The Challenge of Forecasting Metropolitan Growth: Urban Characteristics Based Models versus Regional Dummy Based Models

Abstract

This paper presents a study of errors in forecasting the population of Metropolitan Statistical Areas and the Primary MSAs of Consolidated Metropolitan Statistical Areas and New England MAs. The forecasts are for the year 2000 and are based on a semi-structural model estimated by Mills and Lubelle using 1970 to 1990 census data on population, employment and relative real wages. This model allows the testing of regional effects on population and employment growth. The year 2000 forecasts are for 321 MSAs as they were defined in 1990. Actual year 2000 populations for these MSAs are constructed using the MSA components lists for 1990. Forecast errors are constructed for these "historic" MSAs. The forecast errors for the entire set of cities are examined for regional patterns. A subset of 77 cities is examined more carefully using the State of the Nations Cities (SONC) data base prepared by the Center for Urban Policy Research. SONC contains observations on 2000 demographic and socioeconomic variables for all 77 MSAs in the data set. Selected variables will be used to test a model of forecast areas developed for this project to determine if there are systematic relationships between selected variables and the forecast errors and to determine if a semi-structural model based on urban characteristics variables can improve urban population forecasts.

The Challenge of Forecasting Metropolitan Growth:
Urban Characteristics Based Models versus Regional Dummy Based Models

This paper is about forecasting urban growth. Forecasting urban growth is important for both scholarly understanding and practical use. Several literatures relating to urban growth are delineated by Mills and Lubuele (1995). But, up to the time of their work, as they point out, "no study has succeeded in relating empirical MSA growth to parameter changes of formal models of the determinants of MSA size" (Mills and Lubuele, p. 346). They fill this breach with a semi-structural simultaneous equation model which is used to forecast MSA population, employment, and wages for the year 2000. A similar model developed by Leichenko (2001) has been used to separate growth for cities and their suburbs. Data for the year 2000 are now available to check how close the Mills-Lubuele forecasts came to actual values. The first part of this paper examines the errors in population growth forecasts made by Mills and Lubuele for the set of MSAs delineated in the Woods and Poole *1992 MSA Profile*. These errors are examined for regional patterns. The second section presents a simple model associating the forecast errors for those cities included in the State of the Nations Cities (SONC) data set with a set of plausible explanatory metropolitan characteristics (socio-economic and amenity) variables. A third section presents the results of correlating the growth rates for 1990 - 2000 for the set of cities in the SONC data set with the set of metropolitan characteristics variables. A model constructed on the lines of the Mill-Lubuele model but using metropolitan characteristic variables in place of regional dummies is presented in the fourth section of the paper. In the final section conclusions and some directions for further research are discussed.

I. Forecast errors in the Mills-Lubuele model

In a nation like the United States which is already highly urbanized, the percent urbanized increases only slowly between censuses. Thus, growth of particular MSAs can not be well predicted by growth in the national urbanized population. Mills and Lubuele use a semi-structural model to forecast specific MSA populations, employment and wage levels. The forecasts for the year 2000 are published in Table 4 of their article. The Mill and Lubele model has the following structure for each MSA (i) in census years (t):

$$1)\ P_{it} = a_0 + \sum_{j=1}^{7} a_j R_j + a_8 W_{it} + a_9 E_{it} + a_{10} P_{it-1} + a_{11} P_{it-1}^2 + u_{Pit}$$

$$2)\ E_{it} = b_0 + \sum_{j=1}^{7} b_j R_j + b_8 W_{it} + b_9 P_{it} + a_{10} E_{it-1} + a_{11} E_{it-1}^2 + u_{Eit}$$

$$3)\ W_{it} = c_0 + \sum_{j=1}^{7} c_j R_j + c_8 P_{it} + c_9 E_{it} + c_{10} W_{it-1} + u_{Wit}$$

where P is population, E is employment, W is earnings per worker, and the R_j's are regional dummies for each of the census regions except New England which is taken to be the default region. The lagged squared terms allow for the possibility of either increasing or diminishing returns to urban size. The coefficients are estimated using census data from 1970, 1980, and 1990 employing Box-Cox estimation. These coefficients are then used to estimate the dependent variables for the year 2000. The model incorporates both "jobs-follow-people" and "people-follow-jobs" relationships and the notion that while high wages attract population they deter job growth. The calculated coefficients confirm all except the last contention. Mills and Lubuele

contend that this relationship may be masked by the fact that city size and wages are positively correlated. It is clear that this model omits many influences on relative urban growth rates. The regional dummies should pick up the influence of regional patterns in these omitted variables. If the regional dummies do this successfully then, whether large or small, the forecast errors should not exhibit a regional pattern. The fact that the errors do have a strong regional pattern motivates the research presented here. That is, there must be omitted variables with strong regional patterns themselves whose inclusion would have improved the forecast over simply using dummies for regional effects.

The Data

The first step required in calculating the forecast errors for the year 2000 was recreating the structure of the MSAs of 1990. The Bureau of Census always presents MSA data with backward compatibility but not with forward. That is, the census data for the year 2000 is for MSAs as revised successively in 1993, 1996 and 1999 but the Mill-Lubuele forecasts were for metropolitan areas as defined before the 1993 revisions. *Metropolitan Areas and Components, 1990 with FIPS Codes* and *Metropolitan Areas and Components,1999 (*Metropolitan areas as defined by the Office of Management and Budget) were used to exclude counties (towns and centers in New England) added by revisions in the 1990s and to include counties that had been deleted in those revisions. In some cases counties had been moved from one MSA to another. These counties were restored to their 1990 MSAs. The result was a set of 319 MSAs with year 2000 population estimates for 1990 configured boundaries[1]. The appendix presents the population for year 2000 of the reconstructed MSAs that existed in 1990 (even if they were no longer in existence in 2000), the corresponding Mills and Lubuele forecast populations, and the

forecast errors for those MSAs and PMSAs. It excludes MSAs that came into existence during the 1990's. Table 1 summarizes the population changes.

Table 1: MSA Growth Data Summary

	1990 MSA Population	2000 using 1990 boundaries	2000 using 2000 boundaries	2000 Mills and Lubuele Forecast
Population	193,072,813	219,946,375	227,330,909	206,233,500
Growth from 1990		26,873,562	34,258,096	13,160,687
Percent growth rate		13.92	17.74	6.82
Net difference due to boundary change			7,384,534	
Due to adding counties or towns			8,517,694	
Due to dropping counties or towns			-1,133,160	
1990 boundary forecast error				-13,712,875

PMSA and MSA Population in 1990 totaled 193,072,813. The 'same name' set of cities had grown to 227,330,909 in the 2000 census (there were also several totally new MSAs created in the 1990s that are omitted). Of the increase in these MSAs populations, 8,517,694 was due to adding counties (towns and centers in New England) while 1,133,160 people resided in counties that lost metropolitan status. The net change due to boundary changes was 7,384,534 leaving 219,946,375 as the population living within the boundaries of MSAs as defined in 1990. Mills and Lubuele forecast that these cities would grow a total of 6.82 per cent. They actually grew by 13.92 per cent. The raw underestimate was 13,712,875. The growth rate extremes inside the 1990 boundaries were +85.5% (Las Vegas, Nevada) and -21.7% (Battle Creek, Michigan - now part of the Kalamazoo MSA). Table 2 shows the average growth rates for cities for each census region inside 1990 boundaries.

Table 2. Regional MSA growth rates

Region	MSA Count	Average MSA growth rate
New England	16	3.4%
Mid-Atlantic	36	4.8
South Atlantic	54	17.4
East South Central	23	11.5
West South Central	45	15.3
East North Central	61	7.5
West North Central	25	13.2
Mountain	20	30.4
Pacific	39	18.6

Mills and Lubuele recognize that their forecasts for population growth were conservative compared to other forecasts. The appendix indicates just how conservative the forecasts were. The reported forecast error is defined as the forecast value minus the actual value expressed as a percent of the forecast value. Nearly all of the forecast errors are negative and the average forecast error is -10 percent. However, the forecast errors are not randomly distributed by region While Mills and Lubuele use dummy variables to incorporate regional effects in their forecasts, it appears that there is still a strong regional pattern in their forecast errors. Table 3 presents a regression of the forecast errors on a set of regional dummies using New England as the default region. Heteroskedasticity persists even though the dependent variable is a percent error thus weighted least squares were used. An examination of the standard errors of the coefficients for the South Atlantic, East South Central, West South Central, East North Central, and West South Central suggests these coefficients are not significantly different from each other. Thus, a second regression was run making these regions collectively the default region. This regression is shown as Table 4. The adjusted R-squared and the Schwarz criterion both change in the desired direction but not significantly.

Table 3: Regional Pattern of the Population Forecast: All 1990 MSAs

Dependent Variable: Percent forecast error (PCERROR)
318 observations
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.	Mean % Error
C	13.183	4.743	2.780	0.0058	13.183
Mid-Atlantic	-17.424	4.821	-3.614	0.0004	-4.241
South Atlantic	-22.672	4.857	-4.668	0.0000	-9.489
East South Central	-26.314	5.150	-5.110	0.0000	-13.131
West South Central	-26.561	4.897	-5.424	0.0000	-13.378
East North Central	-27.797	4.925	-5.644	0.0000	-14.614
West North Central	-24.422	4.933	-4.951	0.0000	-11.239
Mountain	-39.429	5.683	-6.938	0.0000	-26.246
Pacific	-17.888	4.931	-3.628	0.0003	-4.705

R-squared	0.3877	Adjusted R-squared	0.3718
S.E. of regression	9.5124	Sum squared residuals	27960.12
Durbin-Watson stat	1.6323	Schwarz criterion	7.4774
F-statistic	24.4537	Prob(F-statistic)	0.0000

Table 4: Revised Regional Pattern

Dependent Variable: Percent forecast error
Included observations: 319
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.	Mean % error
All other Regions	-12.446	0.621	-20.030	0.0000	-12.446
New England	26.861	4.620	5.8136	0.0000	14.415
Mid-Atlantic	8.206	1.060	7.7436	0.0000	-4.24
Mountain	-13.800	3.173	-4.3495	0.0000	-26.246
Pacific	7.741	1.478	5.2372	0.0000	-4.705

R-squared	0.3861	Adjusted R-squared	0.3783
S.E. of regression	9.6383	Sum squared resid	29169.41
Durbin-Watson stat	1.6157	Schwarz criterion	7.4439
F-statistic	49.37165	Prob(F-statistic)	0.0000

Regional variation account for about 39% of the variation in the errors of population forecasts The mean forecast error for the entire sample is -10.092%, confirming the conservative nature of the Mills and Lubuele population forecasts However, the regression results show that their model systematically overestimated population growth in New England while substantially underestimating growth in the Mountain States. Whether additional variables could have reduced

the size of those errors is explored in the next section of the paper.

II. Modeling Forecast errors using an urban characteristics model

The regression results imply that some important regional effects are not being picked up by the regional dummies in the Mill-Lubuele model. The strong regional variation in the forecast errors implies that there are variables with strong regional variance that are associated with growth rates. These variables may be socioeconomic such as racial, age, or human capital characteristics of the populations; industrial structure of the local economy and the local business climate; or they may be amenity variables such as climate, crime rates, and traffic congestion. In order to be useful, a variable must not only be associated with economic growth but also to have a high regional variation. This section presents a simple linear model associating the absolute percentage error in the Mills-Lubuele forecast with a set of urban characteristics including employment, human capital, distribution of income, housing segregation, business climate and amenities. Each of these variables may explain part of the actual growth of the MSAs. The hypothesis tested is that inclusion of these variables would have improved the forecast.

The data used for this part of the study was drawn from the State of the Nations Cities Data Base (CUPR, 1998). This data set includes a variety of demographic, economic, political, and amenity variables for 74 metropolitan areas with at least one drawn from each state. The regression on dummy variables was repeated for this set of MSAs to see how they differed from the population of MSAs presented in Table 3. Most importantly, the Mills-Lubuele forecasts for the New England and Pacific regions MSAs that remain in the data set are much more accurate (the forecast errors are not significantly different from zero). Forecast errors in the remaining regions are still significant with the largest errors are in the Mountain Region. Table 5 presents

the results.

Table 5 Regional affects in the SONC cities.

Dependent Variable: Percent forecast error (PCERROR)
Included observations: 74

Variable	Coefficient	Std. Error	t-Statistic	prob.
Constant	2.754	3.873	0.7112	0.4795
Middle Altantic	-11.497	5.745	-2.0013	0.0495
South Atlantic	-8.693	4.815	-1.8055	0.0756
East South Central	-10.976	5.745	-1.9106	0.0605
West South Central	-11.146	4.899	-2.2752	0.0262
East North Central	-8.825	5.124	-1.7224	0.0898
West North Central	-11.414	5.124	-2.2277	0.0294
Mountain	-27.475	5.124	-5.3626	0.0000
Pacific	-4.967	4.683	-1.0607	0.2927
R-squared	0.3702		Adjusted R-squared	0.2927
S.E. of regression	9.4870		Mean dependent var	-7.5933
Durbin-Watson stat	1.8498		Schwarz criterion	7.7315
F-statistic	4.7755		Prob(F-statistic)	0.00012

III. Urban Characteristics and Patterns of growth from 1990 to 2000

There are a multitude of variables that have been associated with urban growth. More than 3000 of these are available in the SONC data set. Those chosen for inclusion here represent several classes of variables– employment patterns, human capital, labor climate, and amenities. For example, the production sector is represented by the percent of the central city workforce employed in manufacturing (CCPCMANU). This percentage runs from a high of 16.6% in the East North Central region to a low of 7.9% in the Mountain region. Given the transformation of the economy to a service one, the expected correlation of CCPCMANU with urban growth is negative. Including this variable would have lowered forecast growth in the East North Central and New England States where it is highest and raised forecast growth in the Mountain region. On average it would be expected to reduce the absolute value of the forecast errors. In general, a variable with strong regional variation that is positively related to economic growth and has its

highest value in the Mountain States and lowest in the New England would be expected to reduce the absolute value of forecast errors. To see if this is true two simple linear models of the form

4) $PC9020 = \alpha X + \varepsilon$, and

5) $APCERROR = \beta X + \varepsilon$

where PC9020 is the growth of population in the 1990 boundaries between 1990 and 2000, APCERROR is the absolute value of the percent error in the Mill-Lubuele forecast and X is a vector of the variables listed in Table 6. The regression results are shown as Table 7.

Table 6 Description of the variables

Variable	Description & Notes
PCERROR (Dependent)	100*(Forecast Population.- Actual Population)/Forecast Pop.
PC9020 (Dependent)	Percent population growth inside 1990 boundaries from 1900 to 2000
CC9COLL	1990 Percent of the central city workforce with college degrees.
CC9HSCH	1990 percent of CC workforce with a high school degree.
CCPCMANU	1990 percentage of the central city workforce employed in manufacturing
CS9PCIR	1990 Suburban Per capita income divided by central city personal income.
DWB9	1990 Dissimilarity index for Whites and Blacks measures the degree of housing segregation (0 = no segregation)
ED91CEXP	1991 Education expenditure per pupil in public schools
EVL9093	Employment volatility between 1990 and 1993
FC95VCRI	Violent crimes per 100000 population
GMPX809	Gross Metropolitan Product Growth rate from 1980 to 1990
M9STID	Metropolitan area Stress Index. [2]
NOCCDD	Number of cooling days – average year
NOCHAUG	Humidity at noon averaged over the month of August
NOCHDD	Number of heating days – average year
ST83UNCOV	1983 Percent of labor force covered by Union contracts.

Table 7 Regression results for absolute forecast errors and 1990-2000 growth

Included observations: 56					Simple Correlations
Dependent Variable:	APCERROR		PC9020	PC9020	with growth
Variable	Coefficient	Prob.	Coefficient	Prob.	
Constant	-21.108	0.4060	-13.601	0.6774	
CC9COLL	-0.4292	0.2571	-0.7627	0.2453	0.1432
CC9HSCH	0.5776	0.0345	0.8904	0.0360	0.2660
CCPCMANU	-0.1016	0.7515	-0.5227	0.2367	-0.3733
CS9PCIR	-4.8549	0.3521	3.1511	0.6800	-0.3109
ED91CEXP	-0.0002	0.8423	-0.0019	0.1645	-0.3588
EVL9093	-2.5974	0.5238	0.9305	0.8591	0.0770
FC95VCRI	0.0023	0.1130	0.0036	0.1175	-0.1062
GMPX809	1.9260	0.0456	4.7553	0.0024	0.5388
M9STID	-8.1087	0.1807	5.1519	0.5372	0.1766
NOCCDD	0.0010	0.5447	0.0010	0.7613	0.4369
NOCHAUG	-0.4118	0.0083	-0.7624	0.0006	-0.5567
NOCHDD	0.0005	0.6257	-0.0014	0.3741	-0.3674
ST83UNCOV	0.1004	0.5335	-0.2577	0.3136	-0.2366
DWB9	0.2691	0.0156	0.3466	0.0572	-0.3031
R-squared	0.4654		0.7005		
Adjusted R-squared	0.2828		0.5982		
F-statistic	2.5492		6.8483		
Prob(F-statistic)	0.0100		0.0000		
Durbin-Watson stat	2.1310		2.2133		

Interpretation

The simple correlation coefficients between growth and the explanatory variables have the expected signs, However, percent of the Central City population that is college educated, expenditures per pupil, violent crime rates, and employment volatility are not statistically significantly correlated with growth in this period. The percent of central city residents with high school educations is highly correlated with the percent with college educations so it could be that the impact of variation in college educations is masked. But the regional pattern of CC9COLL also contributes to its inability to improve Mills-Lubuele type growth forecasts. With the exception of the Middle Atlantic states and the East North Central States (which have significantly lower values) there is no significant variation in the percent of college educated

people living in central cities of MSAs. While it is true that both of these regions exhibited slow growth, New England, with the second highest percent of college educated central city residents had the absolute slowest rate of population growth. In contrast, the Mountain Region has the highest proportion of highschool graduates residing in central cites (again the Middle Atlantic and East North Central have the lowest percentages). Thus, CC9HSCH, is both significantly correlated with growth and appears to have the right regional pattern but is positively associated with the absolute size of the forecast errors.

The non-amenity variable that best fits the hypothesis of this paper is the ratio of suburban per capita income to central city per capita income (CS9PCIR). It is the most highly regional of the non-amenity independent variables and is strongly negatively associated with economic growth. The regional pattern also is such that the highest ratios are in the slowest growing regions. The regional pattern for this variable is shown in Table 8.

Table 8. Regional pattern of the ratio of suburban to central city per capita income

Dependent Variable: CS9PCIR
Included observations: 73
Excluded observations: 1 (Anchorage, Alaska)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
New England	1.2774	0.1122	11.382	0.0000
Middle Atlantic	0.2806	0.1665	1.6856	0.0967
South Atlantic	-0.1655	0.1395	-1.1864	0.2398
East South Central	-0.0806	0.1665	-0.4841	0.6299
West South Central	-0.2985	0.1420	-2.1027	0.0394
East North Central	0.1562	0.1485	1.0523	0.2966
West North Central	-0.2054	0.1485	-1.3836	0.1713
Mountain	-0.3094	0.1485	-2.0840	0.0412
Pacific	-0.0993	0.1374	-0.7227	0.4725
R-squared	0.3115	Mean dependent var		1.1696
Adjusted R-squared	0.2254	S.D. dependent var		0.3123
Log likelihood	-4.5084	F-statistic		3.6190
Durbin-Watson stat	2.3626	Prob(F-statistic)		0.0016

Not surprisingly the amenity variables show very strong regional variation. The greatest regional variance is shown by average humidity at noon in August. However, nearly all of the variance occurs because of the aridity of the mountain states. Given that those were the fastest growing states it is also not surprising that NOCHAUG is both highly correlated with economic growth and negatively correlated with the absolute forecast error. But, NOCHAUG appears to mask the effects of heating and cooling days as measures of amenities. Removing NOCHAUG from the equation results in both NOCCDD and NOCHDD becoming significant above the 5 per cent level with positive signs in the growth equation, but also positive associations with the absolute forecast error. The Mountain State Region stretches from the Canadian Border to the Mexican border so it is difficult to compare heating days and cooling days with regions to the east of it. It is also clear that the mountains provide other amenities not measured by these variables. Nonetheless, it appears that the inclusion of amenity variables with high regional variation would improve the forecasts of Mills- Lubuele type models.

IV. Forecasts with a semi-structural model with urban characteristics variables.

This section presents a simultaneous equation model of population, employment and wages for a subset of the cities in the State of the Nation's Cities database. The objective is to develop a model parallel to that of Mills and Lubuele but with the regional dummies replaced with seven metropolitan characteristic variables from the SONC database. The form of the model is

$$6)\ P_{it} = a_0 + a_1 E_{it} + a_2 W_{it} + a_3 P_{it-1} + \sum_{j=4}^{10} a_j S_{ijt-1} + u_{Pit}$$

$$7)\ E_{it} = b_0 + b_1 P_{it} + b_2 W_{it} + b_3 E_{it-1} + \sum_{j=4}^{10} b_j S_{ijt-1} + u_{Eit}$$

$$8)\ W_{it} = c_0 + c_1 P_{it} + c_2 E_{it} + c_3 W_{it-1} + \sum_{j=4}^{10} c_j S_{ijt-1} + u_{Wit}$$

where S_{ijt-1} is the value of the jth metropolitan characteristic variable in the ith city in period t-l.

The squared values of population and employment are not used in this model. They were not found to be significant in the Mill-Lubuele results. This confirms other research which also indicates that most cities are not affected by diminishing or strong increasing returns to scale. The SONC data set is for metropolitan areas as defined in the 1996 BEA revisions. Thus, population, wage, and employment data are not directly comparable to those used by Mill and Lubuele. Seven metropolitan characteristic variables were chosen from the SONC data set listed in Table 6 to represent racial disparity, employment patterns, human capital, income disparity between suburb and central city, and amenity values. In order to be included in the data set a variable had to have values for 1970, 1980, and 1990. This put restrictions on which variables could be used and limited the number of cities that could be included in estimating the model. Values of these variables at the beginning of each decade are used to explain population and employment at the end of each decade. Complete and consistent data were available for 58 of the 77 cities in the data set. The regression results in Table 9 are for those 58 cities. There are, then, two end of decade observations for population, employment and wages. A pooled data set was created in which each city is a separate cross-section. Two-stage least squares was used to estimate each of the equations using the beginning of the decade values of the urban characteristic

variables as cross section specific instruments. Autocorrelation was not found to be a significant problem so no transformation seemed necessary[3]. As indicated above, heteroskedasticity is present in the data. The system of equations were first estimated using Generalized Least Squares with cross section weights. The resulting residuals were then checked for contemporaneous correlations and significant correlations were found between the residuals for the population and employment equations for both decades. Thus, the system was re-estimated as periodwise seemingly unrelated regressions. These results are posted in Table 9. The asterisks in the variable names are decade place holders. The signs of the significant variables in the population equation are as expected. The significance of some of the coefficients are sensitive to the weighting used to correct heteroskedasticity. In particular, the coefficient of the percent of the central city labor force in manufacturing hovers around 10% in all specifications. The positive sign for manufacturing may suggest a watershed break between the 1970-1990 period and the 1990s during which the percent employment in manufacturing is negatively correlated with population growth The significant negative sign for heating days reflects the southward and southwestward movement of population after 1970. However, heating days has no corresponding relationship to employment. The strong negative relationship between employment and the lagged value for employment is a mystery. It and the non-relationship between heating days and employment seem to imply that "jobs follow people" is the dominant force in the two decades examined.

Table 9: Simultaneous Equation Regression results

	Equation					
	Population		Employment		Wages	
Variable	Coefficient	t-Statistics	Coefficient	t-stats	Coefficient	t-stats
Constant	71089.28	0.416	-59416.98	-0.363	-906.398	-0.796
POP			0.591740	9.757***	-0.000225	-0.552
EMP	0.896	13.101***			0.000644	0.725
WAGE	-1.468	-0.145	26.13713	2.563***		
	Period Specific Instruments					
POP(-1)	0.599663	18.522***				
EMP(-1)			-0.308380	-2.112**		
WAGE(-1)					1.7130	8.390***
	Cross Section Specific Instruments					
CPCMANU	3771.52	1.571	-1375.89	-0.579	-12.387	-0.759
CC*HSCH	3914.09	1.556	-2375.44	-0.959	18.677	1.014
NOCHAUG	-587.76	-0.423	-131.72	-0.097	15.842	1.770*
NOCHDD	-40.40	-4.602***	7.90	0.831	-0.0021	-0.038
CS*PCIR	-100234.70	-1.476	-43532.83	-0.652	924.528	1.963**
DWB*	-1617.56	-1.855*	127.85	0.137	-6.545	-0.686
FC*VCRI	38.61	0.639	-42.51	-0.687	0.181	0.384
R^2	0.996		0.983		0.952	
D-W Stat	1.864		1.942		2.625	

Table 10 shows forecasts for population, employment and wages for selected cities using the metropolitan characteristic variable model. The percentage forecast errors for population are given. These are calculated as the forecast population minus the 2000 actual population in the metropolitan area as defined in 1999 as a percent of the forecast population. The forecast errors from the Mill-Lubuele model are reproduced for comparison purposes. Because of boundary changes these are not all directly comparable. However, Baltimore, Denver, and Los Angeles were defined with the same components in 1990 and 2000, thus, these pairs of errors are directly comparable. The metro-characteristic model provides more accurate forecasts for two of these cities. The results suggest that the socioeconomic model tends towards positive forecast errors whereas the errors for the Mill-Lubuele model are preponderantly negative. The large error for Albuquerque indicates that there is much room for improvement for this type of modeling. Albuquerque has a low number of heating days and one of the most integrated housing markets in

the United States and the model weights both of these factors heavily. Data for employment and wages for year 2000 inside the 1990 metropolitan boundaries were not available for this study so comparisons of errors of the two models for these variables is not possible

Table 10: Forecasts of the Socio-economic model for selected cities.

	Urban Characteristic Model				
City	Population Forecast	Error	Mills-Lubuele Model Error	Employment Forecast	Wage Forecast
Albuquerque	1,411,562	49.5%	-6.08%	1,073,815	$24,798
Atlanta	3,790,927	-8.5%	-16.7%	2,225,047	31,448
Austin-San Marcos	1,798,541	30.5%	-20.6%	1,296,974	26,853
Baltimore PMSA	3,080,243	17.1%	0.1%	1,944,684	31,665
Cleveland	2,596,874	13.3%	-3.5%	1,615,128	28,069
Denver	2,389,373	11.7%	-19.1	1,640,123	30,924
Las Vegas	1,787,195	12.5%	-44.9%	1,338,813	27,587
Los Angeles	9,776,328	2.6%	4.8%	4,852,143	30,849
Minneapolis	3,151,914	5.8%	-4.3%	1,921,474	31,697

V. Conclusions

The semi-structural model used by Mills and Lubuele created forecasts of Metropolitan Area populations that were short of actual population by an average of 10 per cent. More importantly, the forecast errors have a distinctly regional pattern. Regional variation in the relationships of employment to population to wage differentials plus the use of regional dummies by themselves do not seem to pick up all of the regional influences on growth. If they did, the regional pattern of forecast errors should be random. Their model clearly omits variables that may be important in explaining urban growth. Including any one of these variables could

improve the overall forecast. But, unless it had significant regional variation, it would not change the pattern of forecast errors. This paper suggests several variables that are both correlated with urban population growth and have strong regional variation. Such variables can be either positively (high school education levels) or negatively (per cent of employment in manufacturing) associated with growth and still improve the regional forecast pattern. This paper only begins the process of creating a methodology for finding such variables using data from the State of the Nation's Cities database. This database is not a random sample of the nation's metropolitan areas because of the desire to have at least one MSA from every state (probably creating a downward bias in the average size of included cities. Furthermore, it allows comparison of PMSAs in only three of the nations consolidated MSA's. Nonetheless, the included variables can shed light on the regional patterns of urban growth. The fact that the urban characteristics based model generated smaller errors for several cities suggests that this is a fruitful area for further research and could improve urban growth forecasts.

Endnotes.

1.New England MSA's pose a difficult problem as they are constructed of towns and centers. Woods and Poole constructed MSAs that were slightly different from census NECMSAs with one except which was New Bedford, MA. New Bedford was dropped from this study because the Woods and Poole concept was not compatible with the census concept.

2.The urban stress index indicators provide measures of the relative hardship experienced by urban regions. Each index is a composite of variables in five categories: employment (unemployment rate, labor force participation rate); demographic (single-parent family households as share of all households; dependency ratio); education (high-school, college graduation rates); housing (ratio of median rent to median household income, rate of excessive housing expenditures); social (death and crime rates); and income and poverty (city-to-suburb per capita income ratio, poverty rate, and gini coefficient of household income inequality).

3.That is, these results are equivalent to assuming $\lambda = 1$ in a Box-Cox transformation. Since there is no objective criteria for choosing λ, Mill and Lubuele choose the value that maximized R-squared. For the cities in this data set choice of λ had little effect on R-squared.

References

Center for Urban Policy Research, 1998, ***State of the Nations Cities: a Comprehensive Database on American Cities and Suburbs***, Rutgers University, New Brunswick, NJ

Leichenko, Robin, 2001, "Growth and Change in U. S. Cities and Suburbs," ***Growth and Change***, 32, No. 3, 326-354

Mills, Edwin S. and Lubuele, Luan' Sende, 1995, "Projecting Growth of Metropolitan Areas," ***Journal of Urban Economics***, 37, 344-60

Office of Management and Budget, 1990, ***Metropolitan Areas and Components, 1990, with FIPS Codes***, http://www.census.gov/population/estimates/metro-city/90mfips.txt

Office of Management and Budget, 1999, ***Metropolitan Areas and Components, 1999, with FIPS Codes***, http://www.census.gov/population/estimates/metro-city/99mfips.txt

Woods and Poole, 1992, ***1992 MSA Profile***, Woods and Poole, Inc. Washington D.C.