Creating Communities of Place

Tomorrow's
NEW JERSEY

PROJECTING MUNICIPAL ROAD COSTS UNDER VARIOUS GROWTH SCENARIOS

The Relationship between Municipal Residential Density, Municipal Road Density and Density Change

Document #109

NEW JERSEY OFFICE OF STATE PLANNING
FEBRUARY 1996

Herbert Simmens, *Director*
New Jersey Office of State Planning

Project Manager:

Robert A. Kull
Assistant Director

Prepared by:

James Reilly
Senior Research Planner

This Report is Printed on Recycled Paper

PROJECTING MUNICIPAL ROAD COSTS UNDER VARIOUS GROWTH SCENARIOS

The Relationship between Municipal Residential Density, Municipal Road Density and Density Change

Acknowledgments

The principal researcher and author of this report was James Reilly. Teri Schick provided valuable assistance by providing legal insights, assistance interpreting the statistical findings and in editing the report. Additional comments and editing assistance were provided by Kathleen Kelly, Robert Kull, Carlos Rodrigues and Antoinette F.Seymour. Steven Karp prepared the GIS road maps from Intergraph files provided by NJDOT.

PROJECTING MUNICIPAL ROAD COSTS UNDER VARIOUS GROWTH SCENARIOS

The Relationship between Municipal Residential Density, Municipal Road Density and Density Change

CONTENTS

Abstract

This paper serves several purposes. First, the paper revises the statistical equations used in the OSP Sewer and Road Impact models. Municipal, county and State planners using the updated OSP computer model will be able to more accurately project infrastructure costs associated with alternative long-range growth scenarios.

Second, this paper examines policy and design implications that arise from the research findings. The paper's main statistical finding is that the (linear) supply of municipal roads is strongly related to municipal residential density. This finding supports the State Development and Redevelopment Plan's policy recommendation that higher density development (into centers) would result in infrastructural savings.

The statistical equations documented in this report might be used to demonstrate the "rational nexus" required to support the imposition of development impact fees. It also is proposed that a new type of impact fee, based on the concept of a municipal density standard, could be established. With this system the statistical equations could be used to more equitably distribute recurrent municipal infrastructure related costs.

The paper reports two design implications. The inferential research suggests that the geometry of local road networks directly affects the amount of roadway that gets built. It recommends that this issue, especially in the design of residential neighborhoods in centers, should be studied more closely. The paper also determines a threshold for residential densities of 10 dwelling units per acre above which transit services and/or off-street parking should be required.

Summary

Purpose

This paper provides a more detailed update of the statistical study Projecting Costs for Roads under Various Growth Scenarios (Gottlieb 1990), which first reported the relationship between population density (number of persons per square mile) and the density of municipal roads (the number of center-line lane miles of municipal road per square mile). While this relationship might sound academic, it is the mathematical basis used in the New Jersey Office of State Planning (OSP) Impact models [1] to estimate both road costs and sewer collector costs related to growth. These relationships also were used by Rutgers University Center for Urban Policy Research (CUPR) to evaluate collector sewer and water distribution costs in their report Impact Assessment of the New Jersey Interim State Development and Redevelopment Plan (Burchell, et al., 1992). The advantages of the OSP statistical approach are:

1. it is based on a large sample;
2. it is specific to New Jersey municipalities; and,
3. it is sensitive to changes in residential density.

In addition to assisting OSP in evaluating growth management alternatives, the OSP models are intended to assist county planners, MPO's (Metropolitan Planning Organizations) and planners in other State agencies to forecast infrastructure needs that could result from alternative growth and policy options which they might be investigating.

Method

New housing, population and employment data were collected, including data from the *1990 Census* and the New Jersey Department of Labor's *1990 ES202* files, neither of which were available when the original study was performed. The latest road inventory files were obtained from the New Jersey Department of Transportation (NJDOT). OSP also used municipal area data developed from New Jersey Department of Environmental Protection (NJDEP) geographic information system (GIS) files.

The State's 567 municipalities were grouped (K-Means Clustered) into 5 categories to control for the possible effect that larger municipal areas or more densely developed municipal areas might need disproportionately more (or less) roads simply because of their size or level of development. Municipal, county and State road densities for the municipalities in each of these clusters then were compared to a total of 17 independent variables including: residential densities (1970, 1980 and 1990); employment density (1980 and 1990); residential and employment growth (1970 to 1980 and 1980 to 1990); employment agglomeration (1980 and 1990); and, various demographic measures associated with density or design, including income, housing age groups

[1] See Appendix A for a brief description of the OSP Growth Simulation Models.

and the percentage of multi-household dwelling units. Many of these variables were selected to allow the study to determine if other municipal attributes (such as the existing design of the developed portions of the municipality, or large concentrations of employment mixed into the land use pattern of the municipality) also were associated with changes in the supply of roads in the municipality. To control for demographic factors the principal 1990 variable, population density, was replaced with the variable dwelling unit density.

The report is organized into five chapters. In the first two, the data used in the study are described and a preliminary statistical analysis is performed. In chapter 3, the initial statistical findings are closely examined and the regression equations simplified. In chapter 4, all significant variables are updated to use 1990 data (to simplify the use of these equations for forecasting) and the revised equations are tested. Finally, in chapter 5, the policy and design implications of the statistical findings are presented. Also in chapter 5, inferential methods are used to provide a possible explanation for part of the statistical results. These inferential examinations produce their own findings and design recommendations.

Statistical Findings

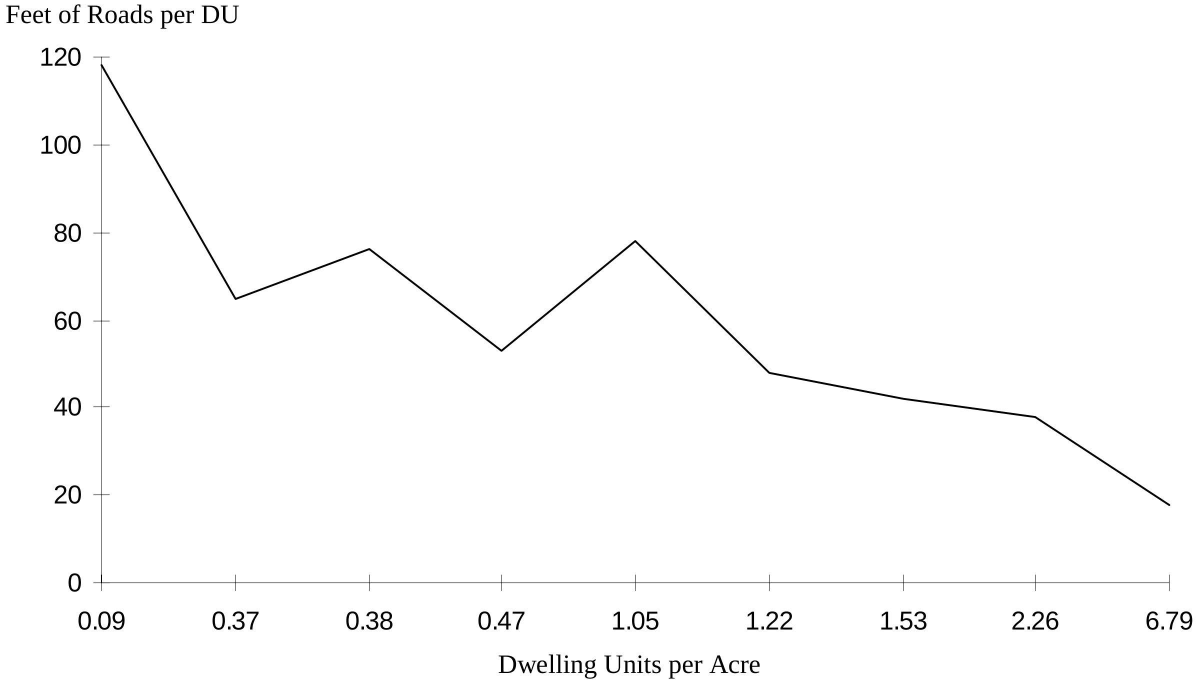
Several important findings, identified by using statistics, are documented in this report.

1. Municipal road densities are related to municipal-scale variables. County and State roads are not correlated to the municipal-scale variables tested in this report.

2. There is a strong, positive correlation between municipal road density and municipal residential density (municipalities with more houses have more roads).

3. If the supply of municipal roads is divided by the number of dwelling units in the municipality, then the relationship is (generally) inverse (in less dense municipalities the number of feet of municipal road per dwelling units is greater than that found in more densely developed municipalities). This finding means that if 10 dwelling units are built on 5 acre lots they would require more municipal road than would the same 10 dwelling units if they had been built on smaller lots. This finding partially refutes the assumption that more municipal roads need to be constructed to provide access in the less densely developed municipalities (the marginal cost of new roads is unavoidably high). This finding also suggests that municipal zoning decisions directly affect the amount of municipal roads that need to be built, and that these residential zoning decisions affect both the amount of roadway that the municipality needs to maintain and the vehicle miles traveled (VMT) on the roads. These observations have direct implications for State air and water quality programs.

Municipal Roads per Dwelling Unit as a Function of Residential Density



4. In the least dense municipal cluster, household income is an important secondary explanatory variable and is inversely related to road density.

5. Surprisingly, employment density is not significantly related to municipal road density, except mildly in the medium density municipal cluster.

6. The most unexpected finding was that only in clusters 2 and 5 was DUDENS90 (dwelling unit density per square mile in 1990) the most explanatory independent variable on the relationship with MUNRDDEN (the 1990 supply of municipal roads per square mile). In cluster 3 and 4 DUDENS70 (municipal dwelling units in 1970 per square mile of municipal area) demonstrated the strongest relationship with MUNRDDEN and in cluster 1 DUDENS80 (dwelling unit density in 1980) was the most significant variable. This relationship to a time lagged variable supports the hypothesis that a dynamic process was occurring in those municipalities.

7. The following table lists the equations that are the most suitable for use in forecasting. However, the results of these equations are likely to over-estimate the future need by some small amount, especially in rapidly growing municipalities, because the municipal road data was cross sectional (1990 only) while the statistical analysis clearly suggests a dynamic system. Despite this problem, these new equations should be incorporated into the OSP Impact models, replacing the 'Hudson Shift' simulation of Plan used in the original models.

Recommended Regression Equations for Modeling Future Roads

	Model Equation	Adjusted R^2
Cluster 1	DUDENS90*.380007+INCOME90*.654875-8.405293	.73269
Cluster 2	DUDENS90*.509907-1.5515163	.84382
Cluster 3 (all)	DUDENS90*.473801+GROWRATE*-.133787-1.182371	.87949
Cluster 4 (all)	DUDENS90*.625104+TRANS*.05695+GROWRATE*-.094558-2.251854	.85906
Cluster 5 (all)	DUDENS90*.551298+GROWRATE*-.200179-1.462126	.90520

where:

DUDENS90 = municipal dwelling units in 1990 (or future year) ÷ municipal area
INCOME90 = 1990 (or future year) municipal median household income
GROWRATE = numeric increase in residential density
TRANS = percentage of municipal multi-household dwelling units

8. The statistical analysis demonstrates that in some municipal groupings the relationship between residential density and municipal road density is improved by the addition of a variable (GROWRATE) which represents the number of new dwelling units recently constructed in the municipality. This variable appears to be significant only in municipal groupings where the change in residential density increased by 100 (or more) dwelling units per square mile during the past 10 to 20 years. The effect of this variable is to reduce the amount of municipal roadway per dwelling unit.

Inferential Findings

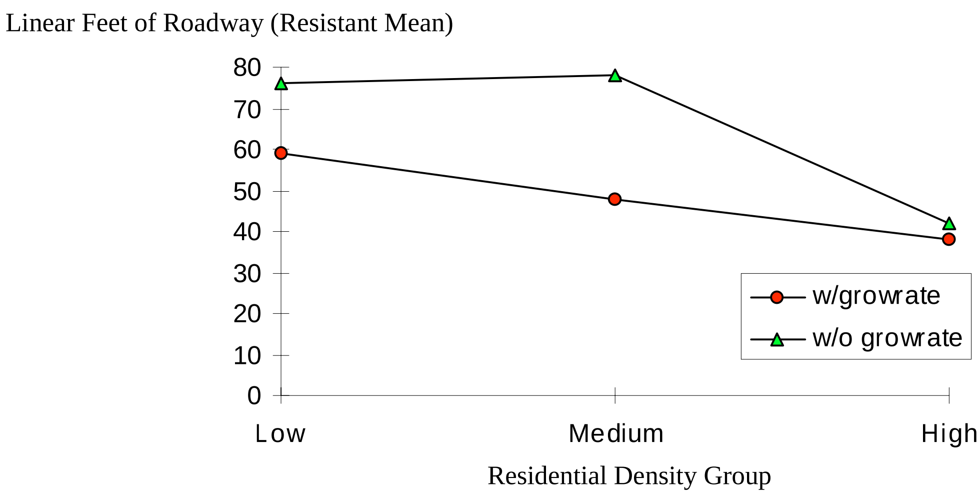
In addition to statistical methods, inferential analysis also was used and produced the following findings.

1. Inferential evidence is presented that strongly suggests that the difference in road density, in municipalities where the GROWRATE variable is significant, is not a lagged deficit but represents an efficiency. Initially it was thought that this variable was reporting a deficit of municipal roads - that the construction of municipal roads lagged the development of substantial numbers of new residential units. However, by estimating both the percentage of land developed into dwelling units and the likely lot geometries, a ratio of the length of municipal roadway (linear feet of road per dwelling unit) divided by the likely residential lot frontage (in feet) has been used to compare the various municipal groupings (as shown in the following diagram). Note that those cluster-subgroups of municipalities that use the variable GROWRATE also have lower resistant mean linear feet of roads. Of particular interest is the finding that this lower road ratio holds true regardless of residential density.

Ratios of Municipal Road per Dwelling Unit to the Estimated Frontage

Residential Density Group	Cluster - Subgroup	Frontage Min.	LF of Roads	Ratio	Significant Variables
Lowest Density	1	257	118	.46	DUDENS90+lnhhi
Low Density	3 - 1	223	65	.29	DUDENS90 - growrate
	4 - 3	223	76	.34	DUDENS90
	5 - 2	223	53	.24	DUDENS90 - growrate
Medium Density	4 - 1	76	78	1.03	DUDENS90
	5 - 1	76	48	.63	DUDENS90 - growrate
High Density	2	48	42	.88	DUDENS90
	4 - 2	48	38	.79	DUDENS90 - growrate + trans
Highest Density	3 - 2	22	18	.82	none

Comparison of Linear Feet of Roadway (per DU)



2. This paper proposes that incremental changes to the geometric pattern of the municipal street network may be the underlying cause for this 'GROWRATE-linked' municipal road efficiency. Visual inspection of municipal aerial photographs suggests that three street network patterns can be identified. The grid and rounded grid, typical of the State's older urban areas and the early post World War II suburbs, exhibit the highest road-to-frontage ratios. The historic rural pattern formed by a grid-based town with radial projecting into the countryside has a much lower ratio. In areas where GROWRATE, which is to say substantial amounts of recent growth have occurred, a branching pattern of linear access roads ending into cul-de-sacs is evident. The following GIS generated images displays the three typologies as found in parts of the listed municipalities.

Municipal Roads - Network Pattern Typologies

Grid / Rounded Grid | Historic Rural | Branching

Statistical Policy Implications and Recommendations

1. In general, municipal road density is inversely related to residential density. The assumption that more municipal roads need to be built to support rural development is true (density is inversely related to municipal road predictions). The statistical relationship between residential density and municipal road length has several policy implications.

a. This finding suggests that municipal zoning decisions directly affect the amount of municipal roads that need to be built and that residential zoning decisions affect: the amount of roadway that the municipality needs to maintain; the capital cost of supplying utilities that follow local roads; the costs of police and school busing; and, the vehicle miles traveled (VMT) thereby having direct implications for State air and water quality programs.

b. The relationship demonstrates that lower density results in the need for more infrastructure per dwelling unit. The statistically derived equations presented in this report might be used as a method to calculate infrastructure costs associated with residential development. This methodology might be used to demonstrate the "rational nexus", required to support the imposition of exactions, that is required by New Jersey law.

c. The relationship could be the basis for establishing an impact fee related to residential density. Findings in this report could be used to establish a new method of allocating recurrent municipal infrastructure-related costs. Under such a system, a publicly beneficial density might be determined as a standard. OSP research has shown that a density of 6 dwelling units to the acre could provide a mix of housing that can be affordable and largely subsidy free for a cross section of New Jersey householders (Reilly 1993). Other standards might also be developed. Under such as system, persons wishing to build less dense housing would be allowed to proceed, but would be assessed an impact fee (which might be an annual charge) to reimburse for the added VMT's, roadway and other infrastructure impacts resulting from their desired lower density (compared to the municipal density standard). Development built at densities higher than the municipal standard might receive a municipal tax reduction (a type of negative impact fee).

d. The relationship demonstrates that the high density and clustering recommendations included in the State Development and Redevelopment Plan produce tangible public infrastructure benefits, as required by the legislative intent in the State Planning Act.

Inferential Policy Implications and Recommendations

This study suggests that using a branching municipal road pattern to serve residential portions of centers could result in substantial cost benefits. Community development boundaries for centers may also be more readily maintained using branching road networks in residential areas within the boundary. However, it is recognized that branching road networks likely lack sufficient collateral circulation and might have to be augmented with circumferential roads and pedestrian paths.

As this analysis only addresses the supply and cost of roads, case studies are appropriate to develop designs that also address these issues of capacity and circulation. A design that increases travel demand (VMT) or creates congestion bottlenecks on arterial roadways serving the development does not meet State planning objectives. These issues may be resolved through design guidelines for providing roadways and pathways traversing branches that provide and promote collateral (and non-vehicular) circulation within and among developments without increasing demands on arterial roadways. Such roadways and pathways can be built and maintained at substantially less cost than collector streets and enhance opportunities for social interaction. These guidelines should be considered in the context of other State planning objectives and incorporated among the OSP design guidelines for centers (New Jersey Office of State Planning, 1993).

2. The high densities (10 to 18 du/area) and the narrow lot geometries (1:5 and 1:4) now found in the older urban municipalities do not provide sufficient roadways to accommodate on-street parking (the resistant mean frontage was 18 feet, as shown in Table 5-4; a distance very close to the length of an automobile). This finding confirms the empirical experience that there is insufficient parking on city streets, and suggests that the lot sizes and densities found in these places require off-street parking or enhanced opportunities for alternative modes of travel such as transit, bicycle or walking.

Introduction

The original statistical study, Projecting Costs for Roads under Various Growth Scenarios (Gottlieb 1990), was part of a larger work effort[2] of the New Jersey Office of State Planning to develop a series of statistical models which could be used to evaluate the fiscal impacts of alternative regional growth management policies. This model was developed to forecast the 'need' for new lane miles of roadway, assuming that levels of service and mode splits remain unchanged. The model was viewed as a broad brush tool to estimate relative cost differences; not a model to compete with the more detailed simulations performed by the New Jersey Department of Transportation. The major findings of that study were:

1. that there is a strong relationship (adjusted R square of .8553) between municipal population density and the municipal density of local roads (the sum of municipal and county roads divided by the municipality's area); and,
2. that there was a moderate relationship (adjusted R square of .7767) between the population density of counties and their density of State roads.

The report also proposed a method to simulate the possible effects of a State Plan, by using a coefficient based on Hudson county's relationship between population density and local roads supply. The report suggested that Hudson County's more compact design was suggestive of the anticipated higher densities that were thought to be associated with the State Plan. The report proposed the use of a Hudson County based factor to adjust, or shift, trend municipal road mile forecasts into 'plan' municipal forecasts. The hedging about the specifics of the State Plan reflects the historic fact that the State Development and Redevelopment Plan (New Jersey State Planning Commission 1992) was being developed at the same time that the research and modeling was being performed.

This report re-investigates and expands the original pioneering research. Four reasons for undertaking this investigation at this time are:

1. the original research relied on estimated 1987 population forecasts. The new study will use the 1990 Census for its demographic, housing and income data;
2. the original research relied on New Jersey Department of Transportation road supply information that reflected a variety of base years for its observations. This study will use the latest available inventory information from NJDOT;
3. the State Development and Redevelopment Plan is published and OSP now has sophisticated computer simulations of the Plan. The revised Road impact model will be designed to interface with this Plan simulation model; and,
4. the original research used 1985 ES 202 employment data. This study will use 1990 ES 202 data.

[2]Studies and models also were prepared that evaluated public school capital costs, public sewer system costs and the operating costs of local, county and State governments.

I. Data and Procedure

Data

Population, income and housing information were collected from the US Censuses of 1990, 1980 and 1970. Municipal employment information (municipal-located jobs in 1980 and 1990) was collected from the New Jersey Department of Labor's ES 202 data files[3].

Areal data was transcribed (for the original 1990 study) from the 1985 New Jersey Department of Community Affairs publication, Municipal Data Book. During the course of this work effort, the municipal area information reported by DCA was compared to new municipal area calculations prepared by OSP using municipal boundary files developed by the New Jersey Department of Environmental Protection and based on corrected USGS topoquads. Municipal discrepancies as large as 30 square miles were identified, although the match for most municipalities was reasonably close. This study uses the municipal areas, in square miles, taken from the NJDEP GIS data set.

The latest available road inventory information was supplied by the New Jersey Department of Transportation. Lane miles of municipal, county and State maintained roadways were provided for the years 1990 and 1994. Initially, this study intended to use the 1994 data. However, when the NJDOT 1990 municipal road inventory was compared to the NJDOT 1994 inventory, changes were noted only in Morris and Essex Counties. Furthermore, the inventory differences in these two counties only reflect changes in the categorization of road ownership, not in the supply of roads (total roads remained constant, but state road miles might have increased while county road miles might have correspondingly decreased). This finding caused the validity of the 1994 data to be questioned since aerial photographs at OSP show that new local roads were built in most counties between 1986 and 1991, suggesting that some difference in the supply of municipal roads should have occured between 1990 and 1991. NJDOT's Bureau of Statewide Planning investigated this matter and reported that a new inventory was being conducted, but would not be finished for a considerable period of time.

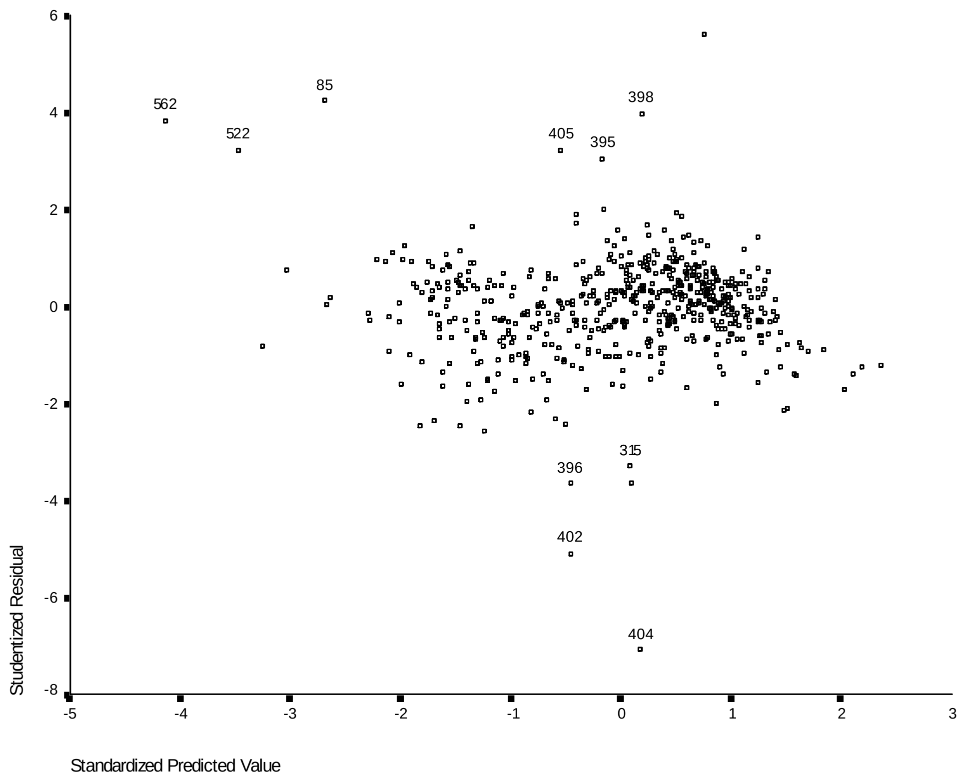
The question then became, which of the two road inventories was more reliable? To determine this issue, the linear regression reported in the original Roads report was performed using both sets of road inventory data. Then a scatterplot of the residuals was plotted. Outlying data was identified by case number, which then was identified with the county and municipal identity (FIPS code). It was discovered that those municipalities which had their inventories altered in the 1994 data set became data outliers. Diagram 1-1 displays the resulting 1994 road data scatterplot and shows the case number identified outliers. These same municipalities were not outliers in the 1990 data sets. This finding suggests that although it was possible that roadways in

---

[3]Although 1970 ES 202 data is available, it was not used in the study due to legal changes in the applicability of Workman's Comprehensive Insurance made in 1979. Prior to 1980, part time employees were not covered and therefore not included in this record. This change made the 1970 data inconsistent with the more recent records.

Morris and Essex counties did change jurisdiction during the period 1990 to 1994 (e.g. some municipal roads in 1990 became county roads in 1994), it was unlikely that the roadway changes in these counties were structural (i.e. county roads were torn up and replaced, mile for mile by new municipal or State roads). For that reason the decision was made to use the 1990 Road inventory data since its categorization appeared to be more internally consistent with the data set for the other 19 counties.

Diagram 1-1
Scatterplot of Residuals Using 1994 Road Data
Outliers from Essex and Morris Counties are Identified by Case Number



Procedure

The initial data analyses performed in this study rely principally on two statistical procedures. On almost all data, linear regression was performed. Linear regression is a mathematical procedure used to study relationships between different sets of data (variables). For example, in the 1990 study, linear regression revealed that changes in population density were symmetrical (related) with changes in the supply of local roads (municipalities with higher population densities had more local roads than did municipalities with lower population densities). The regression program expresses these relationships as equations which are then incorporated into models used to predict future requirements for road miles (based on forecasted increases or decreases in municipal population density). Because linear regression only evaluates *linear* relationships, a fair amount of time was spent coaxing data into linear relationships (the technical term for this mathematical process is ‘transformation”). Converting a value to its logarithmic form is a technique commonly used in this report to transform data into linear relationships.

The other main statistical procedure used in this section is called Clustering (K-Means). As the name implies, this a method to group data into categories. Normally, data is first clustered into groups and then each of the groups is analyzed using linear regression (or some other technique). The K-Means procedure is used to group very large data files, when the number of groups cannot be determined by other analytical methods (because of the large file size). It requires that the user estimate the number of groups. Given the number of groups, the procedure calculates the center value of each group and then assigns each case to one of the groups. To determine the ‘right’ number of groups, one simply guesses the number and takes a look at the characteristics of each group to see if they make sense. One then repeats the procedure using different group numbers and evaluates the results.

Once data was grouped and regressed, several tests were performed to discover if the relationships were valid and if a model developed from the equation would be stable. These tests will be discussed in the report as needed. The resulting regression equation then was used to ‘backcast’ the data which went into producing the analysis. This process is a little like testing an addition by subtracting the additive from the product to see if you end up with the starting number. In statistics, this testing against real data is very important, since the regression equation produces only the *most likely result*, and may not do a very good job of predicting the actual number for a specific municipality. It is very useful to graph this reality testing, because it sometimes demonstrates tendencies in the regression equation that can lead to a more detailed statistical study and a better fitted model.

II. Initial Statistical Analysis

Re-Examining the Original Local Road Regression

It seemed reasonable to begin this analysis by finding out if the new municipal population and road supply data would change the original study findings. In effect, if the relationship was still strong then the work could focus on revising that portion of the model that evaluates the impact of Plan. The results of this linear regression are shown in Table 2-1.

Table 2-1
Regression of Population Density and Local Road Inventory

```
where:
      LNPOP90 = the municipal population divided by the municipal area,
expressed in natural log form

Variable(s) Entered on Step Number
   1..    LNPOP90D
Multiple R           .86973
R Square             .75644
Adjusted R Square    .75600
Standard Error       .80198

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           1          1128.58110       1128.58110
Residual           565           363.39068           .64317

F =    1754.71843       Signif F =  .0000

------------------ Variables in the Equation ------------------
Variable              B        SE B       Beta          T  Sig T
LNPOP90D        .921882     .022008    .869733    41.889  .0000
(Constant)    -4.710614     .162040              -29.071  .0000
```

The regression based on the revised data still showed a good relationship between the density of population and the supply of local road miles, confirming the original report findings. However, the strength of this relationship (adjusted R Square)[4] decreased from .8553, in the 1990 report, to the current finding of .756. This decline in the adjusted R square means that the linear equation produced by the statistical procedure does not 'fit' the actual data as well as it used to. For this reason, a detailed statistical analysis of the relationship between road inventories and other municipal attributes was conducted.

[4]OSP uses the adjusted R square, rather than the R square or the Pearson coefficient (R), because it is generally considered to be a more stable value and a better predictor of how well a model would fit the data.

Variables[5] for Analysis

Perhaps the most subtle finding in the 1990 report was that municipalities with the same population densities can have different supplies of local roadways. This difference might be due partially to chance (there is no absolute road supply standard) or topographical differences (hills might require switchback-like roads, whereas a straight road would serve the same purpose on flatland). However, the 1990 report hypothesized that this road supply range was partially due to conscious human decisions in the built environment, which result in changes in the design and density of development from one municipality to another.

Independent Variables

Examples of design and/or density attributes that might affect the need for local roads are numerous. The original Plan simulation used the “Hudson shift”, which argued that Hudson County municipalities have relatively fewer miles of local roadway than do municipalities of the same population density in other counties, because of the “PCTPRE40 form of urban design - one that is not so oriented to the automobile” (Gottlieb 1990). This PCTPRE40 urban form suggests that the grid street pattern is more efficient than the meander of suburban roads. In another section of the report it is argued that “ if two municipalities have the same population density (i.e., position on the horizontal axis) then their relative positions on the vertical axis may be a good measure of the amount of sprawl found within their borders”. This argument suggests that clustering of single family houses onto part of a large development tract would likely require less roads than the same number of dwelling units built on large lots on the same tract. A similar form of clustering would occur with the construction of multi household dwelling units. Anti-clustering (if there is such a word) might result from the construction of numerous large estates in wealthy communities. The concentration of employment into a community, perhaps by the construction of an industrial park, also might affect the amount of roads needed to serve the municipality. The variables listed in Table 2-2 were selected for testing in this study.

The original variable, population density, was not selected for use in this study. The reason is that population density was replaced by dwelling unit density on the assumption that roads are built to provide access to dwelling units. This change also corrects the problem of how to account for vacant units. Under the original population density model, a municipality with a high degree of vacancy (such as a rapidly growing suburban area with lots of new units newly constructed but not sold at the time of the Census) might appear to have ‘too many’ road miles. It should also be noted that population and the number of dwelling units are highly collinear, which means the substitution of one variable for the other is more in the nature of a technical adjustment and should make little difference in the regression.

[5]Independent variables are attributes that are thought to have an effect on the dependent variables. The Dependent variable is the attribute that is being controlled or that is to be described by the effect(s) of the independent variables. In a model, the dependent variable is the value the model is trying to predict.

Table 2-2
Independent Variables Chosen for Testing

<u>Measures of Residential Development</u>

DUDENS90 - the natural log of (the number of municipal dwelling units in 1990/municipal area)
DUDENS80 - the natural log of (the number of municipal dwelling units in 1980/municipal area)
DUDENS70 - the natural log of (the number of municipal dwelling units in 1970/municipal area)

<u>Measure of Employment Development</u>

JOBDEN90 - the natural log of (the number of municipal based jobs in 1990/municipal area)
JOBDEN80 - the natural log of (the number of municipal based jobs in 1980/municipal area)

<u>Measures of Development Growth</u>

DUDIF89 - the natural log of the numeric change in municipal dwelling units 1980 to 1990.
DUDIF78 - the natural log of the numeric change in municipal dwelling units 1970 to 1980.
JOBDIF89 - the natural log of the numeric change in municipal based jobs 1980 to 1990.
JOBDIF78 - the natural log of the numeric change in municipal based jobs 1970 to 1980.

<u>Measures of Employment Agglomeration</u>

JOBTOT90 - the total municipal based jobs in 1990.
JOBTOT80 - the total municipal based jobs in 1980.

<u>Measures of Density or Design</u>

PCTMULTI - the percentage of multiple household dwelling units in each municipality in 1990.
INCOME90 - the natural log of median municipal household income in 1990.
PCTPRE40 - the percentage of total municipal dwelling units constructed before 1940.
PCTDUS45 - the percentage of total municipal dwelling units constructed during the 1940's and 50's.
PCTDUS60 - the percentage of total municipal dwelling units constructed during the 1960's.
PCTDUS78 - the percentage of total municipal dwelling units constructed during the 1970's and 80's.

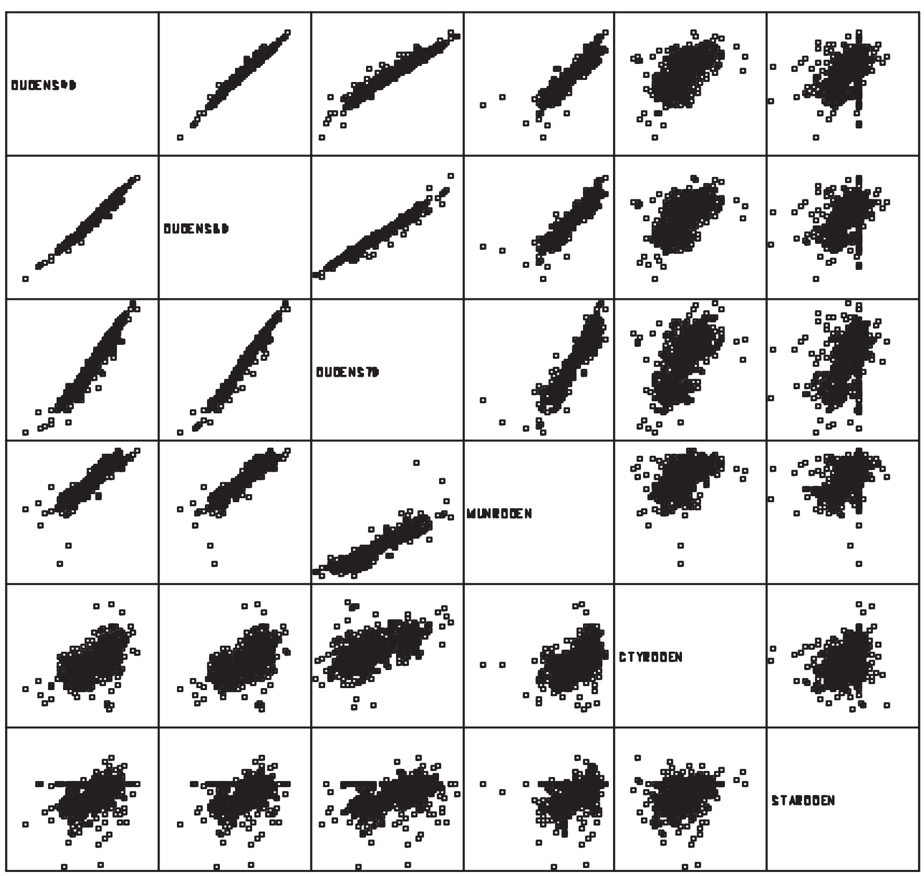
Some other variables also need an additional word of explanation. The last four variables identified as measures of density or design are the percentages of total dwelling units, by municipality, built during specific time periods. The idea behind this grouping was that in general, houses constructed during these time periods represent differing densities, with the most compact development dating from the period prior to World War II. Housing built after W.W.II and through the 1950's is assumed to consist largely of small lot, relatively modest sized (800 to 1200 square feet) units. Since the 1960's the size of residential units has increased, and it is assumed that lot sizes have also increased. The last category consists of those houses built between 1970 and 1990 (PCTDUS78).

Choosing the Dependent Variables

In the 1990 study, municipal and county roads were combined into a single group termed 'local' roads. State roads were evaluated separately. With the decrease in the explanatory ability of the model (the R square declined), it was thought prudent to re-examine the issue of which category of roadway is most responsive to changes at the municipal level.

Perhaps the easiest way to perform this analysis is by using a scatterplot. If the variables are related, a pattern will be formed on the chart. If no relationship is found, then a random pattern will be seen. Diagram 2-1 displays the scatterplots of the variables MUNRDDEN (lane miles of municipal roads divided by municipal area, expressed as the natural log), STARDDEN (lane miles of State road divided by municipal area), CTYRDDEN (lane miles of county road

Diagram 2-1
Scatterplot Matrix Various Road Miles and Dwelling Unit Densities



divided by municipal area), DUDENS90, DUDENS80 and DUDENS70. The more linear and compact the grouping of points, the more significant is the relationship.

Diagram 2-1 contains several interesting findings. First, a good relationship is displayed between all of the dwelling unit densities and municipal road density. Perhaps just as importantly, poor relationships are displayed between all types of road densities, suggesting that the supplies of these roads systems serve different needs. Moderate to weak relationships appear to be exhibited between county road density and dwelling unit densities and between State road densities and dwelling unit densities. These findings suggest that the supply of municipal roads is far more responsive to municipal attributes than are the supplies of either county or State roads. This estimate was tested by performing regression analysis. The analysis found that the density of county roads and the density of dwelling units produced a weak correlation with an adjusted R square of .33639. The regression of State roads found a weak correlation (.27) with 1990 employment density. Because of this finding, only the supply of municipal roads will be studied in this report. It appears unlikely that a reliable model to predict the changes in either county of State roads can be based on municipal changes.

Clustering the Data to Control for Municipal Area

The premise of the 1990 analysis was that there was some relationship between the density of residential development (population density) and the municipal inventory of local roads. The paper also noted that the less developed municipalities "need to build more miles of local roadway per new resident than do densely populated municipalities" and that this phenomenon "may be a good proxy for deficient road net (high marginal costs to growth) in rural areas". It also could be the case that very large municipalities might *need* to build more roads than smaller municipalities to provide the same level of service to identical population densities. It also could be the case, as stated earlier, that design and density decisions effect the range of roadways, as might topography. If one were to construct a single model to be used to predict municipal miles for all places in the state, the fit of that model would have to suffer due to the diversity of local conditions (beyond those being tested in this analysis) that might affect the need for roads. Perhaps more importantly, the model might result in misleading forecasts. For example, many municipalities in the State had substantial portions of their area developed at a time when higher density urban forms were common. To include this data in a Trend model intended for use in the more rural, less developed municipalities could result in an under-estimation of the road miles needed to accommodate growth in these places.

To attempt to control for some of these underlying conditions, the municipal data (sets of all the variables for each municipality) were clustered (grouped) by two variables: the size of the municipality and the amount of municipal roadway in the municipality in 1990. New Jersey's municipalities are very diverse. Diagram 2-2 shows the cluster groups as areas in the chart with distinctly different patterns of municipal road miles (MUNI) and municipal area (AREA). Table 2-3 and Diagram 2-3 displays the ranges among the 567 Municipalities, with respect to area and their inventories of municipal roads.

Diagram 2-2
Municipal Clusters as a Function of Municipal Area and Road Supply

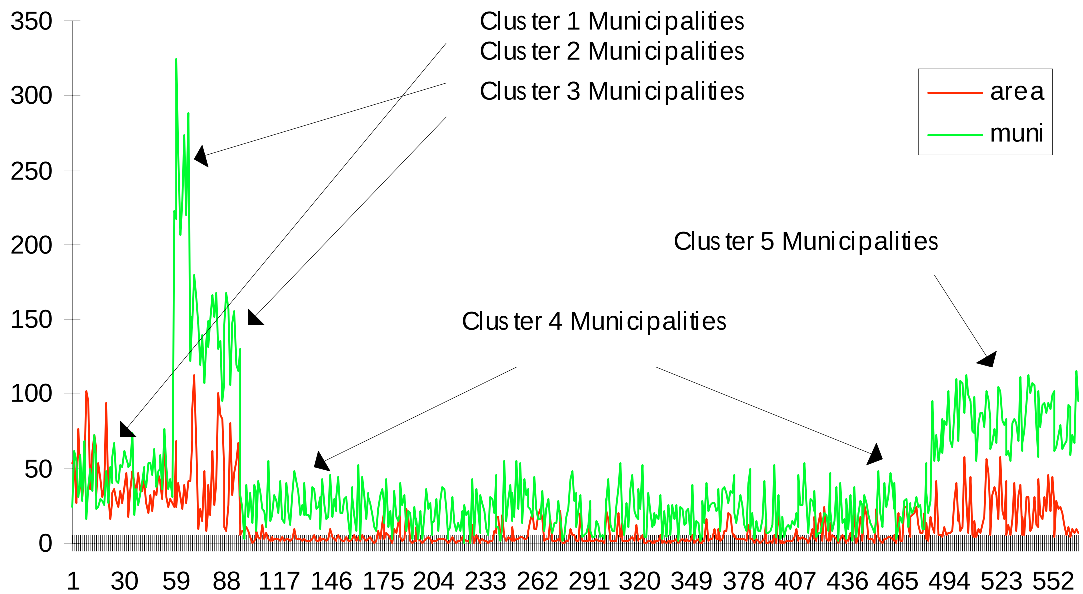
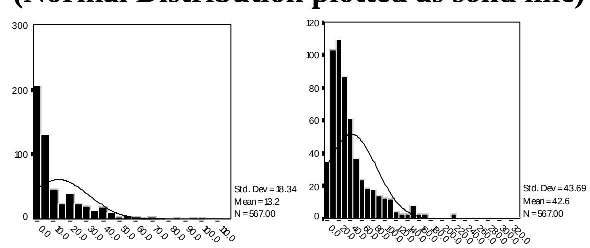


Table 2-3
Description of Municipal Area and Road Inventory Characteristics

	Municipal area in square miles	Miles of Municipal Roads
average	13.241	42.62443
max	112.7908	325.02
min	0.097515	0.01
median	4.390685	28.6

Diagram 2-3
Histogram of Municipal Area and Road Inventory
(Normal Distribution plotted as solid line)



Municipal Area Miles of Municipal Roads

Compared to these general areal and municipal road characteristics, Table 2-4 displays characteristics for each of the five groups identified in the Cluster Analysis. Map 1 shows the cluster identity of each municipality in the State.

Table 2-4
Characteristics of the Five Cluster Groups
(Center Points Displayed in the first chart)

Cluster	Number of Cases	Area (square miles)	Total Municipal Road Miles	Average Road per sq. mile of Area
1	57	41.4831	44.8358	1.073
2	9	35.3309	249.3344	7.114
3	29	46.2870	139.2700	3.022
4	389	4.8519	21.8502	5.250
5	83	19.2217	82.2871	4.315

Cluster 1 - 57 Municipalities

Variable	Mean	Std Dev	Minimum	Maximum
AREA	41.48	18.65	16.16	101.68
Muni. Roads	44.84	14.52	16.54	76.69

Cluster 2 - 9 Municipalities

Variable	Mean	Std Dev	Minimum	Maximum
AREA	35.33	14.57	23.40	69.00
Muni. Roads	249.33	40.34	206.12	325.02

Cluster 3 - 29 Municipalities

Variable	Mean	Std Dev	Minimum	Maximum
AREA	46.29	30.19	7.58	112.79
Muni. Roads	139.27	21.40	95.24	179.08

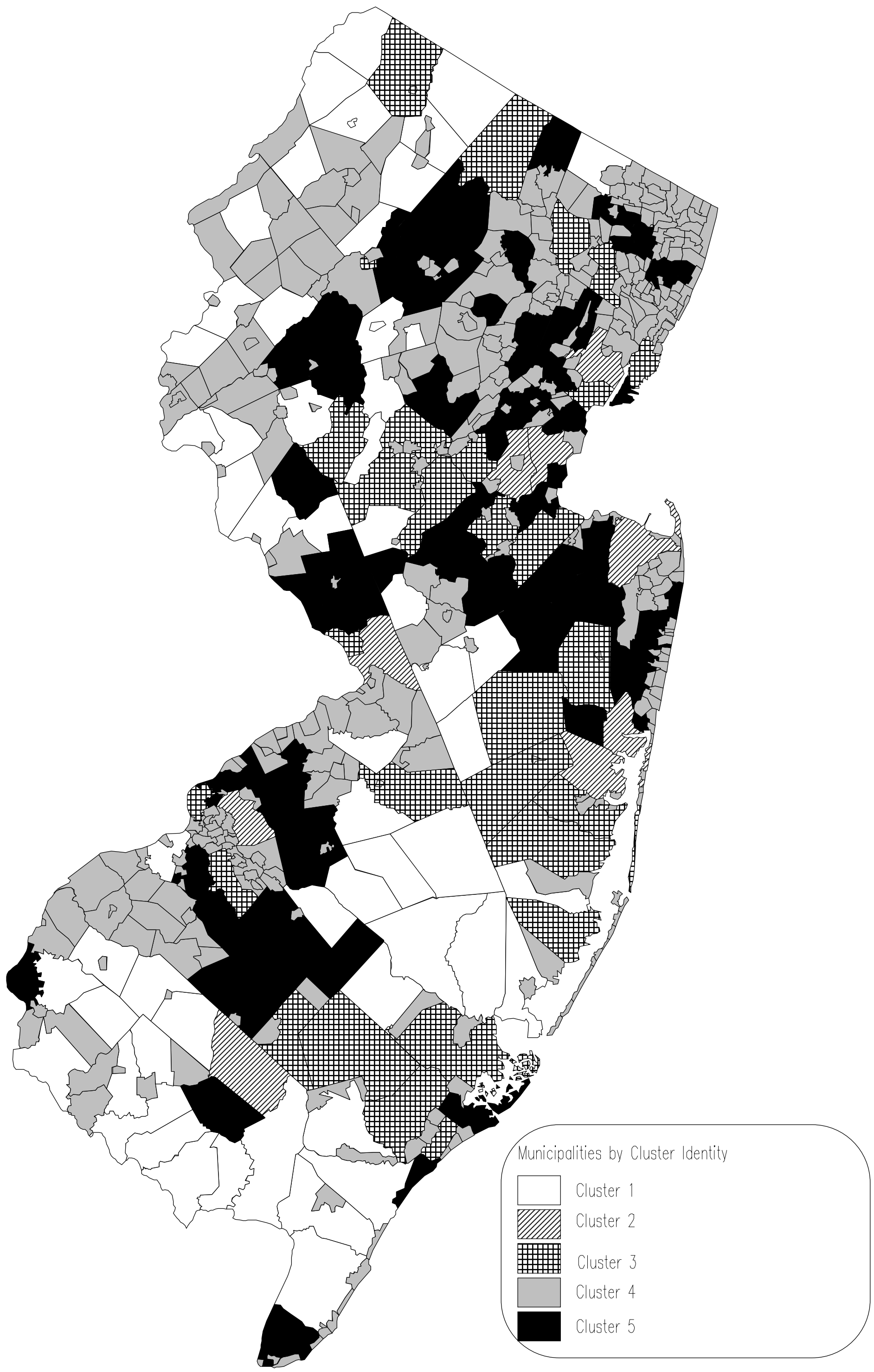
Cluster 4 - 389 Municipalities

Variable	Mean	Std Dev	Minimum	Maximum
AREA	4.85	6.02	.10	25.09
Muni. Roads	21.85	13.01	.01	55.22

Cluster 5 - 83 Municipalities

Variable	Mean	Std Dev	Minimum	Maximum
AREA	19.22	14.73	3.69	58.26
Muni. Roads	82.29	16.73	54.84	115.76

Map 1
Municipalities Identified by Cluster



While the cluster groups organize the State's municipalities with respect to area and municipal road mileage, it is useful to learn if these groupings help to organize other variables used in this study. Table 2-5 displays the average 1990 dwelling unit, employment and income characteristics for the municipalities in each of the clusters. The order of the clusters in the Table has been altered to more clearly show the relationships. It can be seen that both dwelling unit and job densities are symmetric and that they appear to be related to road density. Another interesting finding is that the relationship of dwelling units divided by jobs appears to be inversely related to development or road density. In cluster 2 which has the highest job, dwelling unit and road densities, the dwelling unit to job ratio is the lowest. Conversely, in cluster 1 which has the lowest dwelling unit, job and road densities, the du/job ratio is the highest[6].

It also can be observed that cluster 1 communities have substantially fewer dwelling units per road mile than do the more densely developed municipal clusters. This finding supports the 1990 report speculation that "...sparsely populated municipalities need to build more miles of local road per new resident than do densely populated municipalities" (Gottlieb 1990).

Table 2-5
Cluster Characteristics for Other Variables

Cluster	Mean Jobs	Mean DUs	DU/ Jobs	DU / Sq. Mile	Jobs / Sq. Mile	Median 1990 HH Income	Roads / Sq. Mile	DU / Roads
2	49660	37398	.75	1060	1419	$42,008	7.10	149
4	3490	3125	.89	620	698	$45,690	5.20	119
5	12321	9869	.80	526	648	$51,064	4.30	122
3	19806	19212	.97	413	430	$41,527	3.02	136
1	1841	2577	1.40	63	45	$46,787	1.07	58

Regression Analysis

Table 2-6 displays for each cluster, the variables found to be related to the density of municipal roads (MUNRDDEN); the regression coefficients for each variable; and, the Adjusted R Square for the resulting linear models.

It should be noted that the coefficients in Table 2-6 cannot be used to evaluate the relative importance of any of these variables, because the variable values are not standardized. (For example, the variable PCTPRE40 refers to a percentage ($0 \leq 1$) of units that were built before 1940 while the variable DUDIF78 actually refers to the natural log of the numeric change in dwelling units between 1970 and 1980.) To allow one to judge the explanatory value of each of

[6] The Pearson Coefficient of this relationship is -.1807 for municipalities. Using the clusters described in the following chapters the following relationships were discovered: cluster 1, .0718; cluster 2, -.1165; cluster 3, -.2706; cluster 4, -.2452; and, cluster 5, -.2227.

Table 2-6
Regression Results

Variable	Regression Coefficients				
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
DUDENS90		.679521			.650961
DUDENS80	.489596				
DUDENS70			.446806	.741519	
INCOME90	.773266				.368541
PCTMULTI		-.775132		-.788241	-.511280
PCTPRE40	.990876		-1.434624		
JOBDEN90				-.044447	
JOBDIF89				-.109081	
JOBDIF78			-.200735		
DUDIF78				.453164	.646203
DUDIF89				.299862	
constant	-10.163436	-2.48922	-.502141	-2.383663	-6.151111
Adj. R Square	**.78384**	**.93347**	**.91816**	**.84584**	**.95584**

the variables, Table 2-7 shows how each variable improves the 'fit' of the regression equation. The effect of adding each variable can be interpreted by noting the difference in the R square of the regression equation.

Table 2-7
Explanatory Value of Regression Variables

Cluster 1

Variable	Adjusted R Square	R Square Improvement
DUDENS80	.63484	.63484
INCOME90	.76458	.12974
PCTPRE40	.78384	.01926

Cluster 2

Variable	Adjusted R Square	R Square Improvement
DUDENS90	.84382	.84382
PCTMULTI	.93347	.08965

Cluster 3

Variable	Adjusted R Square	R Square Improvement
DUDENS70	.88421	.88421
JOBDIF78	.90329	.01908
PCTPRE40	.91816	.01487

Cluster 4

Variable	Adjusted R Square	R Square Improvement
DUDENS70	.80926	.80926
PCTMULTI	.82578	.01652
DUDIF78	.83742	.01164
JOBDEN90	.84138	.00396
DUDIF89	.84394	.00256
JOBDIF89	.84584	.00190

Cluster 5

Variable	Adjusted R Square	R Square Improvement
DUDENS70	.90245	.90245
INCOME90	.92451	.02206
DUDIF78	.94277	.01826
PCTMULTI	.95584	.01307

General Findings

The regression of clustered cases produced models which exhibit better fit of the actual data than the fit exhibited by the model generated using all of the cases.

Measures of Residential Development and Growth

1. Consistent with the 1990 findings, this analysis documents the important and positive relationship between the supply of local roads (lane miles of municipal roads divided by municipal density) and the density of residential development (as dwelling units increase the road supply increases, and vice versa). This variable alone explains between 63% (cluster 1) and 90% (cluster 5) of the road supply in the State's municipalities.

2. While the density of dwelling units was the most explanatory variable in each of the cluster groups, the specific time period when residential density was most significant changed from one group to another. In cluster 2, the current (1990) dwelling unit density is significant, while in cluster 1 the 1980 density is used and in clusters 3, 4 and 5, the 1970 residential density was significant. Since the average number of dwelling units has increased in each of the clusters, one possible way to interpret the changing date of the residential density would be to conclude that municipalities in clusters 1, 3, 4, and 5 are exhibiting a *deficit of municipal road miles*.

3. The change in the number of dwelling units was found to be significant, but not very explanatory in clusters 4 and 5.

Measures of Employment Development and Growth

1. The study found that employment agglomeration is not related to road density.

2. The density of employment was not significant. Employment change was not significant in estimating road density in clusters 1, 2 and 5, which represent 149 municipalities. However, in clusters 3 and 5 (418 municipalities), employment change was significant but improved the equation by less that 2% in cluster 3 (29) municipalities) and less than 1% in the 389 municipalities in cluster 4. Where employment change was significant, the signs of the coefficients always were negative, which means that if employment increased (new jobs added) the road density decreased slightly, and vice versa.

3. The failure of employment to exhibit a relationship with road density is interesting. In Table 2-5, it was noted that both the employment density and the dwelling unit densities were symmetric with road density. The regression confirmed the relationship between dwelling units and municipal roads, but did not confirm the relationship with employment. Two explanations for the failure of employment density to be significant are: 1. Employment density is highly autocorrelated with dwelling unit density (this explanation is suggested by the relatively close dwelling unit to job ratios shown in Table 2-5); and, 2. The possible fact that when new dwelling units are built, new circulation roads are built by the developer and then deeded to the municipality but when new employment facilities are built, they tend to either link directly with existing local roads or retain any new circulation roads in private ownership.

Measures of Density or Design

1. Income (INCOME90), the percentage of multi-household dwelling units (PCTMULTI) and the percentage of housing units built before 1940 (PCTPRE40) were explanatory variables in each of the clusters.

2. Income was significant and substantially improved the ‘fit’ of the regression equation in clusters 1 and 5, which are the clusters with the highest average incomes. In cluster 1, income improves the regression equation by almost 13%, while in cluster 5 the improvement is a little over 2%. In both cases the coefficient sign was positive, indicating that if one were to compare municipalities of identical residential density, where incomes were higher the road density was higher, and vice versa.

3. PCTMULTI was significant and negative in clusters 2, 4 and 5. (Perhaps not surprisingly, clusters 2, 4 and 5 have the highest dwelling unit densities.) The negative sign means that where the percentage of multi household units was high, there was a smaller supply of municipal roads and vice versa. This suggests that multi household units, when constructed, do not generate the same level of access road construction as do single family units, or that access roads built to serve the development remain in private ownership.

4. PCTPRE40 is a mysterious variable to interpret. It is significant, but not very explanatory, in clusters 1 and 3. In cluster 3, the sign of the regression coefficient is negative, which means that where the percentage of dwelling units built before 1940 is high, the municipal road supply decreases and vice versa. This is the finding one might expect as a result of the Hudson shift. However, in cluster 1, the sign of the regression coefficient is *positive*. To determine if the effect of PCTPRE40 is being masked by an autocorrelation, a partial correlation

was performed for the municipalities in cluster 1. This analysis is displayed in Table 2-8 and demonstrates that controlling for PCTPRE40 and INCOME90 have very little effect. The analysis also shows that the effect of PCTPRE40 is *negative*. The explanation to this dilemma may be in understanding the cluster characteristics.

Table 2-8
Correlation and Partial Correlation Analysis
Cluster 1 - Significant Dependent and Independent Variables

	DUDENS80	MUNRDDEN	PCTPRE40	INCOME90
DUDENS80	1.0000	.8008	-.5701	.3366
MUNRDDEN	.8008	1.0000	-.3843	.6112
PCTPRE40	-.5701	-.3843	1.0000	-.3193
INCOME90	.3366	.6112	-.3193	1.0000

Controlling for.. INCOME90

	DUDENS80	MUNRDDEN	PCTPRE40
DUDENS80	1.0000	.7985	-.5184
MUNRDDEN	.7985	1.0000	-.2522
PCTPRE40	-.5184	-.2522	1.0000

Controlling for.. PCTPRE40

	DUDENS80	INCOME90	MUNRDDEN
DUDENS80	1.0000	.1985	.7670
INCOME90	.1985	1.0000	.5584
MUNRDDEN	.7670	.5584	1.0000

Cluster 1 is the least densely developed. The average number of municipal dwelling units (and jobs) is the lowest of the five clusters. The overall average area is the second largest. Income also is the second highest. Cluster 1 contains many of the more larger, more rural townships in the State. The high percentage of houses built before 1940 reflects the existence of traditional villages and towns. The resulting regional development pattern then is one with small centers placed in a largely rural setting.

Cluster 3 municipalities have much larger numbers of dwelling units, the lowest average median household income and municipal areas are just slightly larger that cluster 1. This cluster contains the older developed municipalities in the state, suggesting that the Hudson shift was correct in principle. It also demonstrates that in two municipalities with identical dwelling unit densities and median household incomes, the development pattern of small centers separated by largely undeveloped tracts of land requires more roadway than would the development of larger centers.

III. Refining the Model Equations

The regression analysis in the preceding section identified which variables were significantly related to the density of municipal road miles. For example, the density of 1970 residential development, the 1990 median household income and the percentage of housing units built before 1940 were used in an equation whose line passed close to a majority of the actual 1990 municipal road density values found in cluster 1 municipalities. In this section, the equations are simplified in an attempt to make the equations more robust and stable.

While it is easy to assume that the use of all of these variables would produce the 'best' model to predict road density in cluster 1 municipalities (that model would by definition produce the highest adjusted R Square), this strategy may not be true. For example, equations based on a large number of variables tend to be unstable and tend to produce results that closely fit the data that they were based on primarily as a result of their polynomial form and not as a result of the appropriateness of the selected variables. When these same variables are used to predict different data sets, then the model results may not be very accurate. In most equations fewer variables produce better models. Also, it is necessary to closely examine the relationship between the significant regression variables in a attempt to understand how these variables relate to each other and more importantly how these variables relate to the dependent variable.

Three statistical methods are used to determine the best model, given the regression analysis' significant variables. The first two techniques, partial correlation and three dimensional scatterplot mapping, are used to identify relationships between the independent and dependent variables. The third technique is to use a variety of charts to visualize the nature of the errors that would result from the use of alternative models.

The principal analytical tool used in this examination is partial correlation, a statistical procedure used to define the nature of the relationships between groups of variables. In all of these tables, the first line shows the Pearson Correlation coefficients (expressed as values between 0 and 1, with 1 being a perfect fit) which express the relationship between all of the independent variables and MUNRDDEN, the dependent variable. Then, the analysis proceeds, line by line in the table, to control (or remove the effect caused by) one or more of the independent variables. This controlling can result in changes to the correlation coefficient of the remaining independent variables. When a variable is controlled and it substantially reduces the coefficient of any of the remaining variables (or causes a coefficient sign to change), it is usually the case that the controlled variable related most strongly to affected independent variable and not to the dependent variable. Such variables then are removed from the model.

Where a controlled variable causes little change to the coefficients of the remaining independent variables, it is likely that this variable does not modify or magnify any of the other independent variables. This controlled variable may relate directly to the dependent variable, or curiously it might closely relate to another variable which was not included in the regression analysis, but that might be very explanatory. To sort out the nature of these relationships, visual examination using three dimensional (3D) scatterplots is used. If a second variable directly relates to the dependent variable, the municipal data will array in a spiral climbing the Y axis. Since 3D plots cannot be readily seen on the two dimension plane of this page of paper, multiple views of the scatterplot are shown - one view of the XY axis and the second view of the ZY axis. (The view of the XY axis shows the relationship between the dependent variable on the Y axis and the primary explanatory variable on the X axis. The ZY view shows the relationship between the dependent variable and the secondary variable on the Z axis. If more than one secondary variable is left for analysis additional ZY views are shown.) If no relationship to the dependent variable is shown, the variable is dropped form the equation. Conversely, if a strong relationship is demonstrated the variable is retained. Sometimes, a mixed relationship is evident. In these cases clustering is used to further subdivided the municipal group.

The final analysis uses several type of charts to visualize the characteristics of the model(s) produced by the remaining independent variables. A primary assumption of regression is that the distribution of error is random and that the variance of these errors is dependent on (proportionate to) their distance from the mean. To test this assumption, two scatterplots of the errors are produced. Both of these plots really view the same data. The difference between the plots is the emphasis of the diagramming procedure. In the first scatterplot, the errors from the regression are represented on the Y axis and the predicted (ideal) error results are represented along the X axis. The resulting plot should array in a U shaped form, with the smallest error centered on the mean of the data. This plot is intended to highlight the variance of errors as the data moves away from the mean. The second test uses a Normal Probability Plot of the residuals. In this type of diagram, the broken line at a 45 degree angle from the XY axis represents the ideal random distribution of errors. The plot consisting of data circles represents the error for each of the municipal cases in the cluster set. If the error distribution is normal, a line drawn through these municipal specific points should closely approximate the 45 degree ideal distribution. This plotting method illustrates the degree to which the errors are random. If the errors do not closely match the 45 degree broken line, the use of *Linear* Regression may not be appropriate.

Cluster 1

Table 3-1 displays the correlation coefficients for the statistically significant independent variables used in the regression equation to estimate municipal road density data for municipalities in Cluster 1.

Table 3-1
Cluster 1 - Partial Correlations with MUNRDDEN

DUDENS80	**INCOME90**	**PCTPRE40**
.8008	.6114	-.3843
.7670	.5588	controlled
.7983	controlled	-.2527
.8065	controlled	controlled

The Table (3-1) also shows that controlling for either income and/or the percent of dwelling units built before 1940 has little effect on the principal explanatory variable, DUDENS80. However, controlling for INCOME90 reduces the coefficient for PCTPRE40. It appears that PCTPRE40 relates to INCOME90 in the regression equation. (A negative relationship between PCTPRE40 and INCOME90 suggests that median household incomes tend to be lower in municipalities with higher percentages of older housing, and vice versa.) Therefore, the model equation to predict MUNRDDEN might exclude the variable PCTPRE40.

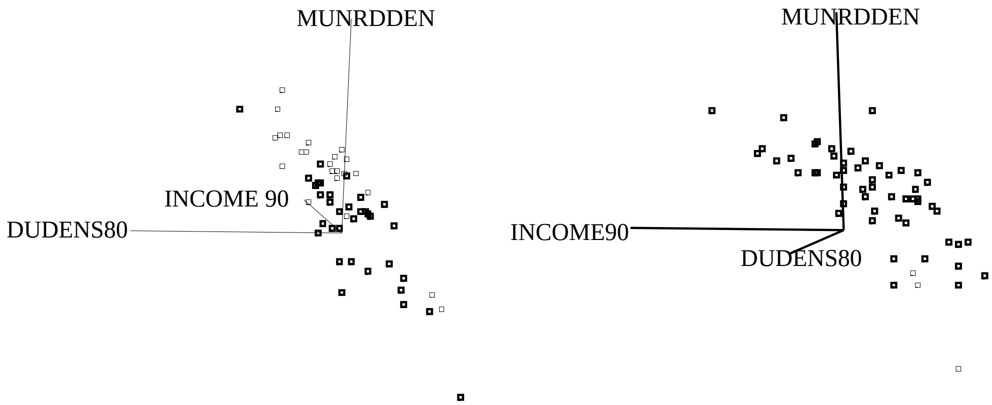
To test this hypothesis, a second regression equation was run which excluded PCTPRE40. Several tests of the regression assumptions then were performed with the only unusual result displayed in Table 3-2. The table shows that the constant and INCOME90 are highly interdependent. This means that to a large extent, the effect of the constant and INCOME90 to improve the ‘fit’ of the data is similar. However, it also demonstrates that the *independent* variables really are independent (near-dependency between DUDENS80 and INCOME90 is not exhibited). This finding suggests that small changes in municipal income can produce large changes in the model’s prediction of the supply of municipal roads.

Table 3-2
Collinearity Diagnostic

			Variance Proportions		
Number	**Eigenvalue**	**Condition Index**	**Constant**	**DUDENS80**	**INCOME90**
1	2.96637	1.000	.00008	.00504	.00007
2	0.03330	9.438	.00377	.91252	.00286
3	0.00033	94.237	.99615	.08244	.99706

The question then becomes, ‘How can an equation improve its predictability (the R Square improves) when the results of the equation appear to become more unstable?’. What is at issue is the strength of the direct relationship between each of the independent variables and the dependent variable. This question was explored in Diagram 3-1, which provides two views of the 3D scatterplot of the variables: MUNRDDEN, DUDENS80 and INCOME90. The diagram shows that both of the independent variables exhibit direct relationships to MUNRDDEN, although the relationship between income and road density is weaker than is the relationship between residential density (1980) and road density. Therefore, this finding tends to argue for using both equations (although a bifurcated relationship between MUNRDDEN and INCOME90 is exhibited in diagram 3-1).

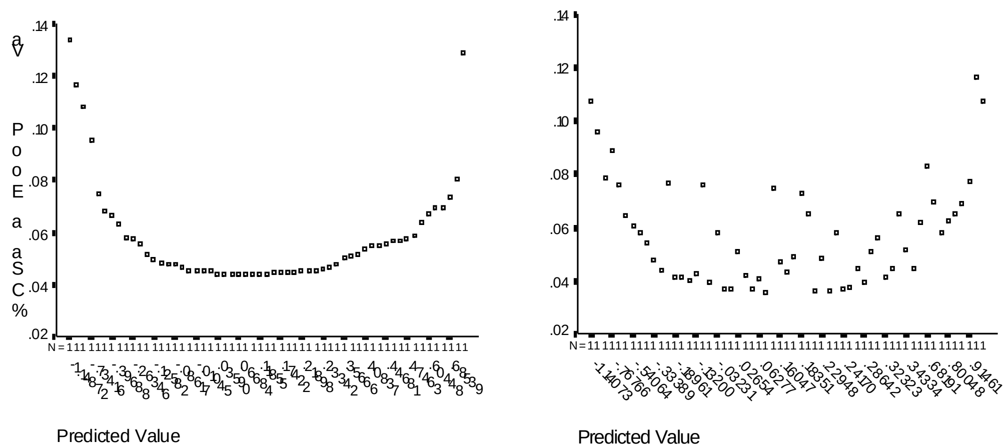
Diagram 3-1
3D Diagrams of Cluster 1 Variables



Because of the high condition index, an additional analysis was performed to determine the model’s volatility. A scatterplot of the predicted values (X axis) and the standard error of each predicted value (Y axis) is used for this analysis. Diagram 3-2 shows the resulting error scatterplots for the one variable equation and for the two variable regression equation. The chart demonstrates several findings. First, errors tend to increase as the distance from the predicted mean increases (this is normal and typical of regression equations). Second, error variation increases as more variables are added to the equation, but the size of the error for most municipalities decreased with the addition of more variables. This can be most clearly seen by comparing the error location on the Y axis of the one and two variable equations (this improvement accounts for the higher Adjusted R Square). This also suggests that the addition of INCOME90 improves the prediction of some cluster 1 municipalities, but for other municipalities the prediction error increases.

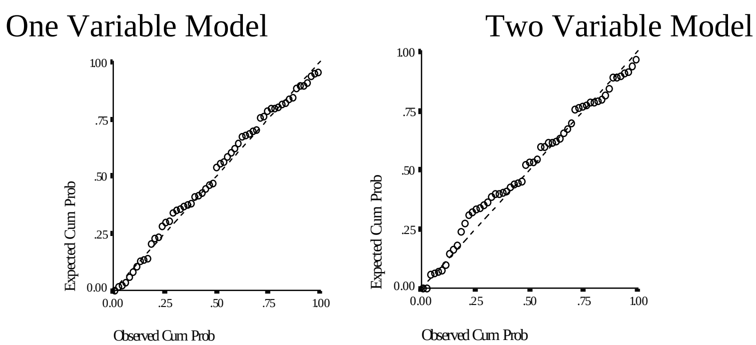
Diagram 3-2
Scatterplot of Predicted Values (Road Density) and Errors for Cluster 1 Models

One Variable Model Two Variable Model



To further test the cluster 1 models, regressions of the two models were performed and Normal Probability plots (NPP) of the residuals prepared (see Diagram 3-3). The one variable model produced an adjusted R square of .63484 and the two variable model produced an adjusted R square of .76444, an apparent substantial improvement in the ability of the equation to model the data. Although the one variable equation produced a lower adjusted R square, its NPP was slightly better fitted to the ideal distribution than was the two variable model. This finding confirms the findings shown previously in Diagram 3-2.

Diagram 3-3
Normal Probability Plot of Cluster 1 Models



It is unusual for an model with a higher R Square to produce a less well fitted probability plot. It is possible that this result is caused by inappropriately expressing the three dimensional mathematical relationship between the dependent and independent variables using a one dimensional line equation (regression). In any event, the use of either model appears to be warranted.

Cluster 2

The principal problem with the equation for Cluster 2 is that this set consists of only 9 municipalities. With such a small number of cases, the addition of any secondary or lesser variables can lead to misleading regression results.

To decide which of the two variables might be eliminated, Partial Correlation was used to sort out the effects of the significant variables, as shown in Table 3-3.

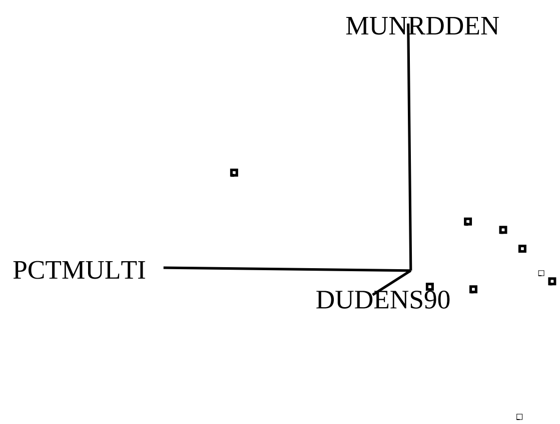
Table 3-3
Cluster 2 - Partial Correlations with MUNRDDEN

DUDENS90	PCTMULTI
.9292	.4936
.9664	controlled
controlled	-.7968

The chart shows that controlling PCTMULTI had little effect on the correlation between DUDENS90 and MUNRDDEN. However, controlling for DUDENS90 not only substantially changed the coefficient of PCTMULTI, it also changed the sign. In other words, with DUDENS90, the relationship between municipal road inventories and the percentage of multi-household dwelling units is positive (as the density of 1990 housing increases, the percentage of multi-household dwelling units increases and the supply of municipal roads increases), yet when PCTMULTI alone is related to MUNRDDEN the sign is negative (when the number of multi-household units increases the municipal road density decreases).

Intuitively, it makes sense for the increased presence of multiple household dwelling units to result in fewer municipal road miles (the sign of PCTMULTI is negative), since the interior circulation roadway, required to provide access to the individual lots of a single household development, is not really required in a multi unit structure. In fact, what is occurring is a type of spurious relationship between PCTMULTI and MUNRDDEN, in the presence of DUDENS90. What the partial correlation is showing is that when the density of residential development is increased, the percentage of multi-household units also is likely to increase. Incidental to this relationship is the fact that when the residential density increases (and along with it the likelihood of increased presence of multi-household dwelling units) the supply of municipal roads also increases. The three dimensional analysis, shown in Diagram 3-4 supports this conclusion. It can be seen that only in one case (Newark) is a relationship between municipal road density and high percentages of multi-household dwelling units substantiated.

Diagram 3-4
Three Dimensional Diagram of Cluster 2 Variables



The result of this analysis is to drop the variable PCTMULTI from the model for Cluster 2 municipalities. The resulting error for the single variable model is shown in Diagram 3-5. As with all regression equations, the errors are smaller at the mean of the predicted values (approximately 2.0 on the X axis) and largest with those values farthest from the mean.

Diagram 3-5
Scatterplot of Predicted Regression Values and the Standard Error of the Prediction

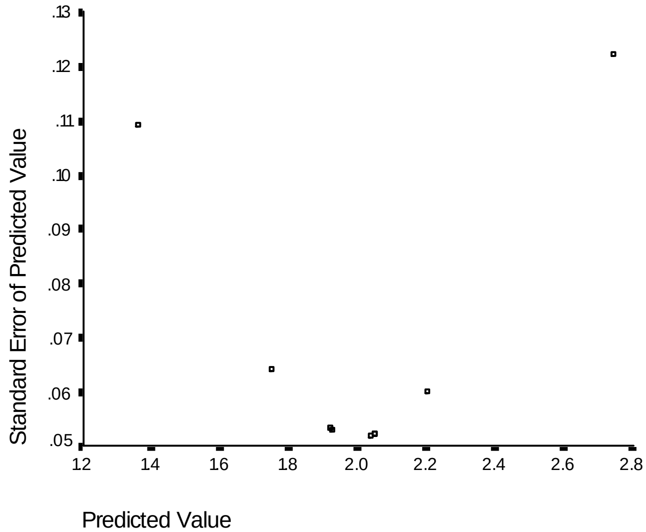
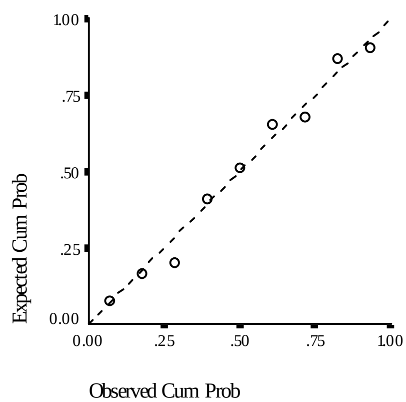


Diagram 3-6 shows the NPP of the final cluster 2 model. The municipal residuals group close to the ideal 45 degree broken line, but do not exactly follow it. The model results might be termed ‘approximately normal’. However, if one assumes that the chart should be exactly interpreted, then S shape of the data (above the ideal value then through the 45 degree line near the mean of the data then below the line) sometimes is indicative that the distribution that is fractal. (The variance is constant with respect to its distance from the mean, except for a scaling factor, which also is a constant.) If this were the case, then re-expressing the equation, would substantially improve the ‘fit’ of the model. However, until further research is conducted, the model appears to produce reasonable results.

Diagram 3-6
NPP of Cluster 2 Model



Cluster 3

Regression analysis for the 29 Cluster 3 municipalities found three variables to be significant: DUDENS70; JOBDIF78 and PCTPRE40. Table 3-4 displays the coefficients that result from Partial Correlations with different controls.

Table 3-4
Cluster 3 - Partial Correlations with MUNRDDEN

DUDENS70	JOBDIF78	PCTPRE40
.9425	-.6270	.7301
.8723	-.1511	controlled
.9231	controlled	.4983
.9164	controlled	controlled

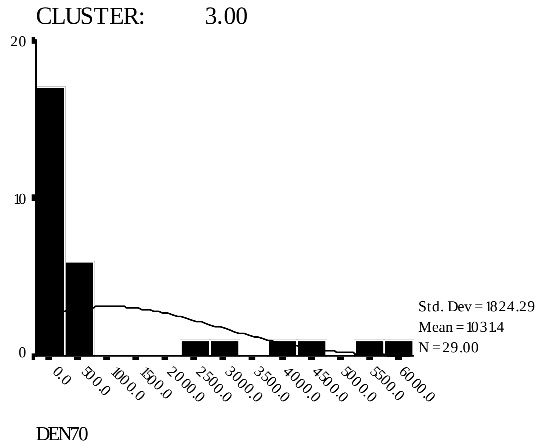
The Table demonstrates that controlling for PCTPRE40 affects the coefficient for JOBDIF78, but has little effect on DUDENS70. This finding supports the theory that the variable PCTPRE40 is helping to refine the effect of the variable JOBDIF78, and may not be directly explanatory of the municipal road supply. In effect, the relationship between these variables suggests that those municipalities which lost employment between 1970 and 1980 also have larger percentages of dwelling units that were built before 1940. The Table also shows that controlling for both JOBDIF78 and PCTPRE40 has very little effect on the regression equation's principal explanatory variable, DUDENS70. Because of these findings, the variable PCTPRE40 was dropped from the model for cluster 3.

Diagram 3-7 shows two views of the 3D scatterplot of the remaining independent variables DUDENS70, JOBDIF78 and dependent variable MUNRDDEN. In the left view the positive linear relationship between MUNRDDEN, on the Y Axis, and DUDENS70, on the X Axis, is shown. The second view is the same diagram rotated to show the weaker, negative relationship between JOBDIF78 and MUNRDDEN.

Diagram 3-7
Rotated 3 Dimensional Analysis of Cluster 3 Variables

It can also be seen in this second view that not all of the municipalities appear to be influenced by the change in employment between 1970 and 1980. This finding suggested that clustering of these municipalities might produce better relationships. Initially, JOBDIF78 was used to create two new subgroups. However, this only produced two new subgroups each of which displayed a mixed relationship with MUNRDDEN. The actual identity (names) of the cluster 3 municipalities then was examined. This cursory analysis suggested that clustering by density might prove useful. A histogram (Diagram 3-8) of Cluster three municipalities arrayed by residential density was prepared and clearly demonstrates the density bifurcation.

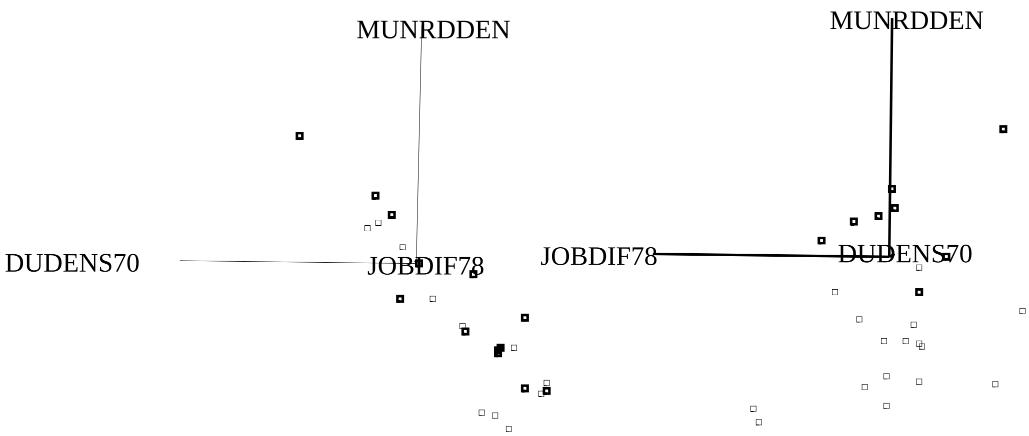
Diagram 3-8
Histograms of Residential Densities for Cluster 3 Municipalities



It was discovered that Cluster 3 divides into two subgroups, one very low density municipalities and the other very high density municipalities. The relationships between variables using three dimensional Scatterplots, is shown in Diagram 3-9. It can be seen that DUDENS70 and (weakly) JOBDIF78 appear to be influential for the set of low density cases. Neither DUDENS70 nor JOBDIF78 appear to exhibit any relationship to the road supply in the high density municipalities, although this finding is seriously compromised by the very small number of cases in the high density subgroup (5). This finding was confirmed in the actual regression analysis, with DUDENS70 and JOBDIF78 resulting in Adjusted R Squares of .86358 in the low density municipal set and no relationship in the high density subgroup.

Diagram 3-9
3D Scatterplot Views of Cluster 3 Relationships

Scatterplot of Cluster 3 Low Density Municipalities



Scatterplot of Cluster 3 High Densities Municipalities

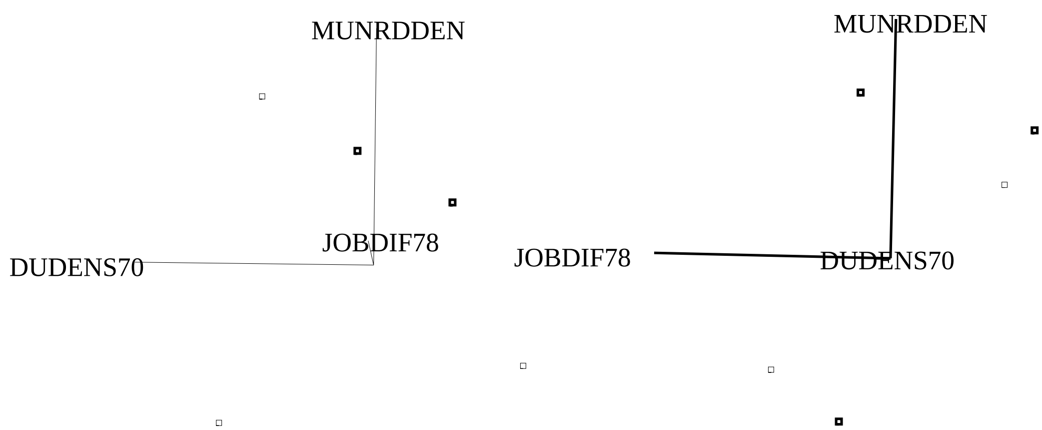


Table 3-5 displays the density and the income characteristics of clusters, for both cluster 3 subgroups. Because of these differences, both the High Density and the Low Density subgroups will be used to present cluster 3 results.

Table 3-5
Cluster 3 Subgroups (classified on municipal density)

Subgroup	Number	Mean (Tukey) Density (dwelling units/mile)	Mean (Tukey) Income (Household 1990)
Cluster 3 High Density	5	4,688 (1970)	$41,527
Cluster 3 Low Density	24	56 (1970)	$56,211

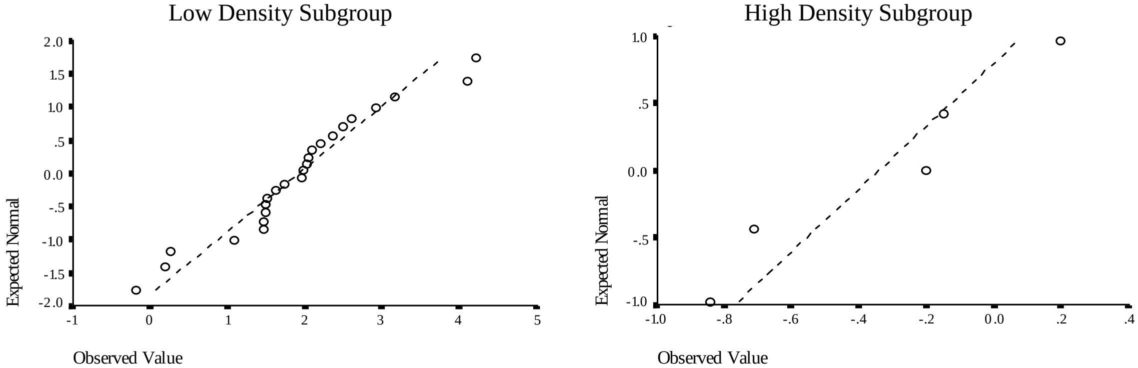
Because the cluster 3 equation uses the variable JOBDIF78, the resistant mean for both subgroups also had to be calculated. Table 3-6 shows the changing values for JOBDIF78 for both subgroups in cluster 3 and Diagram 3-10 demonstrates that the two density-based subgroupings also produce reasonably normal employment change distributions.

Table 3-6
JOBDIF78 Values for the Cluster 3 Subgroups
(Tukey's Mean)

Grouping	JOBDIF78 Value
Entire Cluster 3	5.32
Low Density Subgroup	1.895
High Density Subgroup	-.3398

Diagram 3-10
Normal Probability Plots of JOBDIF78 for Cluster 3 Subgroups

Low Density Subgroup High Density Subgroup



The resulting errors from the two models (entire cluster 3 regression and subgroup one regression) are presented in Diagram 3-11. In the first view to the right of the page, the errors for the regression using the variable DUDENS70 and all cluster 3 cases is presented, while in the view to the left, the regression using DUDENS70 and JOBDIF78 was used for the low density subgroup. The findings suggest that models using either one or two variables, or all cases, or the subgroup 1 (low density) cases would produce similar results.

Diagram 3-11
Comparison of Regression Errors in Cluster 3 Models

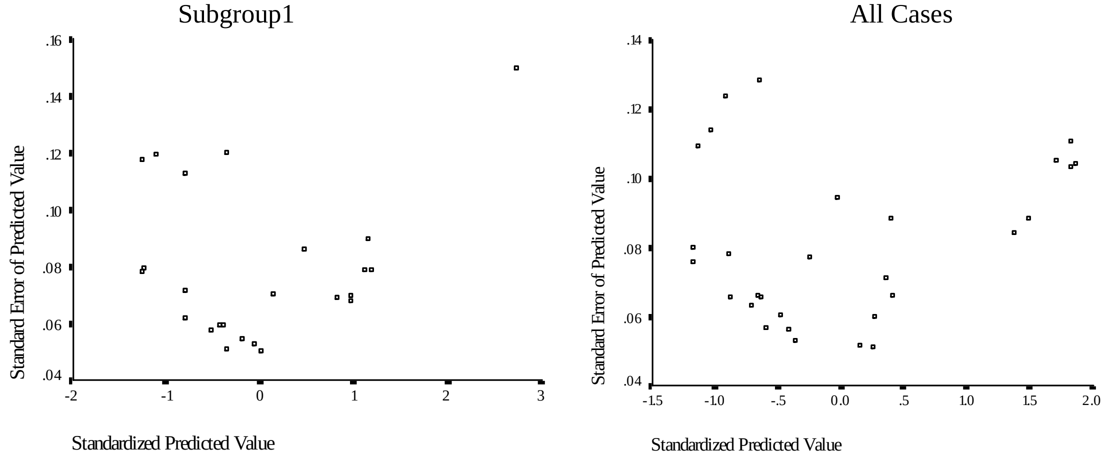
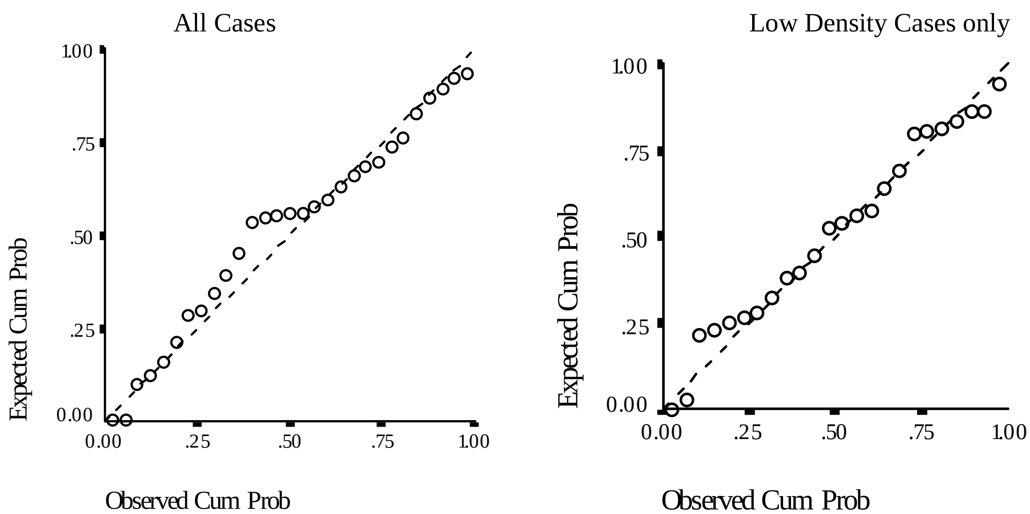


Diagram 3-12 displays the Normal Probability Plots for the regression using all Cluster 3 cases (on the left) and for the regression only using subgroup 1 (low density) cases. As suggested in the error analysis, the models for cluster 3 appear to produce similar results.

Diagram 3-12
NPP of Cluster 3 Models



Cluster 4

Table 3-7 displays the Partial Correlation coefficients for the Cluster 4 variables. Several findings are embedded in this table. First, when the effects of both DUDIF89 and JOBDIF89 are removed, there is little difference in the coefficients of the remaining variables. However, the small improvement in the regression 'fit' that results from the inclusion of these variables supports the strategy of removing these variables from the model equation. Second, controlling for

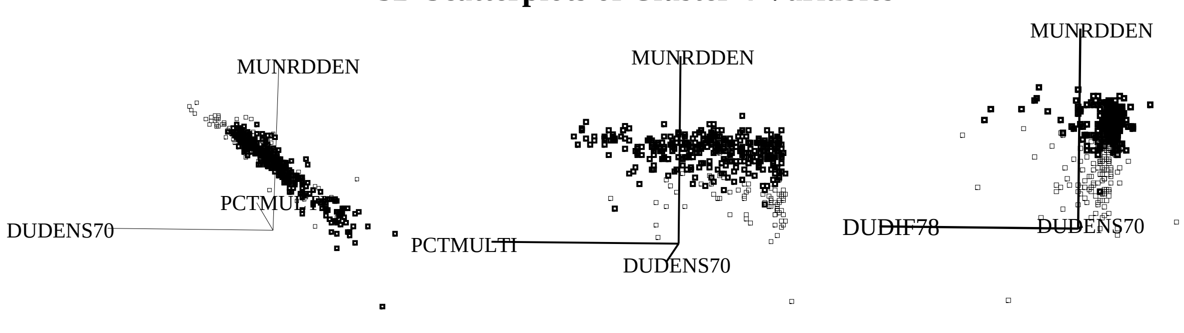
JOBDEN90 significantly reduces the coefficients of the three remaining variables. This suggests that this variable is an important modifier of the remaining variables and that its effect might duplicate the effect of these variables. This finding supports the strategy to remove JOBDEN90 from the model. Third, controlling for the effects of both PCTMULTI and DUDIF78 has little effect, arguing for a more detailed examination of these variables.

Table 3-7
Partial Correlation Results for Cluster 4 Variables

DUDENS70	PCTMULTI	DUDIF78	JOBDEN90	DUDIF89	JOBDIF89
.8999	.3424	-.1537	.6861	-.1775	-.0597
.8983	.3372	-.1185	.6763	controlled	-.0342
.8994	.3366	-.1142	.7122	controlled	controlled
.7854	.0065	-.0332	controlled	controlled	controlled
.7801	.0054	-.0367	controlled	-.0230	controlled
.8979	controlled	-.1528	.6665	controlled	controlled
.9015	.3501	controlled	.7080	controlled	controlled
.9037	controlled	controlled	.6570	controlled	controlled

The relationship between the remaining variables was examined using 3D scatterplots, shown in Diagram 3-13. Three views are presented. In the first view a strong linear relationship between municipal road density and 1970 residential density is shown. In the second view (the middle scatterplot), a more complex relationship is shown between the presence of multi-household dwelling units and municipal road density. What is shown is that as the density of roads increases, the variance of PCTMULTI increases. This suggests that a weighted relationship might be more appropriately used in the equation. The final diagram (right side) shows some municipalities appear to be affected by the change in dwelling units 1970 to 1980 and that many do not appear to be affected.

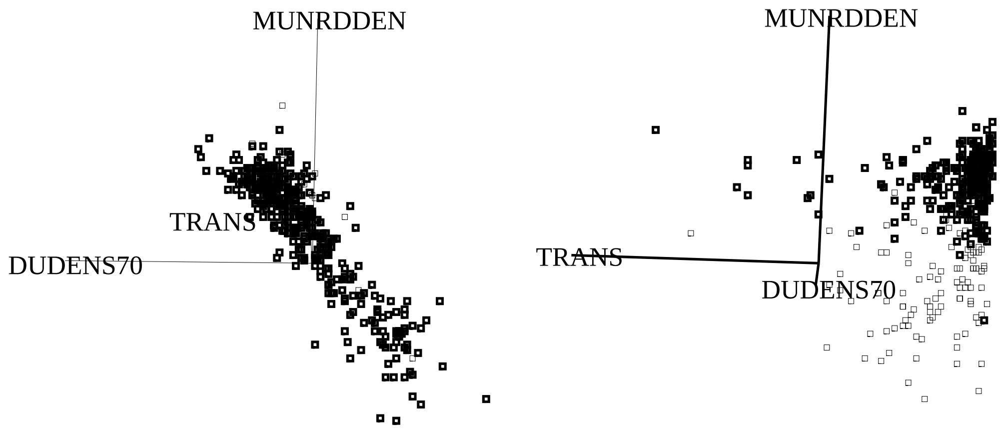
Diagram 3-13
3D Scatterplots of Cluster 4 Variables



The decision to include or exclude the variable PCTMULTI from the regression was greatly influenced by the discovery that the result produced by this variable failed to produce a normal error distribution (this NPP diagram is not shown). To coax the equation towards normality, a new variable TRANS was created from PCTMULTI by using the SPSS procedure

Regression Weight Estimation. The transformed value (TRANS = $PCTMULTI^{-.5}$)produced a more normal distribution (also not shown). Next, a 3D scatterplot, shown in Diagram 3-14, showing the relationship between TRANS, DUDENS70 and MUNRDDEN. As was the case with the variable DUDIF78, a very mixed picture emerges, where some municipalities appear to be affected by TRANS and some are not affected.

Diagram 3-14
3D Scatterplots of Cluster 4 Variables



Creation and Testing of Cluster 4 Subgroups

Because of the mixed relationship between municipal road supply and TRANS (the supply of multiple household dwelling units), the decision was made to use TRANS to create statistically distinct subgroups out of Cluster 4 municipalities. Cluster 4 was divided into 3 subgroups, the characteristics of which are displayed in Table 3-8. As demonstrated in the Table, these areas are very different form one another, with respect to density, income as well as the percentage of multi-household dwelling units that are built in the subgroup. (All values are Tukey's Resistant Mean for the subgroup.) Just as interesting the combination of significant variables for each of these subgroups varied. As shown in Table 3-9, in subgroup 1 only DUDENS70 was explanatory. In subgroup 2 all the variables were used, while in subgroup 3, DUDIF78 was not significant. Diagram 3-15 shows that subgroup 2 exhibits a problem with heteroscedasticity (the mean variance is not proportionate to the Standard Deviation of MUNRDDEN).

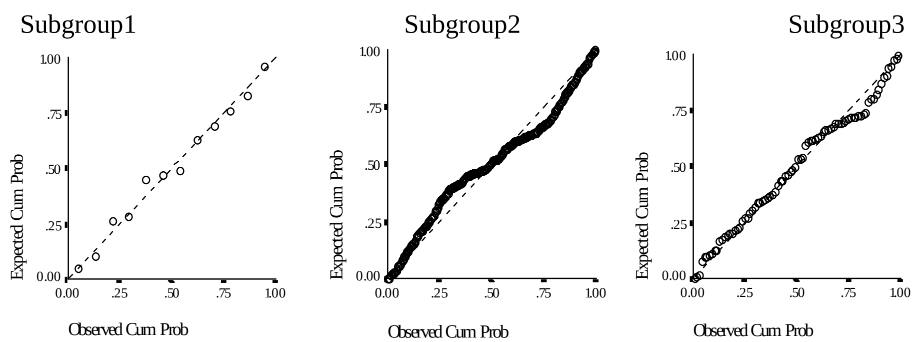
Table 3-8
Cluster 4 Subgroup Characteristics

Subgroup	Number of municipalities	DU's (in 1970) per square mile	PCTMULTI (1990)	Income (1990)
1	12	609	.0107	$68,220
2	293	1085	.3821	$40,019
3	82	165	.0523	$49,553

Table 3-9
Cluster 4 Subgroup Regression Equations

	Regression Equation	Adj. R^2
1	DUDENS70*.538416-1.199373	.69889
2	DUDENS70*.613009+TRANS*.255886+DUDIF78*.40468-2.550177	.82652
3	DUDENS70*.685896+DUDIF78*.559764-.2382176	.90374

Diagram 3-15
Normal Probability Plots of Cluster 4 Subgroups



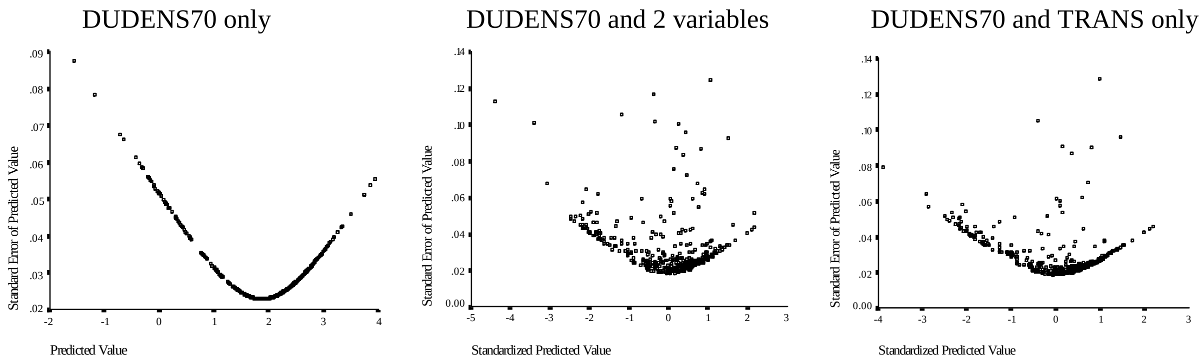
Testing Alternative Equations Using All Cluster 4 Cases

Linear Regressions also were run using *all the cluster 4 cases* and DUDENS70 by itself and in separate combination with both DUDIF78 and TRANS. The resulting equations are shown in Table 3-10. Finally, scatterplots of the errors resulting from these regression equations were plotted and are shown in Diagram 3-16. The center and right diagrams show the effects of including both forms of the second variable in the model equation.

Table 3-10
Cluster 4 Regression Equations (all cases)

cases	Regression Equation	Adj. R^2
all	DUDENS70*.604787-1.951031	.80926
all	DUDENS70*.594214+TRANS*.041966-1.977605	.84898
all	DUDENS70*.613453+TRANS*.049155+DUDIF78*.371994-2.19038	.86020

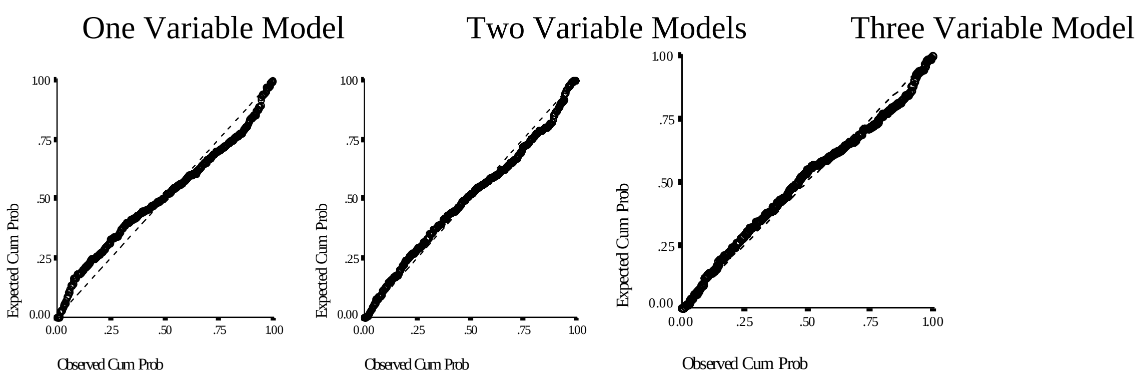
Diagram 3-16
Scatterplots of Errors From Cluster 4 (all cases) Regression Equations



It can be seen that a tradeoff is involved in selecting the final model. Towards the mean of the predicted values, the single variable model appears to produce fewer errors than either of the other models. However, the size of the errors away from the mean is larger in the single variable model and the number of cases clustered near the mean appears to increase with the addition of other explanatory variables. Somewhat off-setting this result is the tendency for some cases near the mean to produce very large errors with the multivariable models.

Diagram 3-17 displays the normal probability plot of the residuals from the Cluster 4 models using all cluster 4 cases. It is evident that the multivariable models produces errors that are more normally distributed than those produced by the one variable model. Interestingly, all of the models exhibit an S shaped error distribution, suggesting that the correct dimension of the equation might be fractal. Exploring this finding, however, is beyond the scope of this research project.

Diagram 3-17
NPP of Cluster 4 Models



The conclusion of this work is that models using DUDIF78 and/or TRANS produce equations that minimize the variance assumption necessary to use regression. Division of Cluster 4 municipalities into 3 subgroups provides insight into the density diversity of the cluster 4 municipalities. It also demonstrates why the variable DUDIF78 produced a mixed relationship with MUNRDDEN. However, it is possible that a single cluster 4 equation might produce results that are equally reliable as results produced by subgroup determined equations.

Cluster 5

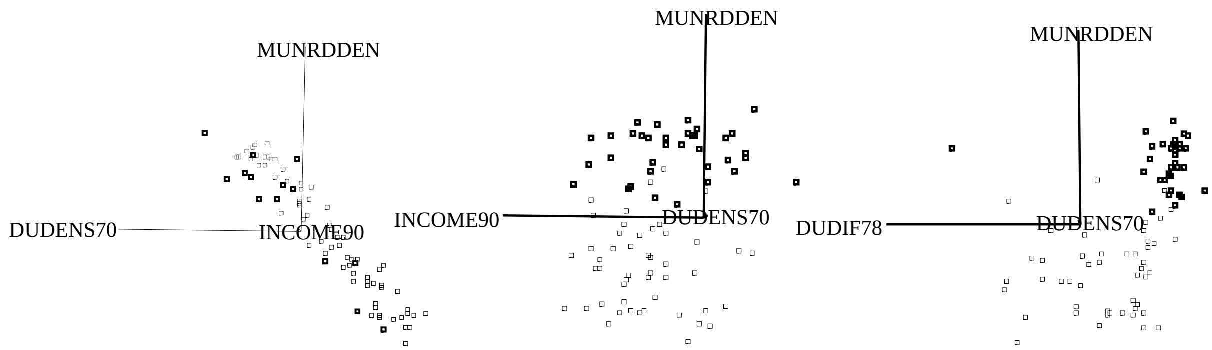
Table 3-11 displays the Partial Correlation results for the cluster 5 regression variables. The table displays two findings. First, controlling for PCTMULTI affects the coefficient for INCOME90. This suggests that PCTMULTI and INCOME90 are negatively related, which means that when the percentage of total dwelling units that are multi-household dwelling units increases, the household income of the municipality tends to decline, and vice versa. This finding allows PCTMULTI to be dropped from the model for cluster 4. No other inter-variable relationships are noted.

Table 3-11
Partial Correlation Results for Cluster 5 Municipalities

DUDENS70	INCOME90	PCTMULTI	DUDIF78
.9506	-.2303	.3012	-.5195
.9356	-.2853	.3431	controlled
.9577	-.0677	controlled	-.5399
.9603	controlled	.2102	-.5400
.9621	controlled	controlled	-.5446
.9550	-.1083	controlled	controlled

The next analysis, shown in Diagram 3-18, uses 3D scatterplots to determine the relationships between the remaining independent variables and MUNRDDEN. The strong positive linear relationship between residential density (1970) and the municipal road density is clearly displayed in the first (left side) diagram in the series. The weaker, less linear but still clearly negative relationship between municipal road density and the change in dwelling units between 1970 and 1980 is shown in the diagram to the right side of this series. (As was the case in the Cluster 4 analysis, it appears that some municipalities are affected by DUDIF78 and some are not affected.) The center illustration shows that no relationship exists between municipal road density and median household income. For this reason, INCOME90 is dropped from the model equation for cluster 5.

Diagram 3-18
3D Scatterplots of Cluster 5 Variables



Creation and Testing of Cluster 5 Subgroups

The mildly negative, but apparently inconsistent relationship between DUDIF78 and MUNRDDEN suggested that subgroups might be created by clustering using DUDIF78. Cluster 5, using DUDIF78, was divided into two subgroups. Table 3-12 displays the characteristics of these subgroups; Table 3-13 shows the regression equations and their correlation adjusted coefficients and Diagram 3-19 shows the resulting NPP's for the models.

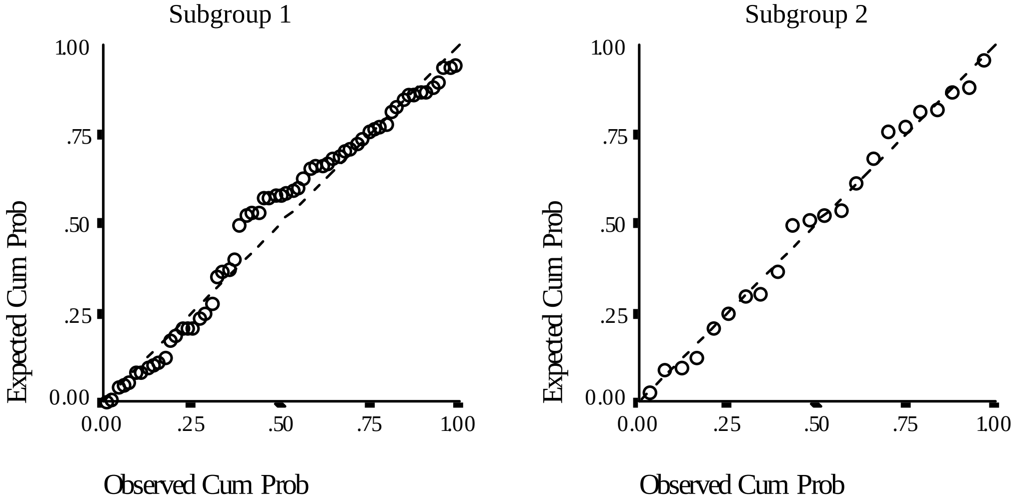
Table 3-12
Cluster 5 Subgroup Characteristics

Subgroup	Number of municipalities	Density 1970 per square mile	DUDIF78	Income (1990)
1	61	630	.1196	$50,869
2	22	127	.5878	$48,296

Table 3-13
Cluster 5 Subgroup Regression Equations

	Regression Equation	**Adj. R^2**
1	DUDENS70*.514035-1.359654	.90433
2	DUDENS70*.598226-1.731513	.82327

Diagram 3-19
Normal Probability Plots of Cluster 5 Subgroups



Testing Alternative Equations Using All Cluster 5 Cases

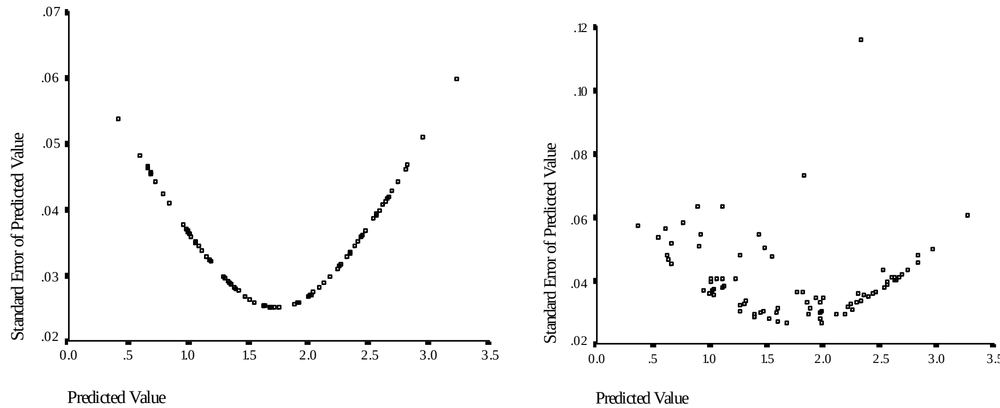
Regressions also were performed using all of the cluster 5 cases. Table 3-14 displays the resulting equations (one equation using DUDENS70 alone and the second using both DUDENS70 and DUDIF78) and their correlation coefficient (adjusted R Square). The left side diagram, in Diagram 3-20, displays the error for the regression equation using MUNRDDEN as the dependent variable and DUDENS70 as the independent variable. The right side shows the error associated with the regression equation using both DUDENS70 and DUDIF78 as the independent variables. In general it can be seen that with the exception of a few municipalities (Brigantine and Ocean City being the notable exceptions), the model using both variables appears to produce smaller errors in most cluster 5 municipalities.

Table 3-14
Cluster 5 Subgroup Regression Equations

cases	Regression Equation	Adj. R^2
all	DUDENS70*.509539-1.309461	.90245
all	DUDENS70*.539699+DUDIF78S*.271657-1.560326	.82327

Diagram 3-20
Scatterplots of Errors From Cluster 5 (all cases) Regression Equations

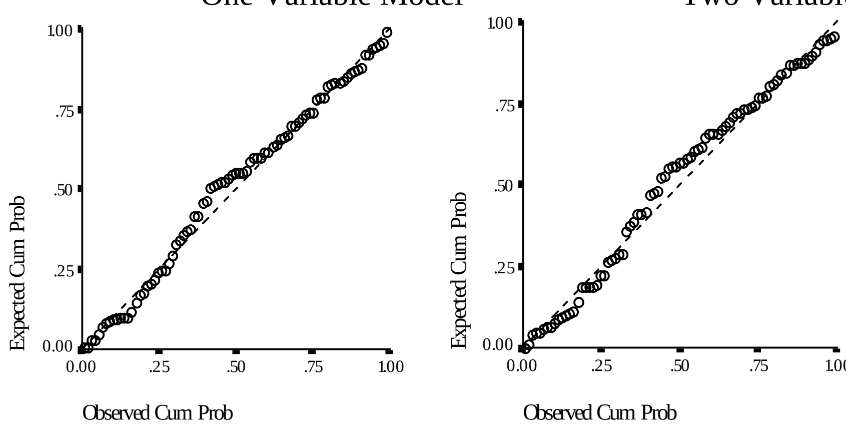
One Variable Model Two Variable Model



The Normal probability plots for the two cluster 5 model are shown in Diagram 3-21. It can be seen that there is little difference between the models, with respect to the distribution of errors. Both equations also have very similar Adjusted R Square values (an adjusted R square of .90245 for the one variable equation and .90677 for the two variable equation). Either of these models could be used to predict municipal road density in cluster 5 municipalities.

Diagram 3-21
NPP of Cluster 5 Models

One Variable Model Two Variable Model



Very similar to the findings for Cluster 4, it appears that the there might be little difference using subgroups or a single equation based on all cases to predict MUNRDDEN.

Findings

1. In the 57 municipalities that comprise cluster 1, the variable PCTPRE40 was dropped

from the model equation to describe municipal road density because it was discovered to be negatively related to the variable INCOME90, and not to the dependent variable, MUNRDDEN. This negative relationship demonstrates that those municipalities with high concentrations of dwelling units constructed before 1940 have lower household incomes than do those municipalities with few PCTPRE40 structures. There are two ways to interpret this finding. First, the data suggests that higher income persons demonstrate a preference to build (or at least live in) new(er) homes. However, this also demonstrates a preference of affluent householders to locate into municipalities with few older houses, even within the municipal set consisting of the relatively high income, rural cluster 1 communities.

2. The variable PCTMULTI was dropped from the model equation to describe municipal road density in cluster 2 municipalities. Analysis of the effect of PCTMULTI found that only in Newark was there a strong relationship between municipal road density and the percentage of multi-household dwelling units.

3. The variable PCTPRE40 was dropped from the model equation to describe municipal road density in cluster 3 municipalities because it was found to relate to the independent variable JOBDIF78 rather than to the dependent variable MUNRDDEN. The negative relationship between employment change between 1970 and 1980 and the presence of PCTPRE40 houses demonstrates that in the most of these lower income, larger municipalities employment declined between 1970 and 1980. This provided the basis for additional clustering based on density. Two subgroups were found, one consisting of very low density municipalities and the second consisting of high density PCTPRE40 municipalities, such as Camden and Jersey City.

4. Several variables were deleted from the model equation to predict municipal road density in cluster 4 municipalities. JOBDIF78 and DUDIF89 were dropped because they added little to the adjusted R square of the equation. JOBDEN90 was deleted since it was related to the other independent variables DUDENS70, PCTMULTI and DUDIF78, and not related strongly to the dependent variable. In other words, municipalities with high residential densities in 1970, high percentages of multi-household dwelling units and losses of dwelling units between 1970 and 1980, tended to be the same municipalities who also had high employment concentrations in 1990, when compared to the other municipalities in cluster 4.

5. PCTMULTI was deleted from the cluster 5 municipal road density equation because it was not strongly related to MUNRDDEN. Instead, PCTMULTI was found to be negatively related to the other independent variable INCOME90. This supports the theory that cluster 5 municipalities with higher concentrations of multi-household dwelling units tended to have lower median household incomes that did municipalities with fewer multi-household dwelling units.

6. It can be seen that cluster 4 and cluster 5, like cluster 3, include a substantial distribution of municipalities, when viewed from the perspective of income and residential density.

7. Although all equations use residential density as the principal explanatory variable, for some clusters and cluster subgroups residential density (DUDENS__) needs to be augmented by an additional variable or variables; most notably DUDIF78. (In cluster 5, the variable DUDIF78

was dropped from the subgroup analysis only after reclassifying the municipalities using DUDIF78.)

IV. Adapting the Equations for Projecting Road Impacts

The initial regression analysis determined which variables were related to the supply of municipal roads. The initial statistical analysis, reported in section 2, defined set of variables that were correlated with municipal road densities. In the preceding section, the number of variables used in these equations was minimized and three of the initial clusters were further subdivided to clarify the relationships of the explanatory variables in the equations. All of this work was performed so that a more comprehensive understanding of the relationships could be developed.

In this section, new regression equations based on this work are produced and tested. The major change to the preceding regression equations is the substitution of DUDENS90 for the time lagged residential density variables. This artificial change is made to facilitate the use of these equations to forecast the supply of municipal roads that would be needed in the future. The idea motivating this new work is that by using 1990 residential densities to predict 1990 municipal road supplies, the forecast year residential densities can be used to predict forecast year municipal road supplies[7]. An important assumption embedded in this approach is that the relationship (the coefficients) between the dependent and the independent variable would be constant, with respect to time.

The second change to the regression equations will be the testing of a new variable GROWRATE[8], which is defined as the percentage of change in the municipality's dwelling units between 1970 and 1990. It is intended that this new variable can substitute for DUDIF78. Again, the motivation for this change is facilitation of the equations for projecting future need for municipal roads. (GROWRATE in forecasting applications would be the percentage of residential change between the base year and the forecast year.) However, GROWRATE has its problems. The most obvious problem is that the variable DUDIF78 is the actual *numeric change*, while GROWRATE is the *percentage of change*.

Regressions Using 1990 Density and GROWRATE

In the remainder of this section, the revised regressions are evaluated for all cluster groups except for cluster 2 (which already used DUDENS90). Several tests are performed on any new equations. First, the assumption that the time lagged residential density and DUDENS90 is examined by using scatterplots (compared to MUNRDDEN). Next, stepwise regression is run to learn if the old (those identified in the preceding section) and new variables are significant. The new R squares are compared to the regression results using time lagged residential density to determine if the equation's predictive abilities have been improved or degraded. Third, Normal Probability Plots for each equation were produced to evaluate the need for weighting or other

[7] This same argument would hold for substituting future density and then lagging them either 10 or 20 years. However, the major complication to this approach is that the Office of State Planning Population and Employment Distribution model does not make linear growth assignments to municipalities. Therefore, to make a municipal road forecast for 2020, the PED would need to forecast municipal densities for both 2020, 2010 (the ten year lag required for cluster 1) and 2000 (the 20 year density lag required for Cluster 3, 4 and 5 municipalities).

[8] In cluster 1 the addition of a new variable related to growth (GROW10, dus90 ÷dus80) also is tested.

transformations. Next, the road predictions produced by the new equations are compared to the resistant mean municipal road length per dwelling unit. In a separate diagram the errors produced by the most promising equations are shown. These last three tests are performed to both demonstrate the characteristics of the equation and to learn if a single equation can be used to predict any subgroups.

Cluster 1 Detailed Analysis

The attempt to introduce GROW10 (dus90 - dus80) discovered that this variable was not significant, perhaps due to the modest increase in residential density. Diagram 4-1 shows scatterplots of residential density for 1980 and 1990 compared to the 1990 supply of municipal roads. While little difference can be seen in the diagrams, the resistant mean changed from 3.7723 in 1980 to 4.0319 in 1990 and the resulting variance changed from .3002 in 1980 to .9538 in 1990. So while the development density only increased by approximately 20%, as some municipalities grew more rapidly than others the variance increased substantially. Despite the large change in the variance, surprisingly the substitution of 1990 residential density for 1980 density resulted in very little loss of predictive ability, as shown in Table 4-1.

Diagram 4-1
Scatterplots of Cluster 1 Residential Densities (1980 and 1990) and Road Densities (1990)

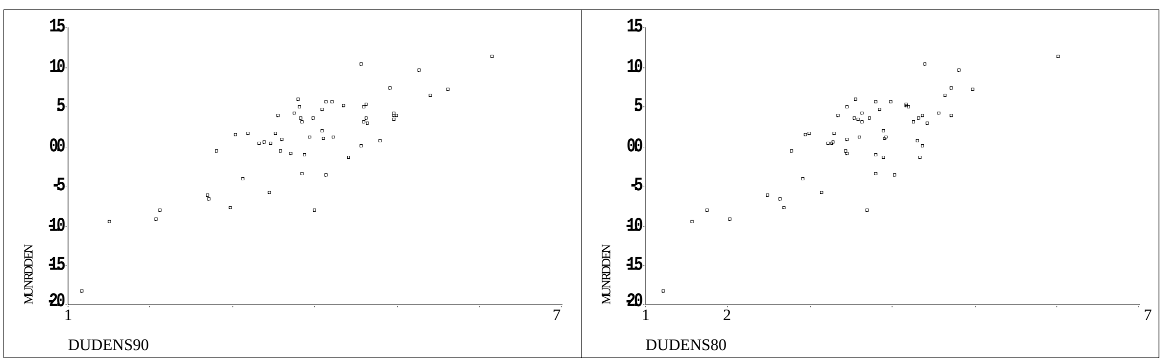
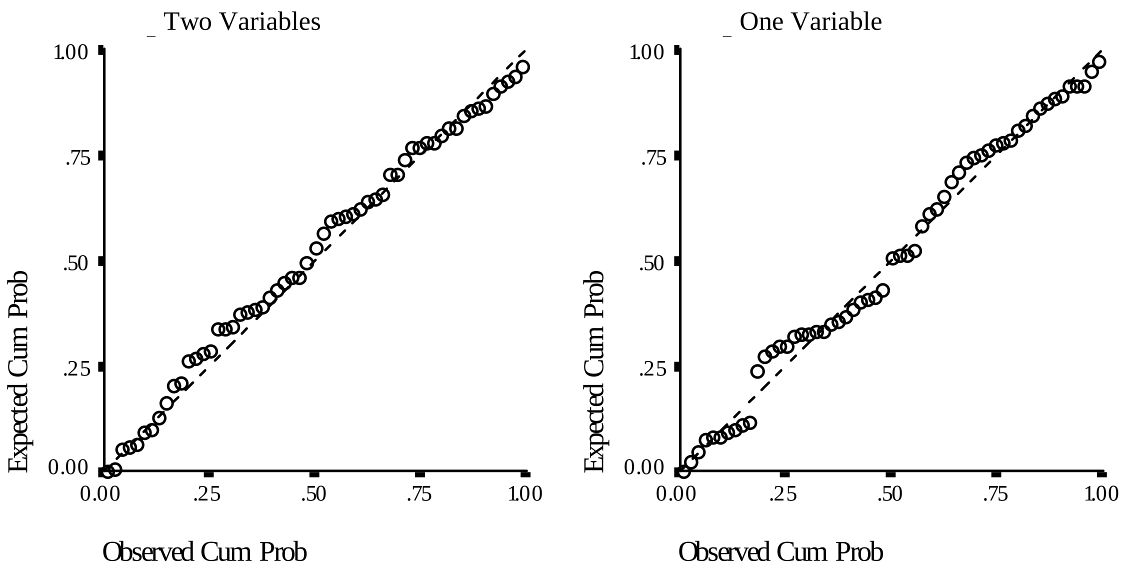


Table 4-1
Regression Results - Cluster 1 Alternatives
(all regressions included DUDENS90)

Alternative	**Adjusted R^2**	**Condition Index**
DUDENS90*.380007+ INCOME90*.654875-8.405293	.73269	96.39
DUDENS90*.459592-1.70599	.63352	8.44
DUDENS80 with LNHHI	.76444	94.24
DUDENS80	.63484	8.73

Diagram 4-2 shows the Normal Probability Plots for the Cluster 1 equations; the diagram on the right is the result of the equation only using DUDENS90 as the dependent variable, while the results diagrammed on the left were produced by an equation that included DUDENS90 and INCOME90. The diagrams demonstrate that both equations are reasonably normal.

Diagram 4-2
NPP of Cluster 1 Alternative DUDENS90 Equations



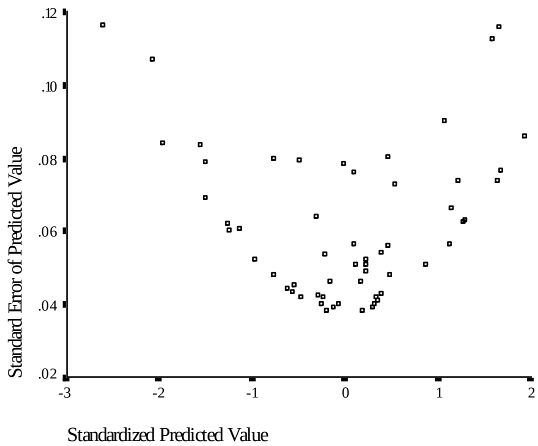
Two tests then were performed on the equations. First, both equations (DUDENS90 with and without INCOME90) were used to predict municipal road per residential unit, using resistant mean values for DUDENS90, INCOME90 and MUNRDDEN. The results of these tests are shown in Table 4-2. It can be seen that both the statistical models consistently underestimate the actual resistant mean. Diagram 4-3 shows the errors that would result from the use of the equation with DUDENS90 by itself or in combination with INCOME90.

Table 4-2
Comparison of Model Predictions to Actual Data

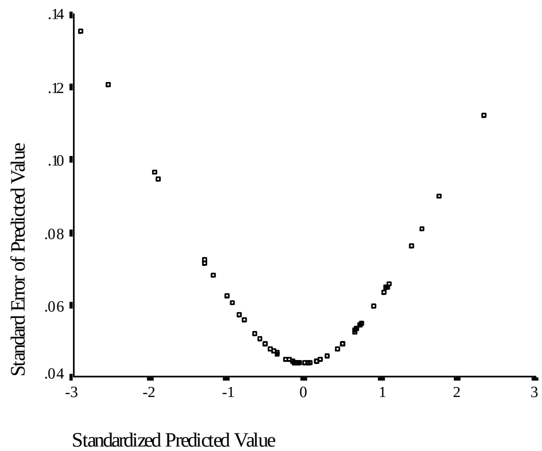
	Linear Feet of Municipal Road per Dwelling Unit
Actual Resistant Mean	118
DUDENS90	108
DUDENS90 and INCOME90	106

Diagram 4-3
Scatterplot of Equation Errors and Predicted Values

DUDENS90 and INCOME90



DUDENS90 only



The result of this analysis is somewhat unclear. GROW10 was demonstrated not to be significant. The variables that had been significant with the time lagged residential density also proved significant using 1990 residential density. A more complicated issue involves the tradeoff between using an equation with less correlation (a lower R square) compared to a more predictive model which exhibits a high degree of Collinearity between INCOME90 *and the constant*[9].

Cluster 3 Detailed Analysis

The regression equation, defined in the preceding section, used JOBDIF78 as an independent variable. Table 4-3 shows the correlations between JOBDIF78, DUDIF78 and GROWRATE for cluster 3 municipalities. Not surprisingly, it can be seen that GROWRATE is not a substitute for JOBDIF78. This finding suggests two things. First, regressions using both GROWRATE (if significant) and JOBDIF78 should be examined. Second, the substitution of a new forecasting variable (like, GROWEMP) might be used for projection purposes.

Table 4-3
Bivariant Correlation of Cluster 3 Independent Variables

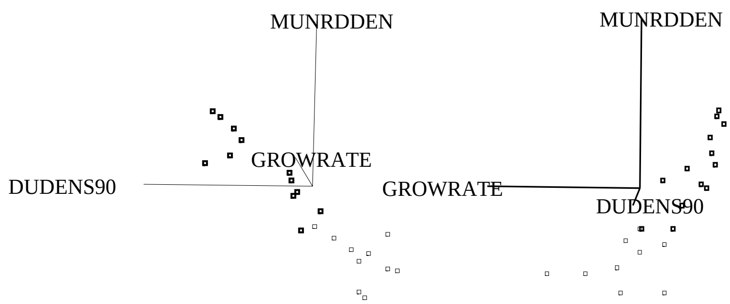
	GROWRATE	DUDIF78	JOBDIF78
GROWRATE	1.0000	.9465	.4700
DUDIF78	.9465	1.0000	.5239
JOBDIF78	.4700	.5239	1.0000

[9]Collinearity between independent variables clearly is a violation of GM assumptions, but this is not so clear when the interrelationship is between an independent variable and the constant.

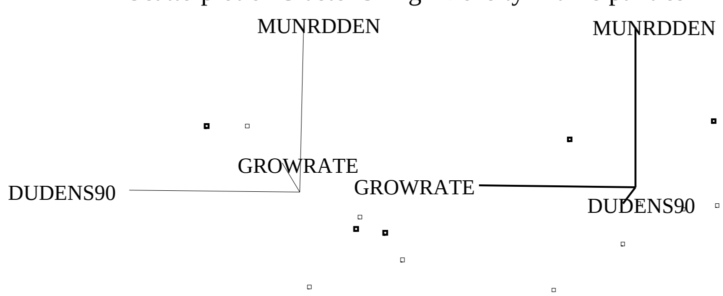
Because Cluster 3 divides into two subgroups, one of very low density municipalities and the other of very high density municipalities, the relationships between DUDENS90 and GROWRATE was first examined using three dimensional scatterplots, shown in Diagram 4-4. It can be seen that DUDENS90 and (weakly) GROWRATE appear to be influential for the set of all cluster 3 cases and for the set of low density cases. Neither DUDENS90 nor INCOME90 appear to exhibit any relationship to the road supply in the 6 high density municipalities. This finding was confirmed in the actual regression analysis shown in Table 4-4. These findings support an approach that continues regressions for both the set of all cluster 3 cases and the set of cluster 3 subgroup 1 (low density) municipalities.

Diagram 4-4
3D Scatterplot Views of Cluster 3 Relationships

Scatterplot of Cluster 3 Low Density Municipalities



Scatterplot of Cluster 3 High Density Municipalities



Scatterplot of All Cluster 3 Municipalities

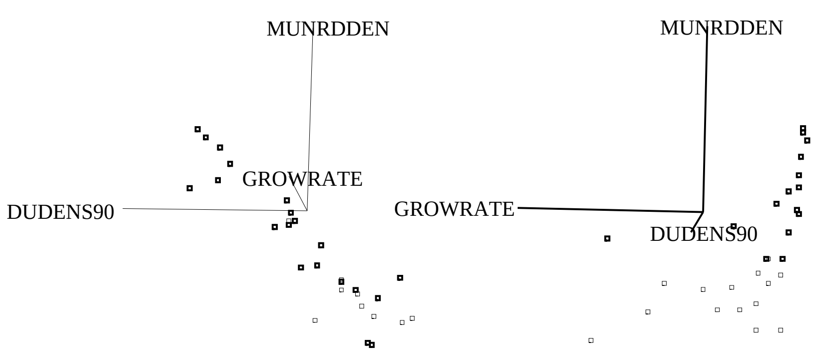


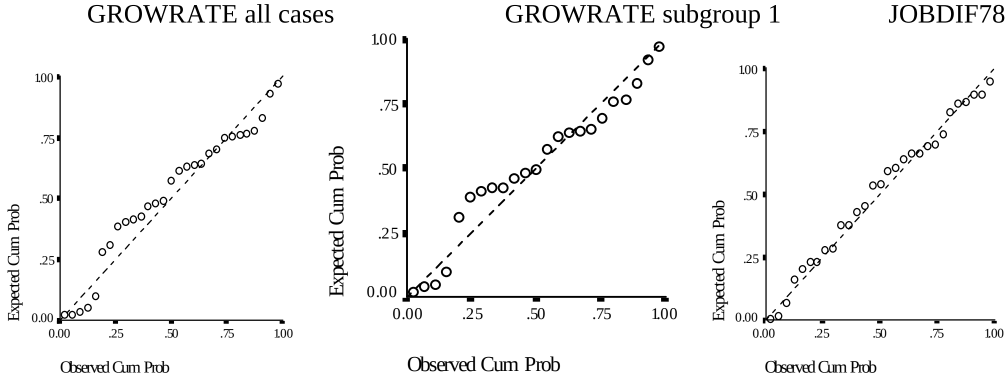
Table 4-4
Regression Results - Cluster 3 Alternatives

Cases	Alternative	Adjusted R^2
all	DUDENS90*.473801+GROWRATE*-.133787-1.182371	.87949
all	DUDENS90*.470957+JOBDIF78*-.158358-1.232526	.88616
1 low density	DUDENS90*.492777+GROWRATE*-.13234-1.286696	.80021
2 high density	no relationship	

Diagram 4-5 displays the Normal probability plots for the two equations using all cluster 3 cases and for the subgroups 1 (low density) equation. The diagrams suggest that all of the variances are ‘approximately’ normal. However, it can also be seen that the equation using employment change (JOBDIF78) appears to be more closely fitted to the ideal Gaussian distribution than are the equations using GROWRATE.

Diagram 4-5
Normal Probability Plot of Cluster 3 Equations

GROWRATE all cases | GROWRATE subgroup 1 | JOBDIF78 all cases



The predictions produced by the preceding models are compared to the resistant mean length (per dwelling unit) in Table 4-5. The surprising, but nice, result appears to be that using either of the equations based on all cases produces predictions that are acceptable for the subgroups. However, it appears that the equation using JOBDIF78 produces better estimations for the low density municipalities. It is possible that this finding suggests that employment changes in these low density places appear to be related to reduced supplies of municipal roadway (the sign of JOBDIF78 is negative).

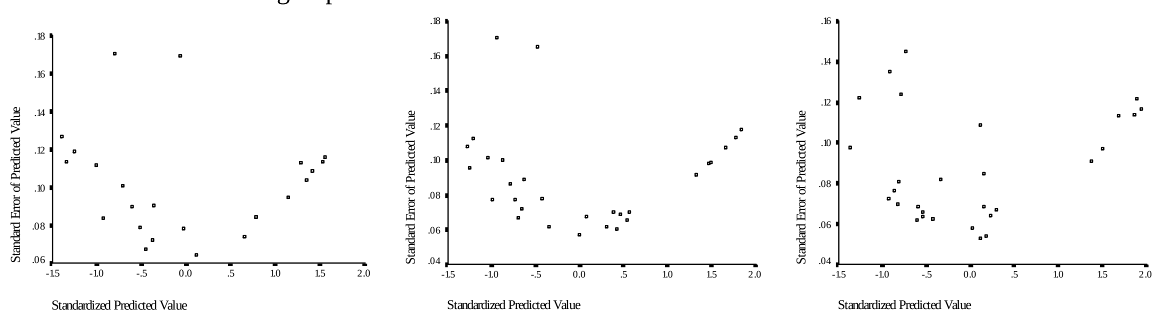
Table 4-5
Comparison of Model Predictions to Actual Data

Cases	Variables (all include DUDENS90)	Predicted Municipal Road per D.U. (feet)	Resistant Mean Municipal Road per D.U. (feet)
subgroup 1	GROWRATE	71	65
subgroup 2	no regression possible	n/a	18
all	GROWRATE	61	60
all	JOBDIF78	57	60
subgroup 1	GROWRATE all cases	71	65
subgroup 1	JOBDIF78 all cases	65	65
subgroup 2	GROWRATE all cases	19	18
subgroup 2	JOBDIF78 all cases	19	18

Diagram 4-6 presents the final analysis of cluster 3 equations. In these diagrams the standard errors for the predicted mean response are compared to the standardized predicted mean values. Given that the size of the error (mean variance) for normally distributed values increase proportionately with the distance from the mean, the shape of the errors should be U or V shaped with the smallest errors at the 0 value on the X axis. It can be seen that the equation developed on all cases and using GROWRATE produces a smoother error curve than does the equation utilizing JOBDIF78.

Diagram 4-6
Scatterplots of Equation Errors

GROWRATE subgroup 1 GROWRATE all cases JOBDIF78 all cases

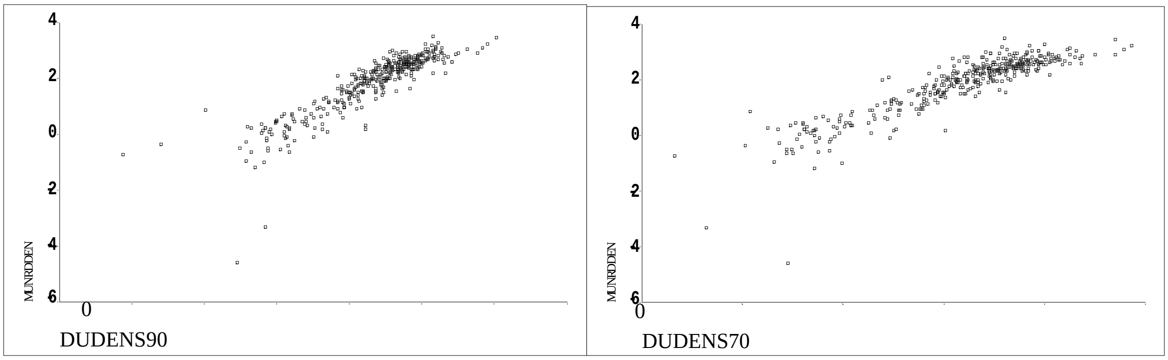


From the preceding analysis it can be seen that the use of the regression equation developed with all cluster 3 cases and using DUDENS90 and GROWRATE can be used to predict municipal road supplies.

Cluster 4 Detailed Analysis

Diagram 4-7 is a scatterplot showing the relationship between DUDENS70 and MUNRDDEN (1990) on the right and DUDENS90 and MUNRDDEN (1990) on the left. It can be seen that there is little difference between the plots. This finding suggests the substituting 1990 residential density for 1970 residential density should not cause major problems.

Diagram 4-7
Scatterplot of Residential Densities (1990 and 1970) and Municipal Road Density



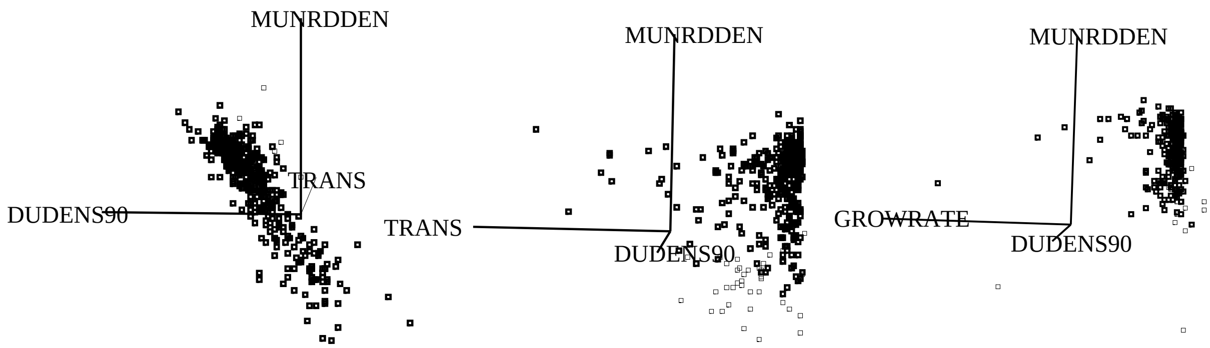
The next analysis estimates the degree to which substitution of DUDIF78, found to be a significant variable in the preceding section, by GROWRATE would cause changes in the regression equation. Table 4-6 contains the Pearson correlation coefficients resulting from this comparison. It can be seen the GROWRATE is similar to DUDIF78, but not the same. The probable cause for the difference is definitional (GROWRATE represents the percent of change while DUDIF78 is the numeric difference).

Table 4-6
Cluster 4 - Correlation Between GROWRATE and DUDIF78

	DUDIF78	GROWRATE
DUDIF78	1.0000	.7838
GROWRATE	.7838	1.0000

The scatterplot of the significant variables, shown in Diagram 4-8, shows three findings. First, the strong relationship between residential density and municipal road density is clearly displayed in the diagram of the left. In the center diagram an inconsistent relationship between the number of multi-household dwelling units and the supply of municipal roads is evident. This finding supports the continued use of TRANS to further subdivide this Cluster. Finally, the diagram on the right shows that there is very little relationship between GROWRATE and the supply of municipal roads. Because of this poor relationship, additional correlation studies were conducted. It was found that both TRANS and GROWRATE were negatively correlated to MUNRDDEN, although as displayed in the scatterplot the relationship between GROWRATE is weaker than that associated with TRANS (GROWRATE = -.2092, and TRANS = -.2342). Partial correlations were also run with the expected finding that the variables were independent of each other.

Diagram 4-8
Three Dimension Diagram of Cluster 4 Significant Variables



Because of the preceding findings, equations including both GROWRATE and TRANS were developed. These models were tested (stepwise regression) for the all cluster 4 municipalities and for the three subgroups. As expected the significance of TRANS is diminished in the subgroups, since they were based on statistical differences in TRANS (a transformed value of the percentage of multi-household dwelling units in a municipality). The following Table displays the resulting model equations and shows their adjusted correlation coefficient.

Table 4-7
Regression Results - Cluster 4 Alternatives

Cases	**Alternative**	**Adjusted** R^2
all	DUDENS90*.625104+GROWRATE*-.094558+TRANS*.050695-2.251854	.85906
all	DUDENS90*.599702+GROWRATE*-.101858-1.937598	.84977
all	DUDENS90*.628828+TRANS*.05505-2.431674	.84868
subgroup 1	DUDENS90*.569983-1.487483	.69307
subgroup 2	DUDENS90*.624666+GROWRATE*-.082576+TRANS*.251677-2.613705	.82818
subgroup 3	DUDENS90*.69998-2.579185	.89376

Diagrams 4-9 and 4-10 show the resulting normal probability plots for each of the cluster 4 (including subgroup) equations. All of the plots exhibit some small degree of heteroscedasticity, despite the development of TRANS as a weighted variable designed to reduce this problem. (In fact it can be seen that equations which include TRANS exhibit a slightly more linear and slightly more normal distribution). Efforts were made to weight TRANS and even DUDENS90, but no improvement was produced. (In fact, DUDENS90 was determined not to need weighting.) It is possible that the distribution ‘error’ is cause by the lack of another variable.

Diagram 4-9
Cluster 4 (all cases) - NPP of Alternative Equations

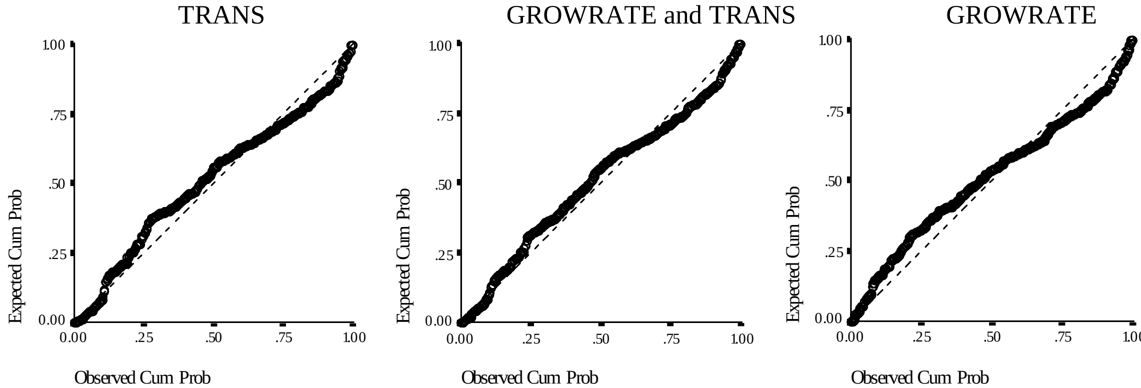
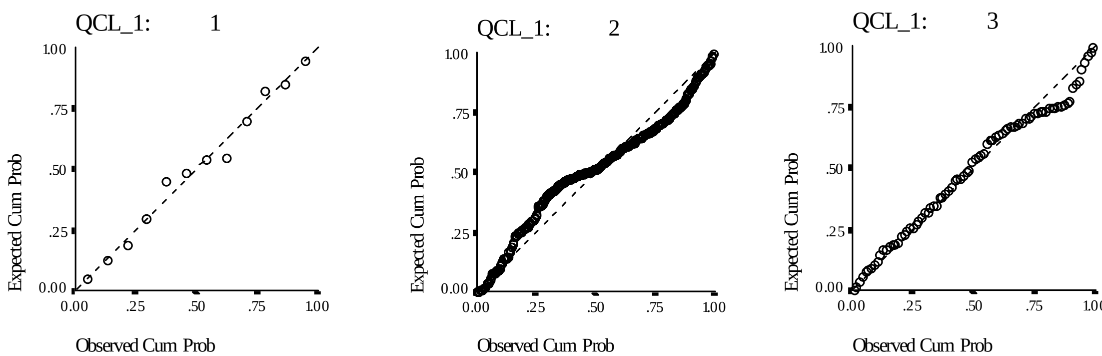


Diagram 4-10
Cluster 4 Subgroups - NPP of Alternative Equations



The second test of the Cluster 4 equations was to use them to predict the amount of municipal road per dwelling unit and to compare the prediction to the actual resistant mean for the cluster or it subgroup. The result of this analysis is shown in the Table 4-8. It appears that the model developed with all cluster 4 cases and including both TRANS and GROWRATE might have done a slightly better job of predicting municipal roads.

Table 4-8
Cluster 4 - Comparison of Model Predictions to Actual Data

Cases	Variables	Predicted Municipal Road per D.U. (feet)	Resistant Mean Municipal Road per D.U. (feet)
all	DUDENS90+GROWRATE+TRANS (all cases)	38	44
all	DUDENS90+GROWRATE (all cases)	40	44
all	DUDENS90+TRANS (all cases)	37	44
subgroup 1	DUDENS90	102	78
subgroup 2	DUDENS90+GROWRATE+TRANS	34	38
subgroup 3	DUDENS90	77	76
subgroup 1	DUDENS90+GROWRATE+TRANS (all cases)	71	78
subgroup 1	DUDENS90+GROWRATE (all cases)	50	78
subgroup 1	DUDENS90+TRANS (all cases)	72	78
subgroup 2	DUDENS90+GROWRATE+TRANS (all cases)	35	38
subgroup 2	DUDENS90+GROWRATE (all cases)	37	38
subgroup 2	DUDENS90+TRANS (all cases)	34	38
subgroup 3	DUDENS90+GROWRATE+TRANS (all cases)	79	76
subgroup 3	DUDENS90+GROWRATE (all cases)	74	76
subgroup 3	DUDENS90+TRANS (all cases)	78	76

This analysis of errors confirms that the model using both TRANS and GROWRATE appears to have a small advantage over the other models. Although the Y axis scales differ in the series, it appears that the equation based on all cases and using both variables reduces the errors in the prediction.

Diagram 4-11
Cluster 4 (all cases) Scatterplots of Equation Errors

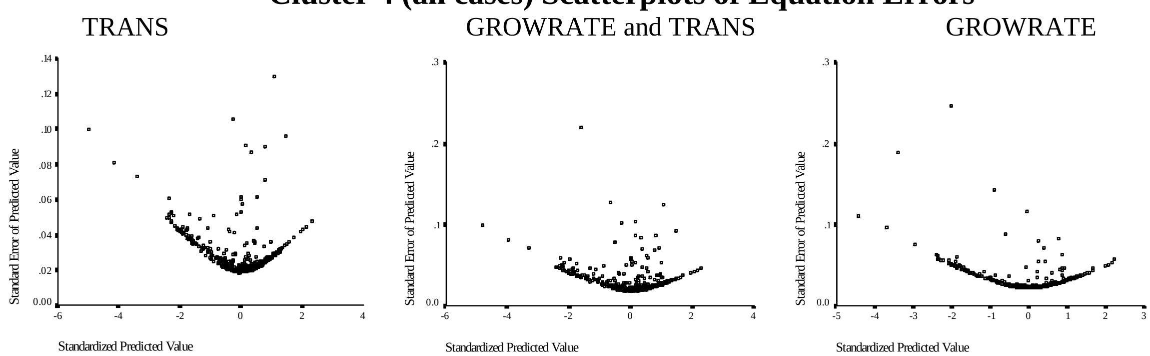
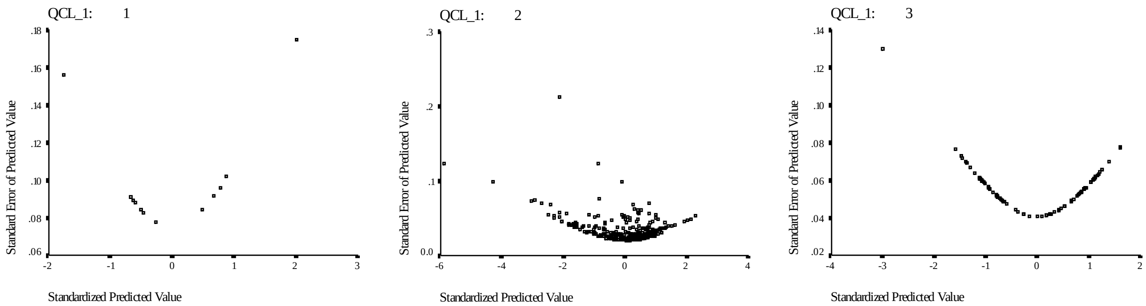


Diagram 4-12
Cluster 4 Subgroups - Scatterplots of Equation Errors



The preceding analysis demonstrates that the 1990 municipal road supply can be adequately predicted for cluster 4 municipalities by using several equations. The best model utilizes both TRANS and GROWRATE and is based on regression using all cluster 4 municipalities. This finding suggests that something about recent growth (1970 to 1990) and the provision of multi-household dwelling units negatively affects the supply of municipal roads. In clusters 1 and 3, neither of these variables appear to be influential.

Cluster 5 Detailed Analysis

Diagram 4-13 and Table 4-9 demonstrate that the substitution of DUDENS90 for DUDENS70 should have little effect and that the new variable GROWRATE is highly correlated with DUDIF78.

Diagram 4-13
Scatterplot of DUDENS70 and MUNRDDEN and DUDENS90 and MUNRDDEN

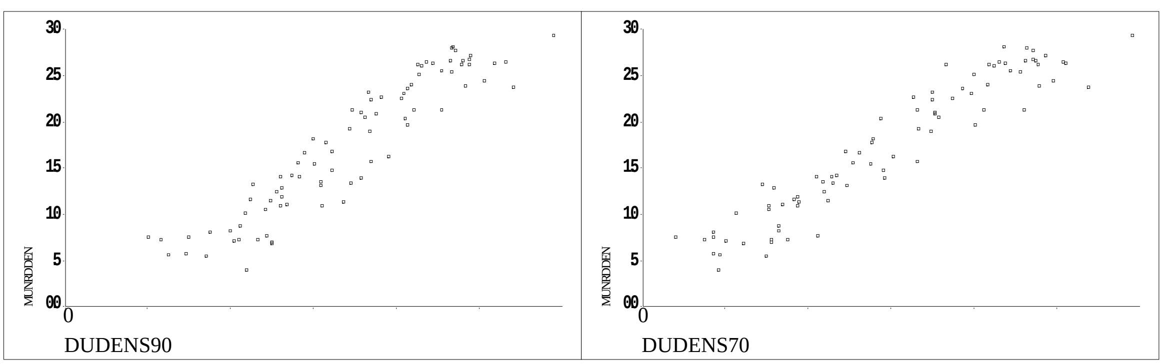


Table 4-9
Cluster 5 Correlations Between GROWRATE and DUDIF78

	GROWRATE	DUDIF78
GROWRATE	1.0000	.9057
DUDIF78	.9057	1.0000

The first scatterplots (all cases) included in Diagram 4-14 demonstrate that the strategy to subdivide cluster 5 based on GROWRATE is sound and very similar to the subdivision that was performed using DUDIF78. This fact was confirmed by comparing group membership of the municipalities based on clustering using the two different variable. Out of a total of 83 municipalities only five changed membership. The diagram also suggests that the relationship between MUNRDDEN and GROWRATE is weaker in the second subgroup.

Diagram 4-14
Cluster 5 Three Dimension Diagrams of Significant Variables

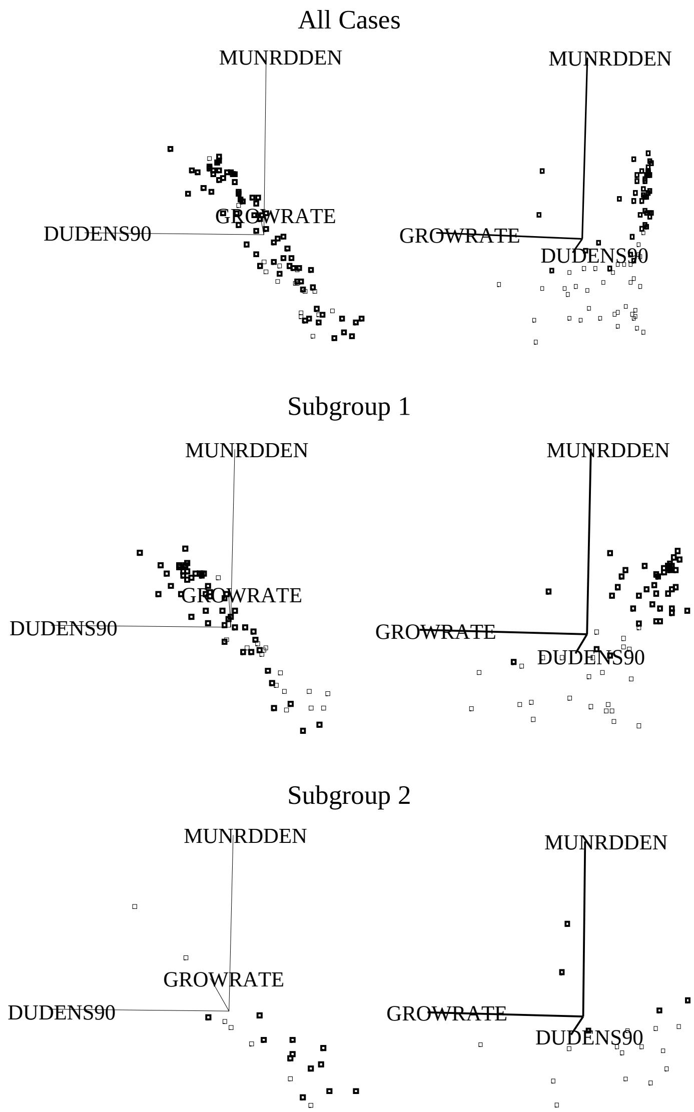


Table 4-10 shows the resulting regression equations for cluster 5 and its subgroups. As expected, the effect of GROWRATE is weaker in subgroup 2 as evidenced by the lower value of its coefficient. Perhaps because of this fact, the equation is less successful, only producing an adjusted R Square of .83145.

Table 4-10
Regression Results - Cluster 5 Alternatives

Cases	**Alternative**	**Adjusted R^2**
all	DUDENS90*.551298+GROWRATE*-.200179-1.462126	.90520
subgroup 1	DUDENS90*.523728+GROWRATE*-.332808-1.112702	.89582
subgroup 2	DUDENS90*.632801+GROWRATE*-.285143-1.673335	.83145

The normal probability plots, the error plotting and the comparison of predictions to actual data, confirm that the model developed using all cluster 5 municipalities is a more versatile and reliable model.

Diagram 4-15
Cluster 5 Normal Probability Plots

All Cases　　Subgroup 1 Cases　　Subgroup 2 Cases

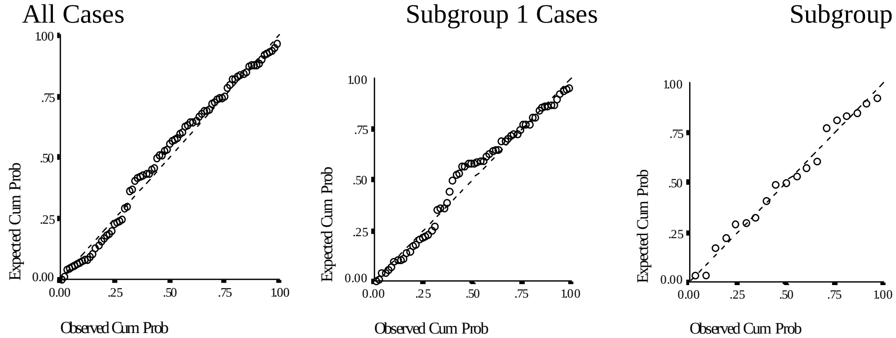
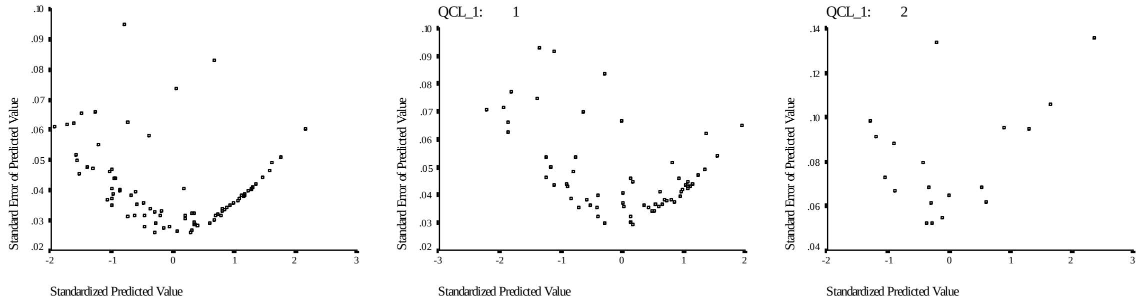


Table 4-11
Cluster 5 Comparison of Model Predictions and Actual Data

Cases	**Variables**	**Predicted Municipal Road per D.U.**	**Resistant Mean Municipal Road per D.U.**
all	DUDENS90+GROWRATE	52	49
subgroup 1	DUDENS90+GROWRATE	48	48
subgroup 2	DUDENS90+GROWRATE	54	53
subgroup 1	DUDENS90+GROWRATE (all cases)	49	48
subgroup 2	DUDENS90+GROWRATE (all cases)	53	53

Diagram 4-16
Cluster 5 - Scatterplots of Equation Errors



Findings

From the preceding analysis, the model equations, using 1990 data, shown in Table 4-12 have been demonstrated to produce reasonable estimates of 1990 municipal road supplies. Although specific subgroup determined equations might produce slightly better forecasts, the recommended equations are more robust since they were produced using a broader data set.

Table 4-12
Recommended Regression Equations for Modeling Future Roads

	Model Equation	**Adjusted R^2**
Cluster 1	DUDENS90*.380007+LNHHI*.654875-8.405293	.73269
Cluster 2	DUDENS90*.509907-1.5515163	.84382
Cluster 3 (all)	DUDENS90*.473801+GROWRATE*-.133787-1.182371	.87949
Cluster 4 (all)	DUDENS90*.625104+TRANS*.05695+GROWRATE*-.094558-2.251854	.85906
Cluster 5 (all)	DUDENS90*.551298+GROWRATE*-.200179-1.462126	.90520

However, it should be noted that using these 1990 density equations would result in a slight over-estimation of the forecasted Municipal roads. The preceding models were an attempt to produce a dynamic forecast based on cross sectional data. In less technical language, the independent variables DUDENS__ and GROWRATE represent data series over a period of 20 to 30 years. However, the dependent variable is 1990 data. The time lagging of residential density demonstrates that as the residential density changed over time, the correlation with municipal road density slightly diminished. This shows that the relationship between residential density and municipal road density changes over time (the direction of the change is, that over time, fewer road miles are required). Use of the 1990 based equations freezes the relationship to their 1990 correspondence not allowing for any added decline to occur. Without having some sort of a municipal road time series, it is impossible to study or model this changing relationship.

V. Policy and Design Implications

The purpose of the preceding research was to improve OSP's capability to project infrastructure costs associated with alternative growth management policies and growth scenarios. For example, the equations developed in chapter 4 would be used in the Office of State Planning Fiscal Impact Model to predict the supply of municipal roads required by a particular Trend or Plan growth scenario. This prediction could then be compared to the model's prediction of road miles (cost) required by another growth scenario and the cost difference could be used to an element to evaluate the two scenarios.

However, during the course of this study it became evident that the statistical findings themselves had policy and design implications. This section describes the analysis used to identify these implications and proposes several land use planning and transportation recommendations which are supported by the statistical research reported in sections three and four of this report. Also in this section, inferential analysis is used to deduce policy and design recommendations.

Statistical Analysis and Findings

Municipal Road Density Inversely Related to Residential Density

In Table 5-1, the resulting significant v ariables for each cluster and cluster subgroups (from chapter 4) are shown with their time lagged residential density and their 1990 residential density. The clusters and cluster subgroups are arrayed by their time lagged densities. This table shows the diversity of residential density grouping for clusters 3, 4 and 5 that were discussed in the preceding two sections.

Table 5-1
Regression Equations Using 1990 Density

Cluster - Subgroup	Number	Mean (Tukey) Density (dwelling units/sq. mile) Time Lagged	1990	Change	Significant Variables
Cluster 1	57	42 (1980)	56	14	DUDENS90+INCOME90
Cluster 3 - 1	24	56 (1970)	239	183	DUDENS90 -growrate
Cluster 5 - 2	22	127 (1970)	299	102	DUDENS90 -growrate
Cluster 4 - 3	82	165 (1970)	240	75	DUDENS90
Cluster 4 - 1	12	609 (1970)	670	61	DUDENS90
Cluster 5 - 1	61	630 (1970)	782	152	DUDENS90 -growrate
Cluster 2	9	942 (1990)	979	37	DUDENS90
Cluster 4 - 2	293	1085 (1970)	1448	363	DUDENS90 -growrate+trans
Cluster 3 - 2	5	4,688 (1970)	4344	-344	none

By reorganizing these municipal grouping by their 1990 residential densities, as shown in Table 5-2, five 1990 density categories can be generalized. These density groupings will be used in the remainder of this section.

Table 5-2
Clusters and Cluster Subgroups Organized by 1990 Residential Density

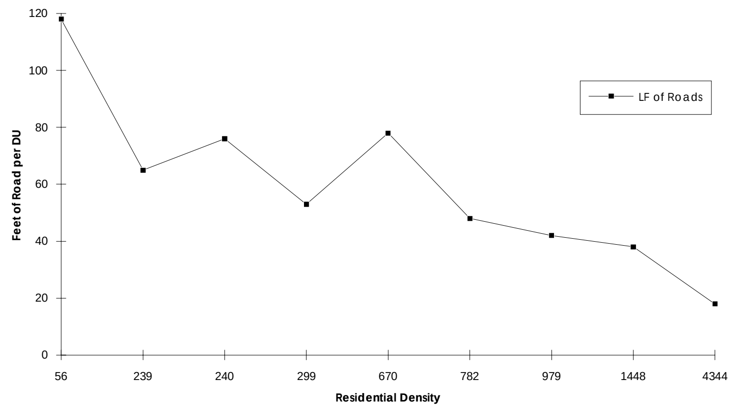
Density Categories	Cluster - Subgroup	#	1990 Density	Significant Variables
Lowest Density	1	57	56	DUDENS90+INCOME90
Low Density	3 - 1	24	239	DUDENS90 - growrate
	4 - 3	82	240	DUDENS90
	5 - 2	22	299	DUDENS90 - growrate
Medium Density	4 - 1	12	670	DUDENS90
	5 - 1	61	782	DUDENS90 - growrate
High Density	2	9	979	DUDENS90
	4 - 2	293	1448	DUDENS90 - growrate + trans
Highest Density	3 - 2	5	4344	none

In Table 5-3, the density categories introduced in Table 5-2 are continued. The new information titled “Feet of Roads per DU” was produced by dividing the resistant mean linear feet of municipal road miles per square mile (in the cluster-subgroup) by the 1990 (residential) density (also per square mile). This produced a resistant mean estimate of the number of linear feet of municipal road for each dwelling unit found in the cluster subgroup. Gottlieb (1990) speculated that the amount of municipal road constructed for each dwelling unit was inversely related to density. Table 5-3 and Diagram 5-1 demonstrates that this speculation was (generally) correct .

Table 5-3
Municipal Roads as a Function of Residential Density

Density Group	Cluster - Subgroup	Density 1990	Feet of Roads per DU	Density Change (DU/Sq mile)	Significant Variables
Lowest Density	1	56	118	14	DUDENS90+INCOME90
Low Density	3 - 1	239	65	183	DUDENS90 - growrate
	4 - 3	240	76	75	DUDENS90
	5 - 2	299	53	102	DUDENS90 - growrate
Medium Density	4 - 1	670	78	61	DUDENS90
	5 - 1	782	48	152	DUDENS90 - growrate
High Density	2	979	42	37	DUDENS90
	4 - 2	1448	38	363	DUDENS90 - growrate +trans
Highest Density	3 - 2	4344	18	-344	none

Diagram 5-1
Municipal Roads per Dwelling Unit as a Function of Residential Density



This is an interesting finding for two reasons. First this finding demonstrates that the State Development and Redevelopment Plan's recommendation to channel growth into higher density centers would produce infrastructure cost savings. Second, this finding demonstrates that the need for more (or less) municipal roads is primarily a result of local zoning [10]. If the only implication of large lot zoning was the need for more municipally supplied and paid for roads, then the only policy implication (and an important one at that!) would be that municipal capital and maintenance savings would accrue if higher residential densities were zoned. However, lower densities, with their longer access (municipal) roads, also mean higher municipal costs to police the roads, longer (more costly) utility service connections (cable TV, telephone, water, sewer, gas), higher school costs associated with longer bus routes, and longer work trips with more air quality discharges and potentially more water quality impacts that affect the region and the State as a whole[11].

While the analyses presented in this paper were conducted at a municipal scale, it is likely that the relationship between residential density and municipal roads would hold at the sub-municipal scale. Therefore, it is reasonable to propose that the equations presented in this paper might be used as a method to calculate the infrastructure costs (e.g., municipal roads, sewer collectors) associated with residential development. This methodology might be used to demonstrate the "rational nexus" required to support the imposition of exactions (on a major development) required by New Jersey law.

However, a more appropriate use of the statistical relationship between municipal residential density and municipal roads, would be to establish a system where municipal infrastructure-related costs would be allocated in accordance with a municipal density standard. Under such a system, a publicly beneficial density might be determined by a jurisdiction as its standard. For example, OSP research has shown that a density of 6 dwelling units to the acre

[10] The histogram, diagram 2-3, suggests that Gottlieb's speculation that places with very low densities have high marginal costs associated with the initial provision of roadways.
[11] Although not documented in this paper, longer county roads also are correlated with lower density development.

could provide a mix of housing that can be affordable and largely subsidy free for a cross section of New Jersey householders (Reilly 1993). Other standards might also be developed. Under such as system, the statistical equation could be used to calculate an impact fee, which would be used to to redistribute public costs that result from inter-municipal density preferences. Persons wishing to build less dense housing would be allowed to proceed, but would be assessed an impact fee (which might be an annual charge) to reimburse for the added VMT's, roadway and other infrastructure impacts of their desired lower density (compared to the impact which would result from the municipal density standard). Persons developing at a density higher than the municipal standard would receive a tax break (a negative impact fee) to insure that they were only paying for the infrastructure impacts associated with their density preference.

Inferential Analysis and Finding

Road Design Affects the Amount of Roadway Needed for Access

While it is generally true that residential density and municipal road density are related, residential density (DUDENS__) alone is not sufficient to explain the supply of municipal roads. If this were the case then LNDUS__D by itself, or in combination with characteristics of the home owners (variables such as income INCOME90) or housing type (such as MULTIDUS or TRANS) would suffice. Instead, in the analysis using the time lagged residential densities, the numeric change in residential housing units (DUDIF78) frequently is significant and this is especially true in clusters where DUDENS70 was used as the principal explanatory variable. In the analysis using 1990 densities, the variable GROWRATE frequently is significant and, like DUDIF78 (used in chapter 3), its sign was always negative. It is also known from the analysis in the preceding section that DUDIF78 and GROWRATE, while not identical, are closely related to each other. These facts lead to the hypothesis that something about the nature of the recent development may be affecting the supply of municipal roads.

By looking at Table 5-3, it can be seen that the variable GROWRATE was significant for municipal groupings that substantially increased their residential density. This is an interesting finding, since one might expect that the rate of growth and not the quantity of growth would be the important factor. And yet the data demonstrates that the rate of growth probably is a misleading indicator. In cluster 1, the rate of growth was approximately 133%, while in cluster 5-1 the rate of growth was 124%. However, in cluster 1, which increased its residential density by 14 dwelling units per square mile, GROWRATE was not significant. In cluster 5-1, which increased its residential density by 152 dwelling units per square mile, GROWRATE was significant. (An even more striking comparison is found by looking at the medium density subgroups.) This finding suggests that the effect causing the reduction in municipal roads per dwelling unit becomes significant only in municipalities that have experienced a substantial increase in the number of new dwelling units (residential density increases by at least 100 dwelling units per square mile). This, in turn, suggests that the cause of the municipal road effect might be less associated with the development of individual lots into houses, than in the development of larger residential subdivisions.

Initially, this report (in section 2) speculated that the reason residential density was time lagged was that municipal roads were being built *after* houses were constructed. In effect, municipal governments were playing infrastructure catch up with residential growth. However, the findings associated with the subgroups of cluster 3, 4 and 5 offer a powerful argument that this observation may not be correct. If rapid growth taxed the municipal capacity to ensure an appropriate supply of municipal roads, then one might expect that this phenomenon would primarily occur in lower density municipal grouping. (The idea behind this assumption is that municipalities with fewer taxpayers would have a more difficult time paying for public improvements caused by sudden substantial growth.) However, the data suggests that GROWRATE is significant regardless of the base year taxpaying population (i.e. it is significant in some of the subgroups in low, medium and high density categories). The second argument against the idea that municipalities are playing infrastructure catch up with local roads is the fact that very few municipal roads actually are built by municipalities. Most of these roadways are constructed by developers and then deeded for public use. Because of these arguments, it is proposed that the time lagged variable (DUDIF78) should be understood not to represent a deficiency of roadway construction, but to indicate a *real difference* in the amount of municipal roads that is being constructed. In fact, the data suggest that this phenomenon is relatively recent (since 1970 in most places and since 1980 in cluster 1).

The question then becomes, what change, or changes, related to the development of a substantial number of dwelling units, perhaps occurring in tract housing development and not found in scattered lot development, might be causing a reduction in the amount of municipal roads being developed. Four possible explanations seem plausible: development is occurring along existing municipal roads; the new development is more dense; the new development has smaller frontage requirements; and, the pattern of the newer roads, built as part of these new developments, somehow differs from the road pattern built previously and somehow is providing access to homes by allowing for the construction of shorter roadways.

The first hypothesis is that new development must be occurring along existing municipal (or county or State) roads. If this were true, then the residential density would increase while the total length of roads would remain constant, resulting in fewer feet of municipal roadway per dwelling unit. There are two logical problems with this idea. First, if this infrastructure advantage is available, why is it only true in municipalities where the density has increased by a substantial amount? In other words, one might expect that if such access ready lots exist, then some homeowners in every municipality would chose to develop on these sites; thereby eliminating any effect when comparing communities. A second argument against this hypothesis is that the coefficient of GROWRATE should decline in higher density municipalities, as the supply of access-ready building lots is diminished by past development. Development along existing roads does occur (as evidenced from aerial photographs). So one might expect that in lower density clusters and subclusters the effect of GROWRATE, measured by the numeric value of the variable's coefficient, would increase. (More homes proportionately would be developed along existing roads since the supply of access-ready lots would be unexpended, therefore the effect of GROWRATE would increase.) However, when the coefficients of subgroup 1, cluster 5, with its density of 782 (1990) is compared to the coefficient of subgroup 2, cluster 5, with its density of 299, the reverse is true. The coefficient for subgroup 1 is -.332808, while the coefficient for subgroup 2 is -.285143, which is less of a dampening effect on residential density. These arguments suggest that the hypothesis, that the reduced need for municipal roads is the result of development on lots fronting existing roadways, does not appear to be supported.

The second hypothesis is that existing residential development in these rapidly suburbanizing municipalities might be less dense than found in those residential portions of the municipality developed since 1970. Since this newer development is denser, then the new development would need less roadway according to the statistical relationship evident in this report between residential density and municipal road density. If this hypothesis were true then municipal density would increase, but the amount of roadway, on average, would decrease. The argument against this idea is while it works in theory, it does not present a very realistic picture of what is happening in practice. The fact is that in most New Jersey municipalities, the suburbanizing areas typically are zoned for lower densities than those densities found in the older established towns or hamlets. Therefore it is unlikely that the residential density in the rapidly developing areas is higher than the development that preceded this recent development [12].

The third possible explanation is related to design standards promulgated for these rapidly developing areas. Even if the zoning demands large lots, if the zoning requires relatively narrow frontages, shorter roads would be needed to provide access than would have been the case with lots with larger frontage requirements. Again this works in theory, but it is doubtful that it holds in practice. Most municipal zoning codes which require larger lot sizes also either are silent on lot geometries or require substantial frontages, that probably have not diminished since 1970. (Again, local zoning might result in slightly longer work trips and associated environmental impacts.)

The final hypothesis is that the design of the road system serving the newer residential areas of a municipality might be different from the design of the road pattern serving the older residential areas or the less intensely developed new residential areas. To test this concept, the ratio of the resistant mean length of municipal road per dwelling unit divided by the estimated residential lot frontage was prepared for each cluster and cluster subgroup. (For example, one might intuitively expect that for each foot of residential lot frontage, there would be one foot of municipal roadway; a ratio of 1.) If there is a real difference in the design of the road system, differences in these ratios would be evident. It would be expected that the ratio for subgroups using GROWRATE would be lower than would the ratio of subgroups not using GROWRATE.

- Estimating Municipal Road to Residential Frontage Ratios

To begin the process, it must be remembered that the result of the regression equation (MUNRDDEN) is in the form:

$$\text{MUNRDDEN} = \log_e \frac{municipal.roads}{municipal.area}$$

[12] Multiple household structures are separately accounted by the use of the variable PCTMULTI or TRANS.

This product is then converted into total miles of road in the municipality by raising the result to the power of *e*, and then dividing by the mean area of the cluster (mean values will be shown using the standard notation of having a bar drawn above the value). Finally, this result, which is the number of linear miles of municipal road per square mile of municipal area, is multiplied by the number of feet in a mile (5,280), thereby producing the number of feet of municipal road in each square mile of municipal area. Given that the regression equation produces the result MUNRDDEN, the following equation summarizes the process of converting the regression forecast into linear feet of municipal road in each square mile of municipal area. The result of this calculation for each cluster and subgroup is shown in Table 5-4.

Linear Feet of Municipal Road per square mile of municipal area = $\frac{LNMUNID^{e}}{\overline{cluster.area}} \times 5280$

Table 5-4
Resistant Mean Linear Feet of Municipal Road per Dwelling Unit

Density Category	Cluster - Subgroup	#	LF of Roads per DU	Significant Variables
Lowest Density	1	57	118	DUDENS90+INCOME90
Low Density	3 - 1	24	65	DUDENS90 - growrate
	4 - 3	82	76	DUDENS90
	5 - 2	22	53	DUDENS90 - growrate
Medium Density	4 - 1	12	78	DUDENS90
	5 - 1	61	48	DUDENS90 - growrate
High Density	2	9	42	DUDENS90
	4 - 2	293	38	DUDENS90 - growrate + trans
Highest Density	3 - 2	5	18	none

The next step in the evaluation process estimates both the density of residential development and the frontage of the residential lots. Gross residential density would simply be the total number of dwelling units divided by the area of the municipality. However, very few municipalities are entirely devoted to residential development. In the more rural areas, farming may be the dominant land use. In most areas, there are also various types of businesses, schools and parks. Therefore, to properly estimate residential density, one has to estimate the percentage of total municipal area that is developed for residential use. The following equation produces estimated residential density per acre. Table 5-5 displays the assumptions about the percentage of municipal area devoted to residential development and the resulting residential densities for each of the density categories.

Residential Density = $\frac{\overline{dwelling.units}}{\overline{area} \times 640 \times per.res}$

where:

per.res = percentage of municipal area developed residential

Table 5-5
Estimated Residential Densities for Density Categories
(Dwelling units per acre)

Percent of Municipal Area Developed Residential in 1990.

	20%	30%	40%	50%	60%	70%	80%	90%
Lowest Density	.21	.32						
Low Density	.43	.28	.21					
Medium Density				2.5	2.0	1.8	1.5	1.4
High Density			3.6	2.9	2.4	2.0	1.8	1.6
Highest Density			17.9	14.3	11.9	10.2		

Residential density is then converted into frontage (feet along the access road[13]) by estimating lot dimensions. For example, if the lot were a typical 1:3 ratio, that means that for every 3 feet of depth, the lot has one foot of frontage. The following formula then was used to calculate the residential frontage for each of the densities assumed in each of the policy groups. (Note: the sum of the lot ratios is the sum of the lot dimension. A 1:5 lot would have a sum.of.lot.ratio of 6.) The resulting frontage estimates are shown in Table 5-6.

$$\text{Residential Frontage} = \sqrt{\frac{Lot.size}{sum.of.lot.ratios}}$$

Table 5-6
Estimated Residential Frontage by Density Categories
(Feet of Frontage)

Percent of Municipal Area Developed Residential

	20%	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**
Lowest Density								
1:2 lot	257	315						
1:1 lot	364	446						
Low Density								
1:2 lot	223	273	315					
1:1 lot	315	386	446					
Medium Density								
1:3 lot				76	83	90	96	102
1:2 lot				93	102	110	118	125
High Density								
1:5 lot			48	54	59	64	68	73
1:4 lot			54	60	67	71	77	82
Highest Density								
1:5 lot			22	24	27	29		
1:4 lot			24	27	30	32		

[13]It is assumed that the narrow part of the lot fronts the street.

The final step in the analysis of the road to frontage ratio is to divide the mean resistant feet of municipal roads per dwelling unit by the estimated frontage for each density group. The result of this calculation is shown in Table 5-7.

Table 5-7
Ratios of Municipal Road per Dwelling Unit to the Estimated Frontage

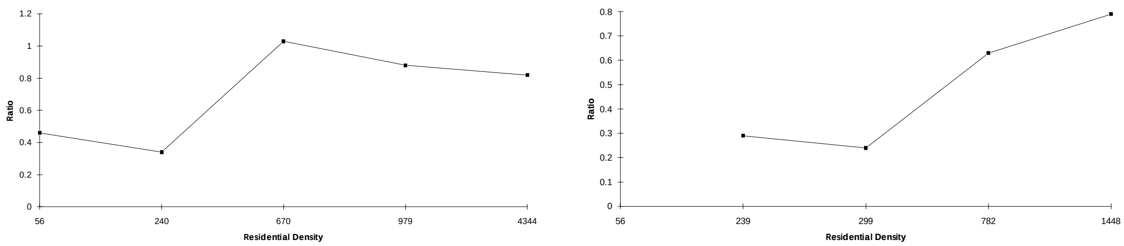
	Cluster - Subgroup	**Frontage Min.**	**LF of Roads**	**Ratio**	**Significant Variables**
Lowest Density	1	257	118	.46	DUDENS90+INCOME90
Low Density	3 - 1	223	65	.29	DUDENS90 - growrate
	4 - 3	223	76	.34	DUDENS90
	5 - 2	223	53	.24	DUDENS90 - growrate
Medium Density	4 - 1	76	78	1.03	DUDENS90
	5 - 1	76	48	.63	DUDENS90 - growrate
High Density	2	48	42	.88	DUDENS90
	4 - 2	48	38	.79	DUDENS90 - growrate + trans
Highest Density	3 - 2	22	18	.82	none

The road to frontage analysis suggests that by controlling for density and lot geometries, there is a real difference between the amount of municipal road needed to serve municipalities that recently have grown by an amount sufficient to increase their density by at least 100 dwelling units per square mile and those municipalities that have grown more slowly. These rapidly growing (in numbers, not percentages) municipalities appear to be able to provide access using less municipal roadway, than is needed to accommodate growth or existing populations in municipalities that had fewer new homes built. This finding supports the hypothesis that there has been a change in the design of these roads over the past 20 (since 1970) years.

The frontage to municipal road length per dwelling unit ratios also demonstrate a remarkable difference in the provision of municipal roads, whether one compares clusters of subgroups that include GROWRATE or those that exclude GROWRATE. This comparison is shown in Diagram 5-2. In both charts, as the residential density approaches 700 dwelling units per square mile, the ratio sharply increases. For example, at the low densities of approximately 200 to 300 dwelling units per square mile, the ratios range from .34 for subgroup 4-1 (which does not use GROWRATE) to .29 and .24 for subgroups 3-1 and 5-2, both of which use GROWRATE. At 670 dwelling units per square mile, the ratio for subgroup 4-1 rises to 1.03 and then declines gently to .88 at the density of 979 dwelling units per square mile. GROWRATE dependent ratios are .63 at density 782 and .79 at 1448 DU/square mile. At the highest density, the ratio declines somewhat. Also at the lowest density, the ratio increases slightly.

Diagram 5-2
Relationship Between Ratio and Residential Density

Grouping without GROWRATE Groupings with GROWRATE



Special Design Provisions Should be Established for Redevelopment of Older Urban Areas

The high densities (10 to 18 du/area) and the narrow lot geometries (1:5 and 1:4) now found in the older urban municipalities do not provide sufficient roadways to accommodate on-street parking (the resistant mean frontage was 18 feet, as shown in Table 5-4; a distance very close to the length of an automobile). This is exacerbated by the fact that over 80% of these existing subgroup 3-2 municipal housing units accommodate more than one household. The limited amount of parking space might be desirable if the intent is to limit car ownership and presumably car usage in favor of transit, bicycle and pedestrian modes of travel. The lot sizes and densities found in these places may not be well suited to auto reliant communities unless off-street parking is provided.

A Municipal Road Network Typology May Explain Municipal Roadway Differences

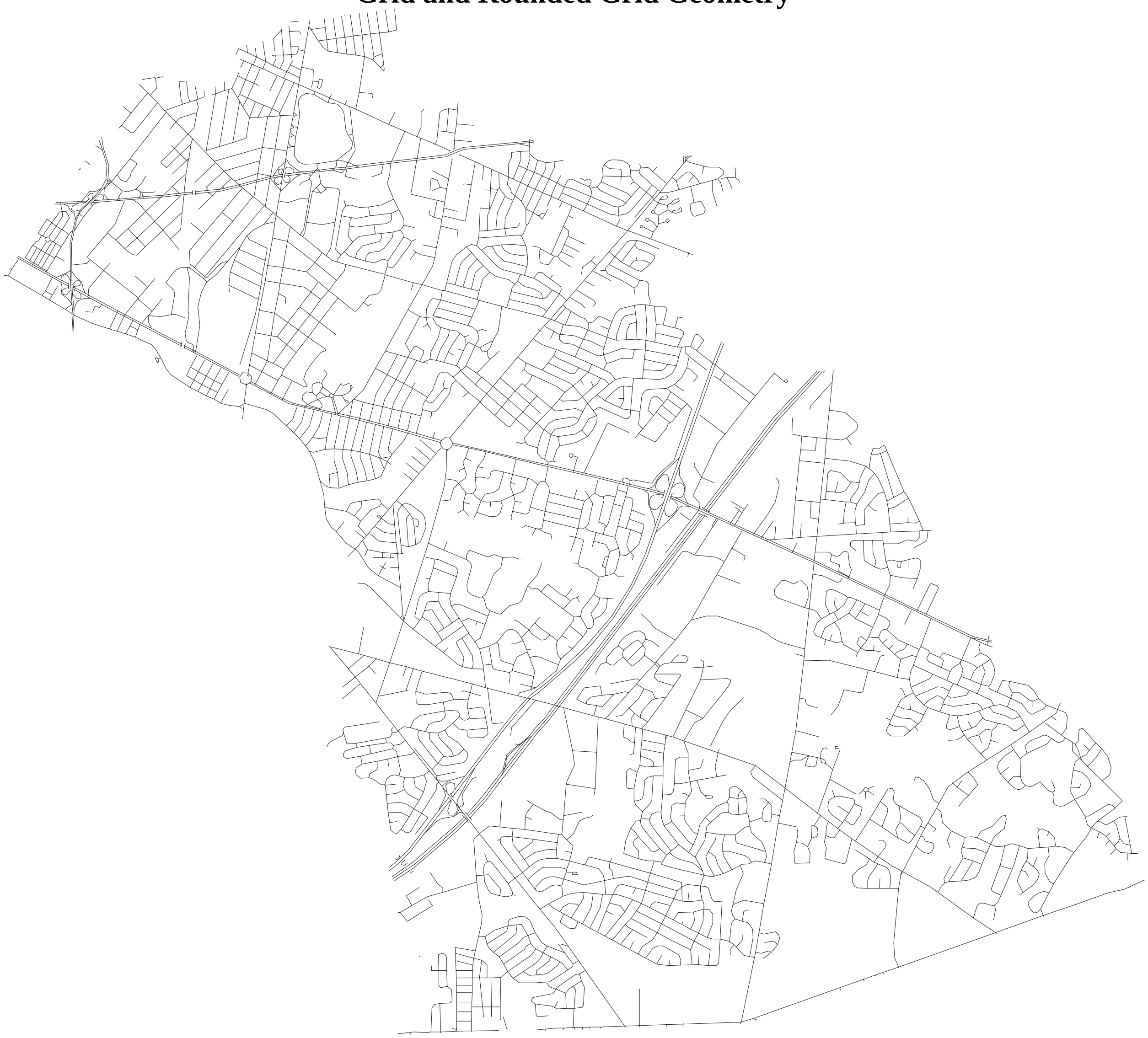
The statistical analyses performed in this study do not p rovide the basis for identifying the cause of the remarkable differences in road to frontage ratios. Therefore a non-random visual analysis of several photoquads was performed. Based on these empirical observations, it is proposed that three municipal road network designs exist, which I will term the historic agricultural network (which is clearly exhibited in cluster 1 communities), the grid based residential road patterns (which are exhibited in medium and high density subgroups), and, a new suburban road pattern which uses branch-like access roads and cul de sacs.

Given the preceding categorization, one can begin to associate the frontage to road ratios to these different road network prototypes. In attempting to associate frontage to road ratios to any subgroup, it is important to remember that pure examples of these three road network prototypes are very hard to find at a municipal wide scale. What is important is the extent to which one type or another dominates in any municipality.

- Grid Based Municipalities

The easiest place to begin this ratio to network design association is with the highest density municipalities. Two types of grid pattern are evident: grids made of straight lines and grid-like shapes produced by rounded circulatory streets. The older compact municipalities, those laid out prior to the automobile, utilize a grid of linear streets. This pattern of city development, made popular during the industrial revolution (in Europe, especially England), had the major intentions of both providing adequate worker housing and minimizing walking distances through increased residential density and narrow frontages (row homes). The rounded grid development patterns can be seen in municipalities designed since the advent of the automobile and up to fairly recent times. In these communities single family detached houses tend to be developed in limited access pods. Internal circulation is provided with rounded grids of access and circulatory municipal streets.

Diagram 5-3
Grid and Rounded Grid Geometry



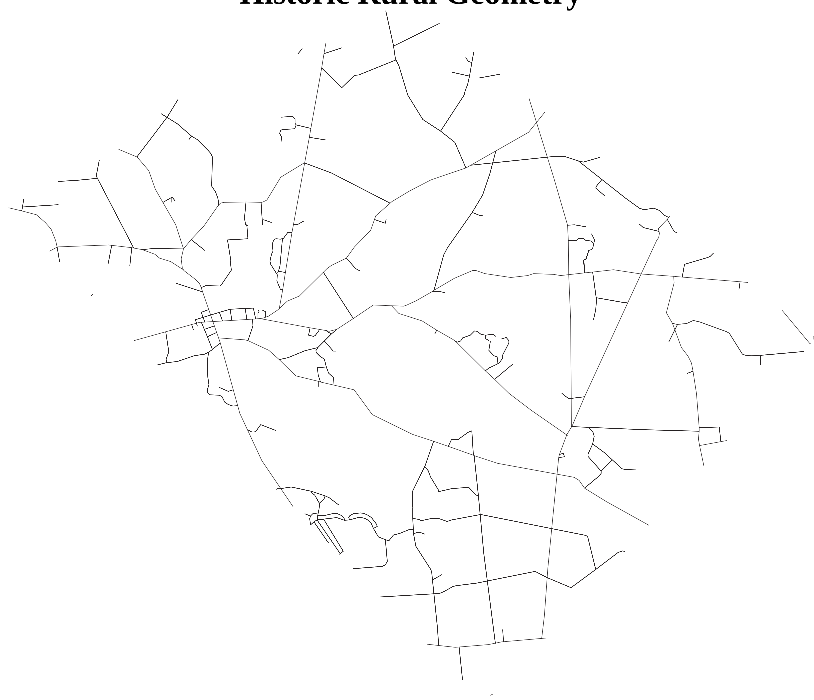
Part of the Municipal Road System of Cherry Hill, Camden County, displaying 'grid and rounded grid' road geometry.
Source: OSP GIS translated from NJDOT Computer File

High density older industrial municipalities, like Camden, Trenton, Paterson and Jersey City, are associated both with a linear based grid road system and a street to frontage ratio of .82 (for cluster subgroup 3-2). This ratio means that for every foot of residential frontage there is just less than a foot of municipal roadway. Curiously, the cluster 2 communities, like Cherry Hill, were built by designers who made a conscious attempt to create suburban communities that were adapted to the automobile. And yet the results of this analysis argue that more municipal roads need to be constructed in these municipalities than are generally constructed in the older industrial grid towns. For example, the street to frontage ratio of cluster 2 is .88 and the street to frontage ratio of municipalities in subgroup 4-1 is 1.03. One possible explanation for this surprising outcome is that the higher ratio of the rounded grid municipalities might be a result of lower density, despite efforts in the ratio methodology to eliminate this effect. A second explanation is that the older grid pattern is more efficient, in the same way that a straight line requires less length to connect two points. That is, by efficient it is meant that less road is required. To test which of these hypotheses were correct, the mean resistant density in the Cluster 3-2 was artificially reduced to 1000 dwelling units per square mile and the lot geometries were changed to 1:4 and 1:3 geometries. With these changes, lot densities and geometries in these older industrial municipalities were very similar to those exhibited by cluster 2 municipalities; the only difference was that the road pattern remained the linear grid. These altered values then were used with the subgroup 3-2 model equation to estimate a new road to frontage ratio. The result of this experiment was that municipal roadway per dwelling unit increased in the older industrial municipalities, but still was less than the 42 feet of roadway required in cluster 2 municipalities. This supports the conclusion that the straight line grid is more efficient than the meandering or rounded grid.

- Historic Rural Municipalities

The lowest density (cluster 1) municipalities tend to be concentrated into the most rural (and probably agricultural) parts of the State. They are grouped in the Pinelands, Cape May, Salem, Cumberland and Sussex counties (see the map in the first section). In these places the rural landscape is served by a thin network of pre-automobile roadways radiating from traditional towns based on grid street patterns. This mixed pattern of grid town and radial roads is referred to as the historic rural pattern. Street to frontage ratios associated with this land development pattern are .46. From this description it can be seen that this historic pattern contains both a linear grid component and a network of radials, along which some access-ready lots have been developed for residential use.

Diagram 5-4
Historic Rural Geometry



Part of the Municipal Road System of AllowayTownship, Salem County, displaying the 'Historic Rural' road geometry.
Source: OSP GIS translated from NJDOT Computer File

- Branching Suburbs

Between the two extremes are the modern suburban residential developments. In these areas, developers have minimized costs by providing access (only) roadways. The pattern of the road network in these areas resembles the pattern of branches on a tree, and will be referred to as the branching pattern for the remainder of this paper. Street to frontage ratios associated with this development pattern range from .24 to .34. It is therefore evident that the branching pattern being built by developers requires substantially fewer linear feet of municipal roads than does either of the grid based patterns or the historic rural pattern (which probably loses its efficiency with its grid based towns).

If the branching pattern serves only a few homes, a single road only will be built. If a larger area is to be served, smaller secondary branches radiate until the road terminates in a cul de sac serving several homes. The resulting radial design only provides ingress and egress connections to dwelling units. The road system lacks collateral circulatory capacity, which means that a break in the road network will severely restrict access to and from the areas distal to the break. However, this system is more efficient (requires less road) than either of the grid based designs.

Diagram 5-5
Branching Geometry

Part of the Municipal Road System of Kinnelon, Morris County, displaying elements of 'branching' road geometry.
Source: OSP GIS translated from NJDOT Computer File

Even within this typology of road networks, the effect of branching can be seen. As noted earlier, municipalities that increased their densities by 100 or more residential units per square mile required fewer linear feet of municipal roads. It now can be understood, that even within municipalities whose cores might be grid based designs, the efficiency associated with GROWRATE is indicative of the development of branching pattern housing developments.

The inferential finding that the branching road pattern requires fewer municipal roads should be approached with some caution. While the geometric efficiency if the branching pattern has ingress and egress advantages, it lacks collateral circulatory capability. Therefore, this street pattern is intrinsically ill-suited to serve development that wants to encourage destination choice, such as the land use pattern of a downtown where traveler access to all uses is desirable. However, branching might be intrinsically suited to be incorporated in the design of residential neighborhoods, especially those at the edges of centers. In this environment the branched roads would minimize nonresidential traffic and would assist in defining development boundaries for the center. The addition of one or two collateral roads or pedestrian paths to the dominant branching road pattern might supply sufficient intra-neighborhood mobility. Design studies should be undertaken to determine the suitability of incorporating the branching road pattern into the design of centers.

Bibliography

Burchell, Robert W. et al. 1992. Impact Assessment of the New Jersey Interim State Development and Redevelopment Plan. New Brunswick, NJ: Rutgers University Center for Urban Policy Research.

Gottlieb, Paul. 1990. Projecting the Costs for Roads under Various Growth Scenarios. Trenton, NJ: New Jersey Office of State Planning.

New Jersey State Planning Commission. Communities of Place: The New Jersey State Development and Redevelopment Plan. Trenton, NJ: New Jersey Office of State Planning.

Norusis, Marija J. 1993. SPSS for Windows: Base System User's Guide, Release 6.1. Chicago, IL: SPSS Inc.

Norusis, Marija J. 1994. SPSS Professional Statistics 6.1. Chicago, IL: SPSS Inc.

Reilly, James. 1993. Revisions to the PED Model: Analysis of Housing Need. Trenton, NJ: New Jersey Office of State Planning.

Sen, Ashish and Srivastava, Muni. 1990. Regression Analysis: Theory, Methods and Applications. New York, NY: Springer-Verlag.

Appendix A - Brief Description of the OSP Computer Models

The Office of State Planning has developed, and continues to refine and revise, a series of computer programs that assign exogenous county-scale forecasts to the municipalities in the counties. In these programs, population and employment are independently and simultaneously assigned using statistical algorithms. The resulting preliminary assignments then are tested to insure that sufficient available, developable land exists to accommodate the municipal assignment. If the land capacity of the municipality is not sufficient, the excess assignment is redirected, by the program, to other municipalities in the county or region. The initial result of the model are Trend estimates of future population, income, households and employment for each of the 567 municipalities in the State. OSP also has developed a program to simulate a wide variety of Plan policies. By altering the attributes of the Plan policies (for example one can change the number, type and density of centers in any municipality), alternative “Plan” municipal forecasts of population, income, households and employment are produced.

In addition to studying the changes in growth patterns that result when Trend and Plan forecasts are compared, OSP also has developed computer programs that estimate the following costs that would result from any set of municipal forecasts:

Public elementary, middle and high school capital costs;
Capital costs for public sewer systems;
Capital costs to build municipal roads; and,
Municipal and State government operating costs.

The purpose of the computer modeling is to test growth management policies and to produce municipal forecasts for use by various agencies, both public and private.

NEW JERSEY OFFICE OF STATE PLANNING
33 WEST STATE STREET
P.O. BOX 204
TRENTON, NJ 08625