Creating Communities of Place

Tomorrow's NEW JERSEY

REVISION TO THE PED MODEL:
A NEW METHODOLOGY TO ASSIGN MUNICIPAL EMPLOYMENT

Document #111

NEW JERSEY OFFICE OF STATE PLANNING

SEPTEMBER 1996

Revision to the PED Model:

A New Methodology to Assign Municipal Employment

Final
September 19, 1996
James Reilly
New Jersey Office of State Planning

Table of Contents

Abstract

This paper describes a behavioral method to estimate total future municipal at-place employment, given an exogenous forecast of total population (and the means to convert this to an forecast of future municipal housing). Both the statistically derived equations and the algorithm schematic needed to incorporate this equation into the Office of State Planning's Spatial Allocation Program are included.

I. Summary

This paper documents the development and testing of a methodology suitable for forecasting total employment within a municipality. There are two major advantages to be gained by incorporating this new methodology into the Population and Employment Distribution (PED) model, the spatial allocation program now used by the Office of State Planning. First, the equation is mathematically defensible, unlike the current OSP employment assignment method. Second, the equation is behavioral, in that it associates employment changes with changes in the number of municipal dwelling units.

This methodology is based on a statistical examination of all 567 New Jersey municipalities using variables suggested by research published by Boarnet (1994) and Putman (1983). The following linear equation expresses the relationship discovered among total 1990 municipal employment and the following independent variables: total 1980 municipal employment; total 1990 miles of municipal (local) roads; and, the change in total dwelling units between 1980 and 1990.

$$LNEMP_{t+10} = .907233 \times LNEMP_t + .0000654855 \times DDUS + .057217 \times LNMUNIM_{t+10} + .703274$$

where:

$$DDUS = dwelling.units_{t+10} - dwelling.units_t$$
$$LNMUNIM = total.municipal.road.miles_{t+10}$$

The equation, however, cannot be simply inserted into the PED. The structure of the PED program had to be changed and modifications made to both calculate and calibrate the linear equation’s forecast.

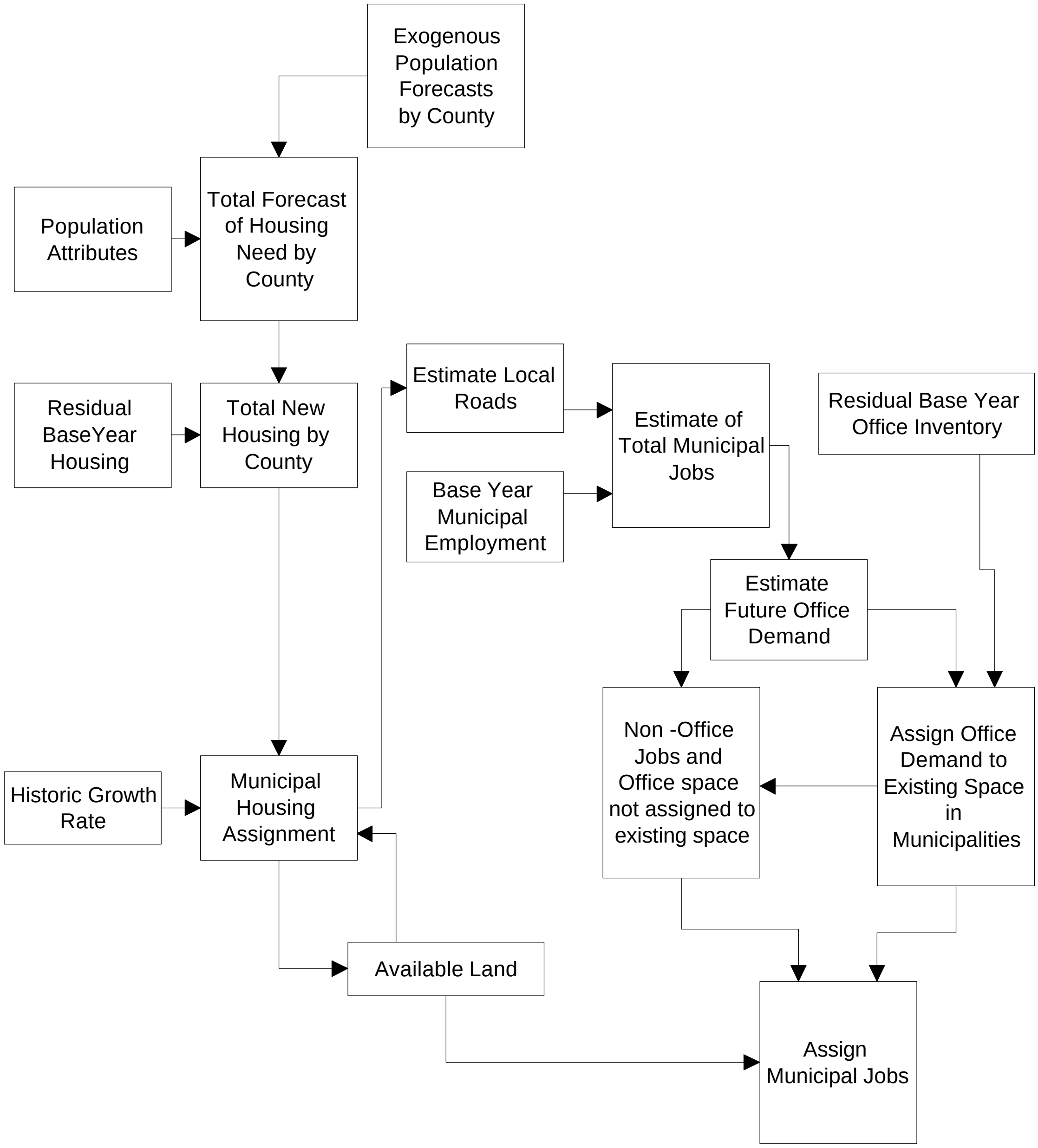
The structure of the PED assignment algorithm was changed to conform with the schematic shown as Diagram 1-1. With this change, municipal housing is estimated and assigned prior to assigning employment . To avoid the problem of over-assigning housing to municipalities with limited supplies of available developable land, the following statistical equation also has been incorporated into the PED to determine the preliminary land requirement for job-related growth. This job-related land is then subtracted from the total available developable supply of land in each municipality, prior to assigning municipal housing.

% of Total Municipal Area, job related development$_{t+i}$ = EXP$(.78818 \times \ln dudens_{t+i} - 7.9301)$

where:

dudens$_{t+i}$ = total municipal dwelling units at time t + i

Diagram 1-1
A Schematic of the Assignment Methodology Making Employment Endogenous



Although the regression equation produced a very high adjusted R^2 of .95453, the following revisions were incorporated into the equation itself and into the PED so that the employment estimates could be modified. These changes were based on testing, described in this report, which demonstrated that the equation had clear limitations.

First, data from the 10 municipalities shown in Table 1-1 had to be excluded because of the atypical nature of the employment found in these places. Independent employment forecasts need to be developed for these municipalities.

Table 1-1
Municipal Outliers Excluded from the Analysis

Case Number	FIPS	Name	Reason
6	34001030	Corbin City	unknown
100	34005035	Chesterfield	Garden State Reception Center and Youth Correction Facility recently constructed
132	34005195	Woodland	New Lisbon Development Center recently constructed
162	34007145	Pine Valley	Primarily a golf course
166	34007165	Tavistock	Primarily a golf course
215	34013070	Newark	Atypically large employment
283	34019125	Union Township	unknown
468	34033020	Lower Alloway Creek Township	Location of a nuclear power plant in a rural Township
553	34041045	Hardwick Township	unknown
562	34041090	Pahaquarry	Part of National Park

Second, this paper demonstrates that the relationship between municipal dwelling units and municipal jobs has been, and is expected to remain, dynamic. Therefore, the use of a time-constrained (1980 - 1990 based) employment estimation equation will underestimate forecast -year municipal jobs. To avoid this problem, municipal forecasts produced using this equation are adjusted using an exogenous regional employment forecast.

Third, because of a minor problem with heteroscedasticity, the actual employment data is positively skewed, although the model results are normally distributed.

Finally, the use of a logarithmic transformation for the dependent variable means that small variances in the forecasted log of future employment may result in fairly large variances in actual municipal jobs. (The relationship only exists at a larger scale of analysis; the real data is more fractal in nature.)

Therefore, the model forecasts need to be reviewed by the county planning agency and calibrated. To accommodate this calibration an additional carrier (with an initial value of 0) is to the equation. The following ‘’rules” should guide the calibration process.

1. In calibrating the model, adjusting the forecast by less than ± 1% per year should be allowed without question. In general, municipalities with less employment than the forecasted total municipal average probably need to have

their forecast reduced, while municipalities with more employment than average need to have their forecast increased.

2. Adjustments larger than ±1% per year need to be documented and substantiated by the requesting agency. It is likely that more municipalities with higher-than-average forecasts will need larger adjustments. Adjustments that change the forecast by more than 3% per year should be discouraged, except in the following cases:

a. Municipalities identified as outliers in this study need to be separately calculated.
b. Municipalities forecast to experience substantial (> two times the mean state change in dwelling units) residential growth.

II. Background of the Project

The original assignment program contained three assumptions regarding employment. First, it assumed that the exogenous regional forecast was correct. Second, it assumed that forecast adjustments relating population growth to employment changes had occurred in the exogenous forecast. Finally, it assumed that both population and employment changes at the municipal scale were affected by the history of growth in the municipality and the supply of available land in the municipality.

The third assumption, that the municipal growth rate could be used to project future growth, is the subject of the revisions described in this paper. Until this research project, the Population and Employment Distribution Model (PED) assumed that the municipal employment growth rate for a specific historic period could be used to identify a preliminary forecast of future employment in that municipality. The resulting raw municipal employment forecast was then constrained by both regional growth, using a proportionate method, and the municipality's supply of available land. This relationship between past growth and forecasted growth was based on "common sense" and not on any mathematical justification. As such, the algorithm was recognized as an assumption and tolerated so that the programming of the initial model could proceed. The intention of the initial model was more to test policy implications than to forecast growth, so that accuracy or justification for the growth assignment equations was not really an issue.

During 1993 and 1994, the model underwent considerable refinement with the sponsorship of the N.J. Department of Transportation. Most of the emphasis of this work was to improve the model's ability to produce reasonable municipal forecasts. Changes were made to both the population and the employment assignment algorithms.

Based on a substantial literature search, the program that assigns municipal population was changed to use only the growth rate of the preceding decade. This preliminary assignment then was again constrained by both the exogenous forecast of regional population growth and the supply of available land in the municipality. With this change, model backcasts were within the acceptable error range of 1% per year; the error range reported for the best population forecast models in the research literature (Smith and Sincich 1987, 1988, 1990, 1991 and Isserman 1984). The initial results of the methodology were presented to county planners for review and comment during 1993 and 1994. In response to their comments, the nature of the assignment algorithm was modified in 1995 from a linear growth equation to a type of logistic equation called a Gompertz curve. Use of the Gompertz curve not only provided a better mathematical simulation of historic growth patterns, it also allowed county planners to participate with OSP in developing a consensus Trend growth forecast by incorporating subdivision activity and known developer interest into the municipal forecasts. (The decay rate in the Gompertz equation is used to calibrate the raw municipal forecasts so that they more realistically reflect recent or proposed subdivision or development activity.)

The employment assignment model was also dramatically changed. As originally designed, all employment growth was assigned to new industrial, commercial or office space, which then was assigned to municipalities in the program. While such a concept might seem reasonable in the rapid economic growth environment of New Jersey in the early 1980's, it was inappropriate for the 1990's conditions of slow employment growth and a substantial over-supply of vacant office space.

Therefore, the employment assignment algorithm was re-written into two subroutines. The first subroutine calculated the demand for office space, using statistical equations, and assigned much of this demand to available existing office space in the region. In the second subroutine all other employment growth and any unfitted office demand were assigned to municipalities based on the previously described methodology. While the employment assignment portion of the PED incorporated a certain level of sophistication in accounting for available land and existing office space, the basic employment assignment algorithm used in the model had not been validated.

Concern about this "hole" in the employment assignment method prompted OSP to secure the consulting services of Dr. Marlon Boarnet in 1994. Boarnet had recently published a employment-estimating algorithm in the *Journal of Urban Economics* (Boarnet, 1994), based on research conducted using 365 municipalities in northern New Jersey. His research finding reported the use of an equilibrium model[1] in which employment was found to lag housing development. In effect, Boarnet was proposing that much employment located to municipalities that had recently experienced housing growth. Boarnet reviewed the OSP assignment model and proposed a multi-step improvement program[2] that would make the assignment algorithm more "behavioral" by making employment growth endogenous to population growth. In effect, Boarnet was proposing that population be assigned first and then that the forecasted population be used as a major variable to determine employment assignment. Other variables that Boarnet reported as being possibly influential in assigning employment were: industry agglomeration; violent crime rates; labor market population changes; municipal land area; transportation access and property crime rates.

---

[1] The prediction of population and employment is inter-related.

[2] This research is the first of the improvement steps proposed by Boarnet. The other main recommendation was to perform submuncipal research to further identify municipal attributes that are related to employment location. This more detailed work is now being planned by OSP.

The lagged employment growth model proposed by Boarnet, appears not to be a new phenomenon. In his book *Integrated Urban Models* (Putman, 1983), Steve Putman noted that "… in newly developing locations, the locations of nonbasic employment, with the exception of a certain amount of speculative activity, seem mostly to follow population location rather than to locate simultaneously"[3]. Putman also identified the importance of including other variables, in addition to population change. He argues for including some measure of existing employment (as does Boarnet) "owing to a kind of inertia which results from the costs of relocation as well as to possible agglomeration economies…"[4], as well as measures of "attractiveness" and travel cost.

[3] Putman, S.H. (1983). *Integrated Urban Models*. Pion Limited, London. P163.
[4] Ibid p164.

III. Developing the New Employment Assignment Model

Variables and Data used in the Analysis

OSP "corrected" the ES 202 data files for 1980 and 1990 so that the employment data identified by 2 digit Standard Industrial Codes (SICs) for each municipality would add up to the municipal total employment reported in the same record[5]. This correction does not necessarily mean that the data is any more accurate -- only that it is internally consistent. (Total employment is equal to the sum of two digit-identified employment for any municipality).

The corrected 1980 and 1990 ES 202 files then were used to determine the change in municipal employment between 1980 and 1990, both in aggregate and by major SIC groupings. Other data used in the analysis came from the U.S. Census and the 1980 and 1990 reports, *Statements of Financial Condition of Counties and Municipalities*, published by the N.J. Department of Community Affairs, Division of Local Government Services.

A total of 70 variables were collected for each of the 567 municipalities in New Jersey. Table 3-1 displays the variables used in this analysis.

Table 3-1
Variables Used in this Study

Measures of Municipal Employment Change[6]

SIC_1_9	- change in total jobs 1980 to 1990 in SIC's 01 to 09 (Agriculture)
SIC_10_1	- change in total jobs 1980 to 1990 in SIC's 10 to 19 (Mining, construction)
SIC_20_2	- change in total jobs 1980 to 1990 in SIC's 20 to 29 (Manufacturing)
SIC_30_3	- change in total jobs 1980 to 1990 in SIC's 30 to 39 (Manufacturing)
SIC_40_4	- change in total jobs 1980 to 1990 in SIC's 49 to 49 (Trans./Com./Util.[7])
SIC_50_5	- change in total jobs 1980 to 1990 in SIC's 50 to 51 (Wholesale)
SIC_52_5	- change in total jobs 1980 to 1990 in SIC's 52 to 59 (Retail)
SIC_60_6	- change in total jobs 1980 to 1990 in SIC's 60 to 69 (F.I.R.E.[8])
SIC_70_7	- change in total jobs 1980 to 1990 in SIC's 70 to 79 (Services)
SIC_80__	- change in total jobs 1980 to 1990 in SIC's 80 to 89 (Services)
SIC_90_9	- change in total jobs 1980 to 1990 in SIC's 90 to 99 (Misc.)
MAN89	- Sum of SIC_20_2 and SIC_30_3
SERVICE	- Sum of SIC_70_7 and SIC_80__
DPEMP89	- change in total municipal private employment 1980 to 1990
DPEMP78	- change in total municipal private employment 1970 to 1980
DEMP89	- change in total municipal employment 1980 to 1990
PCTPVT89	- percentage of private municipal employment change 1980 to 1990

[5] This problem occurred in the employment shown for Atlantic City and Absecon. Some of Atlantic City's hotel employment was incorrectly assigned to Absecon.
[6] "change" refers to numeric change unless otherwise stated in the report.
[7] Transportation, communications and Utilities.
[8] Finance, Insurance, Real Estate.

Measures of Total Municipal Employment and Agglomeration

MAN80	- total municipal manufacturing employment in 1980
SERVICE8	- total municipal service employment in 1980
TCU80	- total municipal TCU employment in 1980
WHOLE80	- total municipal wholesale employment in 1980
RETAIL80	- total municipal retail employment in 1980
PVTEMP90	- total private municipal employment in 1990
PVTEMP80	- total private municipal employment in 1980
TGOVT80	- total government employment in the municipality in 1980
TGOVT90	- total government employment in the municipality in 1990
TEMP80	- total employment in the municipality in 1980
TEMP90	- total employment in the municipality in 1990

Measures of Municipal Population

PDELTA78	- the change in municipal population between 1970 and 1980
DPOP89	- the change in municipal population between 1980 and 1990
PPOP78	- the percentage of municipal population change 1970 to 1980
PPOP89	- the percentage of municipal population change 1980 to 1990
POP1980	- total municipal population in 1980
POP1990	- total municipal population in 1990

Measures of Municipal Density, Development and Land

PDEN90	- municipal population density (sq mile) in 1990
OSP_CUPR	- acres available in the municipality in 1986
SQMILES	- municipal area in square miles
PCTAVAIL	- percentage of total municipal area available for development in 1986
DUS90	- total dwelling units in the municipality in 1990
DUS80	- total dwelling units in the municipality in 1980
DDUS89	- total dwelling units in the municipality built between 1980 and 1990
H8990	- total municipal dwelling units built between 1989 and 1990
H8588	- total municipal dwelling units built between 1985 and 1988
H8084	- total municipal dwelling units built between 1980 and 1984
H70S	- total municipal dwelling units built between 1970 and 1979
H60S	- total municipal dwelling units built between 1960 and 1969
H50S	- total municipal dwelling units built between 1950 and 1959
H40S	- total municipal dwelling units built between 1940 and 1949
PRE40	- total municipal dwelling units built before 1940
SF80	- total single family housing units in 1980
MF80	- total multi-family housing units in 1980

Measures of Municipal Wealth

MEDHHI	- municipal household income in 1980
HHI90	- municipal household income in 1990
HHI8090	- change in municipal household income between 1980 and 1990
ETR80	- Equalized Municipal Tax Rate in 1980

Measures of Municipal Utilities and Access

SEWER08 - percentage of municipal population served by sewers in 2008
SEWER90 - percentage of municipal population served by sewers in 1990
STMILES - centerline miles of State roadway in the municipality in 1990
MUMILES - centerline miles of Municipal roadway in the municipality in 1990
COMILES - centerline miles of County roadway in the municipality in 1990

Transformed Variables

LNEMP90 - Natural Log of TEMP90
LNEMP80 - Natural Log of TEMP80
LNPOP80 - Natural Log of POP1980
LNPOP90 - Natural Log of POP1990
LNDUS90 - Natural Log of DUS90
LNDUS80 - Natural Log of DUS90
LNSTATEM - Natural Log of STMILES
LNMUNIM - Natural Log of MUMILES
LNHHI90 - Natural Log of HHI90
LNDUDEN9 - Natural Log of the municipal dwelling unit density in 1990

Testing the Existing OSP Employment Assignment Algorithm

The first statistical test was to determine the validity of the method now used to make employment assignments. Specifically, the analysis identified correlation (Pearson) between the rate of municipal job change between 1970 and 1980 and the rate of employment change for the period 1980 to 1990. This is a key analysis, since the existing assignment algorithm assumes that the historic growth rate largely dictates future growth or decline. The result of this analysis was a Pearson correlation of .0218. This finding confirmed that the previously used employment assignment algorithm had very little, if any, statistical validity.

Defining the Dependent Variable

Statistical research performed by Boarnet proposes an equation to predict the *change in municipal employment* between 1980 and 1988. Putman's modeling is designed to predict *total employment* in an area, but he implies that distinguishing between basic and nonbasic employment can improve forecasting accuracy. Other research has suggested that forecasting might best be performed by developing algorithms for specific types (SICs) of businesses. These divergent ideas identify the matrix elements confronting any employment forecasting exercise. Should all employment be forecast, or is it better to forecast employment by type? Should total employment be forecast, or should the effort be directed to forecast the change in employment?

Table 3-1
The Forecast Alternatives Matrix

	All Jobs	Jobs by SIC
Total Future Jobs	?	?
Change in Jobs	?	?

Each of these alternatives was investigated in a preliminary way. Correlation tables were prepared using all of the variables. Stepwise regression of the significantly correlated variables then was performed. Several technical findings resulting from this exercise directly affected the selection of the dependent variable used in this study.

It was discovered that all of the regression equations produced highly heteroscedastic forecasts. Transforming the employment variables to their logarithmic values (multiplying the variables by ε) to normalize their distribution tended to reduce, but not totally eliminate, this problem. Efforts also were undertaken to use weighted variables to correct the variance problem, but without success.

Total Jobs or Change in Jobs?

The use of a log transformation created a fatal problem for options that forecast the change in employment. Since employment change might be either a gain or loss, the log transformation could not be used to express integers less than one (e.g., it could not be used to transform negative numbers in the data). While it might be possible to separate municipalities into two sets -- one consisting of municipalities where employment increased and the second consisting of municipalities where employment was stable or declined -- and then to prepare statistical models to forecast the change in employment for each set; such a method would also require that a statistical method be used to forecast which future set (employment gainers or employment stable or loser) municipalities would be at the forecast horizon. In theory this could be accomplished by using a two-stage least square (2SLS) method, but it seems likely that the error resulting from such a procedure would be substantially greater than that which would result from using a well-fitted ordinary least square (OLS) derived equation. Therefore, it was technically much easier to forecast total employment, assuming the value was greater than zero, than to deal with the methodological messiness caused by using the employment change values.

Another problem noted was that methods to predict the changes in employment tend to be less reliable (they tend to have much lower adjusted R^2 values) than do equations intended to predict total employment. For example, the Boarnet equation to predict the change in municipal employment between 1980 and 1988 produced an R^2 of .5315, while Putman reports EMPAL results ranging from .77 to .93. The improved "fitting" of the models predicting total employment is largely due to the strong relationship between the dependent variable (total employment at time t+i) and the independent variable (total employment at time t). This fact leads to a concern about the degree to which employment data for any municipality in the time series (1980 and 1990) are uncorrelated. This possibility was directly tested by regressing total municipal employment in 1990 as the dependent variable with total municipal employment in 1980 as the independent variable and then by performing a Durbin-Watson test. The regression produced an adjusted R^2 of .91056, while the Durbin-Watson test produced a result of 1.93401, demonstrating that serial correlation was not a significant problem[9]. Therefore, while municipal employment in 1990 clearly must encompass much of the 1980 employment, it appears that enough change has occurred during the decade to ensure that these data sets are not mirror images.

Therefore, the decision was made to forecast total future municipal employment, rather than to forecast the change in municipal employment. Such a model produces better-fitting results. Therefore, no analysis of relationships between municipal attributes and the *change* in municipal employment are presented in this paper.

All Jobs or Jobs by SIC?

The decision to forecast total employment rather than employment groups identified by SIC, such as all manufacturing jobs or all retail jobs, was based more on political considerations than on demonstrable facts. There is considerable controversy about the number of jobs located in any municipality, let alone the type of industry group to which these jobs are associated.

Normally State agencies, including OSP, rely on ES 202 data supplied by the N. J. Department of Labor (DOL). ES 202 data is based on unemployment insurance payments and information provided to DOL by employers. NJDOL not only collects the information, it has employees who verify the accuracy of the information. Since the late 1970's, unemployment reports have been mandated for employers of almost all full and part time employees in the state[10].

However, the county planning offices in the Delaware Valley Regional Planning Commission (DVRPC) region (Mercer, Burlington, Camden and Gloucester Counties) have compiled their own estimates of municipal employment that, for many municipalities differ significantly from the ES 202 data. According to planners at DVRPC, these forecasts were prepared using the following methodology. First, the county planning offices were supplied with the

[9] Durbin-Watson tests produce results ranging from 0 to 4. A result equal to or very close to 2 demonstrates that little if any serial correlation is evident between the varaibles.

[10] The major exception to full time employees are parochial school teachers, of which there are an estimated 10,000 in the State; persons earning less than $1,000 per year; and, unpaid family members working in family owned businesses (See Appendix A).

1990 ES 202 municipal employment reports, the U.S. Census Bureau's Travel to Work files (which reports travel to work in 1990 based on a selected sample, from which total municipal employment can be estimated by summing the total trips to any municipality) and a set of municipal employment estimates prepared by the firm of Dun and Bradstreet. The county planning offices then either selected one of these three estimates for each municipality, or prepared an estimate by performing their own study to estimate a "correct" municipal employment[11], or simply substituted their own intuitively derived municipal employment estimate. No attempt was made by DVRPC to verify the accuracy of the base-year employment numbers supplied by the counties.

There likely are some problems with the ES 202 employment numbers produced by the DOL. There are several possible causes of these errors. One cause is that employees are assigned to municipalities based on the mailing address of their employers. Some employers, however, use centralized payroll locations to process unemployment reports for work sites that might be located throughout the State. This can lead the DOL to mistakenly assign employees, who might work at locations throughout the state, to the address of the payroll location. (The DOL believes that most problems of this type have been corrected.) Another cause of error is that public agencies are not as scrupulous in reporting employment locations as are private businesses. This is a particularly difficult problem in Mercer County, because of alleged problems regarding the location of state employees. Another cause is that the DOL uses the zip codes of the reporting businesses to identify municipal location. Zip codes do not conform with municipal boundaries. This makes the process of assigning employees to the correct municipality very difficult, especially for plants that straddle municipal boundaries or for businesses that are located within zip codes that straddle municipal boundaries.

Efforts have been made by OSP to resolve this data problem. In 1994, OSP proposed that DVRPC conduct joint research to identify either which municipal employment report is the most reliable, or to identify a consensus base year employment number. Although DVRPC is an MPO subordinate to the New Jersey Department of Transportation and despite the fact that NJDOT (at the time) was sponsoring OSP's efforts to forecast employment, DVRPC refused to participate in any such resolution exercise. OSP also requested that county planning offices supply copies of any and all studies that were used to justify municipal employment revisions. No county planning offices complied with this request.

Like some other data, such as the crime rate, information about local employment appears to be politically emotional. Until recently, the "Fair Share" affordable housing obligations assigned by the N. J. Council on Affordable

---

[11] For example, the Mercer County Planning Office performed a study to 'correct' employment locational problems between Princeton Borough and Princeton Township.

Housing (COAH)) were based, in part, on employment estimates. (This criteria has now been replaced by the use of commercial and industrial valuations in the COAH methodology to determine Fair Share allocations.) Another contributor to the charged nature of this datum might be a tendency for communities to evaluate themselves based on economic growth. For whatever reason, many municipalities are loath to concur with any set of forecasts that show a real or perceived decline in jobs. Finally, forecasts of employment have been used to determine various infrastructure and transportation needs. Therefore, it is possible that county estimates of employment represent both efforts to correct data errors and to put forward "best face" employment aspirations, rather than actual fact.

Given the difficulties in establishing a baseline municipal employment series, it seemed fruitless to attempt to develop a base line set of municipal employment by SIC or even SIC group. For the purposes of this statistical analysis, the *total* municipal employment information reported in the ES 202 files is assumed to be sufficiently accurate, even if it contains errors, to allow for analysis.

During the next cross-acceptance phase of the State Plan, OSP will attempt to develop a 'justifiable' consensus 1990 municipal employment baseline series for use in the PED. This 1990 baseline series will be forwarded to the DOL as the basis for revisions to their existing ES 202 report. Since ES 202 is the only ongoing employment reporting procedure, data tested in cross-acceptance should be used to validate the accuracy of this information, so that ES 202 can be used with more confidence in the future.

Statistical Analysis of Municipal Employment and Other Municipal Attributes

In the following table the Pearson correlation coefficients relating both the change in municipal employment between 1980 and 1990 (DEMP89) and the total 1990 municipal employment to all of the variables (transformed variables are not presented) are displayed. (Variables that produced high correlations are in boldface.)

Table 3-3
Correlation Coefficients

Measures of Total Municipal Employment and Agglomeration

	TEMP90	DEMP89
MAN80	.8474**	-.0151
SERVICE8	**.9263****	**.2376****
TCU80	.7929**	-.0074
WHOLE80	**.7804****	**.1705****
RETAIL80	**.8437****	**.3422****
PVTEMP80	**.9679****	**.1573****
PVTEMP90	**.9847****	**.4261****
TGOVT80	.7977**	-.0339
TGOVT90	.8191**	.0081
TEMP90	1.0000	.3474**
TEMP80	**.9716****	**.1158****

Measures of Municipal Population

	TEMP90	DEMP89
PDELTA78	**-.4793****	**.2155****
DPOP89	**-.2442****	**.4044****
PPOP78	-.1183**	.0708
PPOP89	-.0494	.1396**
POP1980	.8704**	.0525
POP1990	.8573**	.1193**

Measures of Municipal Density, Development and Land

	TEMP90	DEMP89
PDEN90	**.2208****	**-.1339****
OSP_CUPR	-.0366	.1310**
SQMILES	.0569	.1723**
PCTAVAIL	-.1118**	.0578
DUS80	.8609**	.0449
DUS90	.8471**	.1388**
DDUS89	**-.1003***	**.4897****
PRE40	-.0311	-.2267**
H40S	.0669	-.1369**
H50S	.0615	-.0254
H60S	.0952*	.1472**
H70S	-.0999*	.1039*
H8084	-.0017	.2312**
H8588	-.0426	.1976**
H8990	-.0515	.1010*
SF80	**.5803****	**.2859****
MF80	**.7659****	**-.1136****

Measures of Municipal Wealth

	TEMP90	DEMP89
MEDHHI	-.0795	.1126**
HHI90	-.0795	.1126**
HHI8090	-.1735**	.0763
ETR80	.2634**	-.0804

Measures of Municipal Utilities and Access

	TEMP90	DEMP89
SEWER08	.1684**	.0597
SEWER90	.2291**	.0622
STMILES	**.3555****	**.3096****
COMILES	**.2397****	**.2226****
MUMILES	**.6323****	**.3259****

* - Signif. LE .05   ** - Signif. LE .01   (2-tailed)

" . " is printed if a coefficient cannot be computed

Prior to deciding which variable to use, it must be remembered that to use some of these variables to forecast future employment, the variable itself must be forecast by other subroutines in the PED model. The following table shows whether a specific variable is forecast in the PED model.

Table 3-4
Availability of Variables

Variable	Status
Base year Employment by type	*(Not used - see preceding section)*
DEMP78	*Historic Information*
DEMP89	*Not Available* - circular reference
PVTEMP80	*historic Information*
PVTEMP90	*Not Available* - circular reference
TEMP80	*Historic Information*
Pdelta78	*Historic Information*
DPOP89	*Not Available* - population not determined until end of PED. Algorithm uses municipal income which is derived from employment - circular reference)
PDEN90	*Not Available* (see previous note)
DDUS89	*Available* - Model calculates delta dwelling units prior to employment
SF80	*Historic Information* - but less reliable as forecast period exceeds 10 years
MF80	*Historic Information* - but less reliable as forecast period exceeds 10 years
STMILES	*Not Forecast in PED*
COMILES	*Not Forecast in PED*
MUMILES	*Available* - can be calculated given forecasted dwelling units

Regression Analyses - Predicting Total Future Municipal Employment

Three different models were produced, using stepwise regression, to compare the relationships between sets of the significantly correlated variables. These models were developed by using 1980 data or forecastable data about the changes between 1980 and 1990 to predict total 1990 municipal employment. Two of the models avoid the use of 1980 employment; instead using the strong relationship between population and employment as the principal explanatory variable. The third model directly uses base-year employment to calculate future-year total employment. Each of these models is similar in many ways to those proposed by both Boarnet and Putman. All of the models propose that much future employment results from existing employment inertia -- firms, once located, tend to stay in the same place -- and that the change in employment is associated with population or housing growth. Most of the models also include some measure of transportation access.

The first model uses the independent variables: Pdelta78, POP1980, MUMILES and ETR80. This model proposes that future changes in employment are a result of past population growth, the size of the base-year population (or the relationship between base-year population and base-year employment), a measure of local access (mumiles) and a measure of the municipality's relative tax rate, the base-year equalized tax rate (ETR80). The second model also uses POP1980 as the main predictive variable and it uses the change in the number of municipal dwelling units (DDUS89) between the base year and the forecast year. This model evolved from attempts to substitute DDUS89 for Pdelta78 (the change in municipal population between 1979 and 1980). The third model uses the base-year total employment variable (LNEMP80), a measure of local access (LNMUNIM), and DDUS89. The following table displays selected regression results produced by these three models.

Table 3 - 5
Regression Results for the Three Employment Models

Model Variables	Adjusted R^2	F	Significance of F	Durbin-Watson Coefficient
POP1980, Pdelta78, MUMILES and ETR80	.77685	493.61	.0000	2.05500
POP1980 and DDUS89	.75679	881.59	.0000	2.03794
LNEMP80, LNMUNIM and DDUS89	.91366	1997.50	.0000	1.92955

As demonstrated in Table 3-5, the model using the natural log of base-year employment (LNEMP80), a measure of local access (LNMUNIM) and a measure of residential change (DDUS89) produced the best-fitting simulation of actual events. Because of this finding, this model has been selected to be further refined.

IV. Refining and Testing the Selected Regression Equation

This chapter consists of two parts. The first part explains corrective actions that were taken to improve the model and to ensure that it did not (seriously) violate any of the GM preconditions. The second part presents a detailed examination of the revised model's predictive characteristics.

No regression model is perfect; they are all, by definition, mathematical generalizations. The purpose of this chapter is to document the reliability of the regression equation so that technical users of the OSP PED model can be more fully informed of how the equation was developed and of its predictive shortcomings and strengths.

Refining the Regression Equation

Table 4-1 and Diagram 4-1 presents the results obtained by regressing the natural log of future employment (LNEMP90) with the independent variables: the natural log of 1980 municipal employment (LNEMP80); the natural log of 1990 miles of local municipal roads (LNMUNIM); and, the change in total municipal dwelling units between 1980 and 1990 (DDUS89).

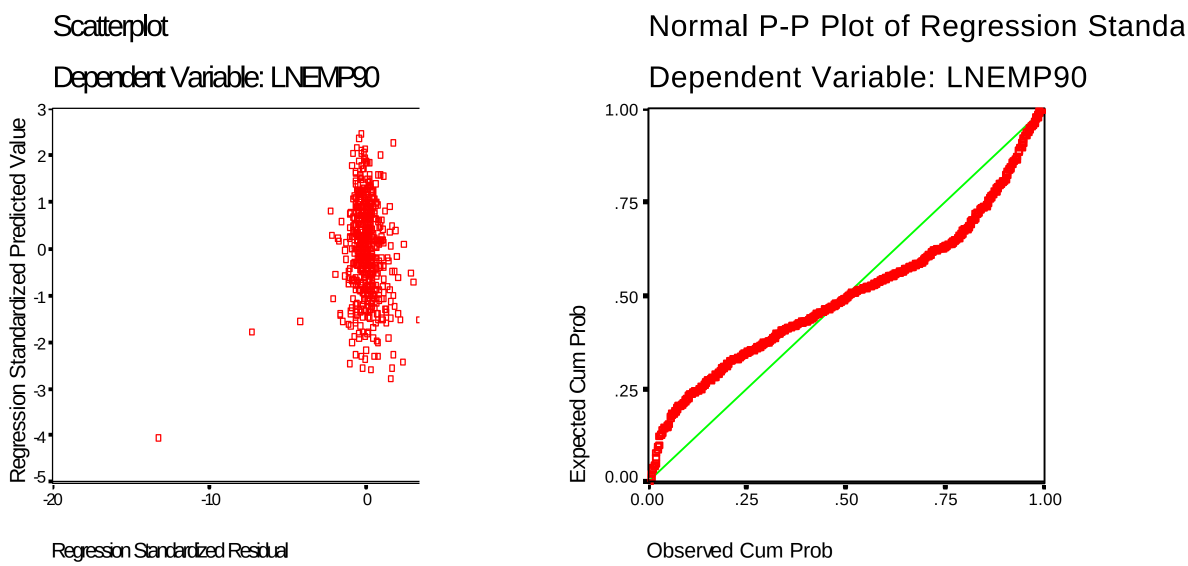
Table 4-1
Employment Regression Results (all cases)

```
Multiple R           .95610
R Square             .91412
Adjusted R Square    .91366
Standard Error       .46925

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           3          1319.51362        439.83787
Residual           563           123.96919           .22019

F =    1997.50208       Signif F =  .0000
---------------------- Variables in the Equation -----------------------

Variable              B        SE B     95% Confdnce Intrvl B       Beta

LNEMP80         .949048     .015341      .918915      .979180    .933555
LNMUNIM         .036840     .022942     -.008222      .081902    .025275
DDUS89      4.95357E-05  1.3105E-05  2.37944E-05  7.52770E-05    .049193
(Constant)      .461827     .096508      .272266      .651387


----------- Variables in the Equation -----------

Variable    Tolerance        VIF         T  Sig T

LNEMP80       .669848      1.493    61.863  .0000
LNMUNIM       .615765      1.624     1.606  .1089
DDUS89        .900606      1.110     3.780  .0002
(Constant)                           4.785  .0000

Collinearity Diagnostics
```

Number	Eigenval	Cond Index	Variance Proportions Constant	LNEMP80	LNMUNIM	DDUS89
1	3.14660	1.000	.00396	.00273	.00594	.02285
2	.78621	2.001	.00241	.00136	.00106	.89653
3	.04901	8.013	.30055	.01435	.77301	.07734
4	.01817	13.159	.69308	.98156	.21999	.00328

Durbin-Watson Test = 1.92955

Diagram 4-1



From the preceding table and diagram some advantages and some drawbacks of the regression equation can be identified. On the plus side, the model produces a good fit of actual results (it has a high adjusted R^2). The model does not suffer from collinearity ; variables in the model are not related to each other. If they were related, it would produce a high variance inflation factor (VIF) value and/or a Durbin-Watson value other than a number close to 2.0[12]. Finally, the model produces a low condition index value, which suggests that small changes in any of the variables will not produce dramatic shifts in the employment forecast. Both of disadvantages are displayed in Diagram 4-1. The scatterplot of the residuals does not produce a random pattern and the normal probability plot of the residuals shows clear signs of heteroscedasticity. Both diagrams suggest a problem with the GM precondition that there be a constant variance of *y* for all values of *x*.

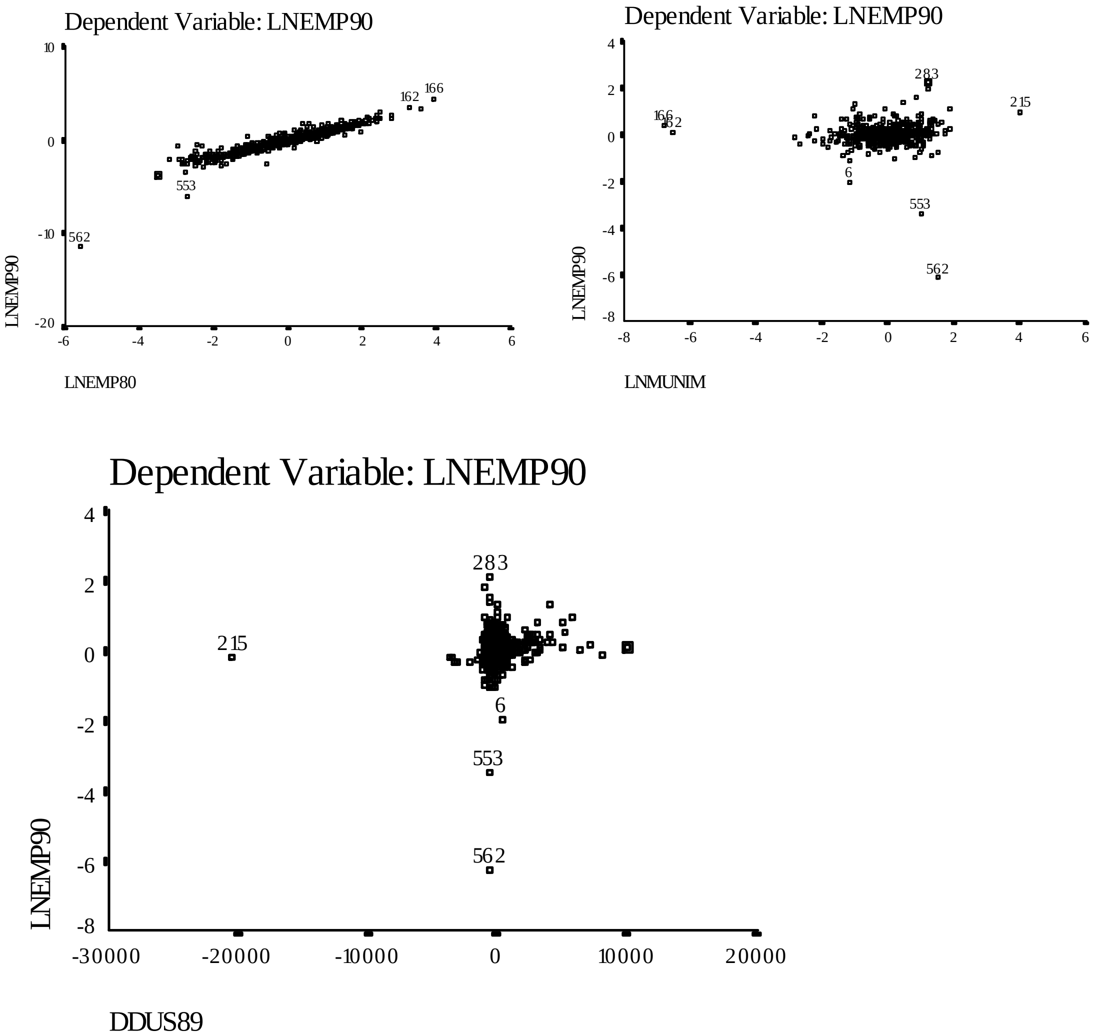
Since variable transformation had already been used to reduce heteroscedasticity, by improving the normality of the data's distribution, a search

[12] In other words, the independent variables are independent of each other and not used to improve the fit of the equation by relating to each other.

for influential outliers[13] was begun by preparing scatterplots of the residuals produced by regressing of each independent variable with the dependent variable. In Diagram 4-2, potential outliers are identified by their case number. The case numbers were then used to determine the FIPS identity and municipal name of the outlier in Table 4-2. This table also provides some explanation, if it is known, of why the municipality's relationship between employment and the independent variables might be atypical.

Diagram 4-2

Scatterplot of Residuals for Each Independent Variable (all cases)



---

[13] The regression line is calculated by a form of distance averaging. Data points which produce large residuals therefore, tend to disproportionately affect the solution of the best fit line's equation. Plotting of these residuals identifies these unusual cases.

Table 4-2
Data Outliers Identified in the Scatterplot of Residuals

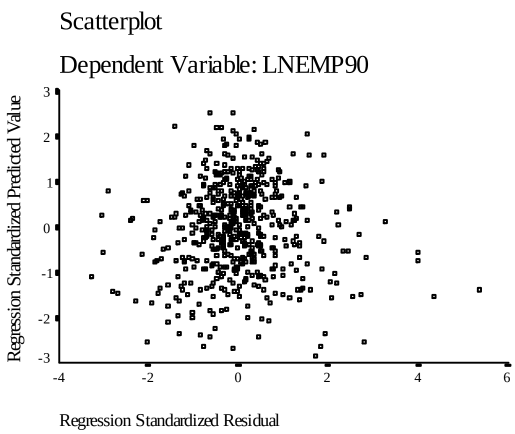
Case Number	FIPS	Name	Reason
6	34001030	Corbin City	Unknown
162	34007145	Pine Valley	Primarily a golf course
166	34007165	Tavistock	Primarily a golf course
215	34013070	Newark	Atypically large employment
283	34019125	Union Township	Unknown
553	34041045	Hardwick Township	Unknown
562	34041090	Pahaquarry	Part of National Park

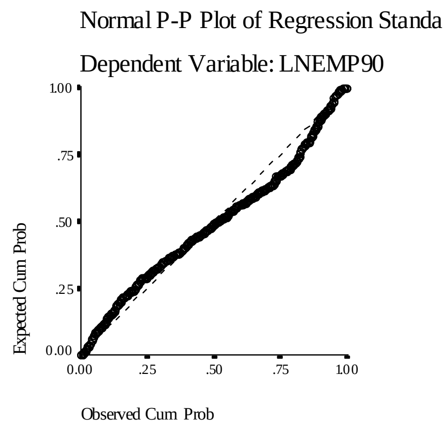
Based on this analysis of influential cases, the seven municipal cases shown in Table 4-2 were deleted from the data set and the regression performed a second time. Table 4-3 and Diagram 4-3 display the results produced by the second regression.

Table 4-3
Employment Regression Results (560 cases)

```
Multiple R           .97409
R Square             .94885
Adjusted R Square    .94858
Standard Error       .33239

Analysis of Variance
                  DF      Sum of Squares      Mean Square
Regression         3          1139.65121        379.88374
Residual         556            61.43041           .11049

F =    3438.28659       Signif F =  .0000

---------------------- Variables in the Equation -----------------------

Variable              B        SE B     95% Confdnce Intrvl B       Beta

LNEMP80         .898743     .011363      .876422      .921063    .933063
LNMUNIM         .063809     .019205      .026086      .101532    .042881
DDUS89     6.28624E-05  1.1579E-05  4.01185E-05  8.56063E-05    .058517
(Constant)      .755216     .072647      .612519      .897912

----------- Variables in the Equation -----------

Variable    Tolerance        VIF         T  Sig T

LNEMP80       .660943      1.513    79.091  .0000
LNMUNIM       .552276      1.811     3.323  .0010
DDUS89        .791806      1.263     5.429  .0000
(Constant)                          10.396  .0000

Collinearity Diagnostics

Number  Eigenval     Cond  Variance Proportions
                    Index  Constant  LNEMP80  LNMUNIM   DDUS89
    1    3.24808    1.000    .00331   .00235   .00410   .02373
    2     .69997    2.154    .00361   .00180   .00090   .79641
    3     .03495    9.640    .40203   .00923   .76904   .17767
    4     .01699   13.825    .59105   .98662   .22596   .00219

Durbin-Watson Test =   1.96005
```

Diagram 4-3
Scatterplot and NPP produced by the Employment Regression (560 cases)





As one might expect, eliminating unusual cases from a data set improves the model's ability to simulate the remaining data. As shown in the preceding table and charts, the fit of the regression equation improved from an adjusted R^2 of .91366 for all cases to an adjusted R^2 of .95458, after seven outlier municipalities were deleted from the data set. Other improvements are a substantially higher F value (1997.5 for all cases and 3895.9 for the reduced data set), higher T values for each of the variables (especially for LNEMP80 and DDUS89) and a somewhat improved Durbin-Watson value. But the really important improvements are displayed in the scatterplot of the residuals and the predicted values and the NPP of the residuals. The scatterplot displays a much improved random pattern, while the NPP shows that deletion of the outlier cases substantially improved, but did not altogether eliminate, the problem with heteroscedasticity.

Because of this continuing (but improved) problem with heteroscedasticity, another examination of residuals scatterplots was performed. The only remarkable results are illustrated in Diagram 4-4, which is the residual plot of the regression of LNEMP90 and LNEMP80. Three new outliers were identified in this diagram. The identity of these new outliers is shown in Table 4-4.

Diagram 4-4
Scatterplot of the Residuals - Regression of LNEMP90 and LNEMP80
(560 cases)

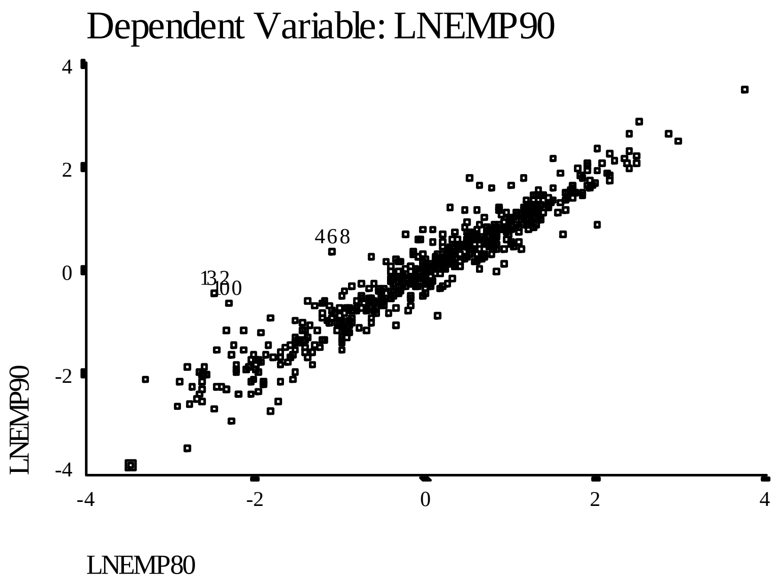


Table 4-4
2nd Set of Data Outliers Identified in the Scatterplot of Residuals

Case Number	FIPS	Name	Reason
100	34005035	Chesterfield	Garden State Reception Center and Youth Correction Facility recently constructed
132	34005195	Woodland Township	New Lisbon Development Center recently constructed
468	34033020	Lower Alloway Creek Township	Location of a nuclear power plant in a rural Township

A third analysis of outliers also was performed. While some additional possible outliers were identified, removing these cases from the data produced marginal improvements in the regression equation. Therefore the total of 10 cases listed in Tables 4-2 and 4-4 encompass all of the outliers removed from the data to improve the performance of the regression equation to forecast municipal employment.

Table 4-5 and Diagram 4-5 show the regression results produced by using this reduced data set. While the fit of the model remained essentially unchanged, the F value increased from 3438.29 to 3891.84. This final equation continues to exhibit a small amount of heteroscedasticity. The analysis presented in the second part of this chapter uses this equation.

Table 4-5
Final Employment Regression Results (557 cases)

Multiple R	.97713
R Square	.95478
Adjusted R Square	.95453
Standard Error	.31329

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	3	1145.96747	381.98916
Residual	553	54.27765	.09815

F = 3891.84130 Signif F = .0000

------------------------ Variables in the Equation -------------------------

Variable	B	SE B	Beta	Tolerance	VIF	T
LNEMP80	.907233	.010785	.938199	.657377	1.521	84.118
LNMUNIM	.057217	.018193	.038392	.548772	1.822	3.145
DDUS89	6.54855E-05	1.0923E-05	.060950	.791224	1.264	5.995
(Constant)	.703274	.068753				10.229

------ in -------

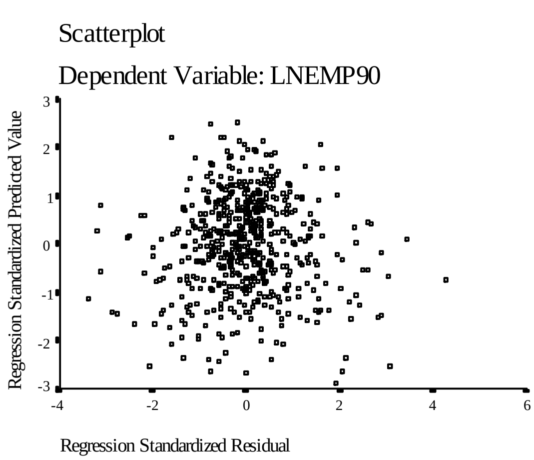
Variable	Sig T
LNEMP80	.0000
LNMUNIM	.0017
DDUS89	.0000
(Constant)	.0000

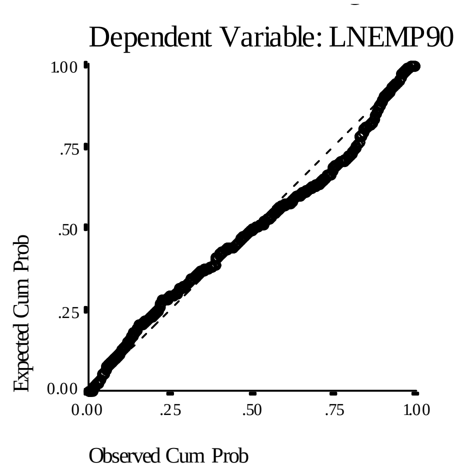
Collinearity Diagnostics

Number	Eigenval	Cond Index	Variance Proportions Constant	LNEMP80	LNMUNIM	DDUS89
1	3.24905	1.000	.00330	.00232	.00407	.02373
2	.69919	2.156	.00361	.00179	.00089	.79574
3	.03494	9.643	.40448	.00846	.76177	.17777
4	.01682	13.900	.58861	.98743	.23327	.00277

Durbin-Watson Test = 1.94008

Diagram 4-5
Residual Scatterplot and NPP produced by the Employment Regression (557 cases)





Performance Characteristics of the Regression Equation

The following section evaluates the model's forecasts so that both its reliability and any weaknesses can be documented. The purpose of this analysis is to better understand how the model performs so that it can be used and calibrated in a more sophisticated, less intuitive, manner.

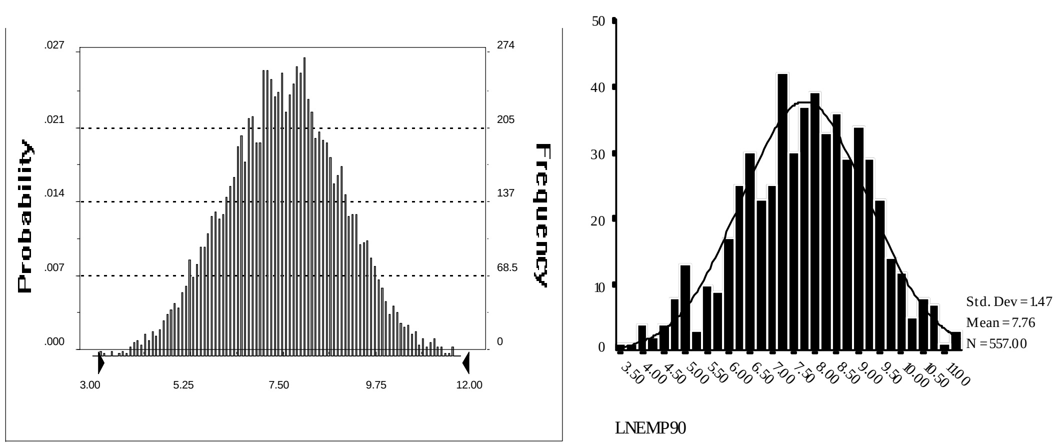
Testing Using a Monte Carlo Simulation

The first test of the model is intended to determine if the equation is mathematically stable and to determine if the results closely mirron the actual distribution of LNEMP90, the value being predicted by the equation. To a certain extent, this testing duplicates several statistical tests previously performed on the model equation. In performing statistical testing, however, real historical values used are always used. In this testing procedure, a special computer program[14] is used to generate values for the independent variables that are random but that conform to both the range of the historic variables and the shape of the distribution of the historic variables. (In more technical terms, the testing program uses a Monte Carlo simulation to test the regression equation.) Although it might seem unusual that using this larger, but bounded, number set should produce results that would differ from using the original number set of 557 cases, this process can identify number combinations that produce results that are erratic or numeric errors.

Table 4-6 and Diagram 4-6 display the result of 9,992 calculations of the employment regression equation and compare these forecasts to the actual descriptive characteristics of the total 1990 municipal employment data.

Diagram 4-6
Diagrams of the Distribution of LNEMP

Regression Forecasts　　　　Actual data



---

[14] Crystal Ball®

Table 4-6
Comparison of Descriptive Statistics
LNEMP90 and Regression Equation Simulation

	Regression Forecast	**Actual**
Mean	7.68	7.7220
Median	7.69	7.7765
Min. value	1.93	-4.6052[15]
Max. Value	11.88	11.9968
Range	9.95	16.6019[16]
Std. Error	.01	.0671
Variance	2.00	2.5503
Std. Deviation	1.42	1.5970
Skewness	-0.06	-1.0416
Kurtosis	2.90	5.8643

Two findings are produced by this analysis. First, the Monte Carlo simulation demonstrates that the regression equation is mathematically stable within the range of variables used to produce the regression. Second, the results demonstrate that the regression equation produces a slightly biased forecasts in certain ways. Most of these findings can be observed by comparing the skewness and kurtosis values produced by the model with those found in the real data. In general, and as expected, it can be seen that the model produces a more normal distribution of results than is actually found in the 1990 employment data. That is, while the historic data approximates a normal distribution, the model results conform more closely to this ideal shape. This means that the real data is a little less regular than are the model results. The historic data is slightly shifted to the right of the high point of the superimposed normal distribution line, while the model data tends to be closer to this ideal distribution. This suggests that the model might produce higher and lower estimates of employment for municipalities with employment just below or above the statewide average than is actually found in the real data. Both of these findings illustrate the problem of heteroscedasticity previously reported in this chapter.

Testing Using Data Subsets

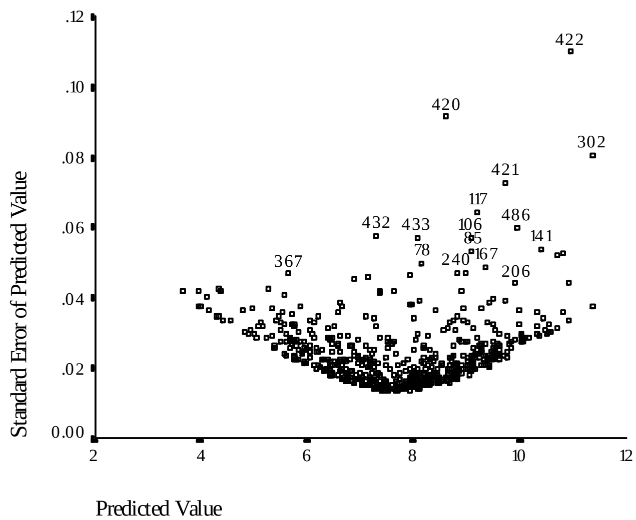
The next analysis uses scatterplots to evaluate model forecast errors using both the entire set of municipal data and subsets of the cases. The purpose of this analysis is to determine if the model forecasts are consistently reliable across the range of municipal values, or if the forecasts are biased in predictable ways that might guide calibration. For example, the previous analysis suggested that the regression equation produces more errors in municipalities

[15] This number appears to be an error, since it represents a value of less than one job.
[16] This value is larger due to the error in the minimum value.

with employment close to the mean. This analysis will verify that proposition and, more importantly, will quantify the size of the error that the model is likely to make. To develop this analysis, the regression equation was used to produce a forecast of municipal employment. This forecasted value then was compared to the actual LNEMP90 value to determine the error produced by the model. This result is displayed using scatterplots showing the forecast error for the y-axis value and the predicted value for the x-axis value. This analysis was conducted using both the entire data set and on quartile subsets.

Diagram 4-7 shows the scatterplots that display regression errors for the entire reduced (excluding outlier municipalities) data set. The plot also shows the case numbers of the municipalities with larger errors.

Diagram 4-7
Scatterplot of Regression Errors



First, the plot of the regression errors demonstrates that the model produces few very large errors. In general, the standard error of the forecast ranges from a little less than 1% to about 5%; for a forecast over a 10-year period. The highest single error in the prediction is between 10% and 12%, or only slightly more than 1% per year.

It can also be seen that the size of model errors tends to increase as the size of the forecast increases or decreases from the mean. There are two meanings to this observation. First, the general concave shape of the errors with respect to the x-axis is a normal anticipated product of OLS regression: the error increases as distance from the mean is increased. Also, as the predicted value increases along the x-axis (and away from the mean prediction), some municipalities exhibit large errors. This appears to be especially true for municipalities with greater-than-average employment. Table 4-7 displays the case numbers, FIPS codes and names of the municipalities in which larger errors occurred. Many of these municipalities experienced rapid residential development during the decade. This suggests that the model many need to be modified (the forecast reduced either positively or negatively) if it is used to forecast employment in a municipality that is expected to undergo substantial growth during the forecast horizon.

Table 4-7
Municipalities where larger errors were produced

Case Number	FIPS Code	Municipal Name	% Change in Dwelling Units 80-90
78	34003275	Rockleigh Township	-5%
86	34003315	Upper Saddle River	3%
106	34005065	Evesham Township	90%
141	34007040	Camden City	-7%
167	34007170	Voorhees Township	103%
206	34013030	East Orange	-7%
240	34015090	Washington Township	62%
302	34023025	Edison Township	35%
367	34025230	Shrewsbury Township	20%
420	34029025	Berkeley Township	96%
421	34029030	Brick Township	39%
422	34029035	Dover Township	49%
426	34029055	Jackson Township	46%
432	34029085	Long Beach Township	222%
433	34029090	Manchester Township	43%
486	34035040	Franklin Township	64%
		State Average	**13%**

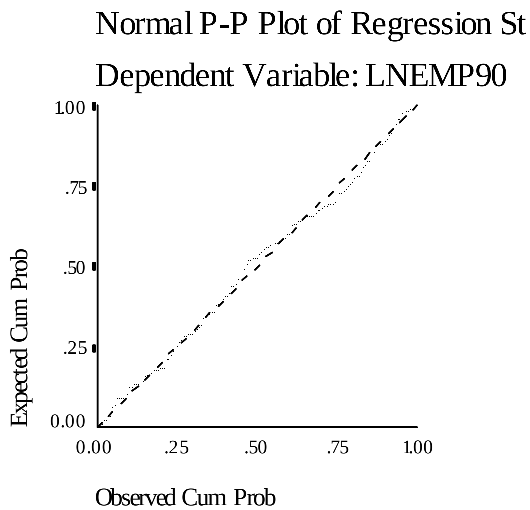
In addition to evaluating the model's overall performance, the actual LNEMP90 values were used to subdivide the data set into quartiles and the regression results developed using each quartile are presented below. Table 4-8 displays various descriptive and regression results for each of the quartiles. Diagram 4-8 showing the NPPs for each quartile and the scatterplot of errors for each quartile.

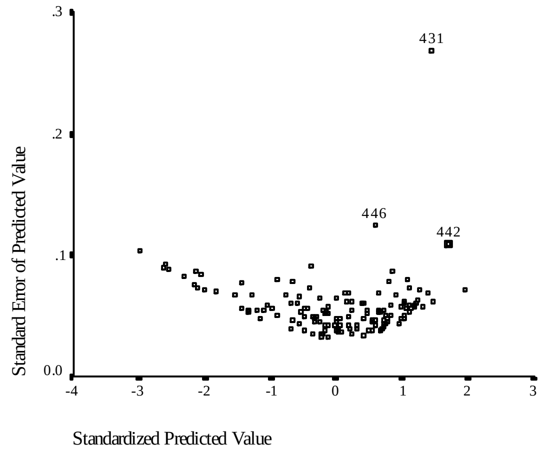
Table 4-8
Statistical Analysis of Forecasts by Quartile

	Bottom Quartile	Lower-Middle Quartile	Upper- Middle Quartile	Top Quartile
Minimum	3.497	6.780	7.832	8.868
Maximum	<6.780	<7.832	<8.868	11.328
Cases	137	138	139	139
Adj. R^2	.75168	.44395	.50105	.85428
F	139.2371	37.7266	47.5280	270.6745
Sig. F	.0000	.0000	.0000	.0000
Durbin-Watson	1.96455	2.00877	1.85405	1.93323

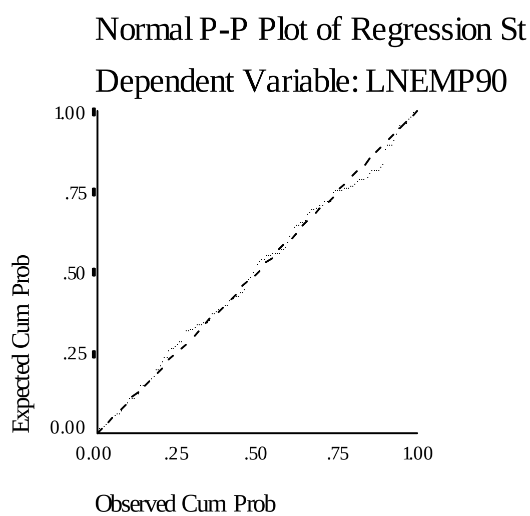
Diagram 4-8
NPP and Error Scatterplots by Quartile

Bottom Quartile





Lower Middle Quartile



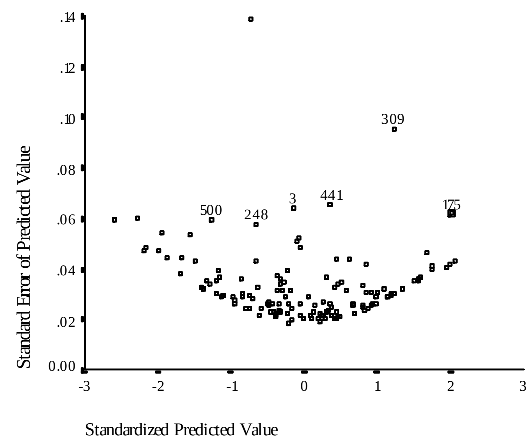
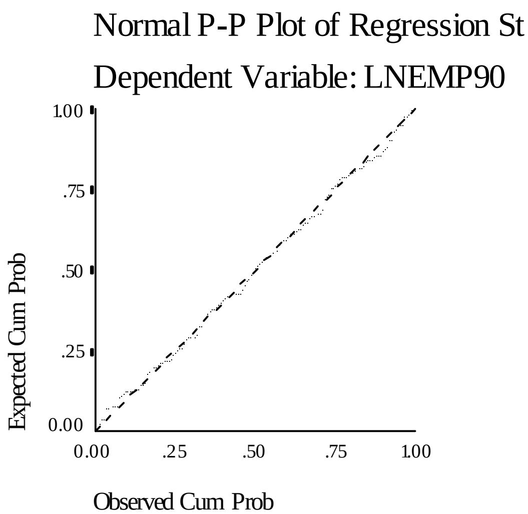
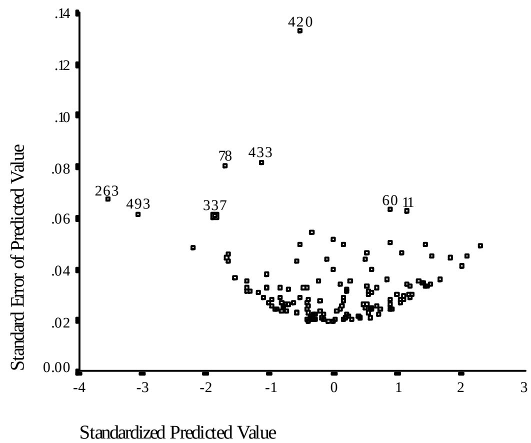


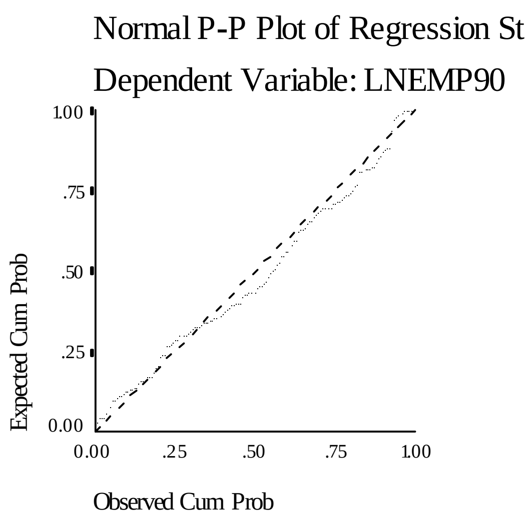
Diagram 4-8 continued

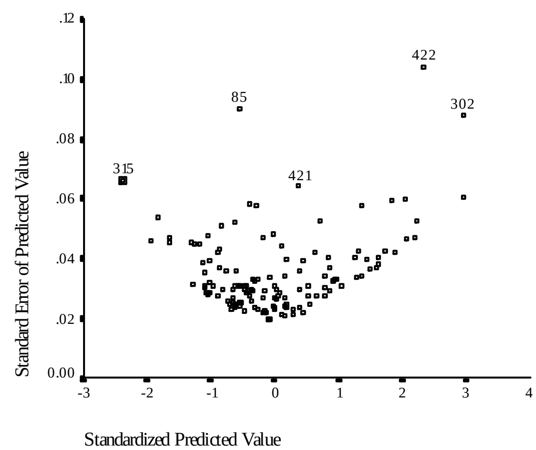
Upper Middle Quartile





Top Quartile





Several findings are evident in this analysis.

1. The distribution of errors is normally distributed in each of the quartiles. This means that predictable results are produced by the model using data applicable for most municipalities.

2. The inability of the model to produce high adjusted R^2 values for the middle quartiles is consistent with the earlier finding that the forecasts are heteroscedastic. For municipalities with average employment, the model is not as consistently accurate; it produces forecasts that are more symmetric to the real data for municipalities with low or high numbers of jobs.

Can the Model Predict Employment Losses?

The next test performed on the regression equation was to determine if it could produce a forecast of employment decline and to determine the variable that was influencing this result. It was discovered that the equation could produce forecasts of employment decline and that the determining variable was DDUS89, the change in the number of dwelling units during the forecast horizon. It also was discovered that DDUS89 had shown a substantial numeric decline in the number of dwelling units (≤-1000) before the model produced a decline in municipal jobs. This finding demonstrates that the equation is not likely to forecast employment decline in places that do not have substantial numbers of existing dwelling units or that do not lose substantial dwelling units during the forecast horizon.

Testing the Model's Municipal Job Predictions

The analysis in the preceding sections of this chapter is based on the regression equation's ability to predict the natural log of the total municipal employment. Logarithms can be confusing, since a small difference in log value might represent a large difference in municipal jobs. To further test the regression equation, historic data was used to predict municipal employment in all of the Camden County municipalities. (The selection of this county was arbitrary.) In Table 4-9 the error between the forecast of jobs and the actual 1990 ES 202 report of jobs in each municipality were compared. Diagram 4-9 displays the actual log jobs and the estimated log jobs on the left and a histogram of the distribution of errors ($\frac{forecast.jobs - ES202}{ES202}$) on the right.

Diagram 4-9

Analysis of Model Predictions - Camden County

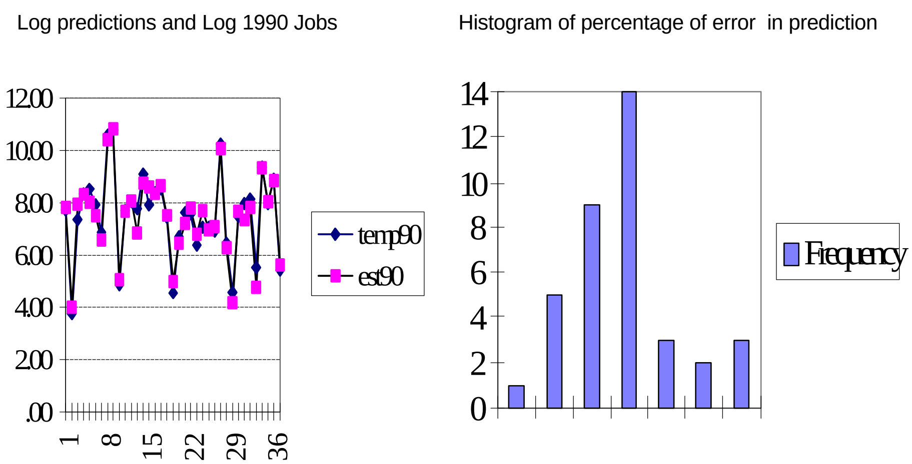


Table 4-9
Jobs Forecasted and Actual for Camden County Municipalities in 1990

	Actual 90 (ln) jobs	lnmunim	ddus89	lnemp80	Forecast 90 (ln) jobs	error # of Jobs	% error
Audubon	7.72	3.08	-22	7.64	7.81	229	10%
Audubon Park	3.74	.73	22	3.56	4.00	13	30%
Barrington	7.36	2.88	-77	7.80	7.94	1229	78%
Bellmawr	8.35	3.60	107	8.15	8.32	-118	-3%
Berlin	8.51	3.11	125	7.84	8.01	-1986	-40%
Berlin Twsp	7.92	3.31	101	7.25	7.49	-947	-35%
Brooklawn	6.85	1.63	-105	6.37	6.57	-228	-24%
Camden	10.58	4.98	-2,421	10.56	10.41	-5979	-15%
Cherry Hill	10.79	5.41	3,046	10.60	10.82	1750	4%
Chesilhurst	4.86	2.64	3	4.58	5.04	26	20%
Clementon	7.68	2.70	89	7.50	7.67	-23	-1%
Collingswood	8.04	3.40	-140	7.89	8.06	57	2%
Gibbsboro	7.77	2.00	-15	6.62	6.83	-1444	-61%
Gloucester Twsp	9.08	4.78	3,761	8.30	8.76	-2433	-28%
Gloucester	7.91	3.35	99	8.50	8.61	2780	102%
Haddon Twsp	8.39	3.61	42	8.17	8.33	-262	-6%
Haddonfield	8.52	3.61	38	8.51	8.64	611	12%
Haddon Heights	7.45	3.13	5	7.30	7.51	111	6%
Hi-Nella	4.54	.89	-15	4.65	4.99	54	57%
Laurel Spring	6.69	2.04	86	6.17	6.44	-182	-23%
Lawnside	7.63	2.38	-28	7.02	7.22	-697	-34%
Lindenwold	7.57	3.45	379	7.57	7.80	504	26%
Magnolia	6.39	2.62	112	6.54	6.81	307	51%
Merchantville	7.06	2.12	20	7.56	7.68	1002	86%
Mount Ephraim	7.06	2.57	-60	6.73	6.97	-99	-8%
Oaklyn	6.90	2.04	68	6.88	7.07	184	19%
Pennsauken	10.25	4.67	609	9.99	10.07	-4613	-16%
Pine Hill	6.45	3.04	485	5.90	6.28	-97	-15%
Pine Valley	4.56	-4.61	0	4.14	4.18	-30	-32%
Runnemede	7.47	3.22	158	7.45	7.66	370	21%
Somerdale	7.98	2.93	72	7.12	7.34	-1372	-47%
Stratford	8.14	3.08	107	7.64	7.83	-912	-27%
Tavistock	5.51	-4.61	6	4.80	4.77	-130	-52%
Voorhees	9.36	3.75	5,033	8.94	9.34	-248	-2%
Waterford	7.98	3.90	992	7.75	8.03	174	6%
Winslow	8.90	4.46	4,001	8.40	8.84	-438	-6%
Woodlynne	5.42	1.62	-60	5.29	5.61	47	21%

The diagram of the log prediction and the log forecast again demonstrates that the model does a very good job for forecasting the natural log of forecast-year municipal employment. But when these logarithmic values are converted into actual jobs, it can be seen that the regression equation produces a more modest result. The actual range of the forecast error is -40% to 102%. Most errors are in the order of 20% over the 10-year period. The forecasts for Camden County are positively skewed[17], leptokurtic and long-tailed. The model tends to underestimate the number of jobs in small municipalities and overestimate job figures for municipalities with greater base-year employment. (The correlation between actual 1990 jobs and the error percentage was -.20496.)

Findings and Recommendations

The results of the preceding analysis are:

1. For use in forecasting, the employment equation should be altered to include an additional dummy variable, whose purpose is to allow for the calibration of the forecast.

2. In calibration of the model, adjusting the forecast by less than ± 1% per year should be allowed without question. In general, municipalities with less employment than the forecasted total municipal average probably need to have their forecast reduced, while municipalities with more employment than average may need to have their forecast increased.

3. Adjustments larger than ±1% per year need to be documented and substantiated by the requesting agency. It is likely that municipalities with higher-than-average forecasts will need larger adjustments. Adjustments that change the forecast by more than 3% per year should be discouraged, except in the following cases:

a. Employment for the 10 municipalities identified as outliers in this study need to be separately calculated.
b. Municipalities forecast to experience substantial residential growth (> two times the mean state change in dwelling units).

[17] This finding agrees with data presented in Table 4-6.

V. Incorporating the Model into the PED Program

Previous Employment Assignment Characteristics

To begin this discussion it is useful to examine, in schematic fashion, the structure of the PED both as initially designed (Reilly, 1990) and as revised in 1993-1994 (Reilly, 1994). With respect to employment, the model required that the user select or enter a forecast of both population and employment for the forecast horizon year. The program then assigned a portion of this exogenous county or regional forecast to any municipality, within the county or region, based on the municipality's actual growth rate during a user-selected historic period. For example, if the job growth rate of a municipality for a selected period of time (say, 10 years) had been 13%, then a growth rate of 13% was assumed to continue in a linear fashion (forever). The sum of all the municipal projections produced in this manner then was scaled back to the exogenous county forecast using a shift-share, or proportionate, method.

These "raw," but scaled, municipal projections were then tested to ensure that sufficient land was available in each municipality to accommodate its growth assignment(Diagram 5-1). Assignments to municipalities with insufficient land were reassigned to other growing municipalities based on each municipality's relative growth rate. The fastest growers got more reassigned growth than the slow growers. Although the schematic diagrams do not show this, the re-assignment and refitting process was performed a total of four times in the program. During these subsequent fitting assignments, any growth re-assigned to a municipality was tested using the residual available land in the municipality. In both the second and third fitting cycles, reassignments were made proportionate to the historic period's growth rate. Again, faster growers got more reassignment than slow growers. In the fourth and last fitting cycle, any residual growth was assigned based on the municipal supply of available land. Municipalities with more land available got more of this residual assigned to them than did those with less land available.[18]

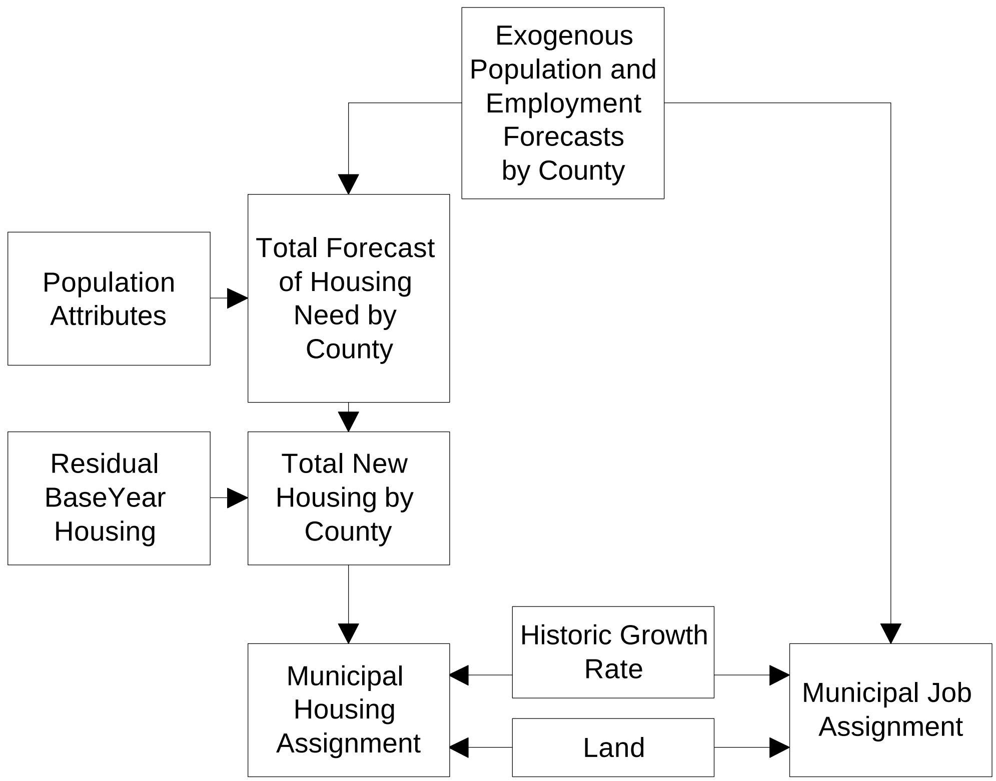
All of the equations and algorithmic structure embedded in this initial model (1990) were based on professional judgment. Clearly, many assumptions are incorporated into the program's code. For example, in this early program it was assumed that past growth rates would continue and that the best way to

---

[18] Testing the model found that most growth assignments were made to municipalities with sufficient land in the first fitting cycle. Further testing showed that very little actual growth was being assigned during the third and fourth assignment cycles, so the effect of assignment rules was not very significant, except in counties where much of the remaining available for development land was concentrated into a few municipalities. The primary role of these third and fourth assignment iterations was to insure that all of the exogenous growth was assigned. Interaction between the housing assignment programming and job assignment programming only occurred during the land testing cycles, when it was assumed that the same proportion of jobs and houses would be assigned as was evident in the base year.

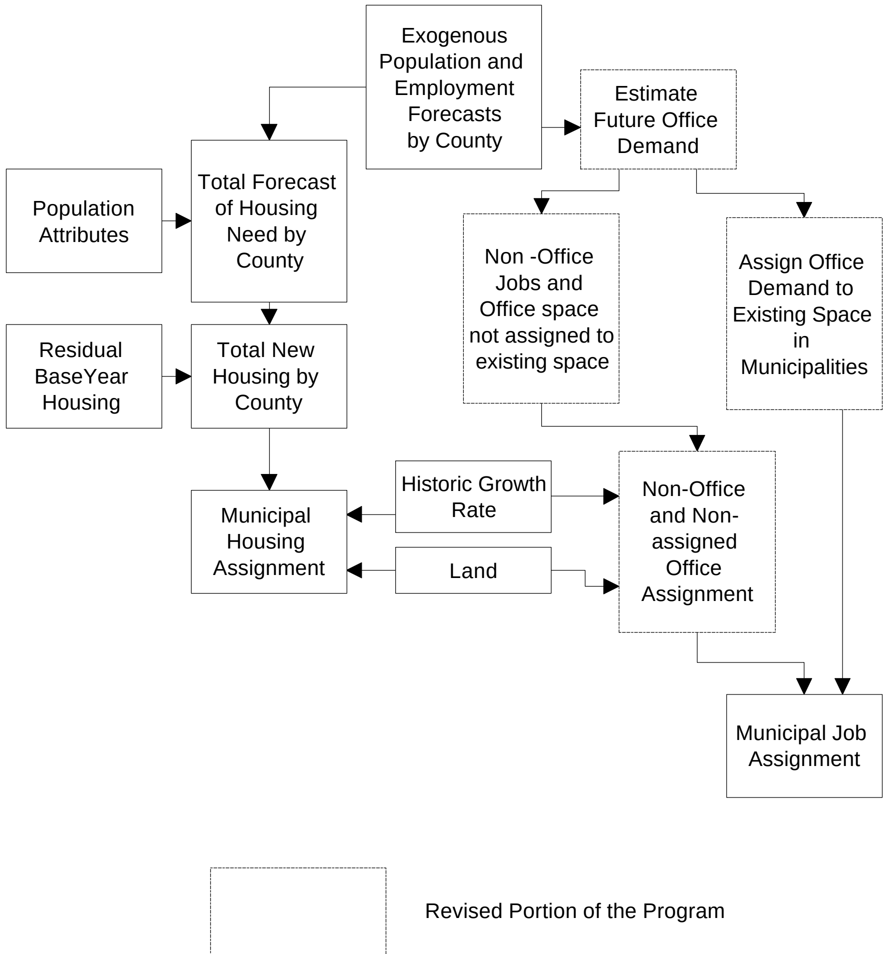
reconcile differences between municipal-based job estimates and exogenous regional forecasts was to scale the municipal forecasts up or down to conform to the regional forecast. While there were a lot of assumptions in the model, its strength was the use of sophisticated information on land availability to influence growth assignments.

Diagram 5-1
A Schematic of the Original Assignment Methodology



In 1994 the model's employment assignment algorithm underwent a major revision. Diagram 5-2 shows the revised structure for this part of the model.

Diagram 5-2
A Schematic of the Revised Assignment Methodology (1994)



The major revision to the employment assignment program was the addition of a subroutine that assigned some of the future estimated demand for office space to vacant residual existing office space[19]. This subroutine first used a statistical equation to estimate the demand for new office space in the county resulting from the exogenous employment forecast. This demand for office space was then assigned to municipalities where job growth had been recorded in the previous decade and where existing inventories of available office space (mostly Class A) were to be found[20]. Any residual demand for office space and all non-office job growth were then assigned using the method originally designed for the model.

The objective of this office assigning subroutine was to simulate more realistically the diffusion of jobs in the region and to improve the calculation of

---

[19] See Reilly, 1994.
[20] Ibid

the amount of land needed to accommodate job growth. But many of the "improvements" to the office space assignment methodology are intuitive or based on professional judgments. For example, while it might seem reasonable to assume that developers carefully site their buildings to ensure that they are built in locations attractive to the market -- the assumption behind assigning office growth to existing vacant office space -- there is no mathematical evidence that this is true. (In fact, the existence of a large supply of vacant office space in some places suggests that developers are quite fallible.)

Therefore, the employment assignment portion of the PED continued to exhibit the following characteristics:

1. The algorithms used to assign employment to municipalities had not been statistically validated.
2. Employment and population were exogenous. All municipal assignments were adjusted to ensure that these exogenous county forecasts were achieved;
3. Municipal estimates of employment and population were calculated independently, but 'fitted' (assigned to land in the municipality determined to be developable and available) in a sort of simultaneous fashion.

Should Employment be Endogenous or Exogenous?

In the previous chapters of this report, the following statistical model has been developed:

$$LNEMP_{t+10} = .907233 \times LNEMP_t + .0000654855 \times DDUS + .057217 \times LNMUNIM_{t+10} + .703274$$

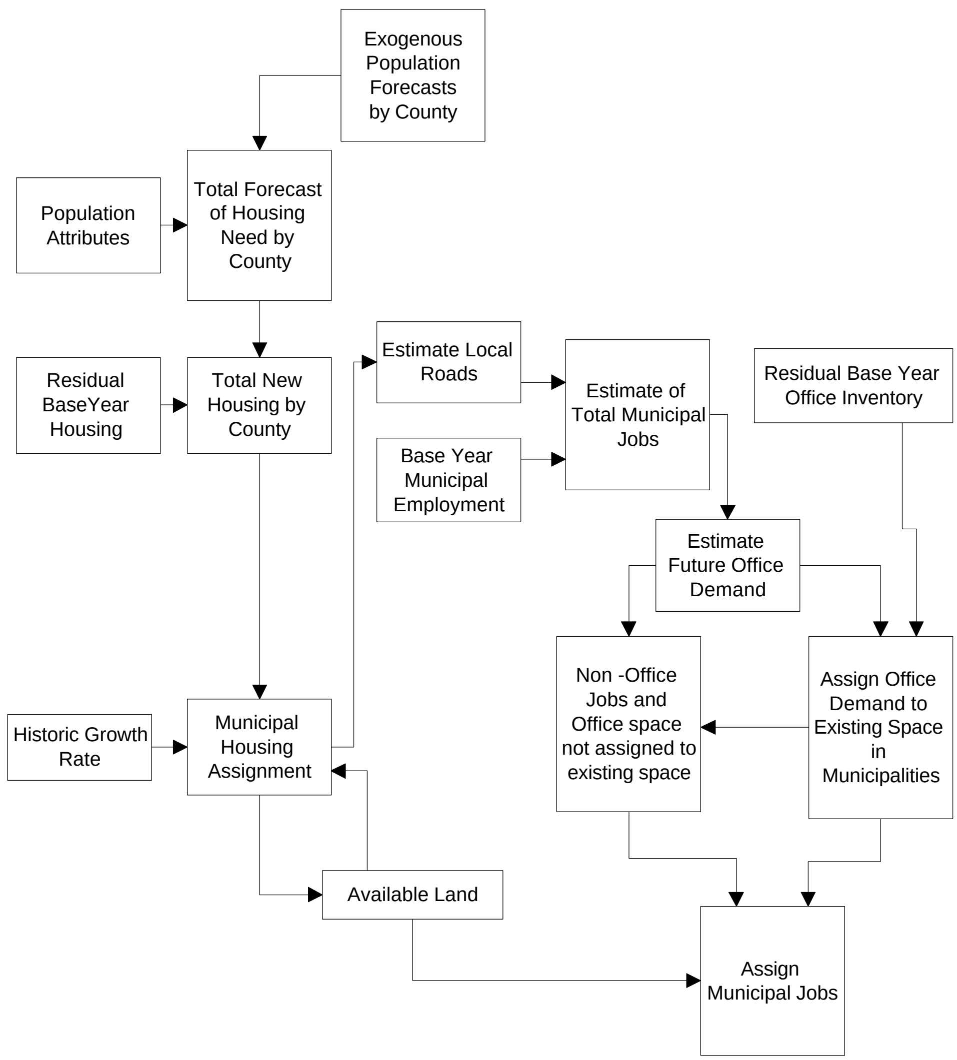
where:

$$DDUS = dwelling.units_{t+10} - dwelling.units_t$$
$$LNMUNIM = total.municipal.road.miles_{t+10}$$

This model produces close approximations of total future municipal employment. It begins to correct the problematic characteristic of the current and past versions of the PED program, in which the employment assignment algorithm lacked mathematical justification. An examination of this new regression equation's carriers demonstrates that the change in municipal dwelling units and local road mileage are the dynamic portions of the local employment model. This suggests that the equation is behavioral, in that employment changes are a result of growth or decline in both the number of municipal dwelling units and in the supply of municipal roads. The equation also suggests that the relationship between existing base-year employment and future employment is static. In other words, given two municipalities with the identical base-year employment, the regression equation tells us that any difference in their future employment is the result of changes (if any) in the number of dwelling units and their miles of local roads between the base year and the forecast year.

If this is true, then the schematic paradigm should be revised so that municipal dwelling units are assigned first, and this housing assignment then used to calculate total municipal employment. This algorithmic structure also conforms to Boarnet's (1994) finding that employment growth lags residential growth. (Other OSP research suggests that this lag, from 1980 to 1990, was on the order of five years. Had a lag of 10 or more years been discovered, than the use of a real lagged dwelling unit variables would have produced statistically significant findings and would have been used in the equation.) Diagram 5-3 shows how the PED program might be structurally revised to make endogenous the calculation of total municipal employment.

Diagram 5-3
A Schematic of the Assignment Methodology Making Employment Endogenous



In this revised schematic the total "fitted" number of municipal dwelling units is determined first. Then, this assignment is used to calculate local roads and to deduce the change in municipal dwelling units. Computation of these carrier values and the addition of base-year municipal employment allows the program to calculate total raw municipal employment, which can be aggregated into total raw county forecasts of employment. This county forecast of employment allows office space to be estimated and assigned to available vacant office space in municipalities. Non-office space is calculated by subtracting the total municipal office assignment from the total raw municipal forecast of employment produced by the employment regression equation. The resulting raw estimate of municipal employment that would need new buildings is then fitted to the municipal supply of residual (less that assigned for new housing) available developable land.

While this schematic produces an endogenous employment forecast, it also raised the following procedural questions.

1. How can housing be assigned to available developable land without having the "knowledge" of the underlying zoning? What is to prevent assigning all -- or too much -- available land in a municipality to housing, given that housing is now assigned before employment?

2. Can coefficients developed based on the period from 1980 to 1990 be used to forecast future employment, or must all or some of these coefficients be adjusted?

3. Given that the employment equation was developed using a 10-year interval, can the equation in its present form be used to calculate forecasts in excess of 10 years, or should be equation be used iteratively?

The Land Fitting Question

In the current (1994) version of the model, housing and employment are fitted to available developable municipal land in a manner that maintains each municipality's 1990 ratio of dwelling units to jobs. For example, if there were 100 jobs and 100 dwelling units in a municipality in 1990, then the model assigns the raw municipal estimate of growth in equal proportions of jobs and dwelling units until either the total raw growth estimate for both jobs and dwelling units is assigned or until the supply of available land is exhausted. The purpose of this approach is to ensure a balance of land uses (i.e., to avoid the problem of assigning all the land to either dwelling units or jobs). The shortcomings of this method are:

1. There was no statistical research to justify the use of the base-year ratio method.
2. The research in this paper demonstrates that employment change lags dwelling unit change. Therefore, the use of a historic proportionate assignment method is not supported.

3. Other research suggests that the use of a linear ratio to express the relationship between dwelling unit density and jobs may not be appropriate[21].

An alternative assignment method was developed using the land-use information derived from the 1986 integrated terrain unit (ITU) GIS maps produced by the N. J. Department of Environmental Protection (DEP). While use of a time series of data would produce a more suitable methodology, it is not yet available. This method therefore relies on a cross-sectional statistical examination of the relationship between 1990 dwelling unit density (number of 1990 dwelling units divided by the total municipal area) and the percentage of total municipal land designated as commercial, service and industrial in the 1986 ITU.

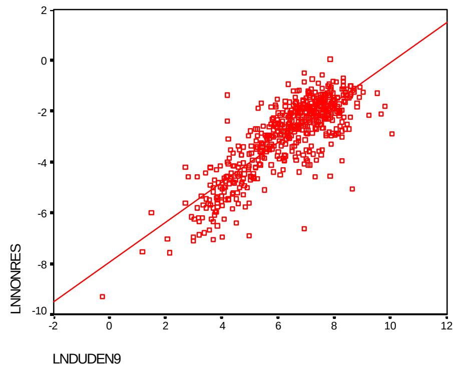
The ITU data set is based on photo interpreted aerials flown in 1986. The photo interpretation was performed by ESRI under contract to DEP. Field checks were performed in each county. The ARC INFO land use/land cover boundary files then were re-aggregated by OSP. This was done so that the ITU land-use coverage in several test quads matched the land-use mapping completed independently by the DVRPC for the counties in its jurisdiction, four of which are in New Jersey. DVRPC's mapping was accomplished by performing a ground checking field survey. This comparison assured that the OSP grouping of ITU land-use categories would conform to generally accepted land-use categories. It also confirmed the that ITU series was highly accurate in identifying and categorizing land uses.

Diagram 5-4 shows the scatterplot which results from this comparison of the percentage of total municipal land developed as job related (LNNONRES) and the municipal dwelling unit density (LNDUDEN9). The scatterplot demonstrates that a strong linear relationship exists between the natural log of municipal population density and the natural log of the percentage of total municipal area developed for job-related uses, excluding agriculture. Examination of the residuals of this linear regression demonstrated that it was relatively homoscedastic. In diagram 5-5 outliers (greater than three standard deviations are identified. Table 5-1 displays the identity of these municipalities, their FIPS codes, and their case numbers, which corresponds to the numbers in Diagram 5-5.

---

[21] See Reilly, 1996b.

Diagram 5-4
Scatterplot of LNNONRES and LNDUDEN9




```
Independent:  LNDUDEN9
  Dependent Mth   Rsq  d.f.       F  Sigf      b0      b1      b2      b3
   LNNONRES LIN  .674   554 1143.24  .000 -7.9301   .7818
   LNNONRES QUA  .702   553  651.96  .000 -10.428  1.7119  -.0798
   LNNONRES CUB  .714   552  458.37  .000 -8.4495   .3718   .1878  -.0164
```

Diagram 5-5
Scatterplot of Residuals (LNNONRES and LNDUDENS9)

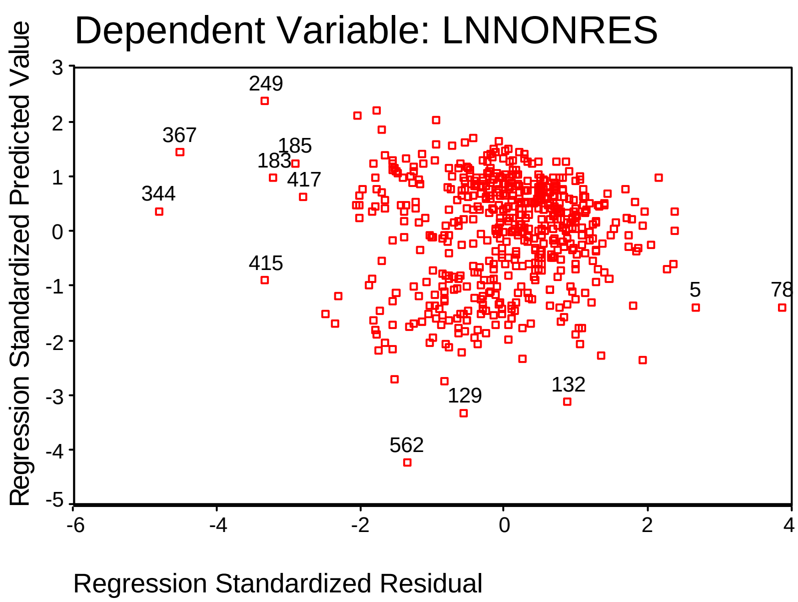


Table 5-1
Outliers in Diagram 5-5

Case Number	Municipal Name	FIPS Code
5	Buena Vista	34001025
78	Rockleigh Borough	34003275
129	Washington Twsp.	34005185
132	Woodland Twsp.	34005195
183	W. Wildwood	34009065
185	Wildwood Crest	34009075
249	Guttenburg	34017015
344	Interlaken	34025100
367	Shrewsbury	34025230
415	Barnegat Twsp.	34029003
417	Bay Head	34029010
562	Pahaquarry	34041090

Deleting the outliers improved the equation's goodness-of-fit value from .821 to .848 and improved the Adjusted R^2 from .673 to .718. The Durbin-Watson value, however, declined from 1.726 to 1.648. This mixed result suggests that deletion of the outliers does not produce sufficient improvement to support the use of the equation derived from the reduced data. Therefore, it is proposed that the following equation can be used to estimate the municipal acres that likely would be developed given a future number of dwelling units.

$$\text{\% of Total Municipal Area job related development}_{t+i} = .78818 \times \ln dudens_{t+i} - 7.9301$$

where:

dudens $_{t+i}$ = total municipal dwelling units at time t + i

Using this equation also simplifies the initial land-fit testing process because it automatically adjusts job-related development density.

Because of the uncertainty of the statistical result, several internal checks should be performed prior to reserving some or all of the available land for job-related development. First, the forecasted total of job-related land should be compared to the base-year (1986) amount of job-related land. If the forecast amount is less than the base-year amount, given a forecast of housing growth, then the base-year dwelling unit and job data should be used to estimate the base-year amount of land developed for job-related land uses. The forecast-year estimate of job-related land should then be subtracted from the base- year estimate to produce a "growth induced" estimate. The growth-induced estimate then should be added to the actual 1986 acreage of job-related land uses to produce the forecast -year demand for job-related land.

Determining the Appropriate Coefficient

Earlier in this chapter, it was proposed that the dynamic coefficients describe growth-induced change, while the coefficient for the static variable defines a fundamental relationship between the number of base-year jobs and the number of future-year (10 year period) jobs. But analysis of historic data and long-term forecasts prepared by other organizations (WEFA, 1992; Urbanomics, 1994) both demonstrate that the coefficient of base-year employment should increase over time. This finding is displayed in Table 5-2, which shows that the ratio of jobs to dwelling units has been increasing and is expected to continue to increase through 2018. (That year marks the forecast horizon, not necessarily the end of the trend to increase).

Table 5-2
Ratio of Jobs to Dwelling Units

	dwelling units	jobs	jobs to du ratio
1960	1992641	1942050	0.974611
1970	2304576	2502423	1.085850
1980	2725057	2966560	1.088623
1990	3075310	3506582	1.140237
2000	2950120	3782960	1.282307
2010	3068960	4197520	1.367734
2018	3142920	4555760	1.449531

As shown in Table 5-2, the number of jobs per dwelling unit has been increasing in New Jersey each decade since 1960. (That year marks the beginning of the data set, but not necessarily the beginning of the trend.) Household size has been declining during the same period, which means that more dwelling units were needed in 1990 than in 1960 to house a given number of people. Both the historic data and the forecasts[22] in the table demonstrate that the use of coefficients based on the 1980 to 1990 time series would underestimate the total number of municipal jobs in the future. Therefore, the equation must be adjusted before it can be used to produce reliable forecasts of future-year municipal employment. The question is, how should the adjustment be made?

Three options present themselves. First, the regression equation could be used to forecast municipal employment, which then would have to be scaled up or down to conform to an exogenous forecast of total employment for the county or region. Second, the exogenous forecast could be used to adjust the equation's prediction using a constant factor. This factor could be developed using information such as that shown in Table 5-2. Third, one could uniformly adjust upward the coefficient for base-year employment . (One could easily argue that this option is very little different that option two.)

Both the use of a proportionate scaling method and the use of a method to adjust to the equation's prediction presume that the dynamic coefficients would not change significantly. That is, the number of jobs would increase

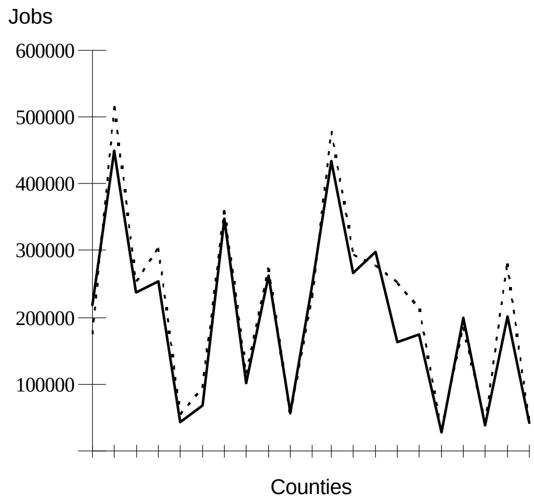
[22] WEFA Long Term New Jersey Forecast Model.

uniformly in the municipalities and would not represent a redistribution of jobs among municipalities. The argument in favor of this position is that locational changes in employment have been caused by changes in the mode of goods shipment, changes in inventory storage and goods handling, structural changes in the type of employment and the diffusion of higher-income households to the suburban and rural part of the state. In effect, one is looking at multiple instances of technology diffusion.

None of these technological changes, however, are of recent vintage. Most interstate highways were built in the 1950's. The DOT does not foresee the construction of a major new roadway network in New Jersey. Even the transition from a predominately manufacturing-based economy to an economy now geared to employment growth in retail and services is not recent. Most of these changes have occurred since the end of World War II, and some argue that they have been occurring since the second quarter of this century. The study of various technology transfers (Banks, 1994) suggests that the adoption of new technologies can be described by the use of logistic relationship. That is, initial acceptance of a new technology is low, but more and more firms adopt it as its advantages become widely understood. Finally, all but a few firms either go out of business or change to the new technology. By using the 1980 to 1990 coefficients for the dynamic carriers, it is assumed that most of the adaptation to these technologies has occurred. That is, it is assumed that most businesses have adopted or adapted to the new technologies prior to 1980.

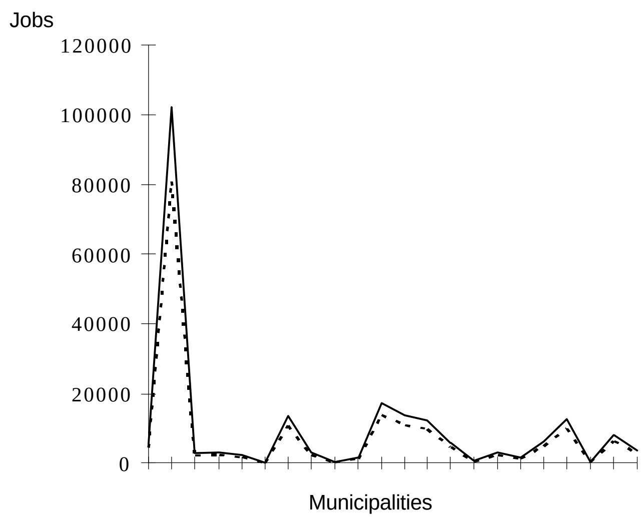
An exogenous forecast and a job-scaling factor have both been incorporated into the PED. But the effect of these controls is relatively small, as displayed in the following diagrams. Diagram 5-6 displays (in the dotted line) the total employment forecast produced by using the employment equation (and the complimentary exogenous population forecast) with the job-scaling factor for each New Jersey county, while the solid line shows the exogenous employment forecast. In most counties the model produces a forecast which is fairly consistent to the exogenous forecast.

Diagram 5-6
Comparison of Model and Exogenous Forecasts



In Diagram 5-7, the effect of revising (proportionate method) model municipal forecasts to conform (produce the same total) to the exogenous county forecast is shown. In this diagram, Atlantic County municipalities are displayed. The solid line represents the revised municipal forecasts and the dotted line presents the actual municipal forecasts produced by the adjusted equation.

Diagram 5-7
Comparison of Adjusted and Controlled Municipal Forecasts



There are two findings to this analysis. First, the municipal employment regression equation can be used to produce an endogenous employment forecast, given model-produced forecasts of the horizon-year dwelling unit density and municipal roads. This forecast, however, will likely be lower than expected because it uses coefficients based on 1980 to 1990 data. Use of either of the methods identified in this section to adjust the municipal forecasts to produce the expected number of future jobs requires the direct or indirect use of an exogenous forecast. So while the computer program can produce an endogenous forecast of employment based primarily on an exogenous population forecast, it still needs to adjust this forecast using an exogenous employment forecast. Second, the municipal employment assignments produced by the regression equation will not be able to simulate shifts in locational preference by major employment group. This fact suggests that some judgment should be exercised in reviewing trend growth forecasts and that calibration of these forecasts is warranted.

Determining the Form of the Equation

This section considers if the regression equation can best be used “as is” to forecast periods in excess of 10 years, or if the equation must be iterated through 10-year cycles to forecast employment more than 10 years into the future.

Both approaches have their flaws. The coefficients developed in the regression equation reflect change during the 10-year period of the analysis. Both methods assume that these coefficients remains constant with respect to the carriers. In fact, there is no evidence to support or to contradict this assumption. The only real difference between the methods is that the iterative one performs compounding and includes this added value into the calculation of the final horizon-year estimate. While this compounding may appear to be a more reasonable application of the equation, iteration is known to produce chaotic results. Therefore, the use of a simple regression equation to calculate all future horizons was adopted in this version of the model.

Bibliography

Banks, Robert B. (1994) Growth and Diffusion Phenomena. Heidelberg: Springer-Verlag.

Boarnet, Marlon G. (1994). *The Monocentric Model and Employment Location*. Journal of Urban Economics, Vol. 36, pp. 79-97.

Isserman, Andrew M. (1984). *Projection, Forecast, and Plan: On the future of Population Forecasting*. Journal of the American Statistical Association, Vol. 50, No. 2, pp. 208-221.

Putman, S.H (1983). Integrated Urban Models. London: Pion Limited.

Reilly, James (1990). Distributing Population and Employment Forecasts to Municipalities. Trenton: New Jersey Office of State Planning.

Reilly, James (1994). Modifications to the PED Model. Trenton: New Jersey Office of State Planning.

Reilly, James (1996b). The Relationship Between Current Expenditures, Valuation and Density. Trenton: New Jersey Office of State Planning. (forthcoming)

Skokowski, Barry (1981) Forty-Third Annual Report of the Division of Local Government Service 1980 - Statements of Financial Condition of Counties and Municipalities. Trenton: Department of Community Affairs.

Skokowski, Barry (1991) Fifty-Third Annual Report of the Division of Local Government Service 1990 - Statements of Financial Condition of Counties and Municipalities. Trenton: Department of Community Affairs.

Smith, S.K. (1987). *Tests of forecast Accuracy and Bias for County Population Projections*. Journal of the American Statistical Association, Vol. 82, No. 400, pp. 991-1003.

Smith, S.K. and Sincich, T. (1988*). Stability Over Time in the Distribution of Population Errors*. Demography, Vol. 25, No. 3, pp. 461-474.

Smith, S.K. and Sincich, T. (1990). *The Relationship Between the Length of the Base Period and Population Forecast Error*. Journal of the American Statistical Association, Vol. 85, No. 410, pp. 376-375.

Smith, S.K. and Sincich, T. (1991). *An Empirical Analysis of the Effect of Length of the Forecast Horizon on Population Forecast Errors*. Demography, Vol. 28, No. 2, pp. 261-274.

Urbanomics (1994) Long Term Population and Employment Forecasts for New Jersey Counties. New York: Urbanomics.

Wharton Econometric and Forecasting Associates (1992). WEFA Regional Forecast of New Jersey. Philadelphia: WEFA

Appendix A

Information Pertaining to ES 202 Data