
Land Use Model

A significant part of the LUTRAQ project has been the integration of location and land use forecasting procedures with the transportation modeling procedures currently in use for the region. To accomplish this the project team recommended making use of the Employment Allocation Model (EMPAL) and the Disaggregated Residential Allocation Model (DRAM), developed by S.H. Putman Associates. These models were integrated into the Metro transportation modeling process, and a series of analyses were completed in 1992. Though successful, the results revealed the need for further analyses. This revised version of *Volume 4: Model Modifications* describes the results of complete new analyses, including tests of revised and extended versions of the models.

n EMPAL Calibration

The EMPAL model of employment location has been successfully calibrated for more than forty metropolitan regions. The largest region is the Southern California Association of Governments - Los Angeles region, with 1980 total employment just under 5.5 million, and a land area of more than 87 million acres. The smallest is the Pikes Peak Area Council of Governments - Colorado Springs region, with total employment just over 200 thousand in 1990. The number of zones, or subareas, used in previous calibration efforts range from 772 for Los Angeles, to the 100 used here for some of the Portland, Oregon analyses.

The quality, or success, of EMPAL calibrations is directly related to the quality of the employment data that is used as an input to the calibration process. Errors or inconsistencies in the employment data can result in calibrations with poor goodness-of-fit and unreliable parameter estimates. One of the purposes of this study is to examine the sensitivity of outcomes in the transportation and land use modelling process to the data inputs for the components of the modelling process. In order to evaluate the sensitivity of the modelling process to the quality of employment data, EMPAL calibrations were examined for three different urban area data sets.

The relationships between data quality and the accuracy of EMPAL calibrations are illustrated here by examining actual data sets and calibration results for three regions: Detroit, Kansas City and Portland. These three regions were chosen for comparison because they have comparable zone systems, but have different amounts of employment and employment densities. These three regions are also examined in the next part of this report, there in terms of the calibration of the DRAM model of residential location.

For Detroit and Kansas City, excellent calibration results were obtained for all employment categories. For Portland, good calibration results were obtained for most employment categories. The inclusion of additional *multiplicative* lagged employment variables in the attractiveness function of the EMPAL model equations improved the calibration results for Portland. Improved calibration results were also obtained for Portland when the zone system was spatially aggregated. The primary reason for the differences between the calibration results for Portland and those for Detroit and Kansas City seems to be the quality of the employment data inputs to the calibration process. Before describing the calibration results for these three regions, it is useful to specify the mathemat-

ical structure of EMPAL, and to characterize the attributes of a successful calibration.

The Mathematical Formulation of EMPAL

EMPAL is a modified version of the standard singly-constrained spatial interaction model. There are three modifications: 1) a multivariate, multiparametric attractiveness function is used; 2) a separate, weighted, lagged variable is included outside the spatial interaction formulation; and 3) a constraint procedure is included in the model, allowing zone and/or sector specific constraints. The model is normally used for 4-8 employment sectors, whose parameters are individually estimated. The development of the model is described in more detail in Putman, S. *Integrated Urban Models* (1983, Pion Limited, London). The equation structure of EMPAL is:

$$E_{j,t}^k = \lambda^k \sum_i P_{i,t-1} A_{i,t-1}^k W_{j,t-1}^k c_{i,j,t}^{\alpha^k} \exp(\beta^k c_{i,j,t}) + (1 - \lambda^k) E_{j,t-1}^k \quad (1)$$

where

$$W_{j,t-1}^k = (E_{j,t-1}^k)^{a^k} L_j^{b^k} \quad (2)$$

and

$$A_{j,t-1}^k = \left[\sum_i (E_{i,t-1}^k)^{a^k} L_i^{b^k} c_{i,j,t}^{\alpha^k} \exp(\beta^k c_{i,j,t}) \right]^{-1} \quad (3)$$

where

$E_{j,t-1}^k$ = employment (place-of-work) of type k in zone j at time t-1,

$E_{j,t}^k$ = employment (place-of-work) of type k in zone j at time t,

L_j = total area of zone j,

$c_{i,j,t}$ = impedance (travel time or cost) between zones i and j at time t,

$P_{i,t-1}$ = total number of households in zone i at time t-1, and

$\lambda^k, \alpha^k, \beta^k, a^k, b^k$ = empirically derived parameters.

Note that it is possible to include other variables in the attractiveness function, in addition to L_j , as it appears in Equation (2).

Evaluating the Calibration Procedure¹

The goal of the calibration procedure is to develop estimates of the parameters of the EMPAL model equations which best fit the general model structure to a specific data set. The success of the calibration procedure can be evaluated by measuring the goodness-of-fit of the model equations to observed employment location, and by examining the signs, magnitudes, and statistical significance of the estimated parameters.

The *likelihood criterion*, ϕ , which is a nonlinear function of the observed numbers and estimated numbers of employees in each zone, is used to evaluate the goodness-of-fit of the EMPAL model. It is analogous to the R^2 statistic used to measure goodness-of-fit in multiple regression analysis. The value of the likelihood criterion is always between zero and one, with higher values representing better goodness-of-fit. For a “perfect” calibration (i.e., the estimated numbers of employees are exactly equal to the observed numbers of employees in every zone), the likelihood criterion would be equal to one. However, it should be noted that it is *never* possible to perfectly fit a model to observed socioeconomic data.

While the likelihood criterion is used to measure the goodness-of-fit of the EMPAL model equations to observed employment location for an **entire** region, *MAPE* (*mean absolute percentage error*) statistics are used to summarize errors in model fit for **individual zones**. For EMPAL, the MAPE statistics are defined as:

$$MAPE^k = \frac{100}{N} \left[\sum_j \frac{|\hat{E}_{j,t}^k - E_{j,t}^k|}{E_{j,t}^k} \right] \quad (4)$$

where

$\hat{E}_{j,t}^k$ = EMPAL estimate of type k employment in zone j,

$E_{j,t}^k$ = observed type k employment in zone j, and

N = number of zones.

For a “perfect” calibration, the MAPE statistic would be equal to zero.

It is often useful to calculate MAPE statistics for a subset of zones. In general, estimation errors will be larger for zones with few employees, and smaller for zones with larger numbers of employees. For all of the calibration results in this report, three MAPE statistics are reported. The first MAPE statistic is defined for all of the zones in a region, the second MAPE statistic is for the smallest zones in the region, and the third MAPE statistic is for the largest zones in the region.

Asymptotic t-tests are used to evaluate the statistical significance of the estimated EMPAL param-

¹ The calibration procedure is described in more detail in the next section of this chapter.

ters. Asymptotic t-tests are similar to standard t-tests, and can be interpreted in a similar manner, although their effectiveness cannot be guaranteed. Smaller values of the asymptotic t-tests indicate that a parameter is not statistically significant. Larger values of the asymptotic t-tests **may** indicate that a parameter is statistically significant, but do not guarantee significance. Asymptotic t-tests approach the effectiveness of the standard t-test as the sample size becomes large. For most of the data sets being used in EMPAL model calibrations, the numbers of observation are large enough to assume that the asymptotic tests are equivalent to the standard tests.

For a successful calibration of the EMPAL model equations, based on results for other metropolitan regions, the likelihood criterion should exceed 0.80 for each employment type, and the asymptotic t-tests should exceed 20.0 for a majority of the estimated parameters. If the likelihood criterion is less than 0.80, then a significant proportion of the variation in employment location cannot be explained by the EMPAL model equations. In most cases, low values of the likelihood criterion suggest that there are errors in the employment data used as an input to the calibration process. A low value of a specific asymptotic t-test indicates that the variable associated with that statistic is, given other variables in the EMPAL equations, uncorrelated with employment location. This absence of correlation may reflect the actual location patterns of employment, multicollinearity among the variables in the model equations, or may be due to errors in the measurement of the variable.

Successful EMPAL calibration procedures should produce parameter estimates that are similar to previous estimates for other regions. We expect the values of the estimated EMPAL model parameters to conform to our theoretical and empirical knowledge of employment location. The value of one or both of the travel time parameters, α or β , should be negative, since employers tend to locate in areas that are accessible to households and other employers. For employment categories with positive agglomeration economies, the employment parameter, a , should be positive. However, this parameter may be negative if employment is dispersed (e.g., retail employment), or if there is strong collinearity between employment and land area. The sign of the land area parameter, b , may be positive or negative. This parameter is especially sensitive to the zone system used to divide a region. If a region has extremely large (in terms of total acreage) rural zones, then it may be difficult to accurately estimate the effect of zone size upon employment location.

Unlike the standard DRAM model, EMPAL has an additive, lagged variable that captures the influence of past employment location upon future employment location. The weighting parameter ($1-\lambda$) measures the magnitude of this effect. For industries with large relocation costs or positive agglomeration economies, the lagged employment parameter *may* be close to one. For industries that benefit less from agglomeration and have little fixed capital, we *might* expect the lagged employment variable to be smaller. In many cases, however, the value of the lagged employment parameter is skewed by errors in the lagged employment data and does not conform to *a priori* expectations. If the estimated lagged employment parameter is either zero or one, then it is likely that there are errors in the lagged employment data, or that the lagged employment is a scalar transform of the current time period employment, as is sometimes the case when simple procedures are used to interpolate the lagged variable.

If the estimated lagged employment parameter is either zero or one, then it is unlikely that EMPAL will produce accurate forecasts of employment location. If there is no weight on lagged employment (i.e., $1-\lambda = 0$), then forecast employment location will be independent of past employment location. If the weight on lagged employment is equal to one, then the EMPAL model collapses to

a linear extrapolation model, and the level of employment in all zones will be equal to a fixed multiple of past employment levels.

The EMPAL Data Sets for Detroit, Kansas City and Portland

The Detroit, Kansas City, and Portland regions are quite different in terms of total employment, employment growth, and employment densities.

Table 4. General Employment Data

Region	Total Employment (Base Year)	Employment Growth between Base Year and Lag Year	Average Employment Density (Employees per Acre)
Detroit	2,172,998	11.6%	6.37
Kansas	644,991	7.8%	3.56
Portland	707,456	15.2%	4.72

For each of these regions, the employment categories used for the EMPAL calibrations are defined as aggregations of Standard Industrial Classification (SIC) categories. The Detroit employment data set was aggregated into eight categories:

- 1) AgMinCon: Agriculture, Forestry, Fishing, Mining, and Construction
- 2) Manufacturing
- 3) TCU: Transportation, Communication, and Utilities
- 4) Wholesale
- 5) Retail
- 6) FIRE: Finance, Insurance, and Real Estate
- 7) Services
- 8) Public

The Portland and Kansas City employment data sets were aggregated into four categories. For Kansas City, these categories are:

- 1) Industrial: Agriculture, Forestry, Fishing, Mining, Construction, TCU, and Wholesale
- 2) Manufacturing
- 3) Services: Services, FIRE, and Government
- 4) Retail

while for Portland the categories are:

- 1) Industrial: Agriculture, Forestry, Fishing, Mining, Construction, Manufacturing, and TCU
- 2) Whl/Ret: Wholesale and Retail
- 3) Office: FIRE and Government
- 4) Services

Usually, employment is aggregated so that employment categories correspond to land use definitions used within the region. Aggregation of SIC categories is also necessary because of the uneven distribution of employment across a metropolitan region.² For example, if fishery employment was defined as an exclusive employment category, it is likely that most fishery employment would be concentrated in a very small number of zones. Given this uneven distribution of employment, it would be very difficult to obtain good calibration results for fishery employment. For this reason, aggregation of employment types is often necessary to overcome unevenness in the employment data. The problems created by an uneven distribution of employment are compounded if the region being examined is divided into a large number of small zones, as is the case for the Detroit (297 zones), Kansas City (377 zones), and Portland (328 zones) data sets.

Calibration Results for the Standard EMPAL Model

The calibration results for Detroit and Kansas City were quite similar. For Detroit (Table 5), excellent calibration results were obtained for all eight employment types. The likelihood criterion measures are all greater than 0.90, and many of the measures exceed 0.95. Only six of the forty parameters have asymptotic t-tests less than 20. Moreover, the signs and magnitudes of all of the EMPAL model parameters conform to our prior expectations. Notice that, as has been true for all of the dozens of EMPAL calibrations done for metropolitan regions across the U.S., none of the lagged employment weighting parameters are equal to zero or one.

² For the Detroit data set, one-half of total employment is concentrated within the 52 largest zones. For Kansas City and Portland, the concentration of employment is even greater: 50% of the employment in Portland is located in the 38 largest zones, while 50% of the employment in Kansas City is located in the 43 largest zones.

Table 5. EMPAL Parameters for the Detroit Region¹
(297-Zone System)

	AgMinCon	Manufact'g	TCU	Wholesale	Retail	FIRE	Service	Public
Alpha - α	0.5394 (18.0)	-0.7477 (74.8)	-0.8070 (29.7)	-0.7989 (19.4)	-0.6570 (49.4)	-0.7920 (26.3)	-0.6638 (54.4)	-0.6659 (25.5)
Beta - β	-0.0878 (77.6)	0.0062 (23.9)	-0.0014 (1.8)	-0.0155 (11.0)	0.0029 (8.6)	-0.0116 (11.8)	-0.0189 (45.7)	0.0181 (29.7)
Employment	0.8542 (138.5)	0.8001 (378.0)	0.8264 (161.2)	1.1811 (205.7)	1.0138 (309.3)	1.1451 (272.1)	1.1041 (379.6)	1.1046 (143.9)
Total Acres	0.6818 (109.8)	0.4077 (152.4)	0.3980 (67.3)	0.6083 (92.3)	0.4980 (143.3)	0.5035 (107.5)	0.5784 (244.5)	0.5150 (77.5)
Lagged Emp. (1-λ)	0.6132 (53.8)	0.7877 (68.5)	0.8161 (29.8)	0.6566 (48.8)	0.7530 (63.0)	0.6677 (65.1)	0.7580 (103.5)	0.8912 (22.4)
R²	0.9089	0.9347	0.9637	0.9569	0.9495	0.9644	0.9670	0.9924
Likelihood Criterion	0.9177	0.9616	0.9538	0.9660	0.9650	0.9706	0.9780	0.9905
MAPE	27.05%	29.39%	22.79%	26.11%	17.39%	19.01%	14.40%	11.08%
MAPE (Smallest)	52.34%	74.16%	45.84%	40.40%	24.51%	29.46%	23.04%	14.40%
MAPE (Largest)	18.21%	14.33%	14.07%	16.18%	11.74%	14.86%	12.15%	9.45%

¹ Asymptotic t-test shown in parentheses.

For Kansas City (Table 6), good calibration results were obtained for all four employment types. The likelihood criterion measures are all greater than 0.80, and only one parameter has an asymptotic t-test less than 20. The signs and magnitudes of all of the estimated parameters are as expected, and none of the lagged employment variables is equal to one or zero.

The most significant difference between the results for Detroit and Kansas City can be seen in the MAPE statistics. The MAPE statistics for Kansas City are significantly larger than the MAPE statistics for Detroit, indicating that prediction error at the zone level is greater for the Kansas City data set. There are two reasons for this difference: 1) the overall goodness-of-fit for the Detroit calibrations is better than for the Kansas City calibrations, and 2) Kansas City has many zones with only a few employees. A large number of small zones (in terms of numbers of employees) can “magnify” the MAPE statistics, since, on average, the percentage errors for small zones will be higher than the percentage errors for large zones. This is apparent if you compare the MAPE statistics for the smallest zones with the MAPE statistics for the largest zones.

Table 6. EMPAL Parameters for the Kansas City Region¹
(377-Zone System)

	Industrial	Manufacturing	Services	Retail
Alpha - α	-0.9399 (40.5)	-.9107 (29.3)	-0.8818 (32.0)	-0.9354 (8.5)
Beta - β	-0.0242 (26.7)	-0.0503 (36.1)	-0.0770 (53.2)	-0.6180 (50.1)
Employment	0.5985 (224.5)	0.1731 (108.1)	0.6021 (222.1)	0.1813 (39.1)
Total Acres	0.5157 (166.3)	0.5050 (135.3)	0.4883 (164.5)	0.4493 (31.2)
Lagged Emp. (1 - λ)	0.7091 (79.5)	0.9322 (50.1)	0.8049 (79.5)	0.8658 (50.8)
R²	0.9074	0.9416	0.9654	0.8365
Likelihood Criterion	0.8969	0.9364	0.9466	0.8640
MAPE	108.34%	162.97%	39.01%	70.45%
MAPE (Smallest)	311.68%	716.67%	78.00%	173.48%
MAPE (Largest)	25.93%	33.72%	16.90%	21.26%

¹ Asymptotic t-test shown in parentheses.

For Portland (Table 7), mixed calibration results were obtained. Only one employment type has a likelihood criterion measure greater than 0.80. (The remaining likelihood criterion measures are close to 0.80.) Most of the estimated parameters have asymptotic t-tests that are greater than zero and reasonable signs and magnitudes. However, for Industrial employment, the estimated weighting parameter for lagged employment is zero, which seems unlikely for this category of employment. The MAPE statistics for all four employment types are significantly larger than those for Detroit, and larger than or equivalent to those for Kansas City.

The primary reason for the mixed calibration results for the Portland region are inconsistencies between the base year and lag year employment data.

Table 7. EMPAL Parameters for the Portland Region¹
(328-Zone System)

	Industrial	Wholesale/ Retail	Office	Services
Alpha - α	0.5869 (42.4)	0.0939 (3.4)	0.5179 (16.9)	0.2693 (7.2)
Beta - β	-0.0614 (116.5)	-0.0599 (58.6)	-0.1805 (61.5)	-0.1170 (40.8)
Employment	0.8198 (508.0)	0.8530 (364.2)	-0.1323 (60.7)	0.0389 (15.7)
Total Acres	0.1601 (107.8)	0.2266 (115.7)	0.0699 (18.8)	0.0489 (13.8)
Lagged Emp. (1 - λ)	0.0000 (123.5)	0.1717 (81.2)	0.8835 (72.1)	0.8546 (77.5)
R²	0.7817	0.7336	0.6847	0.8533
Likelihood Criterion	0.7800	0.7601	0.7218	0.8043
MAPE	147.74%	162.43%	364.92%	115.62%
MAPE (Smallest)	376.77%	524.43%	1327.56%	320.87%
MAPE (Largest)	40.52%	33.04%	52.81%	37.61%

¹ Asymptotic t-test shown in parentheses.

Inconsistencies between the Base Year and Lag Year Employment Data

Since lagged employment appears in *two* places in the EMPAL model equations (see Equations (1) and (2)), it is extremely important that this variable be measured accurately. The degree of correlation between the base year and lag year employment data is an important determinant of goodness-of-fit for the EMPAL model. The cause of the mixed calibration results for Portland becomes apparent if we examine this correlation for each of the three data sets.

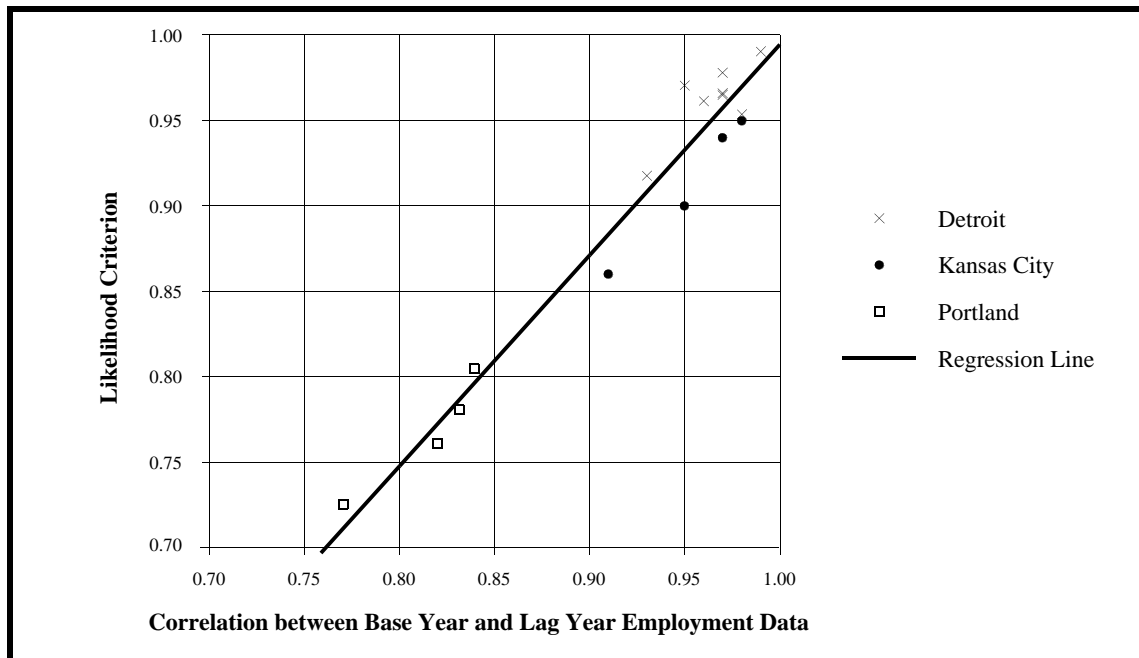
Table 8. Pearson Correlations between Base Year and Lag Year Employment

	Detroit	Kansas City	Portland
Manufacturing	0.93	0.95	0.83
AgMinCon	0.96	0.97	
TCU	0.98		
Wholesale	0.97	0.98	0.82
Retail	0.97		
Services	0.95	0.91	0.77
FIRE	0.97		
Government	0.99		

As Table 8 shows, the correlations between the base year and lag year employment data are significantly lower for Portland than for Detroit and Kansas City.

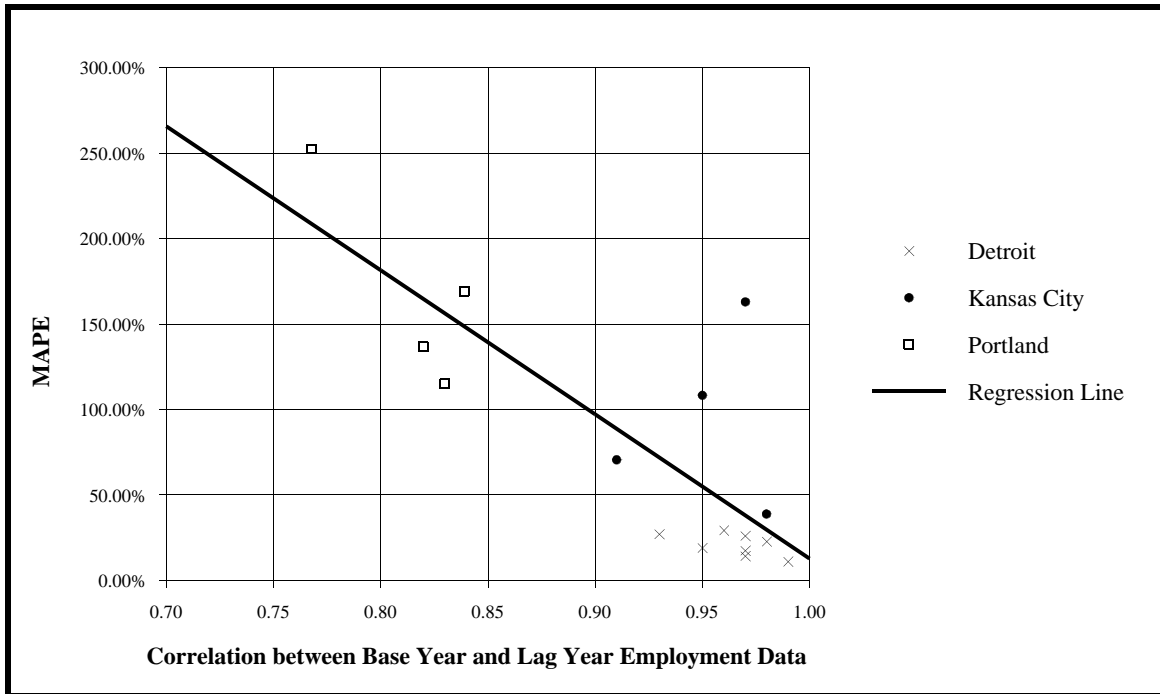
If we plot the likelihood criterion for each employment type and region against the correlations shown in Table 8, one can see a strong positive relationship between goodness-of-fit and consistency in the base year and lag year employment data. See Figure 2.

Figure 2. Likelihood Criterion vs. Base/Lag Years Employment Data



This relationship between goodness-of-fit and consistency in the employment data also holds for the MAPE statistics, which are negatively related to correlation between the base year and lag year employment data. See Figure 3.

Figure 3. MAPE vs. Base /Lag Years Employment Data

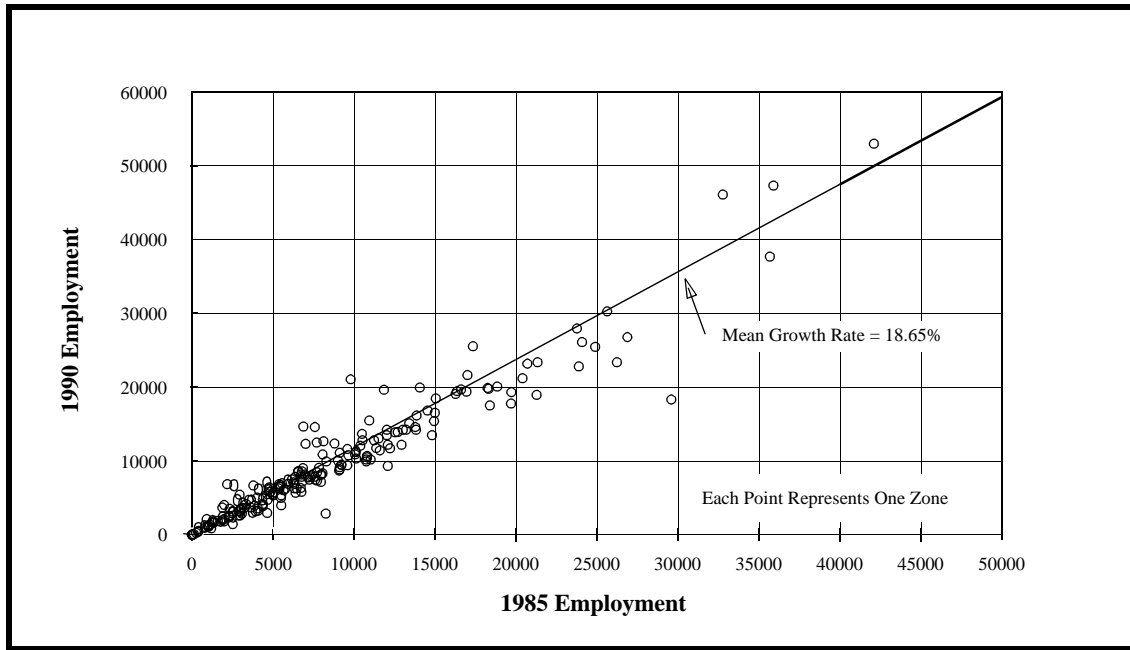


The relatively low correlation between the base year and lag year employment data for Portland is partly a result of rapid employment growth from 1985 (lag year) to 1990 (base year). However, the principal cause of this low correlation is a set of inconsistencies in the 1985 employment data.

If we compare the 1990 and 1985 employment data for Portland, there are 169 zones (51.5% of the total number of zones) in which total employment increased or decreased by more than 40%. While it is possible for employment in a single zone to change by that amount over a five year period, it is unlikely that so many zones would undergo such rapid change. In Kansas City, 83 zones (22.0%) grew or declined by more than 40%, while in Detroit 35 zones (11.8%) changed by more than 40%.

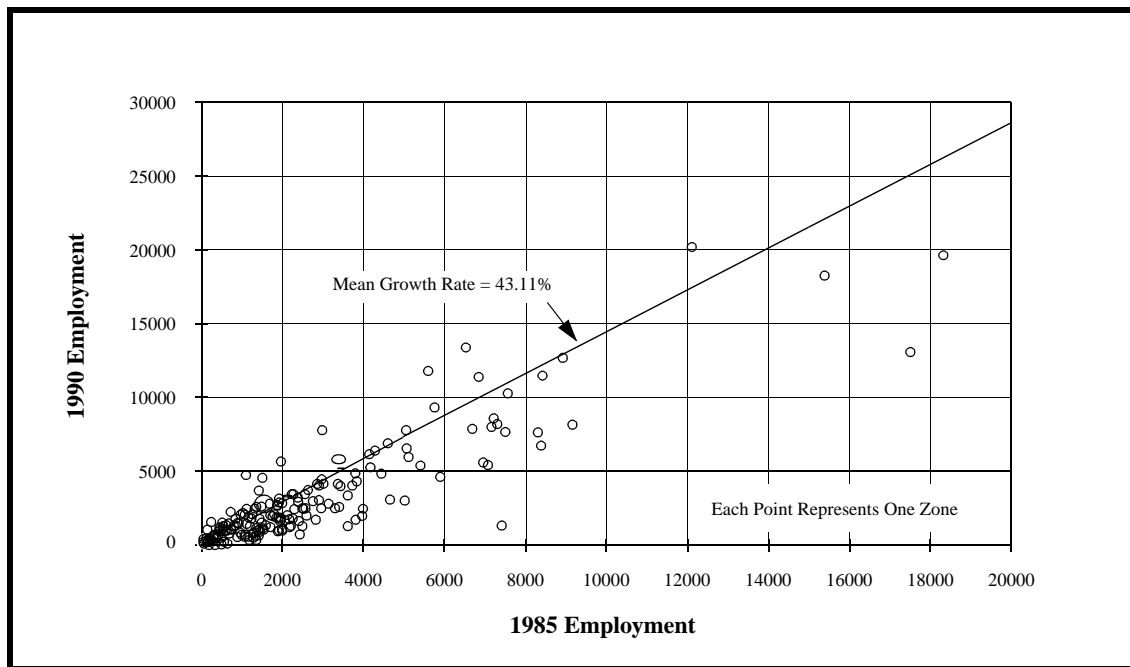
If we plot total employment in each zone for the base year against total employment in each zone for the lag year, then we can quickly identify zones for which there may be errors in either the base year or lag year employment data. Zones with potential errors in the employment data will appear farther from the line representing the average per zone growth rate. Note the relatively tight clustering of the points on the Detroit plot in Figure 4.

Figure 4. Comparison of 1985 and 1990 Employment Data (Detroit)



The greater dispersion of points in Figure 5 for Portland illustrates the relatively low correlation between the base year and lag year employment data.

Figure 5. Comparison of 1985 and 1990 Employment Data (Portland)



The low correlation between the 1985 and 1990 Portland employment data is primarily due to the use of inconsistent geocoding methods. The 1985 employment data was created by matching the addresses on state employment records to census tracts using the DIME data base, while for the 1990 employment data, state employment records were matched to addresses in the TIGER data base. Experience has shown that the DIME and TIGER data bases are not completely consistent. In fact, it is known that the DIME database has many “parity” reverses, where addresses are placed on the incorrect side of a census tract boundary.

Because state employment records are inaccurate (i.e., the addresses provided by the records often do not reflect the actual locations of employees, the numbers of employees are under-estimated, etc.), it is usually necessary to make corrections to the employment data after the address matching has been completed. For the 1990 Portland employment data, many small firms were geocoded by hand, and rigorous imputation methods were used to correct errors created in the address matching process. In contrast, many of the errors created by the address matching process for the 1985 employment data were left uncorrected, and there was no rigorous method used to adjust the allocation of employment to census tracts.

Because of these differences in geocoding methods, the 1985 employment data are not fully consistent with the 1990 employment data, and the calibration results for Portland are not as good as those for Detroit or Kansas City.

Using Lagged Total Employment as a Variable in the EMPAL Attractiveness Term

In the standard EMPAL model, lagged employment appears as a variable in the attractiveness term, and as a separate additive term. For this reason, the EMPAL model is very sensitive to errors in the lagged employment data, as the graphs in the previous section illustrate. In an attempt to improve the calibration results for Portland, a set of calibration experiments were performed in which lagged *total* employment was used as an additional variable in the EMPAL attractiveness term, Equation 2. Although in the case of Portland there is evidence of errors in the lagged employment data, it was hoped that the errors in the total employment data would be less than the errors in the individual employment categories. It was hoped that by aggregating the employment categories, additional information would be included on the prior “size” of each zone, and compensate for errors in the disaggregated employment data.

For Portland, *total* employment data was available for 1980 (the data were divided only into Retail and Other), and was tested as an additional variable in the EMPAL attractiveness term. The hope was that the 1980 total employment data was more accurately measured than the 1985 employment data,³ and that the inclusion of a “doubly”-lagged employment variable would improve the goodness-of-fit of the EMPAL model. Three separate calibration experiments were performed:

Additional Variable(s)

Modified Attractiveness Term

³ The validity of the 1980 employment data has not been thoroughly examined. However, the correlation between 1990 total employment and 1980 total employment is 0.93. The correlation between 1990 total employment and 1985 total employment is 0.94.

$$\begin{aligned}
 \text{1985 Total Employment} \quad W_{j, 1985}^k &= (E_{j, 1985}^k)^a L_j^b (E_{j, 1985}^T)^d \\
 \text{1980 Total Employment} \quad W_{j, 1985}^k &= (E_{j, 1985}^k)^a L_j^b (E_{j, 1980}^T)^d \\
 \text{1980 and 1985 Total Employment} \quad W_{j, 1985}^k &= (E_{j, 1985}^k)^a L_j^b (E_{j, 1980}^T)^d (E_{j, 1985}^T)^e
 \end{aligned}$$

The calibration results for the first experiment (Table 9) are nearly identical to those for the standard EMPAL model (see Table 7). So, it appears that using an aggregated employment variable does not compensate for errors in the 1985 employment data.

Table 9. EMPAL Parameters for the Portland Region (328 Zones)¹
 (1985 Lagged Total Employment Included in Potential Term)

	Industrial	Whl/Ret	Office	Services
Alpha - α	0.7001 (49.7)	0.2676 (9.3)	0.9025 (35.3)	0.2698 (13.0)
Beta - β	-0.0658 (122.5)	-0.0689 (65.0)	-0.3024 (107.6)	-0.1480 (80.8)
Employment	0.8072 (494.2)	0.7004 (289.1)	-0.2175 (96.0)	-0.0498 (19.9)
Total Acres	0.1638 (108.7)	0.2305 (113.9)	-0.0641 (13.3)	-0.0374 (10.2)
1985 Total Emp.	0.0109 (5.5)	0.1584 (64.1)	0.4538 (131.3)	0.6076 (189.8)
Lagged Emp.(1-λ)	0.0308 (122.7)	0.2609 (82.8)	0.8424 (78.0)	0.7241 (92.6)
R²	0.7806	0.7291	0.6788	0.8436
Likelihood Criterion	0.7800	0.7620	0.7250	0.8124
MAPE	147.93%	164.99%	348.75%	110.10%
MAPE (Smallest)	377.29%	536.16%	1247.74%	290.58%
MAPE (Largest)	40.50%	32.73%	50.89%	35.66%

¹ Asymptotic t-test shown in parentheses.

When 1980 total employment is used as additional attractiveness variable (Table 10), there is a slight improvement in goodness-of-fit for three of the employment categories, and a significant improvement for Office employment. For Industrial employment, the value of the weighting parameter for lagged employment is no longer equal to zero. The other parameter estimates are

reasonable, and 20 of the 24 estimated parameters have asymptotic t-tests greater than 20. It is likely that this formulation of EMPAL would produce more accurate forecasts of employment location than the standard formulation.

Table 10. EMPAL Parameters for the Portland Region (328 zones)¹ (1980
Lagged Total Employment Included in Potential Term)

	Industrial	Whl/Ret	Office	Services
Alpha - α	0.9283 (64.2)	0.1884 (5.6)	1.1489 (99.0)	0.2642 (15.1)
Beta - β	-0.0617 (114.4)	-0.0769 (55.5)	-0.3066 (289.2)	-0.1626 (107.0)
Employment	0.4844 (278.2)	0.2483 (97.8)	0.5122 (253.7)	0.0126 (5.2)
Total Acres	0.2246 (131.3)	0.2709 (119.6)	-0.3099 (96.4)	-0.0686 (18.4)
1980 Total Emp.	0.4082 (189.8)	0.6334 (228.1)	0.4815 (200.7)	0.6595 (203.9)
Lagged Emp.(1-λ)	0.2845 (140.5)	0.5059 (109.2)	0.1416 (159.7)	0.6957 (100.1)
R²	0.7934	0.7440	0.8471	0.8529
Likelihood Criterion	0.7957	0.7869	0.7589	0.8196
MAPE	140.25%	141.81%	268.10%	107.34%
MAPE (Smallest)	362.54%	448.14%	923.71%	281.03%
MAPE (Largest)	39.72%	32.59%	51.49%	35.36%

¹. Asymptotic t-test shown in parentheses.

There is little difference between calibration results for the third experiment (Table 11), and the calibration results for the second experiment. For all four employment types, there is a slight reduction in the MAPE statistics, but the signs and magnitudes of most of the parameters are nearly the same as those for the second experiment. However, for Industrial employment, the weighting parameter for lagged employment is close to zero. The behavior of this parameter in these statistics is an important indicator of the data problems we have been addressing. This parameter indicates the extent to which the current location of a specific type of employment is determined by prior location of that same employment type. Industrial employment, particularly, is expected to have a positive parameter value here. The appearance of a 0.0 value for the lagged employment weight (1-

λ) in the standard EMPAL calibration for Portland (Table 7) is a danger signal. Next we note that when we add 1985 total employment to the attractiveness term, there is relatively little change in the value of $(1-\lambda)$, with it shifting from 0.0000 to 0.0308 (Table 9). When the 1980 employment is added to the attractiveness term in lieu of the 1985, the value of $(1-\lambda)$ shifts substantially, from 0.0000 to 0.2845 (Table 10). Also note that the parameters of the lagged total employment variables are virtually 0 when the 1985 variable is used, but 0.482 when the 1980 variable is used. These results, too, imply serious problems with the 1985 input data.

Table 11. EMPAL Parameters for the Portland Region (328 zones)¹
(1980 and 1985 Lagged Total Employment Included in Potential Term)

	Industrial	Whl/Ret	Office	Services
Alpha - α	1.0988 (96.1)	-0.2120 (8.2)	1.0245 (90.6)	0.2772 (16.0)
Beta - β	-0.0565 (134.9)	-0.0709 (46.8)	-0.0057 (12.7)	-0.1629 (108.3)
Employment	0.7840 (491.3)	0.4819 (192.1)	0.4045 (213.8)	0.0313 (13.0)
Total Acres	0.1230 (80.2)	0.2767 (120.4)	-0.3974 (147.1)	-0.0621 (16.9)
1980 Total Emp.	0.5189 (278.1)	0.8670 (317.0)	0.3655 (162.7)	0.6128 (191.7)
1985 Total Emp.	-0.4460 (229.8)	-0.4689 (180.4)	0.2580 (112.4)	0.0443 (15.2)
Lagged Emp.($1-\lambda$)	0.0455 (163.6)	0.4542 (119.4)	0.2322 (142.2)	0.6799 (101.1)
R²	0.8160	0.7662	0.8414	0.8525
Likelihood Criterion	0.8068	0.7940	0.7491	0.8196
MAPE	132.37%	132.85%	267.23%	106.64%
MAPE (Smallest)	343.61%	411.88%	902.18%	277.80%
MAPE (Largest)	40.02%	32.91%	50.56%	35.28%

¹. Asymptotic t-test shown in parentheses.

Spatial Aggregation Tests for Portland

Non-systematic errors in the geocoding of employment data produces over-estimates of employment for some zones, and under-estimates of employment for other zones. Spatially aggregating

the employment data should reduce these errors, since for many of the aggregated zones the over-estimates and under-estimates present in the disaggregated zones will offset. For the Portland employment data, aggregating the employment data from 328 zones to 100 zones does improve the consistency between the 1990 and 1985 employment data sets, as can be seen from the correlations in Table 12.

To examine the effect of spatial aggregation, the standard EMPAL model, as well as the extended EMPAL formulations described in the previous section, were recalibrated using the 100 zone, aggregated data set.

Table 12. Pearson Correlations between Base Year and Lag Year Employment

Employment Type	Portland (100-Zones)	Portland (328-Zones)
Industrial	0.87	0.83
Wholesale/Retail	0.88	0.82
Office	0.95	0.85
Services	0.98	0.93

There were a number of calibrations done overall, and the differences, in terms of level of geographic detail as well as attractiveness variables used, are summarized in the following tabulation.

Table 13. Calibration Summary for Portland

Model Formulation	328-Zone Data Set	100-Zone Data Set
Standard EMPAL Model	Table 7	Table 14
Extended: 1985 Total Employment	Table 9	Table 15
Extended: 1980 Total Employment	Table 10	Table 16
Extended: 1980 and 1985 Total Employment	Table 11	Table 17

Since spatial aggregation improves the consistency between the base year and lag year employment data, the goodness-of-fit of the calibrations for the 100 zone data set are significantly better than for the corresponding 328 zone calibrations. For all four sets of 100 zone calibrations, the likelihood criterion measures exceed 0.80. In fact, for many of the employment types, the goodness-of-fit of the EMPAL model equations (as measured by the likelihood criterion and the MAPE statistics) are as good as the fit achieved for some of the employment categories in Kansas City and Detroit.

Except for the weighting parameters for lagged employment, all of the signs and magnitudes of the estimated parameters are satisfactory. However, for all four sets of calibrations, there is at least one weighting parameter ($1-\lambda$) which is equal to zero, or approximately equal to one. So, although goodness-of-fit is improved, it appears that some systematic residual error remains in the employment data even after it is aggregated to 100 zones.

Table 14. EMPAL Parameters for the Portland Region¹
(100-Zone System)

	Industrial	Whl/Ret	Office	Services
Alpha - α	1.1440 (84.5)	1.0728 (65.4)	-1.4953 (84.3)	-3.1202 (21.3)
Beta - β	-0.0763 (147.9)	-0.0714 (110.6)	-0.3209 (213.0)	-1.2849 (51.6)
Employment	0.8640 (417.2)	0.8583 (346.3)	-1.6812 (412.5)	0.0680 (4.6)
Total Acres	0.1600 (64.8)	0.1466 (47.1)	-0.9326 (174.9)	-0.9767 (31.5)
Lagged Emp.($1-\lambda$)	0.0000 (98.5)	0.0715 (66.6)	0.9841 (27.3)	0.9014 (47.2)
R²	0.8402	0.8372	0.8334	0.9500
Likelihood Criterion	0.8520	0.8823	0.8928	0.9276
MAPE	115.94%	137.39%	252.75%	171.40%
MAPE (Smallest)	346.41%	464.21%	904.07%	740.91%
MAPE (Largest)	27.21%	26.27%	32.76%	21.93%

Table 15. EMPAL Parameters for the Portland Region (100 zones)¹
 (1985 Lagged Total Employment Included in Potential Term)

	Industrial	Whl/Ret	Office	Services
Alpha - α	0.7852 (50.1)	0.4101 (20.6)	0.8492 (34.7)	-0.1142 (8.8)
Beta - β	-0.0756 (120.4)	-0.0561 (72.2)	-0.1889 (119.6)	-0.1311 (129.3)
Employment	0.9306 (448.1)	0.8807 (342.3)	-0.6684 (263.0)	0.6669 (303.1)
Total Acres	0.1556 (62.3)	0.1751 (54.0)	-0.1162 (28.7)	-0.0306 (8.5)
1985 Total Emp.	-0.1066 (53.3)	-0.0512 (19.6)	0.2488 (64.6)	0.2152 (84.9)
Lagged Emp.(1-λ)	0.0019 (106.0)	0.1425 (66.1)	0.9790 (27.4)	0.0000 (81.6)
R²	0.8465	0.8391	0.8315	0.9738
Likelihood Criterion	0.8524	0.8815	0.8933	0.9383
MAPE	116.33%	139.76%	257.71%	152.25%
MAPE (Smallest)	352.19%	476.25%	926.11%	659.43%
MAPE (Largest)	26.44%	27.01%	32.56%	18.23%

Table 16. EMPAL Parameters for the Portland Region (100 zones)¹
 (1980 Lagged Total Employment Included in Potential Term)

	Industrial	Whl/Ret	Office	Services
Alpha - α	1.1416 (105.2)	0.4731 (31.5)	0.8521 (49.5)	-0.1564 (12.4)
Beta - β	-0.0546 (139.1)	-0.0422 (75.3)	-0.0135 (28.4)	-0.1315 (132.4)
Employment	0.6944 (337.3)	0.5891 (253.3)	-0.7289 (315.5)	0.7146 (322.0)
Total Acres	0.1791 (74.1)	0.1419 (47.6)	-0.1325 (36.1)	-0.0474 (13.1)
1980 Total Emp.	0.2392 (127.9)	0.2914 (131.2)	0.6234 (168.1)	0.1553 (64.0)

Continued on page 48

Continued from page 47

Table 16. EMPAL Parameters for the Portland Region (100 zones)¹
 (1980 Lagged Total Employment Included in Potential Term)

Lagged Emp.(1-λ)	0.0002 (102.5)	0.0433 (83.9)	0.9803 (30.3)	0.0000 (83.7)
R²	0.8439	0.8368	0.8321	0.9750
Likelihood Criterion	0.8601	0.8926	0.8949	0.9386
MAPE	109.99%	145.39%	247.51%	155.35%
MAPE (Smallest)	321.37%	531.10%	881.99%	675.62%
MAPE (Largest)	25.79%	25.95%	32.57%	17.22%

Table 17. EMPAL Parameters for the Portland Region (100 zones)¹
 (1980 and 1985 Lagged Total Employment Included in Potential Term)

	Industrial	Whl/Ret	Office	Services
Alpha - α	1.2055 (104.1)	0.7008 (53.8)	0.8546 (51.4)	-0.1148 (8.8)
Beta - β	-0.0597 (141.1)	-0.0384 (78.2)	0.0265 (61.6)	-0.1325 (129.4)
Employment	0.8495 (416.6)	0.8155 (352.2)	-0.8532 (410.1)	0.6594 (294.3)
Total Acres	0.0811 (32.7)	0.0482 (16.1)	-0.1337 (36.2)	-0.0337 (9.2)
1980 Total Emp.	0.5542 (295.1)	0.6683 (300.6)	0.5657 (166.5)	0.0979 (39.7)
1985 Total Emp.	-0.5589 (278.1)	-0.6418 (272.5)	0.5126 (175.4)	0.1242 (48.0)
Lagged Emp.(1-λ)	0.0001 (121.7)	0.0000 (99.6)	0.9811 (28.5)	0.0398 (82.5)
R²	0.8717	0.9084	0.8324	0.9746
Likelihood Criterion	0.8761	0.9102	0.8944	0.9386
MAPE	101.52%	132.36%	245.00%	154.12%
MAPE (Smallest)	307.95%	482.49%	866.99%	668.64%
MAPE (Largest)	22.59%	20.87%	32.57%	17.73%

¹. Asymptotic t-test shown in parentheses.

Controlling for the Urban Growth Boundary

Unlike many other metropolitan regions, the Portland region has a legislatively defined Urban Growth Boundary. The Urban Growth Boundary prevents development beyond the urban fringe, effectively limiting the amount of vacant developable land and constraining employment growth in rural areas. It was thought that the existence of the Urban Growth boundary might be responsible for the mixed calibration results obtained for the 328 zone Portland data set, in particular, the relatively large values of the MAPE statistics. To examine this possibility, a single calibration experiment was performed for Industrial employment, in which a dummy variable was include in the EMPAL attractiveness term to identify zones that fall within the Urban Growth Boundary.

The value of the dummy variable was set equal to one for zones outside the Urban Growth Boundary, and equal to two for zones within the boundary.⁴

$$W_{j, 1985}^k = (E_{j, 1985}^k)^{a^k} L_j^{b^k} (2)^{f^k} \quad - \quad \text{Attractiveness of zones inside the UGB}$$

$$W_{j, 1985}^k = (E_{j, 1985}^k)^{a^k} L_j^{b^k} (1)^{f^k} \quad - \quad \text{Attractiveness of zones outside the UGB}$$

Given this mathematical definition of a zone's attractiveness, the attractiveness of zones outside the UGB are unaffected by the dummy variable. It is important to note that virtually all of Portland's industrial employment (96%) is to be found within the UGB. Our hope was that the inclusion of this variable might help to “control” for the effect of the UGB. In fact, the dummy variable does turn out to be statistically significant, but does not add to the equation's explanation of variance.

⁴ The usual practice is to use 0 and 1 for the values of such a variable, sometimes referred to as a Kronecker Delta (δ). This would not work in a product form equation such as EMPAL's, since the 0 values would force the zone's attractiveness to 0. This is the reason for the use of 1 and 2 for the values of the dummy variable in this experiment.

Table 18. EMPAL Parameters for Portland Industrial Employment (328 Zones)¹
 (Urban Growth Boundary Dummy Variable Included in Potential Term)

	Standard Model	With Dummy Variable
Alpha - α	0.5869 (42.4)	1.0352 (77.0)
Beta - β	-0.0614 (116.5)	-0.0678 (131.4)
Employment	0.8198 (508.0)	0.8041 (488.9)
Total Acres	0.1601 (107.8)	0.1711 (109.1)
Inside UGB		0.7451 (45.3)
Lagged Emp.(1-λ)	0.0000 (123.5)	0.0534 (134.8)
R²	0.7817	0.7772
Likelihood Criterion	0.7800	0.7842
MAPE	147.74%	149.95%
MAPE (Smallest)	376.77%	387.14%
MAPE (Largest)	40.52%	40.67%

¹. Asymptotic t-test shown in parentheses

Conclusions

The accuracy of EMPAL forecasts of employment location depend on the accuracy of the parameter estimates produced by the calibration procedure. If the goodness-of-fit of an estimated EMPAL model equation is poor, then it is not possible to obtain accurate forecasts of employment location. If individual parameters are imprecisely estimated, then the factors which determine employment location may be misidentified. The weighting parameter for lagged employment is especially important, since the value of this parameter determines the degree to which past employment location influences future employment location.

For the Detroit region, excellent calibration results are obtained for the standard EMPAL model. The calibration procedure for the standard model produced excellent goodness-of-fit and precise parameter estimates. Given these results, it is likely that EMPAL would produce good forecasts of employment location.

The calibration results for Kansas City are also quite good. The goodness-of-fit statistics indicate that the EMPAL model equations explain most of the variation in employment location. Unfortunately, the relatively high values of the MAPE statistics for Industrial and Manufacturing employment indicate that at the zone level there are significant prediction errors for these two employment types. However, the standard EMPAL model would probably produce accurate forecasts of employment location for the Kansas City region, since K-factors can be used to compensate for prediction error.

For Portland, mixed calibration results are obtained for the *standard* EMPAL model formulation. For most of the employment types, the goodness-of-fit is adequate, and the parameter estimates are precisely estimated. However, for Industrial employment the weighting parameter for lagged employment is equal to zero, which contradicts our theoretical and empirical understanding of the determinants of employment location. It is likely that this estimate of the weighting parameter is an artifact of errors in the lagged employment data.

A number of experiments were performed in an attempt to improve the fit of the EMPAL model equations to the Portland data set, and to gain more accurate measures the influence of past employment location. In one set of experiments, total lagged employment (for 1985 and/or 1980) was included as an additional variable in the attractiveness function of EMPAL. In the second set of experiments, the employment data was aggregated to 100 zones to compensate for errors in the lagged employment data. Of the eight experiments performed, the most promising results were obtained for an *extended version of EMPAL that uses 1980 total employment as an additional attractiveness variable*. The goodness-of-fit for this version of EMPAL was significantly better when compared to the fit for a standard EMPAL formulation. More importantly, none of the estimated values of the weighting parameters for lagged employment were equal to zero or one. *Spatial aggregation* also improved the fit of both the standard EMPAL model and extended formulations. However, for all of the spatial aggregation calibration experiments, at least one employment type had an unsatisfactory weighting parameter estimate.