changes and negative economic changes. The measures on the economic side did not have to be built into the allocation model and hence were never implemented. On the demographic side, the measures involved shifting the focus of the allocation process from net household changes to gross household changes. Based on an extensive analysis of housing transition over time in the 227 sample counties, relationships were developed to predict the reduction in an area's households due to losses from the housing stock (which average about 0.5% per year across all areas and can exceed 1% in areas with relatively old, low-value housing). Such reductions were netted out of the household changes analyzed in the model calibration process, and the forecasting tableau that later applied the model equations focused similarly on gross household changes rather than net changes. This embellishment made little difference to the forecasted absolute numbers of households, and no use was made of the extra information that it generated because top-down allocations were never conducted below the district level, so no further description is required here.

The second issue involved the ability of the allocation model to reflect public policies. Because of concern that the model would not explicitly reflect most supply-side influences on growth, the available-land weightings included in most predictive variables were designed in such a way that changes in an area's policy regime could be given rough expression via changes in the available-land parameters. The available-land weightings went on to play an important role in the model structure, as discussed below, but their policy aspect was never pursued.

Another issue was whether or not to break down demographic variables by race, which would have allowed the use of race as a predictor in the allocation model. This had been the practice in prior studies, but was a subject of concern. One problem with using race as a predictor was that it often assumed too strong a role. Once thrust into statistical prominence by the existence of racial avoidance behavior, racial variables tended to act as surrogates for growth factors involving density and available land, since black persons traditionally inhabited urban core areas. Another problem was that racial variables created problems of interpretation because the behaviors captured by race-related model parameters could change in the future (having demonstrably done so in the past). As it happened, few large geographic areas in the Charlotte region were extreme in terms of racial mix. Only four of the region's fifteen counties – collectively accounting for less than 8% of its total population – had black population shares under 12% or over 28% in 2000. What this meant for the present study was that including racial variables in the allocation model could not make a great deal of difference to the forecasts regardless of the extent to which the model equations reflected differences elsewhere. Hence race was set aside as a subject of measurement and a potential predictor.

The next question was the need to delete observations from the calibration sample to keep the regression analyses from being overly influenced by individual observations. Numerical dominance of a statistical sample by a few observations can be a big problem in cross-sectional analyses even when the sample size is in the hundreds. The risk of obtaining unreliable results for this reason is elevated by the omission of causal factors (e.g., supply-side influences), but would exist even if the analysis could address all kinds of relationships, because numerical dominance can result from unique events involving a single company. Previous studies had tried to minimize the problem by analyzing large samples, by spreading the predictive burden across many equations, and by weighting observations in a manner to be described; but they had always stopped short of deleting sample observations. The present study took this additional step.

The following three cases illustrate how severely a cross-sectional sample of socioeconomic data can be dominated. The quantities to which the percentages apply are dev-change variables in which an observation for a county equals its 2000 earnings minus its 1990 earnings times the ratio of 2000 earnings to 1990 earnings for the metro area containing the county. The percentage cited to describe the dominance problem for each industry is the share of the sample's total variation (sum of squares) that is supplied by the most-dominant metro area among the 29 in the sample. The percentages cover entire metro areas, since these are the groups of observations eligible for deletion, but nearly all of the variation involves the individual counties noted.

* "Other" retail trade. For some unknown reason, the 1990s brought a massive shift in the distribution of metro Atlanta's home-supply retailing (e.g., Home Depot) from Fulton and DeKalb counties to Cobb County. Abrupt shifts are uncommon in retail trade except when they involve new regional malls, so the home-supply phenomenon caused metro Atlanta to account for 71% of the sample's total variation in "other" retail trade

* Depository and non-depository institutions (i.e., banks and credit unions). The Richmond area profited during the 1990s from the rapid emergence of Capital One, Inc. as a nationwide financial presence. The growth of Capital One occurred at new office-park locations in Henrico County, which wraps around Richmond, and may have involved some withdrawal of functions from the city. The result was an 8% decline in Richmond's constant-dollar earnings from banking while Henrico County increased by 348%. When expressed in dev-change terms, this pattern caused metro Richmond to account for 50% of all variation in the banking variable across the 227 counties. (Despite even greater banking expansion, metro Charlotte supplied only 12% of the total banking variation.)

* Communication. The leading economic driver for metropolitan Kansas City during the 1990s was the explosive growth of Sprint Corporation. Although the city itself may have had some Sprint offices, the corporate headquarters were located across the river in Johnson County, Kansas. Communications earnings rose by 171% in Jackson County (containing Kansas City) and 1063% in Johnson County. The difference between these rates of change caused metro Kansas City to supply 80% of all variation in the communications variable.

These cases of sample dominance were all caused by rapid growth, with absolute declines playing no role except for banking in Richmond. The key points are that: 1) there is almost no way to explain such extreme occurrences statistically with variables that express what actually happened; and 2) when offered dominant observations like these, a regression will grasp at any numerically useful predictors whether they make sense or not. An available predictor that was high for Cobb, Henrico or Johnson County and low for Fulton, Richmond or Jackson County, without having many other extreme values, might receive great explanatory weight in regressions for the above industries whether or not it had any substantive relevance.

The present study determined that only trimming the sample could deal with this problem adequately. The deletions had to involve entire metropolitan areas to preserve the structure of the model. The rule applied in selecting metro areas for deletion was that no metro area should supply more than 25% of total variation in a dependent variable. Deletions were unnecessary for household variables because no metro area accounted for more than 11% of total variation in those cases (given the use of divisors as described momentarily). The deletions for economic variables involved 22 metro areas in 15 of the 32 industry groups, as listed in Table 8.

The selection of metro areas for deletion was accomplished simultaneously with the determination of weightings for sample observations. The issue in that regard was the need to balance variation in the sample. A general characteristic of regression analysis is that all variables on both sides of a regression equation can be multiplied by any constant (which can vary across observations) without imparting bias to the coefficient estimates or measures of statistical significance obtained from the regression. Weightings are commonly employed in cross-sectional analyses to deal with the general problem of heteroscedasticity, or unequal error variances. The objective can be described as creating a level playing field so that "small" areas are not rendered irrelevant by "big" areas. The Charlotte modeling effort used weightings to address heteroscedasticity

Table 8. DELETION OF OBSERVATIONS FROM REGRESSION SAMPLES

Industry Group	Metro Areas Deleted and Number of Observations (Counties) Involved	Sample Size
Farming	San Antonio (4)	223
Agricultural serv., forestry & fish.	Raleigh-Durham-Chapel Hill (6)	221
Mining	None	227
Construction	None	227
Manufacturing	Raleigh-Durham-Chapel Hill (6)	221
Transportation, Commun. & Util.:		
Transportation	Cincinnati-Hamilton (12)	215
Communication	Kansas City (11), Raleigh-D-CH (6)	210
Electric, gas & sanitary service	Atlanta (20)	207
Wholesale Trade	None	227
Retail Trade:		
GAFO (dept.-store-type goods)	None	227
Automotive retailing	None	227
Eating & drinking places	None	227
Other retail trade	Atlanta (20)	207
Finance, Insurance & Real Estate:		
Depository & non-depos. inst.	Richmond-Petersburg (10), Nashville (8)	209
Other finance	Indianapolis (9)	218
Insurance carriers	Indianapolis (9)	218
Insurance agents and services	None	227
Real estate	None	227
Services:		
Hotels & other lodging places	None	227
Personal serv. & private h'holds	None	227
Business services	None	227
Auto repair, services & parking	None	227
Miscellaneous repair services	None	227
Amusement & recreation serv.	Cincinnati (12), Louisville (7), Minn. (13)	195
Health services	None	227
Legal services	Tampa-St. Pete. (4), Raleigh-D-CH (6)	217
Educational services	None	227
Social serv., memb. org. & misc.	None	227
Engineering & mgmt. services	Raleigh (6), Norfolk (13), Atlanta (20)	188
Government		
Federal government (civilian)	Columbus (6)	221
State government	Grand Rapids (4)	223
Local government	None	227

problems arising from both area-size differences and special dominance situations like those cited above. The weightings in a given analysis were numbers that held constant across all counties within a metro area.

In the economic equations, the weightings took the form of divisors based on metro sums of 1990 earnings. The metro sum for each industry was expressed as a ratio to the average sum (i.e., to total 1990 earnings for the sample divided by 29, not 227). This ratio, raised to an exponent, became the divisor for the dependent variable and all independent variables in the regression for the given industry. The study design involved using the same exponent value for all industries and determining this value in the process of dealing with numerical dominance. The 25% rule noted above was applied to dependent (dev-change) variables with divisors included, and the chosen exponent was the value that minimized the cross-industry average share of variation supplied by the most-dominant metro after the deletion of extremes. This exponent value turned out to be 0.81. On the demographic side, similar computations for the household variables yielded an exponent value of 0.90. No observations were deleted from the household regressions since no metros came close to the 25% threshold in those cases.

The last issue involved geographic scale. As already described, data limitations mandated the use of counties and equivalent jurisdictions as the observation units for model calibration, but the equations were intended for use in sub-county (i.e., district) forecasting as well as county-level forecasting. The adopted rule was that forecasting units could range in geographic size down to 50 square miles. There was nothing about the model calibration process that limited the application of the resulting relationships to a particular geographic scale, and targeting areas of 50 square miles would not have extrapolated beyond the range of the sample (which included eight areas that small or smaller.) Nevertheless, given that the observation units had a median size exceeding 400 square miles, there was a need to assure that the estimated relationships would be maximally relevant to areas smaller than a typical county.

This need was addressed, along with objectives involving policy inputs, by relying heavily on proximity variables rather than other types of predictors. As already described, a proximity variable consisted of some relevant quantity, such as earnings in an industry group or households in an income category, summed across the entire metro after being weighted by an inverse function of distance from the subarea for which the variable was being computed. The values of proximity variables were heavily influenced by the amount of activity within the subject subarea itself (especially when they involved relatively high exponents and low values of their other two parameters). But the smaller a subarea's geographic size, the closer it would lie to neighboring subareas, and hence the greater contributions they would make to its values of proximity variables. Thus such variables should be largely invariant to the scale of observation units.

Predictive variables that simply described initial conditions, past changes or current changes in an area implicitly reflected its geographic scale, since big areas tended to feature big numbers while small areas featured small numbers. Simple predictors of this nature had worked well enough in previous studies, but the Charlotte project went further in applying relationships to small forecasting units, so extra care was needed to make the relationships scale-invariant. Hence all of the proximity variables entering the calibration process were weighted by estimates of available land. The only predictors besides proximity variables allowed in the equations were dev-change, dev-share and dev-mean variables for the industry or household sector under analysis, plus dev-share variables for the three household groups (or the two not under analysis). All of the non-proximity variables besides same-sector past dev-change were weighted by available land as well.

Limiting most attention to proximity variables and avoiding racial predictors altogether had the effect of reducing the numbers of explanatory factors found significant in the regressions and retained in the model equations. The final economic equations contained 3.7 independent variables on average, and the final household equations contained an average of 4.7. Past studies using the same approach had yielded averages of 4 to 5 variables in the former case and about 6 in the latter. Restricting eligible predictors and deleting observations also tended to lower R-square values. This was true because extreme observations were often numerically explainable, sometimes to a spectacular extent (though the posited relationships might be ridiculous). The losses of R-square that resulted from setting aside such cases served as a chastening reminder that small-area growth patterns are always a challenge to explain meaningfully.

Incorporation of Available Land

Including a measure of available land in most of the allocation model's predictive variables was intended to provide a crude expression of supply-side limitations on growth and a potential mechanism for registering policy influence (although this mechanism was never utilized). The following paragraphs describe how the measure was derived for the 227 counties in the calibration sample and the areas addressed in forecasting. References to "developable land" and "available land" are understood to mean the following:

Developable land. The portion of a county or other subarea, measured in square miles, that is physically suitable for development in urban land uses, whether or not such uses already exist.

Available land. The developable land in a county or other subarea that remains vacant at a given point in time (or is developed at such low intensity that its conversion to a higher use would be routine).

Technically the Charlotte study lacked data on both developable land and available land, but circumstances allowed total land area to serve adequately as a surrogate for developable land in most of the 227 sample counties. Nearly all eastern metro areas with one to five million inhabitants occupy non-mountainous, maturely eroded landscapes where the required allowances for water bodies and steep slopes are small and predictable. Hence developable land is highly correlated with total land, even though the magnitudes are not identical. The two exceptional cases are Norfolk-Portsmouth and New Orleans, which contain extensive areas too wet for urban use. In the study these areas were identified from maps in the National Wetlands Inventory and subtracted from total county size to yield estimates of developable land.

As for available land, the study had no direct information at all. The only relevant data consisted of demographic and economic variables that could be used to compute density measures. The strategy was therefore to posit a functional form linking available land to density and then to obtain empirical estimates of any parameters involved. This would involve expressing the ratio of available land to developable land as a one-parameter or two-parameter function of development density. The value or values of parameters in this function would be established prior to the model calibration process and assumed to hold constant for all economic and household sectors (implying that available land was the same for all land uses, although its importance to growth could vary). Absolute amounts of available land would be computed from the ratios yielded by this function. The quantity used to multiply other variables in the allocation model would then consist of available land divided by metro average available land, all taken to an exponent. The exponent would be allowed to assume different values in different equations and would be determined in the model calibration process by iteratively finding the value that maximized R-square. Each exponent would then express the relative importance of land availability to the given economic sector or household group.

The first task was to select a measure of development density, preferably one that reflected both population and employment. The chosen measure was based on the facts that: 1) employment is about half as great as population on average; and 2) about 20% of all urban land is used by sources of employment. These circumstances imply that land consumption per employee equals about half of land consumption per resident (since 0.2/0.5 is half of 0.8/1). Hence the density measure simply equaled population plus employment times one-half. This sum was said to express development density in "population/employment" or "pop/empl" units.

The designation of a functional form for available land followed the principle that a model should have interpretable parameters even if the interpretation rests on a highly idealized scenario. The chosen scenario focused on the tendency of an area to develop at progressively higher marginal densities. After some experimentation with functional forms, the choice was a form based on the assumption that marginal development density varied inversely with the share of developable land still available. Letting D = average density, D' = marginal density, A = available land, L = total developable land, and k = a parameter to be determined, this function and its evaluated integral are as shown in the first two lines below. The third line gives the solution for the available land ratio (A/L) as a function of average density in population/employment units.

$$D' = k/(A/L)$$
$$D = -k*\ln(A/L)$$
$$A/L = \exp(-D/k)$$

Figure 6 on page 52 shows the available land ratios yielded by the above relationship at various density levels, given different values of the parameter k. The graph spans the density levels found in the model calibration sample, which range from 32 population/employment units per square mile in metropolitan San Antonio (Wilson County) to 8,594 units per square mile in St. Louis. The parameter k determines how fast the available land ratio approaches zero as observed density rises. This functional form can closely replicate the results of assuming linear and quadratic functions for marginal density, but offers the advantages noted above.

The remaining task was to find an appropriate value of k. This was accomplished by running preliminary versions of the allocation model regressions. As noted above, the quantity that multiplied other independent variables in the allocation model was available land divided by metro average available land, all raised to an exponent. The preliminary regressions involved the use of trial-and-error methods to find best-fitting values of both the parameter k and the overall exponent. With the value of k established in this fashion, only the exponent would be allowed to vary in the model calibration process to follow.

The preliminary analyses addressed the three income-based household groups and three aggregate economic sectors, namely "industrial" activity, producer services and consumer services as defined in the next section. The regressions were run with the full sample of 227 observations, using divisors computed as described above. The term involving available land was incorporated in all independent variables except one (the past dev-change variable for the industry or household group under analysis). Three to five independent variables were found significant at better than the 0.5% level in each regression. The values of R-square ranged from 0.50 to 0.86 for economic sectors and from 0.78 to 0.79 for household groups. These findings are summarized in Table 9, which occupies the lower portion of the next page. (Details for independent variables are omitted because these results are supplanted by the final calibration data.)

As shown by the right-hand columns of Table 9, the best-fitting values of the parameter k in the available land function ranged from 1,500 to 7,500, and the best-fitting exponent values ranged from 0.1 to 0.7. The process of finding these values revealed strong, and expected, positive associations between k and the exponent. Entering higher values of k would weaken the relationship of available land to density and thereby let more weight be placed on the relationship (via the exponent) without a loss of R-square. Available land was found to have almost no importance for producer service activity – not surprisingly, since office buildings can trump other land uses in terms of value per acre – and little importance for the industrial sector.

The value of k chosen for general use in estimating available land was 4,000 population/employment units per square mile. This selection gave the most weight to the k-values of 3,500 obtained in two household analyses because experimentation showed that the higher values could each be lowered to 4,000 at a sacrifice of only 0.002 in R-square. Despite later changes that altered the household equations, the k-value of 4,000 was retained throughout the model calibration process and served well by all indications.

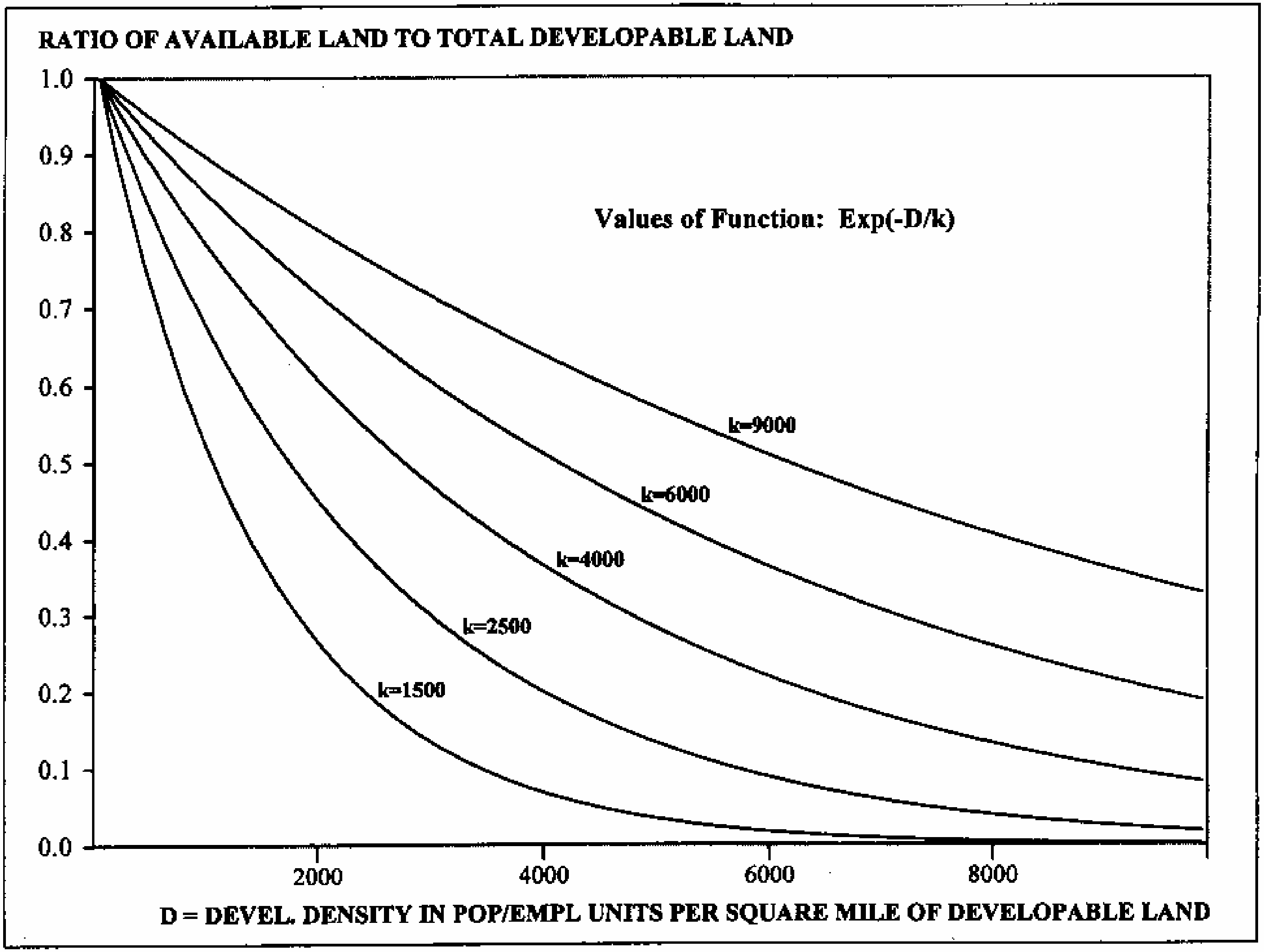
Table 9. SYNTHESIS OF RESULTS FROM PRELIMINARY REGRESSIONS

	R-Square Value for Regression	Number of Significant Independent Variables: Including Avail. Land	Number of Significant Independent Variables: Other (past dev-change)	Best-Fitting Values of Parameters in Avail. Land Index: k	Best-Fitting Values of Parameters in Avail. Land Index: Exponent
Dependent Variable:					
Industrial Earnings	0.499	3	0	1500	0.25
Producer Service Earnings	0.640	4	0	7000	0.1
Consumer Service Earnings	0.857	4	1	6500	0.65
Lower-Income Households	0.792	2	1	3500	0.7
Middle-Income Households	0.794	4	1	7500	0.6
Upper-Income Households	0.775	3	1	3500	0.6

SCENARIO YIELDING "K" VALUE OF APPROXIMATELY 4,000

Residential development at 2.42 dwelling units per gross acre	= 1,549 d.u.s per square mile
Residential devel. with 20% of land reserved for employment	= 1,239 d.u.s per square mile
Population @ 2.584 persons per d.u. (227-county ave. in 1990)	= 3,202 persons per square mile
Employment @ one job per two residents	= 1,601 jobs per square mile
Population/employment units (sum with empl. half-weighted)	= 4,002 pop/empl units per sq.mi.

Figure 6. ALTERNATIVE VALUES OF AVAILABLE LAND FUNCTION



As shown by the computations in Table 9, a situation where k equals 4,000 is a scenario in which an area's residential development starts at roughly 2.4 occupied dwelling units per gross acre. The marginal density of residential development then rises progressively as more land is used. Based on the formulas stated earlier, the marginal density reaches 10 units per acre when half the area's developable land remains available, and 30 units per acre when only one-sixth remains available. The average densities at these points are about 2,800 and 7,200 population/employment units per square mile.

As thus established, the available land function clearly overestimated how much of an area's land could actually accommodate most land uses, particularly non-residential uses. This fact was not problematic in itself because the allocation model's equations were not sensitive to intra-metropolitan scale effects. For example, suppose that the counties in a three-county metro had available land ratios of 0.8, 0.6 and 0.4 according to the general formula, but their viable sites for, say, manufacturing activity accounted for only 8%, 6% and 4% of developable land. Substituting the latter figures for the former in the manufacturing equation would leave its predictions unchanged, because the term used as a multiplier in the equation's independent variables would contain available land divided by mean developable land for the metro. But entering a different pattern of percentages for a different land use – say, 10%, 10% and 5% for wholesale trade – would change the outputs from the equation for that activity, and likewise a different set of manufacturing percentages based on different zoning policies would also produce a change. The model was thus able to accept and reflect land descriptions quite different from what was available during its calibration.

Modeling Sequence

A precedence ordering of variables is required in any forecasting model that is not a simultaneous-equation system (wherein all values of variables would be mutually determinate). The ordering of variables is accompanied by a restriction of explanatory factors in each equation to variables that appear earlier in the sequence than the one being explained. The best ordering is simply the one that yields the greatest overall predictive accuracy when the model is applied.

Following the practice in earlier studies, the present modeling sequence involves an arrangement of variables into four major groups, namely households and three groups of industries. The latter are referenced as "industrial" activities, producer services and consumer services. This economic grouping reflects functional differences in that most industrial establishments are involved in handling physical goods, while producer service establishments provide intangible products to businesses, and consumer service functions deal directly with consumers. The key factor, however, is that the groups have varying needs for proximity to other activities at a sub-regional scale. Industrial establishments generally have the weakest activity linkages because their main site selection criteria involve infrastructure, natural resources and physical land suitability. Consumer service establishments are the most strongly influenced by other development because their competitive success turns upon access to households.

Locational dependence is relevant for allocation modeling because the least dependent functions should be addressed first in the modeling sequence, when no other current changes are available as predictors. The most dependent functions should come last because their equations can benefit most from predictors that pertain to current changes. This leads to a sequence in which the model first addresses industrial activities, then producer services and finally consumer services. The remaining question is where to position the household group. Any choice is a compromise, since household location patterns are linked to all economic functions on a mutually determinate basis by virtue of employment as well as patronage relationships. In some past models calibrated for small metro areas, households were placed between the industrial and producer service groups because households contributed more as predictors of producer services than vice versa. The present study positioned households after producer services, however, because testing showed that producer services would enter two of the three household equations.

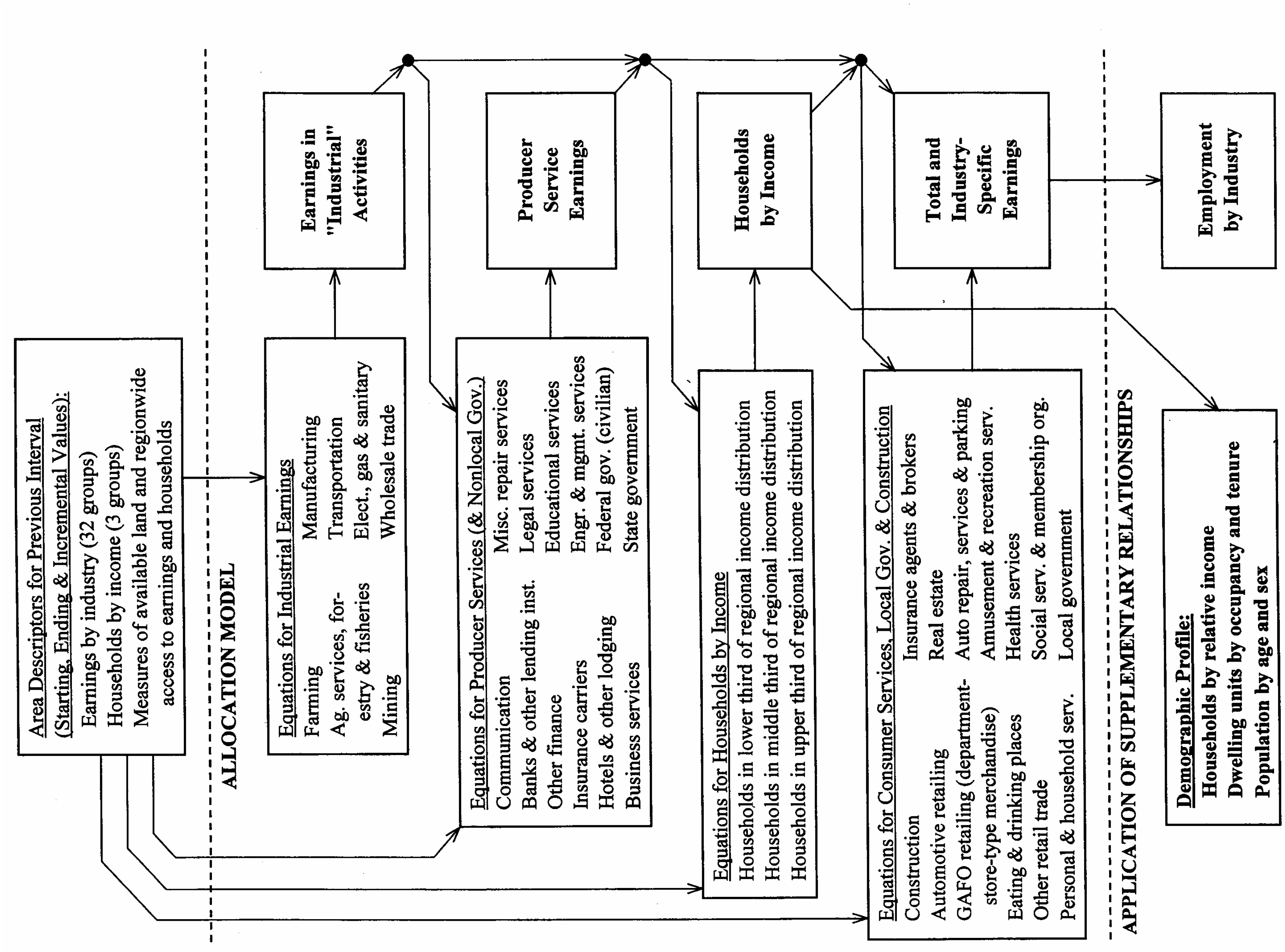
The resultant modeling sequence is depicted graphically in Figure 7 on the next page, which serves to identify the industries contained in the three economic groups. This same ordering of variables was followed in the model calibration process and each round of forecasting for the study area. As noted earlier, the industries in each group were totaled rather than taken individually when computing dev-change, dev-share and dev-mean variables for use as predictors in subsequent equations.

The variables eligible as predictors in each model equation – reflecting the information that would be available in each round of forecasting – consisted of: 1) past changes and initial conditions in the three major economic groups; 2) past changes and initial conditions in the specific industry under analysis (for economic equations); 3) past changes and initial conditions in the three household categories; and 4) current changes in economic and/or household groups already addressed by the modeling sequence.

Constraints and Special Circumstances

The following statements summarize the characteristics of variables in the allocation model, as explained in the preceding subsections. All dependent variables – i.e., variables appearing on the left-hand sides of equations – were expressed in dev-share form (meaning they were "current" dev-share variables relative to the time interval being analyzed or forecasted). Every independent variable was expressed in one of four forms: past dev-change, initial dev-share, initial dev-mean, and current dev-change. Most of the quantities incorporated in these forms were proximity measures equaling distance-weighted sums of attractors. The only predictors that did not incorporate proximity calculations were dev-change, dev-share and dev-mean variables for the industry or household category being analyzed, plus dev-share variables for the three (or two other) household categories. All of the predictors but one – past dev-change for the sector under analysis – were weighted by an index of available land raised to an exponent (with the weighting applied before the conversion to dev-change, dev-share or dev-mean form). And lastly, a divisor based on the metro sum of activity in the given sector was applied to all variables on both sides of each equation. The divisor only affected the calibration process and is not mentioned here in other contexts.

Figure 7. SEQUENCING OF VARIABLES IN MODEL DEVELOPMENT AND APPLICATION



This highly structured format led to the imposition of constraints on the allowable signs of regression coefficients. A constraint meant that a predictor would not be allowed to enter a regression equation with a negative coefficient if its relationship with the dependent variable was intended to be positive, or vice versa, even if it would play a significant role in a statistical sense. The constraints were partly motivated by the inclusion of available-land weightings. In general, the construction of a composite explanatory variable almost always presupposes that the regression coefficient will have one sign or the other, because factors are combined on the assumption that they will all be pulling in a given direction. For example, attraction measures were multiplied by available-land measures in the Charlotte model because both were expected to exert positive influences. Allowing such variables to enter regression equations with negative signs would have negated the logic behind their construction.

The constraints imposed upon regression coefficients are stated and explained below. Because they are special cases, independent variables pertaining to the same industry or household group as the dependent variable are called “same-sector” variables, even if they pertain to households rather than industry groups.

Same-Sector Past Dev-Change: Coefficient Always Positive. Past change in an activity is very often a strong predictor of current change. A negative relationship between past and current change would mean that an industry tends to cycle up and down. Farming seems to be the only case in which this systematically occurs, and farm earnings are unpredictable anyway, so there is little cost in requiring the coefficients for same-sector dev-change variables to be positive.

Same-Sector Initial Dev-Share: Coefficient Always Negative. A same-sector dev-share variable expresses the relative geographic concentration of an activity. A negative coefficient for such a variable says that areas with more than their share of an activity will tend to gain less of it than areas where the activity is initially in short supply. This is the expected pattern. By including same-sector dev-share negatively and same-sector dev-change positively, an equation can describe a situation where growth feeds on itself but is subject to diminishing returns or other countervailing forces. Positive coefficients almost never occur for same-sector dev-share variables, but in any case could not be allowed due to the presence of available-land weightings.

Same-Sector Initial Dev-Mean: Coefficient Always Negative. Dev-mean variables resemble dev-share variables except that they express concentration of activity in absolute rather than relative terms. For example, if a county has a high same-sector dev-mean value for wholesale trade, it exceeds most other counties in its absolute amount of wholesaling, whereas a high dev-share value would mean that wholesaling comprises a relatively large share of its economy whether big or small in absolute terms. Since these variables operate similarly in regressions, negative coefficients have also been required for same-sector dev-mean variables.

Past or Current Dev-Change Proximity Variable: Coefficient Always Positive. Though such situations can be imagined, there is little need to allow for the chance that growth in one activity systematically discourages growth in another. Rich people who don’t like looking at mobile homes, for example, can usually adjust by moving into the next valley rather than the next county. Meaningful negative relationships become even less likely when the descriptors of change are proximity variables that express regionwide gradients of attraction. Hence dev-change variables have always been required to enter with positive coefficients.

Initial Dev-Mean Proximity Variable: Coefficient Always Positive. Dev-mean proximity variables express proximity to static rather than incremental attractors. Positive coefficients have been required for these variables because they include available land and are not designed to express repulsion. Other predictors have been sufficient to capture the forces of urban de-concentration and dispersal.

Initial Dev-Share Variable for a Household Category: Coefficient Always Positive. Household income levels operate in many ways to shape urban growth patterns. In particular, past studies have repeatedly suggested that many kinds of activity – not limited to residential development and consumer services – tend to follow upper-income households. The effects are produced not only by changes in number of households (captured by dev-change proximity variables) but also by the relative proportions of upper-income and lower-income households at each point in time. Hence separate dev-share variables for the three income categories, weighted by available land, have been tested in all regressions. Positive linkages have been required due to the available-land weightings, but negative impacts have been capturable via the assignment of positive coefficients elsewhere, since dev-share variables always sum to a constant for each metro area.

In any given analysis, the same-sector past dev-change variable was the only eligible predictor that did not include a weighting by available land. This was the case because the constraining role of available land should already have been reflected to a large extent in a sector's past growth pattern, making a weighting redundant. Same-sector initial dev-share and dev-mean variables were also special cases because their expected and intended roles involved negative signs. To assure that available land would have a positive influence, the available-land index (i.e., the ratio raised to an exponent) was used as a divisor rather than a multiplier of the quantities entering the dev-share and dev-mean computations.

Regression Results

Finding the best-fitting combination of independent variables in an allocation-model equation is usually not difficult. The process was somewhat simplified in the present case by the restrictions on eligible variables and allowable coefficient signs. The only special twist was the need to find a best-fitting value of the available-land exponent. Since no analytical solution was possible, this had to been done by trial and error on the premise that the "best" exponent was the one that maximized R-square.

Each regression analysis proceeded by entering independent variables one at a time on the basis of correlations with residuals from the previous step. (Technically this approach is not quite as efficient as stepwise regression, which relies on partial-R values, but there are advantages in not automating the process.) The threshold for retention of a variable was 5% significance in a two-tailed test, with a few very minor exceptions, and most predictors were well above this threshold. The median t-statistic for all variables in the final calibrated model was 3.36, denoting slightly better than 0.1% significance with the sample sizes in question.

The analysis spreadsheets were set up so that all variables were recalculated given a change in the assumed value of the available-land exponent. The common procedure was to start each analysis with an exponent value of 0.5 and get fairly close to the best set of predictors before varying the exponent. Its value would then be progressively shifted in whichever direction served to raise R-square, with periodic checks to see if any variables in the equation were losing significance or if any variables outside the equation were looking more viable as candidates for inclusion. Wholesale substitutions of variables were very rare, even with large changes in the exponent, and the additions and subtractions that did occur were usually identifiable beforehand as borderline cases. An unambiguous local optimum could always be reached without any great difficulty. The numerical stability of the regressions led to confidence that these local optimum solutions were also global optima.

The 35 dependent variables in the allocation model were analyzed following the procedures and guidelines discussed here and in the previous subsection. Then the fitted equations were used to "predict" the 2000 values of all variables in the 227 sample counties, based on 1980 and 1990 data and the metro totals for 2000. Though more or less satisfactory for the sample as a whole, these initial results were unsatisfactory for the Charlotte region. Hence the entire model was recalibrated.

The problem involved the divisors used to minimize heteroscedasticity in the regressions. As discussed above, the original divisor for an industry was the ratio of total metro earnings in that industry to the sample average earnings per metro in that industry, all taken to the 0.81 power. The divisor for a household category was computed similarly with the ratio taken to the 0.90 power. These computations had the convenient feature that divisor values were constant for counties in a metro, meaning that all explanatory variables summed to zero for each metro and hence the regression model did not include a constant term. (Nonzero intercepts are always undesirable in an allocation model.) The weakness was that the divisors dealt with numerical dominance on an inter-metropolitan basis but not an intra-metropolitan basis, which turned out to be serious for the Charlotte region. The solution was to re-run all the regressions using divisors of a type employed in the 2000 Charlotte study. For each industry, the divisor value for a county equaled the geometric mean (i.e., the square root of the product) of the county's 1990 earnings in all industries and the metro average 1990 earnings in the industry under analysis. Thus the alternative divisor adjusted for both the scale of the county economy and the size of the industry in the metro area as a whole. The household-equation divisors were computed similarly using total households in the county and category-specific households in the region (metro). The new divisors were the only changes in regression inputs. The recalibration yielded substantially different coefficients and available-land exponents, frequently accompanied by substitutions of independent variables in the final results.

The original divisors are called "exponential" divisors and the alternative numbers are referenced as "geometric mean" divisors. After the model was recalibrated using the geometric mean form, the two sets of equations were compared on the basis of their ability to predict actual 2000 conditions. For each industry and household category, the equation that best replicated 2000 conditions in the study area was selected for inclusion in the final model. The selections were based strictly on outcomes for the Charlotte region and in some cases involved slight reductions in predictive accuracy for the 227-county sample. Equations based on geometric mean divisors were selected for 19 of the 32 economic sectors. However, equations based on exponential divisors were retained for a majority of the region's largest industries and for all three categories of households.

A general circumstance in allocation modeling is that the more level the playing field – in terms of the extent to which scale differences among observations are offset by heteroscedasticity adjustments – the lower the R-square values obtained in model calibration. Shifting from exponential to geometric mean divisors in the present study lowered R-square in all but four of the 35 regressions, with three-quarters of the changes equaling -0.1 to -0.3. The overall average impact on R-square was –0.123. In about half of all cases there were also reductions in the number of predictors found significant and retained, with the average exceeding half a variable per regression.

The final results of the model calibration process are presented in Table 10 on the next four pages. The upper portion of the table's first part shows the notation used in describing the regression results. The right-hand column lists the independent variables entering the equations, some of which are stated as functional forms because they involve proximity to attractors other than the sectors under analysis. The left-hand column lists the descriptors that serve as arguments of the functions. As described earlier, all of the dependent variables subjected to analysis were current dev-share variables and all of the independent variables besides SDC (past dev-change in the sector under analysis) were weighted by the available land index. Lastly, the first part of the table lists the suffixes used to denote different combinations of parameter values in the proximity variables.

The regression results for each economic sector and household category are listed in a separate box. The figure in parentheses following the name of the sector is the number of observations used in the given regression. (See Table 5 and the surrounding discussion.) The text below the sector name indicates whether the regression involved an exponential or geometric mean divisor and lists the intercept value in the latter case. This part also gives the R-square value obtained and the best-fitting value of the available-land exponent. The columns occupying the remainder of the box then present the regression coefficient, the t-statistic and the significance level for each independent variable. The t-statistics shown here have been recomputed to allow for the loss of 29 degrees of freedom due to the manner in which variables were constructed.

Table 10. REGRESSION RESULTS USED IN ALLOCATION MODEL, PART I

Descriptors Used in Access Variables

IE	Industrial earnings
PE	Producer service earnings (including nonlocal government)
CE	Consumer serv. earnings (incl. local gov't & construction)
TE	Total earnings
LH	Households in lower third of regional income distribution
MH	Households in middle third of regional income distribution
UH	Households in upper third of regional income distribution
TH	Total households

Variables

SDC = Same-sector past dev-change
SDS = Same-sector initial dev-share
SDM = Same-sector initial dev-mean
SDM1 = SDM with avail. land exponent = 1
DS( ) = Dev-share variable for households
PDCA( ) = Past dev-change access variable
CDCA( ) = Current dev-change access var.
DMA( ) = Initial dev-mean access variable

Notation for Parameters in Access Var.s

Suffix p: Exponent = 2, intra-area impedance = 5, terminal impedence = 5
Suffix q: Exponent = 2.5, intra-area impedance = 5, terminal impedence = 5
Suffix r: Exponent = 2.5, intra-area impedance = 3, terminal impedence = 3

Results

Earnings in Farming (223)

Type of Divisor: Exponential — No Constant Term
Avail. Land Exponent = 0.00 — R-Square = 0.0331

Independent Var.	Coeff.	t-Statistic	Signif.
SDM	-0.1016	-2.578	1%

Earnings in Mining (227)

Type of Divisor: Geom. Mean — Intercept = -1.9196
Avail. Land Exponent = 0.35 — R-Square = 0.3188

Independent Var.	Coeff.	t-Statistic	Signif.
SDS	-0.2573	-9.469	<0.01%
PDCA(TH)r	28.887	2.662	1%

Earnings in Ag. Services, Forestry & Fish. (227)

Type of Divisor: Geom. Mean — Intercept = 2.4374
Avail. Land Exponent = 0.22 — R-Square = 0.3992

Independent Var.	Coeff.	t-Statistic	Signif.
PDCA(UH)r	143.19	2.950	0.5%
SDS	-0.5733	-8.573	<0.01%
SDM	-0.0670	-3.243	0.5%
DS(MH)	0.9566	2.690	1%
DS(UH)	0.5570	3.591	0.05%
PDCA(TE)p	0.9816	2.720	1%

Earnings in Construction (227)

Type of Divisor: Exponential — No Constant Term
Avail. Land Exponent = 0.63 — R-Square = 0.5553

Independent Var.	Coeff.	t-Statistic	Signif.
CDCA(TH)r	707.49	4.163	<0.01%
PDCA(PE)r	26.280	3.774	0.05%
SDM	-0.0212	-2.775	1%
CDCA(PE)p	15.674	2.604	1%
DS(MH)	4.5147	2.131	4%

Earnings in Manufacturing (221)

Type of Divisor: Exponential — No Constant Term
Avail. Land Exponent = 0.23 — R-Square = 0.5451

Independent Var.	Coeff.	t-Statistic	Signif.
SDM	-0.0491	-7.354	<0.01%
SDC	0.2651	5.208	<0.01%
SDS	-0.1135	-4.459	<0.01%
DS(MH)	14.773	3.056	0.5%
DS(UH)	9.3009	5.886	<0.01%

Earnings in Transportation (204)

Type of Divisor: Geom. Mean — Intercept = 2.7137
Avail. Land Exponent = 0.00 — R-Square = 0.2185

Independent Var.	Coeff.	t-Statistic	Signif.
SDS	-0.1732	-5.278	<0.01%
SDC	0.1864	3.657	0.05%
PDCA(CE)q	16.9399	2.165	3%

Earnings in Communication (210)

Type of Divisor: Geom. Mean — Intercept = -6.7985
Avail. Land Exponent = 0.52 — R-Square = 0.1830

Independent Var.	Coeff.	t-Statistic	Signif.
SDS	-0.1515	-4.654	<0.01%
PDCA(UH)r	258.89	2.127	4%
DS(UH)	0.7067	1.914	6%

Earnings in Electric, Gas and Sanitary Services (227)

Type of Divisor: Geom. Mean — Intercept = 1.6105
Avail. Land Exponent = 1.00 — R-Square = 0.1388

Independent Var.	Coeff.	t-Statistic	Signif.
SDC	0.2269	2.766	1%
SDS	-0.0912	-4.396	<0.01%
PDCA(LH)r	250.83	1.961	5%

Table 10. REGRESSION RESULTS USED IN ALLOCATION MODEL, PART II

Earnings in Wholesale Trade (207)			
Type of Divisor: Geom. Mean		Intercept = 3.9660	
Avail. Land Exponent = 0.24		R-Square = 0.5456	
Independent Var.	Coeff.	t-Statistic	Signif.
SDC	0.5861	8.139	<0.01%
SDS	-0.2913	-7.344	<0.01%
PDCA(TH)r	196.16	2.119	4%
DS(UH)	2.6523	3.328	0.1%

Earnings in GAFO Retailing (Department-Store-Type Goods) (227)			
Type of Divisor: Geom. Mean		Intercept = 5.2046	
Avail. Land Exponent = 0.27		R-Square = 0.4308	
Independent Var.	Coeff.	t-Statistic	Signif.
CDCA(TH)r	170.12	2.289	3%
SDC	0.5120	3.917	0.05%
CDCA(PE)r	15.651	5.132	<0.01%
SDS	-0.2952	-3.299	0.5%

Earnings in Other Retail Trade (207)			
Type of Divisor: Geom. Mean		Intercept = -0.1452	
Avail. Land Exponent = 0.50		R-Square = 0.5038	
Independent Var.	Coeff.	t-Statistic	Signif.
CDCA(TH)r	301.38	9.008	<0.01%
CDCA(PE)r	4.0321	3.534	0.1%
SDS	-0.0643	-1.978	5%

Earnings in Other Finance (218)			
Type of Divisor: Geom. Mean		Intercept = 4.3567	
Avail. Land Exponent = 0.31		R-Square = 0.3970	
Independent Var.	Coeff.	t-Statistic	Signif.
SDS	-0.4929	-8.445	<0.01%
DS(UH)	2.5255	2.773	1%
PDCA(UH)r	375.33	3.758	0.05%
DS(LH)	2.0714	2.060	4%

Earnings in Real Estate (227)			
Type of Divisor: Exponential		No Constant Term	
Avail. Land Exponent = 0.00		R-Square = 0.6827	
Independent Var.	Coeff.	t-Statistic	Signif.
SDS	-0.6702	-5.868	<0.01%
SDM	-0.0941	-4.624	<0.01%
PDCA(TE)r	1.1080	1.888	6%
CDCA(PE)q	9.8803	2.194	3%
PDCA(UH)q	1058.7	2.633	1%
DS(UH)	0.9238	3.603	0.05%

Earnings in Automotive Retail Trade (227)			
Type of Divisor: Exponential		No Constant Term	
Avail. Land Exponent = 0.43		R-Square = 0.5992	
Independent Var.	Coeff.	t-Statistic	Signif.
CDCA(TH)r	237.53	7.040	<0.01%
CDCA(PE)r	2.9393	3.350	0.1%
PDCA(LH)r	211.53	2.051	5%
DS(MH)	1.4568	2.764	1%
DSA(PE)q	1.3538	2.616	1%

Earnings in Eating & Drinking Places (227)			
Type of Divisor: Exponential		No Constant Term	
Avail. Land Exponent = 0.21		R-Square = 0.6749	
Independent Var.	Coeff.	t-Statistic	Signif.
SDS	-0.4379	-9.230	<0.01%
CDCA(PE)r	3.3784	3.488	0.1%
SDM	-0.0269	-2.462	2%
PDCA(UH)r	509.43	5.238	<0.01%
CDCA(UH)r	458.81	5.063	<0.01%
DMA(LH)p	36.076	2.896	0.5%
PDCA(LH)p	347.63	2.859	0.5%

Earnings in Banks & Credit Agencies (209)			
Type of Divisor: Geom. Mean		Intercept = 0.7029	
Avail. Land Exponent = 0.74		R-Square = 0.3465	
Independent Var.	Coeff.	t-Statistic	Signif.
CDCA(IE)r	11.637	5.644	<0.01%
PDCA(PE)r	22.488	3.731	0.05%
DS(UH)	4.9142	3.582	0.05%
DS(LH)	4.3268	2.877	0.5%

Earnings in Insurance Carriers (218)			
Type of Divisor: Geom. Mean		Intercept = -3.2102	
Avail. Land Exponent = 0.00		R-Square = 0.3217	
Independent Var.	Coeff.	t-Statistic	Signif.
SDC	0.3699	4.854	<0.01%
CDCA(IE)r	3.6016	3.880	0.05%
SDS	-0.1302	-2.347	2%
DS(UH)	0.7366	2.687	1%

Earnings by Insurance Agents & Brokers (227)			
Type of Divisor: Geom. Mean		Intercept = 4.2817	
Avail. Land Exponent = 0.50		R-Square = 0.3995	
Independent Var.	Coeff.	t-Statistic	Signif.
SDC	0.2684	3.207	0.5%
CDCA(PE)r	4.1583	4.921	<0.01%
DS(UH)	0.5455	3.183	0.5%

Table 10. REGRESSION RESULTS USED IN ALLOCATION MODEL, PART III

Earnings in Hotels & Motels (227)			
Type of Divisor: Exponential		No Constant Term	
Avail. Land Exponent = 0.24		R-Square = 0.2409	
Independent Var.	Coeff.	t-Statistic	Signif.
SDM	-0.0373	-4.515	<0.01%
PDCA(PE)r	1.5127	2.346	2%
PDCA(CE)r	1.5827	2.587	1%

Earnings in Auto Repair, Automotive Services and Parking (227)			
Type of Divisor: Geom. Mean		Intercept = 1.6105	
Avail. Land Exponent = 0.16		R-Square = 0.1388	
Independent Var.	Coeff.	t-Statistic	Signif.
CDCA(TH)r	54.714	3.214	0.5%
PDCA(TE)r	1.6709	4.906	<0.01%
DMA(LH)p	5.4420	3.385	0.1%
PDCA(LH)q	143.09	2.498	2%
SDS	-0.1508	-1.979	5%

Earnings in Amus. & Recr. Services (195)			
Type of Divisor: Exponential		No Constant Term	
Avail. Land Exponent = 0.00		R-Square = 0.1723	
Independent Var.	Coeff.	t-Statistic	Signif.
SDS	-0.5140	-3.990	0.01%
SDC	0.6659	3.074	0.5%
DS(MH)	2.6478	3.091	0.5%
CDCA(UH)p	166.57	1.871	6%

Earnings in Health Services (227)			
Type of Divisor: Exponential		No Constant Term	
Avail. Land Exponent = 0.26		R-Square = 0.6306	
Independent Var.	Coeff.	t-Statistic	Signif.
PDCA(CE)r	39.922	12.382	<0.01%
SDS	-0.3292	-10.162	<0.01%
DS(MH)	5.3090	3.014	0.5%

Earnings in Educational Services (227)			
Type of Divisor: Exponential		No Constant Term	
Avail. Land Exponent = 0.21		R-Square = 0.5473	
Independent Var.	Coeff.	t-Statistic	Signif.
SDS	-0.2097	-8.783	<0.01%
DMA(PE)r	0.1952	2.588	1%
PDCA(UH)p	263.54	4.183	<0.01%
DS(UH)	0.6570	4.503	<0.01%

Earnings in Personal Services and Private Households (215)			
Type of Divisor: Geom. Mean		Intercept = 1.6235	
Avail. Land Exponent = 0.15		R-Square = 0.2608	
Independent Var.	Coeff.	t-Statistic	Signif.
CDCA(TH)r	51.615	4.511	<0.01%
DS(UH)	0.4358	4.341	<0.01%

Earnings in Misc. Repair Services (227)			
Type of Divisor: Exponential		No Constant Term	
Avail. Land Exponent = 0.62		R-Square = 0.1666	
Independent Var.	Coeff.	t-Statistic	Signif.
PDCA(IE)r	1.4005	3.967	0.01%
SDS	-0.0741	-2.248	3%
PDCA(LH)r	104.09	2.426	2%

Earnings in Business Services (207)			
Type of Divisor: Geom. Mean		Intercept = -6.5940	
Avail. Land Exponent = 0.09		R-Square = 0.2593	
Independent Var.	Coeff.	t-Statistic	Signif.
CDCA(IE)r	9.3372	3.187	0.5%
PDCA(TE)r	5.0508	2.267	3%
SDS	-0.4250	-3.778	0.05%
DS(UH)	3.8031	3.661	0.05%

Earnings in Legal Services (217)			
Type of Divisor: Geom. Mean		Intercept = -0.7745	
Avail. Land Exponent = 0.17		R-Square = 0.1213	
Independent Var.	Coeff.	t-Statistic	Signif.
PDCA(UH)r	97.447	3.883	0.05%
SDC	0.1027	1.858	6%

Earnings in Social Services, Membership Organizations & Misc. Services (227)			
Type of Divisor: Geom. Mean		Intercept = 2.1284	
Avail. Land Exponent = 0.20		R-Square = 0.4940	
Independent Var.	Coeff.	t-Statistic	Signif.
DMA(MH)p	14.810	3.583	0.05%
SDM	-0.0387	-4.078	0.01%
PDCA(TH)r	71.927	2.069	4%
SDS	-0.3446	-6.850	<0.01%
CDCA(UH)r	151.56	3.223	0.5%
CDCA(PE)r	3.0253	2.932	0.5%
DS(MH)	1.7020	2.875	0.5%

Table 10. REGRESSION RESULTS USED IN ALLOCATION MODEL, PART IV

Earnings in Engineering & Mgmt. Serv. (188)			
Type of Divisor: Geom. Mean		Intercept = 0.8605	
Avail. Land Exponent = 0.38		R-Square = 0.2523	
Independent Var.	Coeff.	t-Statistic	Signif.
SDS	-0.1995	-3.971	0.01%
PDCA(TE)r	3.5521	3.177	0.5%
DS(UH)	1.4048	3.316	0.1%

Earnings in Federal Government (Excluding Military) (221)			
Type of Divisor: Exponential		No Constant Term	
Avail. Land Exponent = 0.00		R-Square = 0.2662	
Independent Var.	Coeff.	t-Statistic	Signif.
SDS	-0.0967	-5.210	<0.01%
PDCA(CE)r	5.3421	3.160	0.5%

Earnings in State Government (223)			
Type of Divisor: Geom. Mean		Intercept = 1.5296	
Avail. Land Exponent = 0.06		R-Square = 0.1016	
Independent Var.	Coeff.	t-Statistic	Signif.
SDC	0.3363	4.218	<0.01%
SDM	-0.0105	-2.198	3%

Earnings in Local Government (227)			
Type of Divisor: Exponential		No Constant Term	
Avail. Land Exponent = 0.63		R-Square = 0.7057	
Independent Var.	Coeff.	t-Statistic	Signif.
CDCA(TH)r	1145.9	11.174	<0.01%
SDC	0.3224	3.863	0.05%

Number of Households in Lower Third of Regional Income Distribution (227)			
Type of Divisor: Exponential		No Constant Term	
Avail. Land Exponent = 0.86		R-Square = 0.7972	
Independent Var.	Coeff.	t-Statistic	Signif.
PDCA(TE)r	0.6829	13.135	<0.01%
SDC	0.2483	6.202	<0.01%
SDS	-0.0387	-6.859	<0.01%
DMA(TH)p	0.6416	2.537	2%

Number of Households in Middle Third of Regional Income Distribution (227)			
Type of Divisor: Exponential		No Constant Term	
Avail. Land Exponent = 0.85		R-Square = 0.6820	
Independent Var.	Coeff.	t-Statistic	Signif.
SDC	0.4390	5.176	<0.01%
SDM1	-0.0059	-5.611	<0.01%
PDCA(IE)r	0.4229	1.947	6%
SDS	-0.1550	-3.207	0.5%
PDCA(UH)r	52.040	2.193	3%
SDCA(PE)r	0.4933	2.355	2%

Number of Households in Upper Third of Regional Income Distribution (227)			
Type of Divisor: Exponential		No Constant Term	
Avail. Land Exponent = 0.72		R-Square = 0.6902	
Independent Var.	Coeff.	t-Statistic	Signif.
SDC	0.8143	10.555	<0.01%
SDM1	-0.0217	-8.959	<0.01%
SDS	-0.1626	-7.590	<0.01%
SDCA(PE)p	1.5226	4.253	<0.01%

In these regressions, the structured nature of the analysis and the deletions of extreme observations caused R-square to average only 0.373 in the economic regressions and 0.723 in the household regressions. The economic situation was not as bad as the former figure would indicate, however, because the lowest R-square values were obtained for relatively unimportant sectors (e.g., farming). For the six most important industries, accounting for 55% of the Charlotte region's total earnings, the average R-square was 0.540. The earnings-weighted average for all industries was 0.452.

As typically found when calibrating allocation models of this type, the R-square values obtained for consumer service activities were far higher on average than those for industrial and producer service functions. This was to be expected partly because consumer service activities are inherently more predictable, given their orientation toward local customers, and partly because they were addressed last in the modeling sequence and thus could be linked to current changes in the two other economic groups plus households.