P-Value       0.746

NMCi          -36.8    -34.3    -32.5
T-Value       -4.60    -8.30    -6.35
P-Value       0.136    0.014    0.008

S              4.09     3.14     4.02
R-Sq          98.38    98.10    95.31
R-Sq(adj)     88.69    93.34    89.06
C-p             7.0      5.2      4.9
```

**Figure 4-1 Stepwise regression analysis**

Stepwise regression selected four variables as shown in step 3; they were used for the final regression analysis. The variables were $ADT_i$, $PT_i$, $NMC_i$ and NL. Multiple linear regression analysis was conducted with these four variables. The final regression model equation is given in Equation 4.1.

$$Y_i = -29.1 + 7.40\ NL_i + 0.00382\ ADT_i + 4.46\ PT_i - 32.5\ NMC_i \qquad \text{(Eq. 4-1)}$$

The coefficient of determination R-Sq and R-Sq(adj) were close to 1, which means that most of the variability in $Y_i$ can be explained by the regression model; also the overall P-value was significant ($P < 0.05$). Since the sample size was small and the PRESS statistic was not available, there was a possibility that the model might not be good. A residual analysis was conducted to check the model adequacy, and the residual plots are given in Appendix A-1. The regression values are shown in Figure 4-3.

```
The regression equation is
Yi = - 29.1 + 7.40 NLi + 0.00382 ADTi + 4.46 PTi - 32.5 NMCi

Predictor        Coef     SE Coef          T        P
Constant       -29.05       13.77      -2.11    0.125
NL              7.403       2.720       2.72    0.072
ADTi        0.0038242   0.0007525       5.08    0.015
PTi            4.4570      0.9200       4.84    0.017
NMCi          -32.518       5.121      -6.35    0.008

S = 4.017              R-Sq = 95.3%         R-Sq(adj) = 89.1%
PRESS = *              R-Sq(pred) =    *%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         4      984.46      246.11     15.25    0.025
Residual Error     3       48.42       16.14
Total              7     1032.88

Source       DF      Seq SS
NLi           1       91.12
ADTi          1       47.13
PTi           1      195.49
NMCi          1      650.71

Unusual Observations
Obs        NLi         Yi         Fit      SE Fit    Residual    St Resid
  3       4.00      53.00       53.00        4.02        0.00           * X

X denotes an observation whose X value gives it large influence.
```

**Figure 4-2 Regression analysis results**

The major assumption in the regression analysis was that the errors were normally distributed. Figure A-3 given in the appendix, shows the normal probability plot of the residuals. It can be observed that the residuals are almost normally distributed, as the points lie approximately on a straight line. Figure A.2 shows the residuals were contained in a horizontal band, so there was no obvious defect; which means that the errors are uncorrelated and randomly distributed. This means the statistical assumptions were not violated. The confidence interval on regression coefficients was determined to measure the overall quality of the regression equation. It was found that the width of the confidence interval was large (Figure 4-3), which means there was a large amount of uncertainty in the predicted model coefficients. Therefore, for practical purposes, a planner should be cautious while using this model. In addition, there is a need to refine the model by using a larger sample size.

| Predictor | Coef | SE Coef | 95.0% Coef CI |
|---|---|---|---|
| Constant | -29.05 | 13.77 | ( -60.315718, 2.215718) |
| NLi | 7.403 | 2.720 | ( 2.695400, 12.110599) |
| ADTi | 0.0038242 | 0.0007525 | ( 0.002363, 0.005265) |
| PTi | 4.4570 | 0.9200 | ( 1.35354, 7.560455) |
| NMCi | -32.518 | 5.121 | ( -41.67217, -23.363829) |

**Figure 4-3 Confidence intervals on coefficients**

Among the four variables, number of lanes and nearby major city/ expressway were novel for this model. From Equation 4-1, it can be seen that the percent of through trips increased positively with the number of lanes, average daily traffic and the percent of trucks, implying these factors have a positive impact on through trips. The negative influence of the nearby major city/ expressway can be attributed to the higher percentage of internal-external trips in that direction, dramatically reducing the percentage of through trips.

### 4.3 Model for Distribution of Through Trips Between Stations

A statistical regression analysis was run on the cities to develop a model to predict through-trip distribution. The values of all the variables are given in Table 4-2 stepwise regression analysis was conducted and the insignificant predictors were eliminated.

**Table 4-2 Observed data**

| **ARAB** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Origin** | **Destination** | **$Y_{ij}$** | **$Y_j$** | **POP** | **$NMC_j$** | **$RTECON_{ij}$** | **$ADT_j/\sum ADT_j$** | **$ADT_j$** | **$NL_j$** | **$PT_j$** |
| US231 N | AL69 W | 9 | 52 | 7174 | 0 | 0 | 0.30 | 6620 | 2 | 9.5 |
| | US231 S | 16 | 53 | 7174 | 0 | 1 | 0.30 | 6720 | 4 | 6 |
| | AL69 E | 13 | 37 | 7174 | 1 | 0 | 0.40 | 8950 | 2 | 7.4 |
| AL69 W | US231 S | 11 | 53 | 7174 | 0 | 0 | 0.23 | 6720 | 4 | 6 |
| | AL69 E | 25 | 37 | 7174 | 1 | 0 | 0.31 | 8950 | 2 | 7.4 |
| | US231 N | 16 | 38 | 7174 | 1 | 0 | 0.45 | 13060 | 4 | 4.6 |
| US231 S | AL69 E | 15 | 37 | 7174 | 1 | 0 | 0.31 | 8950 | 4 | 7.4 |
| | US231 N | 30 | 38 | 7174 | 1 | 1 | 0.46 | 13060 | 4 | 4.6 |
| | AL69 W | 8 | 52 | 7174 | 0 | 0 | 0.23 | 6620 | 4 | 9.5 |
| AL69 E | US231 N | 17 | 38 | 7174 | 1 | 0 | 0.49 | 13060 | 4 | 4.6 |
| | AL69 W | 11 | 52 | 7174 | 0 | 0 | 0.25 | 6620 | 4 | 9.5 |
| | US231 S | 9 | 53 | 7174 | 0 | 0 | 0.25 | 6720 | 4 | 6 |
| **Roanoke** | | | | | | | | | | |
| **Origin** | **Destination** | **$Y_{ij}$** | **$Y_j$** | **POP** | **$NMC_j$** | **$RTECON_{ij}$** | **$ADT_j/\sum ADT_j$** | **$ADT_j$** | **$NL_j$** | **$PT_j$** |
| US431 N | AL22 W | 7 | 30 | 6563 | 0 | 0 | 0.30 | 3120 | 2 | 8.2 |
| | US431 S | 11 | 58 | 6563 | 0 | 1 | 0.36 | 3750 | 2 | 12.8 |
| | AL22 E | 5 | 38 | 6563 | 0 | 0 | 0.35 | 3670 | 2 | 7.5 |
| AL22 W | US431 S | 5 | 58 | 6563 | 0 | 0 | 0.28 | 3750 | 2 | 12.8 |
| | AL22 E | 5 | 38 | 6563 | 0 | 1 | 0.27 | 3670 | 2 | 7.5 |
| | US431 N | 20 | 33 | 6563 | 1 | 0 | 0.45 | 5960 | 4 | 7.5 |
| \`US431 S | AL22 E | 5 | 38 | 6563 | 0 | 0 | 0.29 | 3670 | 2 | 7.5 |
| | US431 N | 46 | 33 | 6563 | 1 | 1 | 0.47 | 5960 | 4 | 7.5 |
| | AL22 W | 7 | 30 | 6563 | 0 | 0 | 0.24 | 3120 | 2 | 8.2 |
| AL22 E | US431 N | 24 | 33 | 6563 | 1 | 0 | 0.46 | 5960 | 4 | 7.5 |
| | AL22 W | 8 | 30 | 6563 | 0 | 1 | 0.24 | 3120 | 2 | 8.2 |
| | US431 S | 6 | 58 | 6563 | 0 | 0 | 0.29 | 3750 | 2 | 12.8 |

Stepwise regression analysis selected four variables as shown in step 5; they were used for the final regression analysis (see Figure 4-4). They were $NMC_j$, $RTECON_{ij}$, $ADT_j$ and $NL_j$. Here step 6 was avoided to include a significant predictor from the traffic-engineering point of view, $ADT_j$, so there will be at least one variable which represents the roadway traffic.

```
  Alpha-to-Enter: 0.15  Alpha-to-Remove: 0.15

 Response is   Yij    on  8 predictors, with N =   24


    Step           1        2        3        4        5        6
Constant   -78.3843 -75.9700 -81.0619 -28.8721   1.9736   0.3151

Yj             0.02     0.04
T-Value        0.08     0.27
P-Value       0.934    0.792

POP          0.0113   0.0110   0.0120   0.0050
T-Value        0.96     1.01     1.19     0.67
P-Value       0.354    0.328    0.250    0.510

NMCj           16.3     16.6     15.8     17.9     16.0     13.2
T-Value        2.37     2.97     3.43     4.29     5.12     5.25
P-Value       0.032    0.009    0.003    0.000    0.000    0.000

RTECONij        8.1      8.1      8.1      8.4      7.9      8.2
T-Value        2.90     3.01     3.08     3.23     3.21     3.25
P-Value       0.011    0.008    0.007    0.005    0.005    0.004

ADTj/(Su         30       29       31
T-Value        0.85     0.88     1.02
P-Value       0.409    0.392    0.320

ADTj       -0.00202 -0.00205 -0.00210 -0.00134 -0.00080
T-Value       -1.49    -1.62    -1.72    -1.38    -1.47
P-Value       0.156    0.124    0.104    0.186    0.159

NLj             2.6      2.6      2.6      2.8      3.0      2.1
T-Value        1.54     1.80     1.89     2.08     2.27     1.76
P-Value       0.144    0.090    0.076    0.052    0.035    0.094

PTj             0.1
T-Value        0.08
P-Value       0.936

S              5.61     5.43     5.28     5.29     5.21     5.36
R-Sq          78.45    78.44    78.35    77.01    76.43    73.76
R-Sq(adj)     66.96    69.01    70.71    70.62    71.47    69.83
C-p             9.0      7.0      5.1      4.0      2.4      2.3
```

**Figure 4.4 Stepwise regression analysis**

The regression analysis was conducted once again with the four selected predictors. The final regression model equation is given in Equation 4-2. A residual analysis was also conducted to check adequacy of the model, and the residual plots are given in Appendix A-2. The coefficients of determination R-Sq and R-Sq(adj) were within acceptable range, and the overall P-value was very significant ($P < 0.05$). The regression values are shown in Figure 4-5.

$$Y_{ij} = 1.97 + 16.0\ NMC_j + 2.98\ NL_j + 7.93\ RTECON_{ij} - 0.000803\ ADT_j \qquad \text{(Eq. 4-2)}$$

```
The regression equation is
Yij = 1.97 + 16.0 NMCj + 2.98 NLj + 7.93 RTECONij -0.000803 ADTj

Predictor        Coef     SE Coef          T        P
Constant        1.974       3.647       0.54    0.595
NMCj           16.043       3.132       5.12    0.000
NLj             2.980       1.312       2.27    0.035
RTECONij        7.926       2.469       3.21    0.005
ADTj       -0.0008029   0.0005474      -1.47    0.159

S = 5.211              R-Sq = 76.4%         R-Sq(adj) = 71.5%
PRESS = 1029.42        R-Sq(pred) = 52.97%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         4     1673.05      418.26     15.40    0.000
Residual Error    19      515.91       27.15
Total             23     2188.96

Source       DF      Seq SS
NMCj          1     1213.67
NLj           1       97.01
RTECONij      1      303.95
ADTj          1       58.42

Unusual Observations
Obs       NMCj        Yij         Fit      SE Fit    Residual    St Resid
 20       1.00      46.00       33.08        3.26       12.92        3.18R

R denotes an observation with a large standardized residual
```

**Figure 4.5 Regression analysis value**

Figure A.6, in the appendix, shows the normal probability plot of the residuals. It was observed that the residuals were normally distributed, as the points were approximately on a straight line. Figure A-5 supports that the residuals were randomly distributed with the fitted value of $Y_{ij}$. So the regression analysis assumptions that the residuals are normally distributed and the errors are uncorrelated were accepted, and the developed model was considered as statistically adequate. The P-value for three variables was less than 0.05, and the overall P-value was significant ($P < 0.001$). In this regression model, the Prediction Error Sum of Squares (PRESS Statistic) was used to measure the model quality.

Figure 4-5 contains the squared PRESS residuals, and PRESS statistics. The value of the PRESS = 1029.42 is nearly twice as large as the residual sum of squares for this model, $SS_{Res} = 515.91$. This indicates that the model may show a large variation in predicting future values; at the same time the R-Sq(pred) value of 52.97% makes the validity of the model within acceptable limits.

This model shows a significant relationship between the roadway characteristics and the percent of trips ($Y_{ij}$). Among the four variables, the number of lanes ($NL_j$) and near by major city ($NMC_j$) were the only new factors. The $ADT_j$ and $RTECON_{ij}$ found in the previously published models were found significant for the new model also. From Equation 4-2, it can be seen that the percent of trips is positively increasing with the number of lanes, existence of any nearby cities or expressways, and route continuity from station i to station j. This is consistent with the reality of traffic flow, observed during the course of this study.

### 4.4. Single Model to Predict Percent of Trips Between Two Stations

This model was developed to predict the percent of trips between any two stations in the community. The idea behind this model was to combine both the through-trip percent estimation model and the trip distribution model into a single equation. This would reduce the number of steps that exist in all previously developed models. The same two cities, Arab and Roanoke, were considered for this model development. The values of all the variables are given in Table 4-3. Since the number of samples was larger, it was possible that the model would be universal. A stepwise regression analysis was conducted to find the most significant variables (see Figure 4-6), and five variables were selected for the model: NMCi, RTECONij, ADTj, NLj and IEF. In the stepwise regression, it was observed that the statistical software removed all the variables which were not significant or were highly correlated with the selected variables.

**Table 4-3 Observed data**

| **Arab** | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Origin** | Destination | Yij | ADTi | ADTj | POP | NLi | NLj | NMCi | NMCj | PTi | PTj | RTECONij | IEF |
| **US 231 N** | US 231 N | 62 | 13060 | 13060 | 7174 | 4 | 4 | 1 | 1 | 4.6 | 4.6 | 0 | 1 |
| | AL 69 W | 9 | 13060 | 6620 | 7174 | 2 | 2 | 1 | 0 | 4.6 | 9.5 | 0 | 0 |
| | US 231 S | 16 | 13060 | 6720 | 7174 | 4 | 4 | 1 | 0 | 4.6 | 6 | 1 | 0 |
| | AL 69 E | 13 | 13060 | 8950 | 7174 | 4 | 2 | 1 | 1 | 4.6 | 7.4 | 0 | 0 |
| **AL 69 W** | AL 69 W | 48 | 6620 | 6620 | 7174 | 2 | 2 | 0 | 0 | 9.5 | 9.5 | 0 | 1 |
| | US 231 S | 11 | 6620 | 6720 | 7174 | 4 | 4 | 0 | 0 | 9.5 | 6 | 0 | 0 |
| | AL 69 E | 25 | 6620 | 8950 | 7174 | 4 | 2 | 0 | 1 | 9.5 | 7.4 | 0 | 0 |
| | US 231 N | 16 | 6620 | 13060 | 7174 | 4 | 4 | 0 | 1 | 9.5 | 4.6 | 0 | 0 |
| **US 231 S** | US 231 S | 47 | 6720 | 6720 | 7174 | 4 | 4 | 0 | 0 | 6 | 6 | 0 | 1 |
| | AL 69 E | 15 | 6720 | 8950 | 7174 | 4 | 4 | 0 | 1 | 6 | 7.4 | 0 | 0 |
| | US 231 N | 30 | 6720 | 13060 | 7174 | 4 | 4 | 0 | 1 | 6 | 4.6 | 1 | 0 |
| | AL 69 W | 8 | 6720 | 6620 | 7174 | 2 | 4 | 0 | 0 | 6 | 9.5 | 0 | 0 |
| **AL 69 E** | AL 69 E | 63 | 8950 | 8950 | 7174 | 4 | 4 | 1 | 1 | 7.4 | 7.4 | 0 | 1 |
| | US 231 N | 17 | 8950 | 13060 | 7174 | 4 | 4 | 1 | 1 | 7.4 | 4.6 | 0 | 0 |
| | AL 69 W | 11 | 8950 | 6620 | 7174 | 2 | 4 | 1 | 0 | 7.4 | 9.5 | 0 | 0 |
| | US 231 S | 9 | 8950 | 6720 | 7174 | 4 | 4 | 1 | 0 | 7.4 | 6 | 0 | 0 |

| **Roanoke** | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Origin** | **Destination** | **Yij** | **ADTi** | **ADTj** | **POP** | **NLi** | **NLj** | **NMCi** | **NMCj** | **PTi** | **PTj** | **RTECONij** | **IEF** |
| **US 431 N** | US 431 N | 77 | 5960 | 5960 | 6563 | 4 | 4 | 1 | 1 | 7.5 | 7.5 | 0 | 1 |
| | AL 22 W | 7 | 5960 | 3120 | 6563 | 4 | 2 | 1 | 0 | 7.5 | 8.2 | 0 | 0 |
| | US 431 S | 11 | 5960 | 3750 | 6563 | 4 | 2 | 1 | 0 | 7.5 | 13 | 1 | 0 |
| | AL 22 E | 5 | 5960 | 3670 | 6563 | 4 | 2 | 1 | 0 | 7.5 | 7.5 | 0 | 0 |
| **AL 22 W** | AL 22 W | 70 | 3120 | 3120 | 6563 | 2 | 2 | 0 | 0 | 8.2 | 8.2 | 0 | 1 |
| | US 431 S | 5 | 3120 | 3750 | 6563 | 2 | 2 | 0 | 0 | 8.2 | 13 | 0 | 0 |
| | AL 22 E | 5 | 3120 | 3670 | 6563 | 2 | 2 | 0 | 0 | 8.2 | 7.5 | 1 | 0 |
| **US 431 S** | US 431 S | 43 | 3750 | 3750 | 6563 | 2 | 2 | 0 | 0 | 13 | 13 | 0 | 1 |
| | AL 22 E | 5 | 3750 | 3670 | 6563 | 2 | 2 | 0 | 0 | 13 | 7.5 | 0 | 0 |
| | US 431 N | 46 | 3750 | 5960 | 6563 | 2 | 4 | 0 | 1 | 13 | 7.5 | 1 | 0 |
| **US 431 S** | US 431 S | 43 | 3750 | 3750 | 6563 | 2 | 2 | 0 | 0 | 13 | 13 | 0 | 1 |
| | AL 22 W | 7 | 3750 | 3120 | 6563 | 2 | 2 | 0 | 0 | 13 | 8.2 | 0 | 0 |
| **AL 22 E** | AL 22 E | 62 | 3670 | 3670 | 6563 | 2 | 2 | 0 | 0 | 7.5 | 7.5 | 0 | 1 |
| | US 431 N | 24 | 3670 | 5960 | 6563 | 2 | 4 | 0 | 1 | 7.5 | 7.5 | 0 | 0 |
| | AL 22 W | 8 | 3670 | 3120 | 6563 | 2 | 2 | 0 | 0 | 7.5 | 8.2 | 1 | 0 |
| | US 431 S | 6 | 3670 | 3750 | 6563 | 2 | 2 | 0 | 0 | 7.5 | 13 | 0 | 0 |

```
  Alpha-to-Enter: 0.15  Alpha-to-Remove: 0.15

 Response is   Yij    on 11 predictors, with N =   32


    Step          1        2        3        4        5
Constant     13.708    8.327    6.145   10.307    5.984

IEF            45.3     45.3     47.4     47.3     47.2
T-Value       10.76    14.73    16.15    16.53    17.00
P-Value       0.000    0.000    0.000    0.000    0.000

NMCj                    14.3     14.6     18.5     17.7
T-Value                 5.22     5.80     5.28     5.16
P-Value                0.000    0.000    0.000    0.000

RTECONij                          8.3      8.0      7.8
T-Value                          2.55     2.52     2.51
P-Value                         0.016    0.018    0.019

ADTj                                  -0.00086 -0.00125
T-Value                                  -1.56    -2.13
P-Value                                  0.131    0.042

NLj                                                 2.4
T-Value                                            1.63
P-Value                                           0.115

S              10.3     7.53     6.90     6.73     6.54
R-Sq          79.42    89.39    91.39    92.10    92.83
R-Sq(adj)     78.74    88.66    90.47    90.93    91.45
C-p            34.4      6.2      2.1      2.0      1.7
```

**Figure 4-6 Stepwise regression analysis**

Multiple regression analysis was conducted with the remaining five predictors. The equation obtained for estimation of the percentage of trips from station i to station j is shown in Equation 4-3. The coefficients of determination R-Sq and R-Sq(adj) were very close to 1, and the overall P-value was very significant ($P < 0.001$). The regression values are shown in Figure 4-7.

$$Y_{ij} = 5.98 - 0.00125\ ADT_j + 2.42\ NL_j + 17.7\ NMC_j + 7.76\ RTECON_{ij} + 47.2\ IEF \qquad \text{(Eq. 4-3)}$$

```
The regression equation is
Yij = 5.98 - 0.00125 ADTj + 2.42 NLj + 17.7 NMCj + 7.76 RTECONij + 47.2 IEF

Predictor        Coef     SE Coef          T        P
Constant        5.984       4.124       1.45    0.159
ADTj       -0.0012508   0.0005862      -2.13    0.042
NLj             2.415       1.482       1.63    0.115
NMCj           17.720       3.436       5.16    0.000
RTECONij        7.760       3.094       2.51    0.019
I-E            47.232       2.778      17.00    0.000

S = 6.536               R-Sq = 92.8%         R-Sq(adj) = 91.5%
PRESS = 1793.47         R-Sq(pred) = 88.43%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         5     14386.1      2877.2     67.34    0.000
Residual Error    26      1110.8        42.7
Total             31     15497.0

Source       DF      Seq SS
ADTj          1       398.2
NLj           1       272.1
NMCj          1      1139.0
RTECONij      1       229.1
IEF           1     12347.7

Unusual Observations
Obs       ADTj        Yij         Fit      SE Fit    Residual    St Resid
 21       3120      70.00       54.14        2.68       15.86       2.66R
 27       5960      46.00       33.67        3.72       12.33       2.29R
R denotes an observation with a large standardized residual
```

**Figure 4.7 Regression analysis**

Figure A-9, given in the appendix, shows the normal probability plot of the residuals. It was observed that the residuals were normally distributed, as the points were close to a straight line. In addition, Figure A-8 shows that the residuals were randomly distributed with the fitted values. Therefore, the regression analysis assumptions that the residuals were normally distributed and errors were uncorrelated were accepted and the developed model was considered as statistically adequate. The R-Sq (adj) value (91.5 percent) is very close to the R-Sq value (92.8 percent). For this model the value of the PRESS = 1793.47 is only one and a half times as large as the residual sum of squares, $SS_{Res}$ = 1110.8. By looking at the R-Sq(predic) value of (88.43 percent), it was clear that the model will be good in predicting future values. This means that its performance in predicting new data is better than the models developed in the previous sections.

The roadway characteristics and the average daily traffic show a significant relationship to the percentage of trips ($Y_{ij}$). The main aspect of this equation was that all the predicting factors were related to the destination station j. The route continuity was the only variable which seemed to have a minor connection to the origin station. The three variables $NMC_j$, $RTECON_{ij}$ and IEF were either 1 or 0. This equation was similar to the trip distribution model given in Equation 4-2 except for the IEF factor, the coefficient of which had a significant impact on the overall percent of trips.

### 4.5 Summary of Model Development

This chapter presented models developed to predict through trips at external stations and distribution of through trips between stations, and the single model to predict the percent of trips between two stations.

All three models show reasonable coefficients of determination and P-values.  The residual plots for all three models were used to support their level of statistical validity, and the models were tested for their performance in predicting new data.  The newly developed models were validated to a community other than the communities used for the calibration. The next chapter deals with validation of the model using the other case study cities.

## CHAPTER FIVE
## MODEL VALIDATION AND RESULTS

The models developed in the previous chapter were tested to evaluate the predictive capability to estimate the external trips in other Alabama communities. Two communities, Arab and Roanoke, were used for model development. A third community, Hartselle, was used to validate all three models developed and the other communities were used to validate the final inclusive model. The estimated values obtained from the new models were also compared with the estimated values of respective models published in the Transportation Research Record (TRR) and NCHRP manual. This showed the merits and demerits, if any, of the new model in comparison to the previous models.

Before proceeding with the validation procedure for all three models, a brief understanding of the validation community, Hartselle, is useful. There are two major roadways passing through the community, one having two lanes and the other four lanes. Heavy traffic was observed on the four-lane highway, which connects the community to a nearby major city in the north. During the data collection process, there was a difficulty in capturing the license plates of all cars on both lanes, so emphasis was given to recording one lane clearly and recording the second lane as clearly as possible, while giving preference to the other lane. During the data reduction process, vehicles license plates with insufficient information where removed from the database. The non-intrusive data collection methodology was capable of recording the data with 80 percent to 90 percent accuracy for two lane roadways. Similar results were observed for four lanes roadways, when considering the two lanes of traffic given the emphasis during the data collection.

### 5.1 Validation of Percent Through Trip Model

The newly developed equation for percentage of through trips (TT) is shown as Equation 5-1. The previously published equations for percentage of through trips are taken from the Transportation Research Record (TRR) are shown in Equation 5-2, and the NCHRP Manual equations are shown in Equation 5-3. The average daily traffic and percentage of trucks were common for all three models. Urban population was common with both the TRR and NCHRP models. The TRR equation was developed to find the percentage of through-trip ends at an external station; it was specifically developed for small urban communities. Roadway functional classification and percentage of pick-ups and vans were unique for the NCHRP model. The NCHRP model contained more variables because it was developed to find percentage through-trips at external stations, regardless of whether it is a large or small community (see Section 2.1.2). The novelty of the new model was that it found $NL_i$ and $NMC_i$ as new variables that can define the traffic pattern more precisely than the traditionally used variables.

$$Y_i = -29.1 + 7.40 \text{ x } NL_i + 0.00382 \text{ x } ADT_i + 4.46 \text{ x } PT_i - 32.5 \text{ x } NMC_i \qquad \text{(Eq. 5-1)}$$

$$Y_i = 9.29 - 0.00031\text{x } UP + 0.0026 \text{ x } ADT + 1.48 \text{ x } TRK \qquad \text{(Eq. 5-2)}$$

$$Y_i = 76.76 + 11.22 \text{ x } I - 25.74 \text{ x } PA - 042.18 \text{ x } MA + 0.00012 \text{ x } ADTi + 0.59 \text{ x } PTKS - 0.48 \text{ x } PPS_i - 0.000417 \text{ x } POP \qquad \text{(Eq. 5-3)}$$

where UP = Urban population, TRK and $PTKS_i$ = percent of trucks excluding pick-ups and vans and $PPS_i$ = percent of vans and pick-ups (Modlin, 1982). The values of the variables for Hartselle are listed in Table 5-1.

**Table 5-1 Observed Hartselle data**

| Hartselle | | | | | | | |
|---|---|---|---|---|---|---|---|
| Origin Station | $Y_i$ | NL | $ADT_i$ | $PT_i$ | $PSVP_i$ | POP | $NMC_i$ |
| US 31 N | 73 | 4 | 19790 | 4.4 | 44.4 | 12019 | 1 |
| AL 36 W | 58 | 2 | 7220 | 5.6 | 52.6 | 12019 | 0 |
| US 31 S | 35 | 4 | 14410 | 4.1 | 49 | 12019 | 0 |
| AL 36 E | 25 | 2 | 10580 | 4.8 | 50.2 | 12019 | 1 |

By using all three models, the values of the percent through trips were estimated and matched with observed values of percent through trips ($Y_i$). The percent through trips were represented by $Y_i$ (TT) for the new model, and $Y_i$ (TRR) and $Y_i$(NCHRP) for the models published in TRR and NCHRP manuals, respectively. Table 5-2 shows the calculated values using all three models and compares the difference between the three estimated percentages and the actual observation. Figure 5-1 shows the 95 percent and 99 percent prediction intervals for the data observed in this project.

**Table 5-2 Validation & comparison of models**

| **Hartselle** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Origin Station** | **$Y_i$** | **$Y_i$ (TT)** | **Yi (TRR)** | **Yi (NCHRP)** | **Yi-Yi(TT)** | **Yi-Yi(TRR)** | **Yi-Yi(NCHRP)** |
| US 31 N | 73 | 63.3 | 63.5 | 29.7 | 9.7 | 9.5 | 43.3 |
| AL 36 W | 58 | 38.3 | 32.62 | 24.9 | 19.7 | 25.38 | 33.1 |
| US 31 S | 35 | 73.9 | 49.1 | 26.6 | -38.9 | -14.1 | 8.4 |
| AL 36 E | 25 | 15.1 | 40.2 | 26 | 9.9 | -15.2 | -1 |

```
New Obs     Fit     SE Fit        95.0% CI            95.0% PI
US31 N    63.33       7.75   (   38.68,   87.99)  (   35.56,   91.11) XX
AL36 W    38.32       4.00   (   25.60,   51.05)  (   20.28,   56.36) X
US31 S    73.94       6.73   (   52.52,   95.36)  (   49.00,   98.88) XX
AL36 E    15.09       6.82   (   -6.60,   36.78)  (  -10.09,   40.27) XX

New Obs     Fit     SE Fit        99.0% CI            99.0% PI
US31 N    63.33       7.75   (   18.08,  108.59)  (   12.36,  114.31) XX
AL36 W    38.32       4.00   (   14.96,   61.68)  (    5.21,   71.43) X
US31 S    73.94       6.73   (   34.63,  113.25)  (   28.16,  119.72) XX
AL36 E    15.09       6.82   (  -24.72,   54.90)  (  -31.12,   61.30) XX
X  denotes a row with X values away from the center
XX denotes a row with very extreme X values

Values of Predictors for New Observations

New Obs       NLi      ADTi       PTi      NMCi
US31 N       4.00     19790      4.40      1.00
AL36 W       2.00      7220      5.60      0.00
US31 S       4.00     14410      4.10      0.00
AL36 E       2.00     10580      4.80      1.00
```

**Figure 5-1 Prediction intervals for observed data**

By looking at the prediction intervals, half of the observed data lay within a 95 percent prediction interval (PI), and all the observed data lies within 99 percent PI. This means the model is good is predicting 50 percent of the observed data with a significant level $\alpha = 0.05$, and it is good in predicting all the observed data with a level of significance $\alpha = 0.01$. From this it is understood that even though the sample size for model development was small, this model is still good in predicting future values.

By looking at estimated values of the percentage of through trips by all three models and comparing these values with observed percentage of through trips ($Y_i$), it is observed that 50 percent of the estimated values by the new model (TT) lie within the limit of 10 percent of the observed values. This is true for the NCHRP model also. At the same time 75 percent of the estimated values lie within the limit of 20 percent of the observed trips for the new model, as well as well as the TRR model. This means the new model was providing results similar to the other two models.

The major difference in the observation values is for US 31 South, which could be attributed to the error in the data collection process. Examining the percentage of missed vehicles during the editing process, this section of the roadway had the highest percent (42 percent), which is much larger than other approaches in Hartselle. Figure 5-2 shows the relationship between the percent error in the data and percent difference between the estimated and observed values of through trips obtained by using the new model. It can be inferred that these factors are closely related to each other. In other words, smaller error in the observed data leads to more accurate estimates.

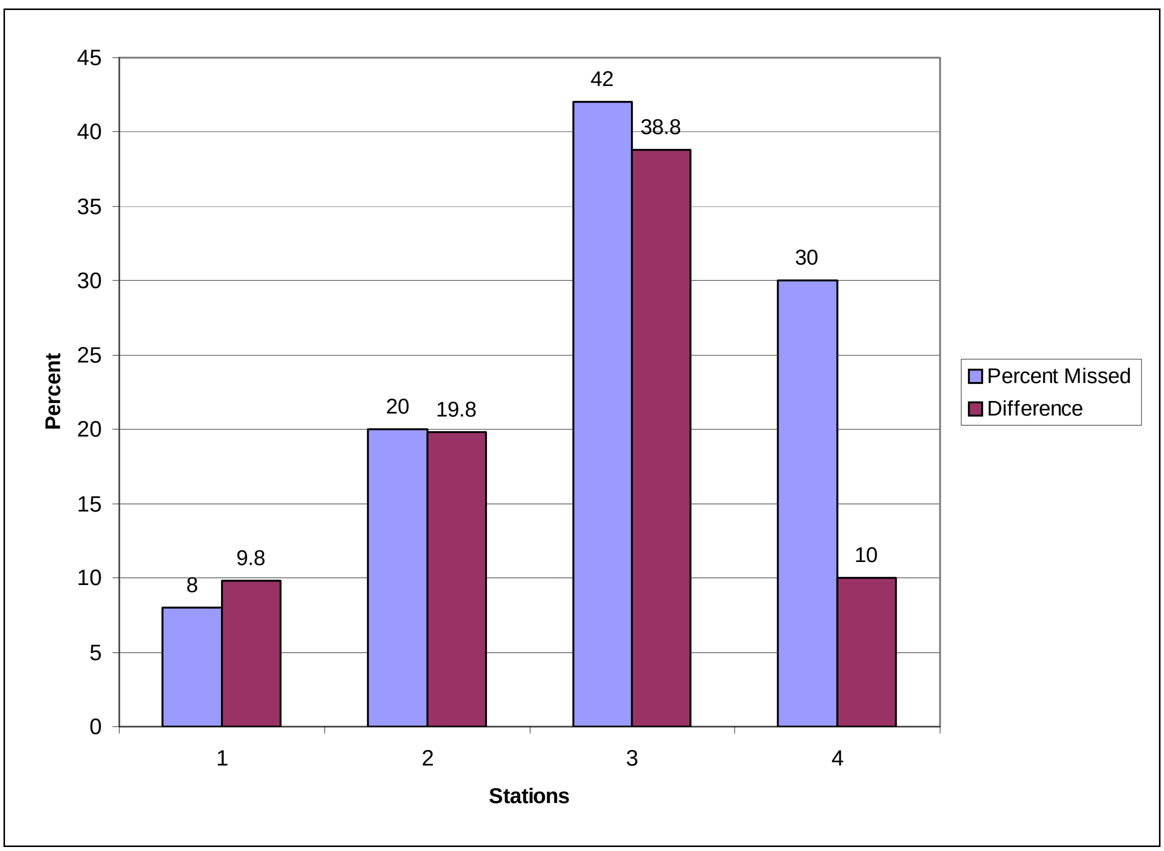


**Figure 5-2 Comparison chart**

## 5.2 Validation of Distribution of Through-Trip Model

The newly developed through-trip distribution equation (DT), see Equation 5-4, and the through-trip distribution equations published in TRR (Equation 5-5) and NCHRP report (Equation 5-6) are given below. Through-trip distribution is considered as the necessary second step in obtaining a matrix of through trips among stations. In the case of Hartselle, there are four roadways entering to the community, so there are twelve samples for the validation purpose; this is a positive aspect from a statistical point of view. The values of the variables are given in Table 5-3. Since all the roads are either U.S. or State highways, they are considered as principal arterial.

$$Y_{ij} = 1.97 + 16.0 \text{ x } NMC_j + 2.98 \text{ x } NL_j + 7.93 \text{ x } RTECON_{ij} - 0.000803 \text{ x } ADT_j \qquad \text{(Eq. 5-4)}$$

$$Y_{ij} = -7.4 + 0.55 \text{ x } PTTDES + 24.68 \text{ x } RTECON + 45.62 \text{ x } ADT / \sum ADT_j \qquad \text{(Eq. 5-5)}$$

$$Yij = -7.40 + 0.55 \text{ x } PTTDES_j + 24.68 \text{ x } RTECON_{ij} + 45.62 \text{ x } ADT_j / \sum ADT_j \qquad \text{(Eq. 5-6)}$$

where $ADT/ \sum ADT_j$ = ADT at destination station divided by the sum of ADTs at all stations, and PTTDES = percent of through trips at the destination station (Modlin, 1982). By using all three models, the values of the percent through trips from the origin station i to the destination station j are predicted and compared with the observed values of $Y_{ij}$ (see Table 5-4).

**Table 5-3 Observed Hartselle data**

| | | | | Hartselle | | | | |
|---|---|---|---|---|---|---|---|---|
| **Origin** | **Destination** | **$Y_{ij}$** | **$Y_j$** | **$NMC_j$** | **$RTECON_{ij}$** | **$ADT_j/\sum ADT_j$** | **$ADT_j$** | **$NL_j$** |
| US 31 N | AL 36 W | 12 | 58 | 0 | 0 | 0.22 | 7220 | 2 |
| | US 31 S | 37 | 35 | 0 | 1 | 0.45 | 14410 | 4 |
| | AL 36 E | 24 | 25 | 1 | 0 | 0.33 | 10580 | 2 |
| AL 36 W | US 31 S | 20 | 35 | 0 | 0 | 0.32 | 14410 | 4 |
| | AL 36 E | 30 | 25 | 1 | 1 | 0.24 | 10580 | 2 |
| | US 31 N | 8 | 73 | 1 | 0 | 0.44 | 19790 | 4 |
| US 31 S | AL 36 E | 13 | 25 | 1 | 0 | 0.28 | 10580 | 2 |
| | US 31 N | 15 | 73 | 1 | 1 | 0.53 | 19790 | 4 |
| | AL 36 W | 7 | 58 | 0 | 0 | 0.19 | 7220 | 2 |
| AL 36 E | US 31 N | 5 | 73 | 1 | 0 | 0.48 | 19790 | 4 |
| | AL 36 W | 9 | 58 | 0 | 1 | 0.17 | 7220 | 2 |
| | US 31 S | 11 | 35 | 0 | 0 | 0.35 | 14410 | 4 |

**Table 5.4 Validation and comparison**

| | | | | | Hartselle | | | |
|---|---|---|---|---|---|---|---|---|
| **Origin** | **Destination** | $Y_{ij}$ | $Y_{ij}$(TT) | $Y_{ij}$(TRR) | $Y_{ij}$(NCHRP) | $Y_{ij}$-$Y_{ij}$(TT) | $Y_{ij}$-$Y_{ij}$(TRR) | $Y_{ij}$-$Y_{ij}$(NCHRP) |
| US 31 N | AL 36 W | 12 | 2.1 | 34.7 | 34.7 | 9.9 | -22.7 | -22.7 |
| | US 31 S | 37 | 10.2 | 56.9 | 56.9 | 26.8 | -19.9 | -19.9 |
| | AL 36 E | 24 | 15.4 | 21.3 | 21.3 | 8.6 | 2.7 | 2.7 |
| AL 36 W | US 31 S | 20 | 2.3 | 26.5 | 26.5 | 17.7 | -6.5 | -6.5 |
| | AL 36 E | 30 | 23.4 | 41.8 | 41.8 | 6.6 | -11.8 | -11.8 |
| | US 31 N | 8 | 13.9 | 52.9 | 52.9 | -5.9 | -44.9 | -44.9 |
| US 31 S | AL 36 E | 13 | 15.4 | 19.2 | 19.2 | -2.4 | -6.2 | -6.2 |
| | US 31 N | 15 | 21.9 | 81.4 | 81.4 | -6.9 | -66.4 | -66.4 |
| | AL 36 W | 7 | 2.1 | 33.3 | 33.3 | 4.9 | -26.3 | -26.3 |
| AL 36 E | US 31 N | 5 | 13.9 | 54.54 | 54.54 | -8.9 | -49.54 | -49.54 |
| | AL 36 W | 9 | 10.1 | 57.1 | 57.1 | -1.1 | -48.1 | -48.1 |
| | US 31 S | 11 | 2.3 | 27.7 | 27.7 | 8.7 | -16.7 | -16.7 |

Data for all the variables are relatively easy to obtain. Number of lanes, route continuity, internal-external factor and $NMC_j$ values are always readily available. The only factor that requires some effort is the ADT. The 95 percent prediction intervals for the observed data were given in Figure 5-3. It can be seen that 83.3 percent of the observed data lies within 95 precent PI, which means the model was able to predict 83.3 percent of the observed data within a significance level of $\alpha = 0.05$.

```
New Obs     Fit     SE Fit         95.0% CI              95.0% PI
AL36 W     2.14       2.37    (   -2.82,    7.10)  (   -9.84,   14.12)
US31 S    10.25       5.39    (   -1.04,   21.54)  (   -5.44,   25.95) XX
AL36 E    15.48       3.00    (    9.20,   21.76)  (    2.90,   28.07)
US31 S     2.33       4.95    (   -8.04,   12.70)  (  -12.72,   17.37) X
AL36 E    23.41       3.64    (   15.79,   31.03)  (   10.10,   36.71)
US31 N    14.05       5.75    (    2.02,   26.08)  (   -2.19,   30.29) XX
AL36 E    15.48       3.00    (    9.20,   21.76)  (    2.90,   28.07)
US31 N    21.97       6.22    (    8.96,   34.99)  (    5.00,   38.95) XX
AL36 W     2.14       2.37    (   -2.82,    7.10)  (   -9.84,   14.12)
US31 N    14.05       5.75    (    2.02,   26.08)  (   -2.19,   30.29) XX
AL36 W    10.06       3.04    (    3.69,   16.43)  (   -2.57,   22.69)
US31 S     2.33       4.95    (   -8.04,   12.70)  (  -12.72,   17.37) X
X  denotes a row with X values away from the center
XX denotes a row with very extreme X values

Values of Predictors for New Observations

New Obs      NMCj       NLj  RTECONij      ADTj
AL36 W       0.00      2.00      0.00      7220
US31 S       0.00      4.00      1.00     14410
AL36 E       1.00      2.00      0.00     10580
US31 S       0.00      4.00      0.00     14410
AL36 E       1.00      2.00      1.00     10580
US31 N       1.00      4.00      0.00     19790
AL36 E       1.00      2.00      0.00     10580
US31 N       1.00      4.00      1.00     19790
AL36 W       0.00      2.00      0.00      7220
US31 N       1.00      4.00      0.00     19790
AL36 W       0.00      2.00      1.00      7220
US31 S       0.00      4.00      0.00     14410
```

**Figure 5-3 Prediction interval for observed data**

Figure 5-4 shows the relative performance of each model with the observed values of percentage of trips from station i to station j. Comparing the estimated values of percent through trips between stations with observed values $Y_{ij}$, 83 percent of the estimated values from the new model (DT) lie within 10 percent of the observed values. At the same time only 25 percent of the estimated values lie with 10 percent of the observed values for the TRR and NCHRP models. This means the new model is providing far better results than the other two. Regarding the difference in the observed value for US 31 South, it could be due to the error in the data collection process as was explained in the previous section. The comparison of the three models shows that the new trip distribution model had the best performance in terms of prediction errors. Data for all the variables are easy to obtain (Zhao and Chung, 2001). Number of lanes, route continuity and NMC values are always readily available. The only factor that requires some effort is the ADT; in most cases it can be obtained from traffic maps.

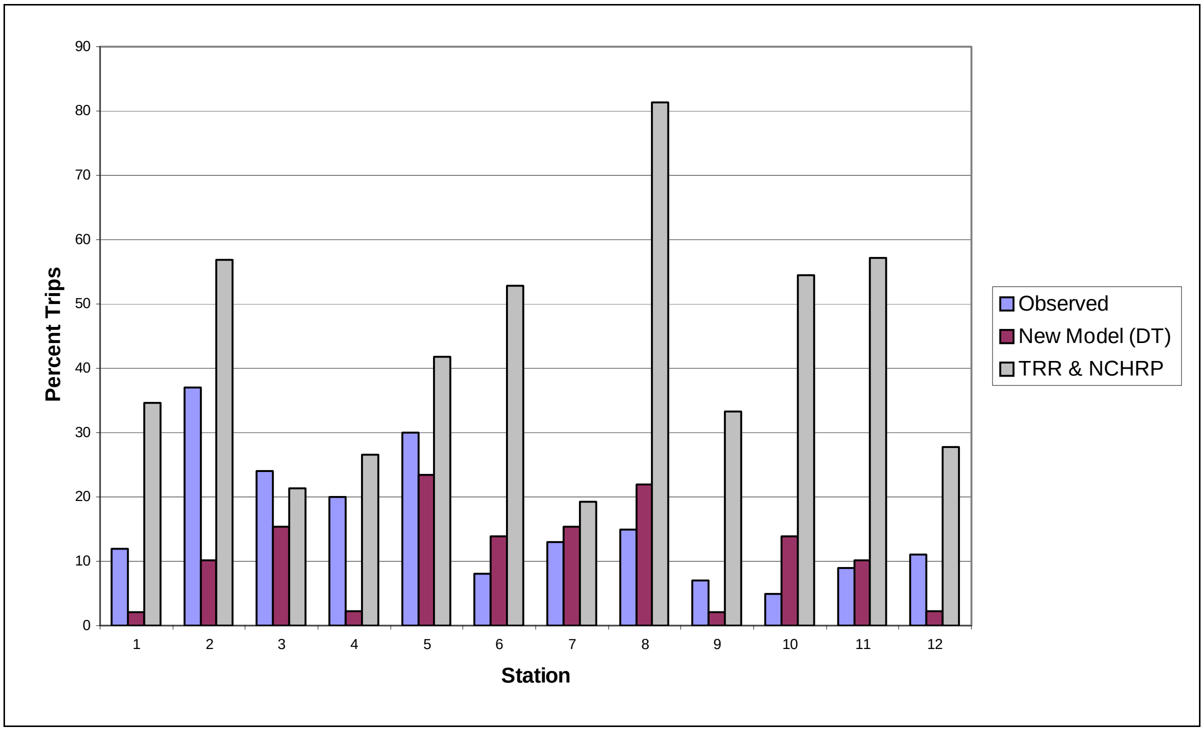


**Figure 5-4 Percent trips comparison**

### 5.3 Validation of Single Model to Predict Percent Trips Between Stations

The single model to predict the percent of trips from any external station to all other stations, including the origin station, is shown in Equation 5-7. The advantage of this model is that the internal-external trips and the external-external trips can be calculated by using the same model, which means the model for estimation of through trips at external stations and the model for the distribution of through trips between stations are joined in a single model.

$$Y_{ij} = 5.98 - 0.00125\ ADT_j + 2.42\ NL_j + 17.7\ NMC_j + 7.76\ RTECON_{ij} + 47.2\ IEF \qquad \text{(Eq. 5-7)}$$

Considering Hartselle as the validation community, there were sixteen samples for validation purposes. The values of the variables are given in Table 5-5. Since there was no single step TRR or NCHRP model, the values obtained from previous calculations were used for comparison purpose.

**Table 5-5 Observed Hartselle data**

| | | | Hartselle | | | | |
|---|---|---|---|---|---|---|---|
| **Origin** | **Destination** | **$Y_{ij}$** | **$ADT_j$** | **$NL_j$** | **$NMC_j$** | **$RTECON_{ij}$** | **IEF** |
| US 31 N | US 31 N | 27 | 19790 | 4 | 1 | 0 | 1 |
| | AL 36 W | 12 | 7220 | 2 | 0 | 0 | 0 |
| | US 31 S | 38 | 14410 | 4 | 0 | 1 | 0 |
| | IAL 36 E | 23 | 10580 | 2 | 1 | 0 | 0 |
| AL 36 W | AL 36 W | 42 | 7220 | 2 | 0 | 0 | 1 |
| | US 31 S | 20 | 14410 | 4 | 0 | 0 | 0 |
| | AL 36 E | 30 | 10580 | 2 | 1 | 1 | 0 |
| | US 31 N | 8 | 19790 | 4 | 1 | 0 | 0 |
| US 31 S | US 31 S | 65 | 14410 | 4 | 0 | 0 | 1 |
| | AL 36 E | 13 | 10580 | 2 | 1 | 0 | 0 |
| | US 31 N | 15 | 19790 | 4 | 1 | 1 | 0 |
| | AL 36 W | 7 | 7220 | 2 | 0 | 0 | 0 |
| AL 36 E | AL 36 E | 75 | 10580 | 2 | 1 | 0 | 1 |
| | US 31 N | 5 | 19790 | 4 | 1 | 0 | 0 |
| | AL 36 W | 9 | 7220 | 2 | 0 | 1 | 0 |
| | US 31 S | 11 | 14410 | 4 | 0 | 0 | 0 |

Comparison of the models shows that the new Single Model to Predict Percent Trips Between Stations (EEI) has the best performance in terms of prediction errors, with 63 percent of estimated percent of trips having an error smaller than 10 percent and 87 percent of estimated percent trips having an error smaller than 25 percent respectively. The performance of the TRR model was better than the NCHRP model, as the TRR model was able to predict 56 percent of data with an error smaller than 20 percent of the observed values. The NCHRP model had the worst performance with only 44 percent of estimated values lying within an error limit of 20 percent from the observed values (see Table 5-6).

**Table 5-6 Validation and comparison**

| | | | | Hartselle | | | | |
|---|---|---|---|---|---|---|---|---|
| **Origin** | **Destination** | **$Y_{ij}$** | **$Y_{ij}$(EEI)** | **$Y_{ij}$(TRR)** | **$Y_{ij}$(NCHRP)** | **$Y_{ij}$-$Y_{ij}$(EEI)** | **$Y_{ij}$-$Y_{ij}$(TRR)** | **$Y_{ij}$-$Y_{ij}$(NCHRP)** |
| US 31 N | US 31 N | 27 | 55.8 | 36.5 | 70.3 | -28.8 | -9.5 | -43.3 |
| | AL 36 W | 12 | 1.8 | 34.7 | 34.7 | 10.2 | -22.7 | -22.7 |
| | US 31 S | 38 | 5.4 | 56.9 | 56.9 | 32.6 | -18.9 | -18.9 |
| | AL 36 E | 23 | 15.3 | 21.3 | 21.3 | 7.7 | 1.7 | 1.7 |
| AL 36 W | AL 36 W | 42 | 48.9 | 67.38 | 75.1 | -6.9 | -25.38 | -33.1 |
| | US 31 S | 20 | 0 | 26.5 | 26.5 | 22.3 | -6.5 | -6.5 |
| | AL 36 E | 30 | 23 | 41.8 | 41.8 | 7 | -11.8 | -11.8 |
| | US 31 N | 8 | 8.6 | 52.9 | 52.9 | -0.6 | -44.9 | -44.9 |
| US 31 S | US 31 S | 65 | 44.8 | 50.9 | 73.4 | 20.2 | 14.1 | -8.4 |
| | AL 36 E | 13 | 15.3 | 19.2 | 19.2 | -2.3 | -6.2 | -6.2 |
| | US 31 N | 15 | 16.4 | 81.4 | 81.4 | -1.4 | -66.4 | -66.4 |
| | AL 36 W | 7 | 1.8 | 33.3 | 33.3 | 5.2 | -26.3 | -26.3 |
| AL 36 E | AL 36 E | 75 | 62.5 | 59.8 | 74 | 12.5 | 15.2 | 1 |
| | US 31 N | 5 | 8.6 | 54.54 | 54.54 | -3.6 | -49.54 | -49.54 |
| | AL 36 W | 9 | 9.5 | 57.1 | 57.1 | -0.5 | -48.1 | -48.1 |
| | US 31 S | 11 | 0 | 27.7 | 27.7 | 13.3 | -16.7 | -16.7 |

From the PI given in Figure 5-5, it was observed that 82 percent of the observed data lay within the limit of 95 percent PI. It should be mentioned that all the observed data that was outside this limit was associated with US 31 South, where an error occurred during data collection. From this information, it can be implied that the model is good at predicting 82 percent of observed values with a significance level of $\alpha = 0.05$. The graphical comparison of the observed percent trips with estimated values of all three models is shown in Figure 5-6. As negative percentages never exist in reality, those figures were considered as zero for practical purposes.

```
New Obs     Fit     SE Fit        95.0% CI            95.0% PI
US31 N    55.84       6.56    (   42.36,   69.32)  (   36.81,   74.88) XX
AL36 W     1.78       2.62    (   -3.61,    7.18)  (  -12.69,   16.26)
US31 S     5.38       5.97    (   -6.90,   17.66)  (  -12.82,   23.58) X
AL36 E    15.30       3.52    (    8.07,   22.54)  (    0.04,   30.56)
AL36 W    49.02       3.17    (   42.51,   55.52)  (   34.09,   63.95)
US31 S    -2.38       5.41    (  -13.50,    8.74)  (  -19.82,   15.06) X
AL36 E    23.06       4.33    (   14.16,   31.96)  (    6.95,   39.18)
US31 N     8.61       6.27    (   -4.27,   21.50)  (  -10.00,   27.23) X
US31 S    44.85       5.71    (   33.11,   56.59)  (   27.01,   62.70) X
AL36 E    15.30       3.52    (    8.07,   22.54)  (    0.04,   30.56)
US31 N    16.37       6.86    (    2.26,   30.48)  (   -3.11,   35.85) XX
AL36 W     1.78       2.62    (   -3.61,    7.18)  (  -12.69,   16.26)
AL36 E    62.53       3.97    (   54.38,   70.69)  (   46.82,   78.25)
US31 N     8.61       6.27    (   -4.27,   21.50)  (  -10.00,   27.23) X
AL36 W     9.54       3.52    (    2.31,   16.78)  (   -5.72,   24.80)
US31 S    -2.38       5.41    (  -13.50,    8.74)  (  -19.82,   15.06) X
X  denotes a row with X values away from the center
XX denotes a row with very extreme X values

Values of Predictors for New Observations

New Obs      ADTj       NLj      NMCj  RTECONij       IEF
US31 N      19790      4.00      1.00      0.00      1.00
AL36 W       7220      2.00      0.00      0.00      0.00
US31 S      14410      4.00      0.00      1.00      0.00
AL36 E      10580      2.00      1.00      0.00      0.00
AL36 W       7220      2.00      0.00      0.00      1.00
US31 S      14410      4.00      0.00      0.00      0.00
AL36 E      10580      2.00      1.00      1.00      0.00
US31 N      19790      4.00      1.00      0.00      0.00
US31 S      14410      4.00      0.00      0.00      1.00
AL36 E      10580      2.00      1.00      0.00      0.00
US31 N      19790      4.00      1.00      1.00      0.00
AL36 W       7220      2.00      0.00      0.00      0.00
AL36 E      10580      2.00      1.00      0.00      1.00
US31 N      19790      4.00      1.00      0.00      0.00
AL36 W       7220      2.00      0.00      1.00      0.00
US31 S      14410      4.00      0.00      0.00      0.00
```

**Figure 5-5  Prediction intervals for observed data**

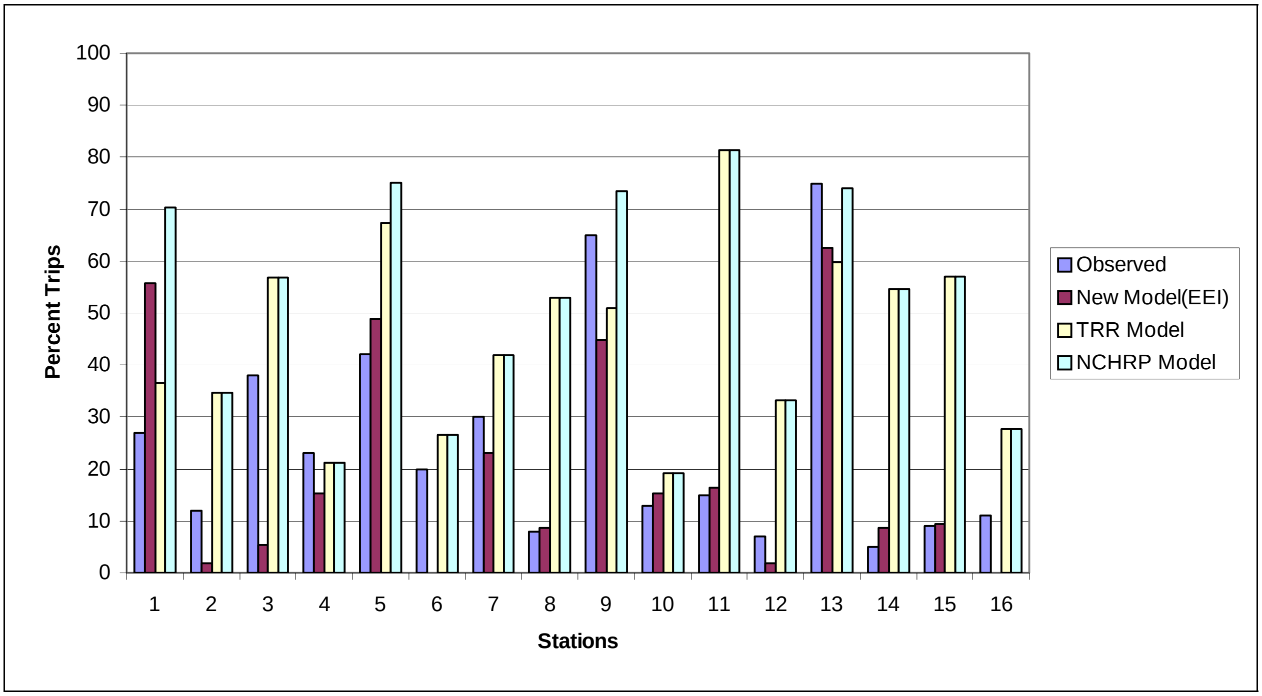


**Figure 5-6 Percent trips comparison**

Data for all the variables are relatively easy to obtain. Number of lanes, route continuity, the internal-external factor and NMC values are always readily available. The only factor that requires some effort is the ADT. Figure 5-6 shows the relative performance of each model with the observed value of percentage of trips from station i to station j.

### 5.4 All Case Study Cities with the Combined Model

At this point, the new single equation capable of developing the percentage of external trips and distribution of trips has been shown successful for Hartselle. In the following section, this model will be applied to the remaining case study cities to see if it has transferability to other communities in Alabama. The methodology employed to test the transferability of the model was to compare the predicted percentage of traffic to the observed traffic using a graphical representation and the Nash-Sutcliffe Model Efficiency Statistic (Nash and Sutcliffe, 1970).

The graphs displaying the predicted percent traffic and actual percent traffic are shown as Figures 5-7 through 5-10. In each figure, a 1:1 line has been added to assist in understanding the relationship between the two values. Examining the series of figures, it is possible to draw the conclusion that there is a decent linear relationship between the prediction and the actual values. This linear relationship supports the notion that the single equation model can be transferred to other communities and remain valid.

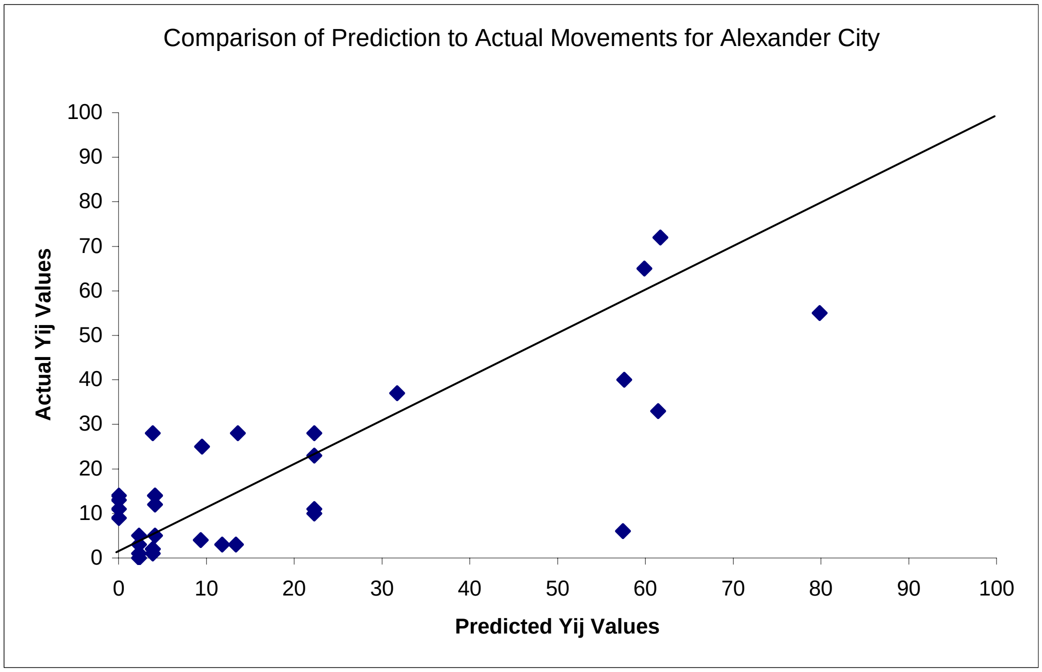


**Figure 5-7 Comparison of data for Alexander City**

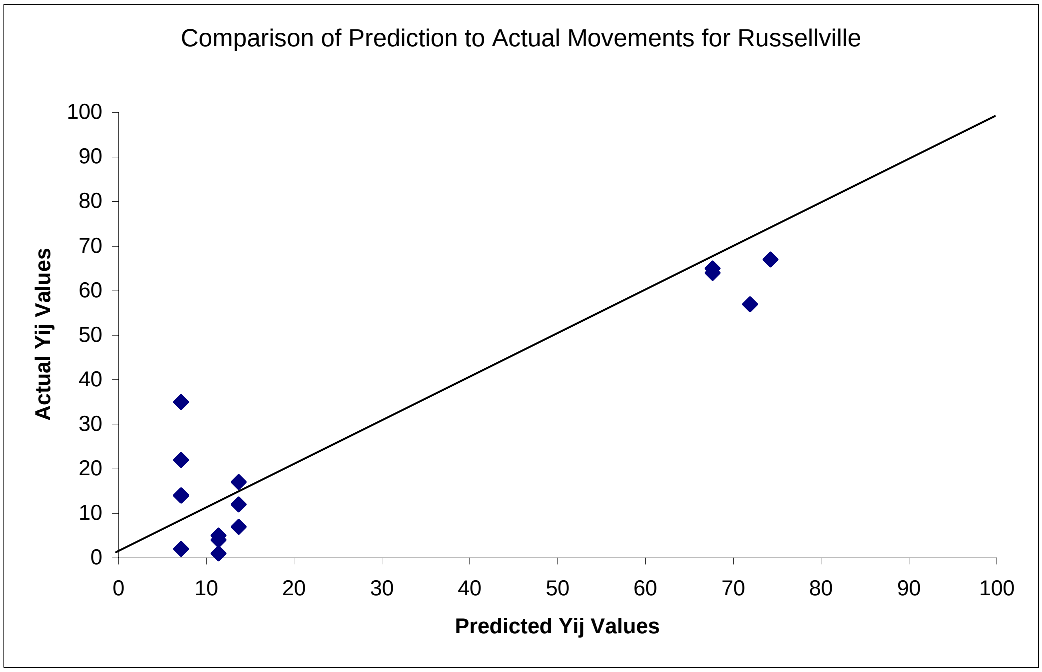


**Figure 5-8 Comparison of data for Russellville**

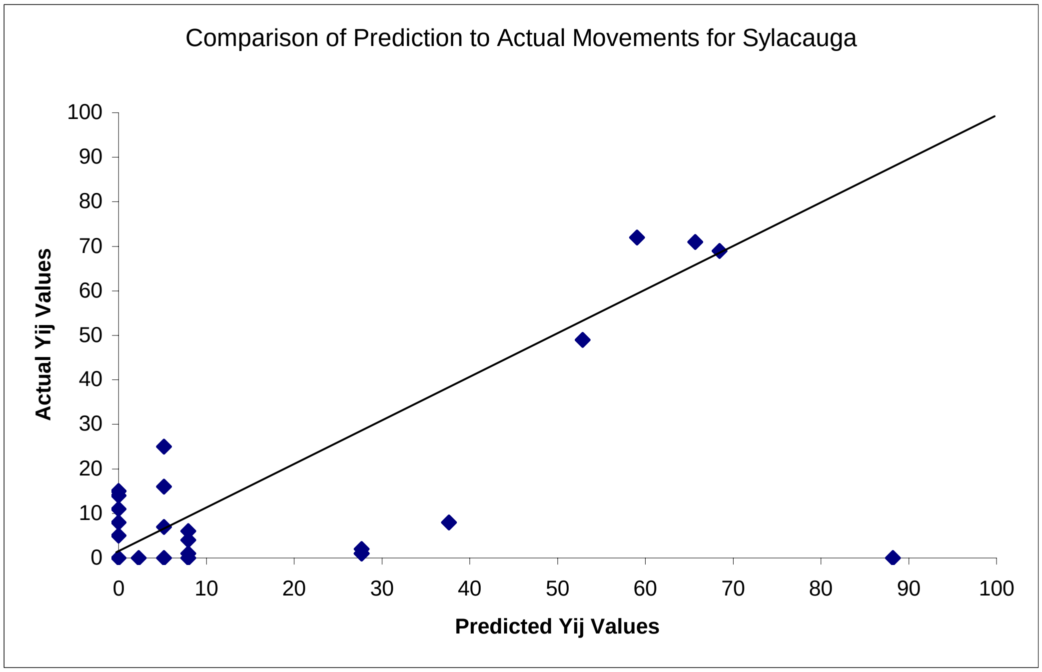


**Figure 5-9 Comparison of data for Sylacauga**

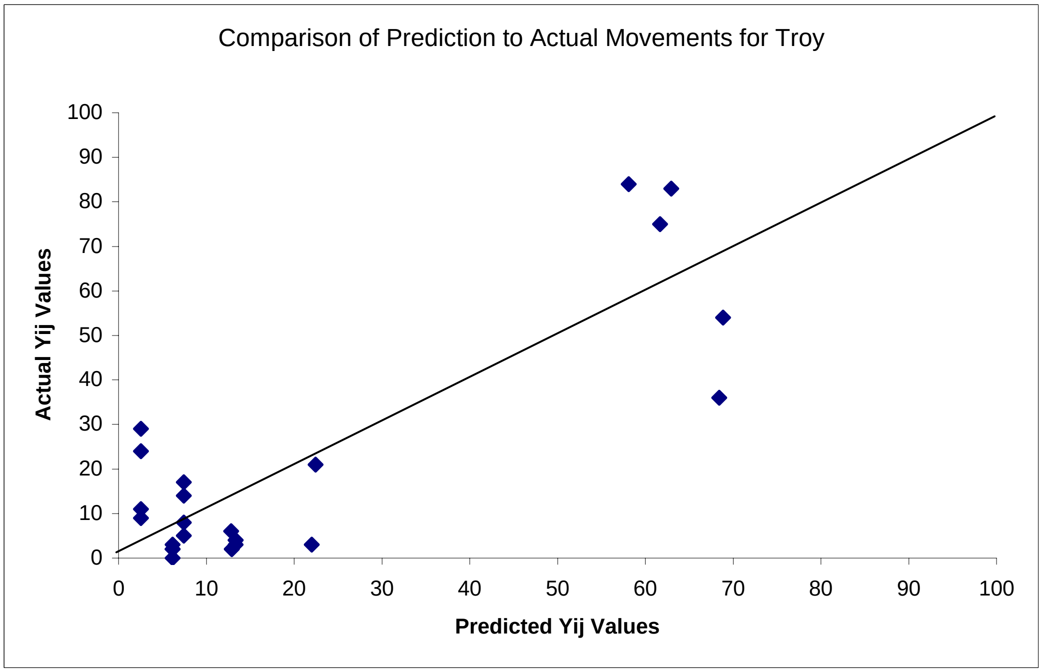


**Figure 5-10 Comparison of data for Troy**

To strengthen the conclusion that the model is transferable to other communities within the state, a Nash-Sutcliffe Model Efficiency Statistic was calculated for each community. The test statistic represents the percentage of the variance in the measured data that is explained by the simulated data (Nash and Sutcliffe, 1970). In other words, the efficiency statistic compares the variance in the model errors with that in the observed data and indicates how closely the plot of the simulated versus observed values comes to the 1:1 line. Thus, it indicates the model's ability to replicate data along a 1:1 line. The statistic is calculated as:

$$\text{Test Statistic} = 1 - [\ \Sigma\ (\text{observed data} - \text{actual data})^2\ /\ \Sigma\ (\text{observed data} - \text{average observation})^2\ ]$$

The statistic can vary from +1.0 to –1.0. Any positive value of this parameter indicates some predictive value in the model above random selection. Based on the data collected and analyzed from the four test case cities, the test statistic was computed to be:

| | |
|---|---|
| Alexander City | 0.60, |
| Russellville | 0.85, |
| Sylacauga | 0.76, and |
| Troy | 0.62. |

All of the values for the test statistic indicate a strong relationship and strengthen the conclusion that the model has transferability to other communities.

### 5.5 Results and Discussion

Three models were developed to predict the percentage of through trips in small urban communities. The validation results showed that the model to predict through trips at an external station showed similar performance to that of the TRR and NCHRP models; and the through-trip distribution model and single model to predict percent trips between stations showed more accurate results than TRR and NCHRP models. The validation for the prediction of through-trip percents could have been better if a larger sample size were available. From a statistical point of view, all three models showed excellent coefficients of determination (R-Sq), and the overall P-values were within a reasonable range ($P < 0.025$). Based on the model adequacy and ability to predict future values, the single model to predict the percentage of trips between stations was considered to be the most valid model. In addition, the application of the model to the additional case study cities supports the application of the model to determine traffic flows at other locations in Alabama.

# CHAPTER SIX
# CONCLUSION

The non-intrusive data collection method used for this study was successful. The video surveillance equipment used was capable of recording the data with 80 percent to 90 percent accuracy on two lane roadways. Difficulty surfaced with four lane roads, where some camcorders where not able to record vehicles in both lanes accurately. This result reduced the accuracy of the data.

Video surveillance can be improved by providing more rigorous training for the data collection staff and providing additional cameras for roadways with multiple lanes to increase the number of vehicles properly recorded. Methods to reduce glare and direct sunlight should be examined to improve the quality of the image.

## 6.1 Model Conclusion

Identification of the effect of nearby cities, expressways or businesses on a small community traffic flow is a new contribution. This factor was never mentioned in any of the previous models. This factor plays an important role in all three models developed during this study. For the purpose of this report, this factor is limited to a city or expressway within reasonable distance. Further studies need to be performed to determine which cities or roadways can be included in this factor. The number of lanes was the other contributing factor, which essentially replaced functional classification. Since most roadways in small communities will be U.S. or state highways; the number of lanes was found to be a determining factor for the traffic volume.

The novelty of these models lies in their dependence on a minimal amount of data to predict trip percentages. Average daily traffic and percent trucks are the only two variables requiring data collection. All other variables used in these models are readily available.

The application of the model to the validation communities demonstrated the ability of the model to accurately predict traffic flow characteristics in a wide variety of communities. Based on the graphs and test statistics calculated during this research, the single equation model also demonstrated transferability to other communities, a vital component in transportation modeling.

Desired future research involves collecting more data, modifying the model using a larger sample size, providing a broader definition for the new variable (NMC), and defining the kind of facilities to be included in this factor.

## 6.2 Closure

Three multiple regression models were developed for estimating the percentage of through trips for small urban communities. A large number of potential predictors were identified and investigated. Six variables were used in the final models. They include Average Daily Traffic (ADT), Number of Lanes on the roadway (NL), Percentage of Trucks (PT), Nearest Major City / Expressway (NMC), Route Continuity (RTECON) and the Internal-External Factor (IEF). The models were able to explain 67 percent to 95 percent of the total variability depending on the variables used.

The models are divided into two sets. The first set consists of the models to predict through trips at an external station and for the distribution of through trips between stations. The second set is the single step model to predict the percentage of trips from one external station to another. The second set of models provided better estimates the percentage trips. Hence the single equation model is recommended.

**CHAPTER SEVEN**
**REFERENCES**


Alabama Municipalities, Total Population 2000. http://www.bhammpo.org/docs/alabamapop_1990-2000.pdf

Anderson, M. D. "Evaluation of Models to Forecast External-External Trip Percentages" ASCE Journal of Urban Planning and Development. Vol. 125, Issue 3, p. 110-120. September 1999.

Anderson, M. D., Souleyrette, R. R. "A Network Demand Model for Rural Bypass Planning" Paper No. 00-1205. 2000.

Army Technology. http://www.army-technology.com/contractors/civil/perceptics/ 2002.

Census population information. 1980 Census of Population, Number of Inhabitants, United States Summary. Report PC 80-1-A1, p. 1-37. Bureau of the Census, U.S. Department of Commerce, Washington D.C., 1983.

FHWA Planning Guide. http://www.fhwa.dot.gov/planning/rural/planningfortrans/ 2ourrts.html#IIAwir. 2002.

Gupta, R., Fricker, J. D., Moffett, D. P. "Reduction of Video License Plate Data" Transportation Research Record 1804, Paper No. 02-2475. 2002.

Khisty, C. J., Rahi, M.Y. "Evaluation of Three Inexpensive Travel Demand Models for Small Urban Areas" Transportation Research Record No. 1283. p. 70-78.

Khisty, C. J., Rahi, M. Y. "Inexpensive Travel Demand Techniques" Final Report WA-RD 125.1. Washington State Department of Transportation. September 1987.

Low, D. A. "A New Approach to Transportation Systems Modeling". Traffic Quarterly, Vol. 26, No. 3, 1972.

Maps from Yahoo Mapservice. www.maps.yahoo.com

Martin, W. A., McGuckin, N. A. "Travel Estimation Techniques for Urban Planning" National Cooperative Highway Research Program, Rep. 365, Transportation Research Board, Washington, DC, 1998.

Modlin, D. G. "Synthesized through-Trip Table for Small Urban Areas" Transportation Research Record 842, Transportation Research Board, Washington, DC, 1982.

Montgomery, D. C., Peck, E. A. "Introduction to Linear Regression Analysis" WILEY. Second Edition, 1992.

Nash, J.E. and J.V. Sutcliffe. "River Flow Forecasting through Conceptual Models Part 1 – A discussion of Principles," Journal of Hydrology, Vol. 10, 1970.

Neumann, E., Halkias, J., Elrazaz, M. "Estimating Trip Rates from Traffic Counts" Journal of Transportation Engineering, ASCE, Vol. 109, No. 4, 1983.

Population archive. http://www.asc.edu/archives/populate/sheet.html, October 7 1998.

Shuldiner, P. W. "Acquisition and Analysis of License Plate Data Using Video and Machine Video Technology" Presented at National Traffic Data Acquisition Conference, Rocky Hill, CT, 1994.

Shuldiner P. W., Hudson J. B. "Acquiring Travel Time and Network Level Origin-Destination Data by Machine Vision Analysis of Video License Plate Images" Presented at National Traffic Data Acquisition Conference, Albuquerque, NM, 1996.

Traffic map. 1996 Alabama Traffic Flow Map, Alabama Department of Transportation, 1997.

U.S. Census Bureau, Census 2000 Housing Units. http://quickfacts.census.gov/hunits/states/01000.html

Walpole, R. E., Myers, R. H., Myers, S. L., Ye, K. "Probability & Statistics For Engineers & Scientists" Prentice Hall. Seventh Edition, 2002.

Washburn, S. S. "Speech Recognition for On-Site Collection of License Plate Data: Exploratory Application Development and Testing" Journal of Transportation Engineering. November/December 2002.

Washburn, S. S. and Nihan, N. L. "Using Voice Recognition to Collect License Plate Data for Travel Time Studies" Transportation Research Record 1593. 1997.

Webopedia definition. http://www.webopedia.com/TERM/o/optical_character_recognition.html 2002.

Zhao, F., Chung. S. " Estimation of Annual Average Daily Traffic in Florida County Using GIS and Regression" TRB Paper 01-3440.

**APPENDIX**
**RESIDUAL ANALYSIS PLOTS**

## A.1 Model to Predict Through Trips at External Stations

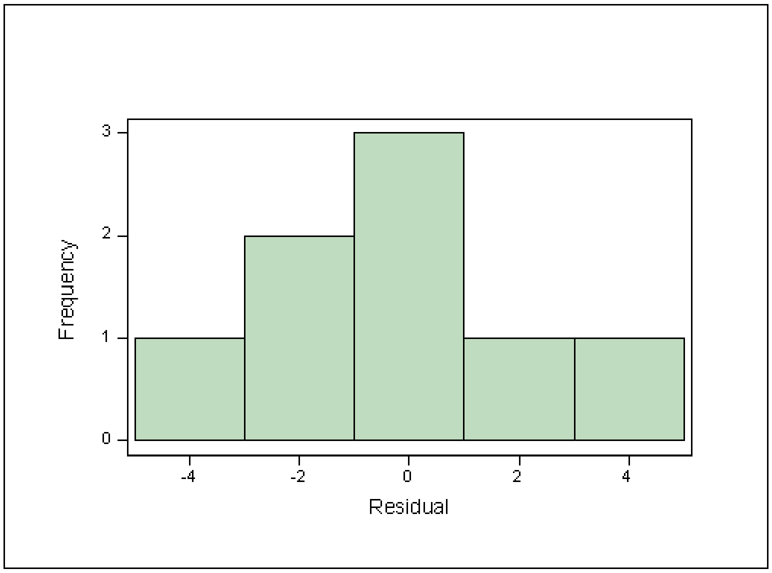


**Figure A-1 Residual histogram for $Y_i$**

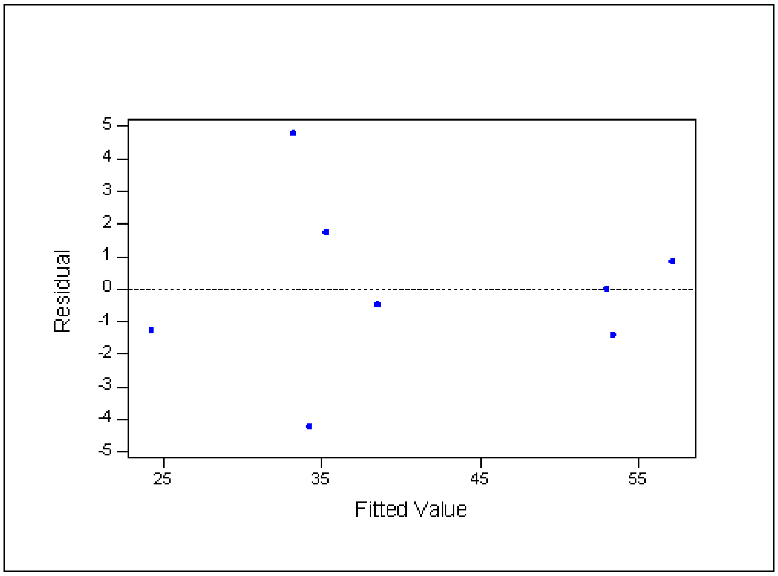


**Figure A-2 Residuals vs. fits for $Y_i$**

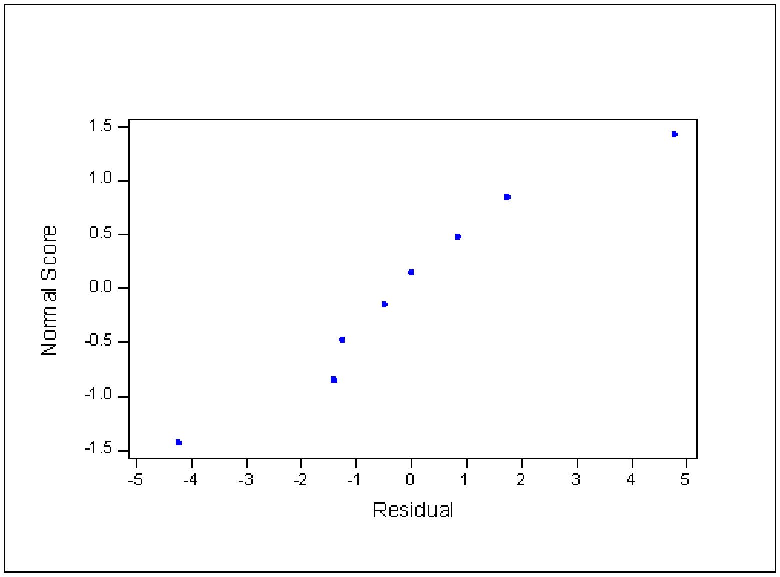


Figure A-3 Normal plot of residuals for $Y_i$

## A.2 Model for Distribution of Through Trips Between Stations

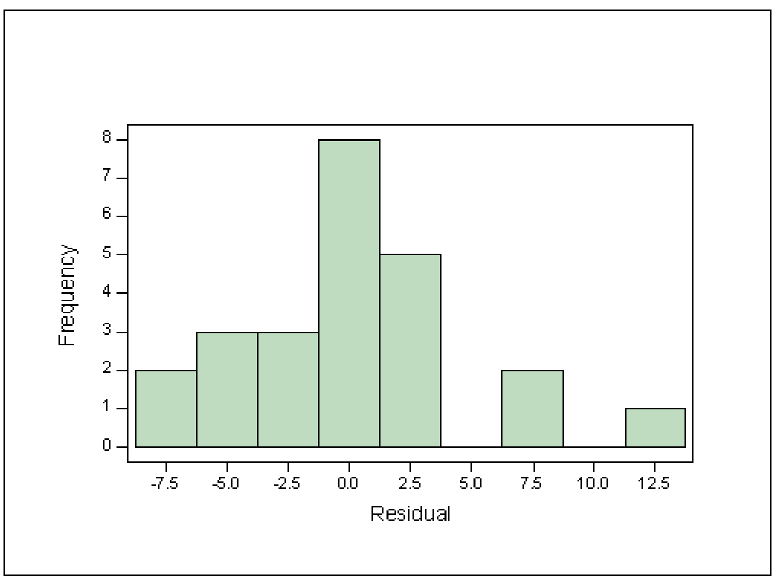


**Figure A-4 Residual histogram for $Y_{ij}$**

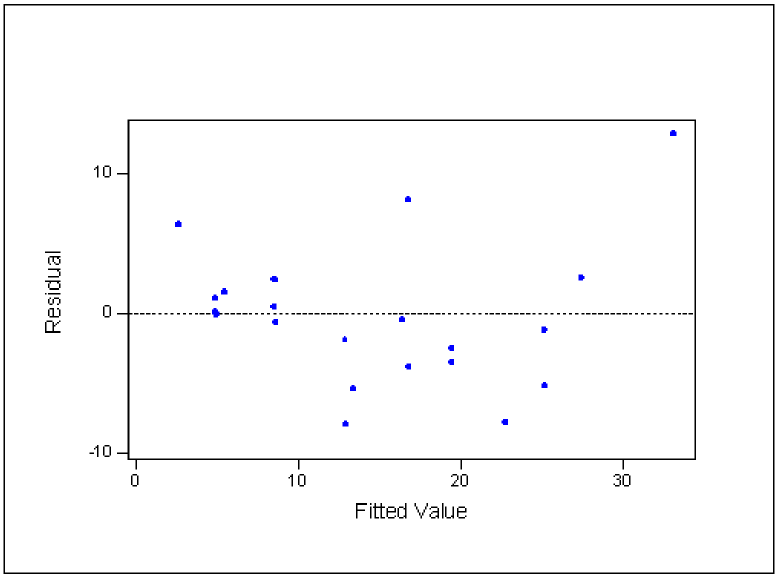


**Figure A-5 Residual vs. fits for $Y_{ij}$**

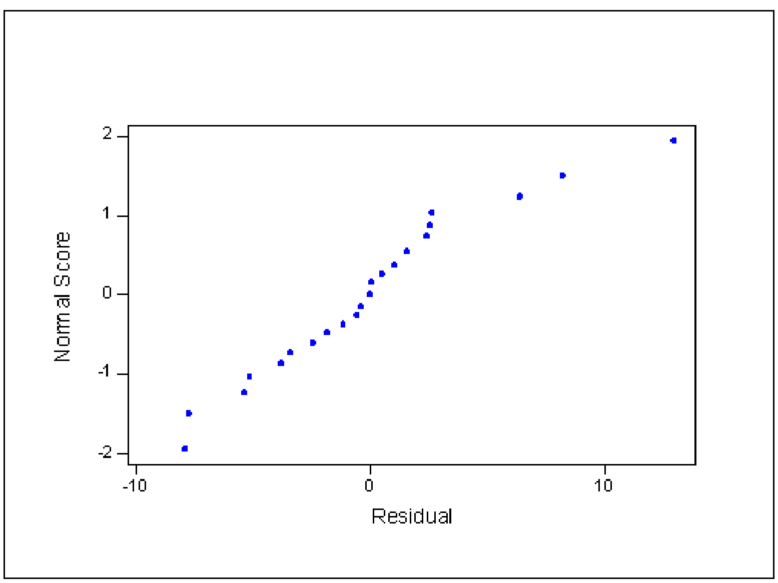


**Figure A-6 Normal plot of residuals for $Y_{ij}$**

### A.3 Single Model to Predict Percent Trips Between Two Stations

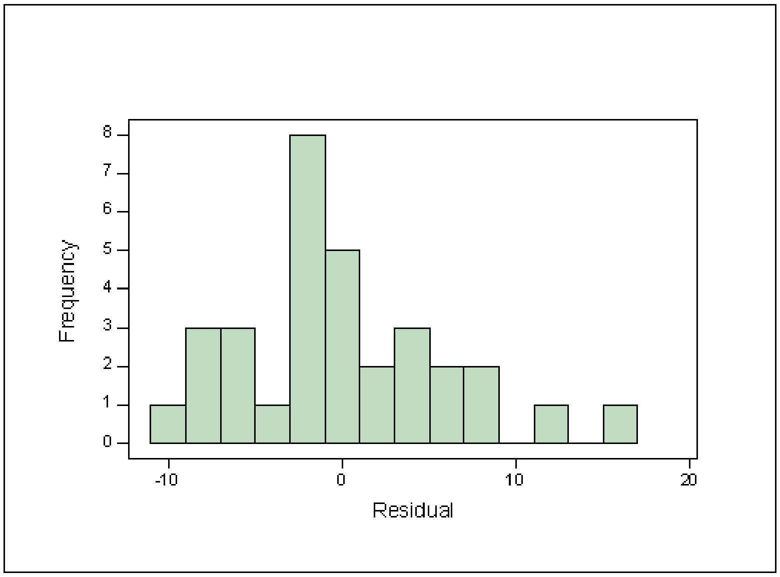


**Figure A-7 Residual histogram for $Y_{ij}$**

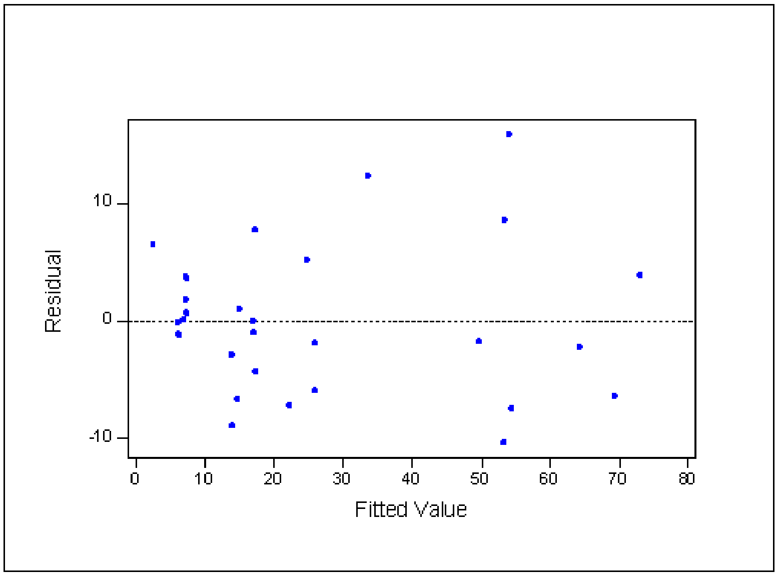


**Figure A-8 Residual vs. fits for $Y_{ij}$**

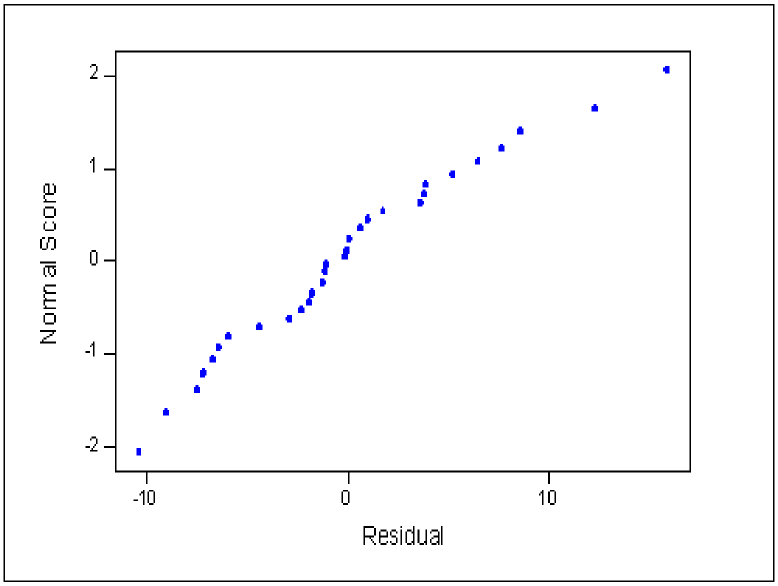


**Figure A-9 Normal plot of residuals for $Y_{ij}$**