Regional Science and Urban Economics 42 (2012) 561–568

Contents lists available at SciVerse ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

Yet even more evidence on the spatial size of cities: Urban spatial expansion in the US, 1980–2000

Kurt Paulsen *

Department of Urban and Regional Planning, University of Wisconsin-Madison, 925 Bascom Mall, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 10 May 2011
Received in revised form 31 January 2012
Accepted 6 February 2012
Available online 18 February 2012

JEL classification:
R14
R52

Keywords:
R14 Land use patterns
R52 Land use and other regulations

ABSTRACT

This paper expands empirical testing of the predictions of the standard monocentric urban model to examine the size of urban spatial expansion for all US metropolitan regions for the years 1980, 1990 and 2000. Until recently, the lack of temporally and spatially-consistently interpreted data on urban extent for U.S. metropolitan areas has limited the ability to estimate panel models of land use change. This paper first describes the creation of a consistent data set on urbanized land area for all U.S. metropolitan areas. It then estimates cross-sectional models of the determinants of urban spatial extent. For purposes of comparison, results are compared to temporally-consistent satellite-derived measures of developed land area from the National Land Cover Database. The paper then presents panel-model estimates of changes in urban land area, finding that both population and income elasticities are substantially less than unity. Tests for the structural stability of urbanized land area determinants across metropolitan size categories and polycentric regions are presented.

1. Introduction

Measuring and understanding the drivers of the spatial expansion of urban land development is of great interest not only to urban and regional economics, but also to fields as diverse as landscape ecology, water quality, transportation, and climate change (Kalnay and Cal, 2003). Climate models, infrastructure spending, transportation models, and water quality models all involve some projection of future land uses, densities, and the spatial patterns of urban development—all without any empirical consensus as to relative elasticities of urban land demand. For example, a recent National Research Council (2009) report presented a number of future urban density and land scenarios in its climate model, with no agreement among experts as to what future urban land use would be, while suggesting the need for additional research to estimate finer-grained changes in urbanized land areas in metropolitan areas.

The traditional monocentric "Alonso–Mills–Muth" (AMM) model of urban spatial structure continues to serve as a core theoretical and empirical concept of urban land development within and across metropolitan areas because of its elegance, simplicity, and empirical tractability (Brueckner, 2001; Malpezzi and Guo, 2001; McMillen, 2006). Because of the difficulty in acquiring historically consistent measures of urbanized land area extent, however, the model has only been empirically tested a handful of times. For the US, Brueckner and Fansler (1983) first estimated the monocentric model with cross-section data on 40 urban counties and urbanized area data from 1970. McGrath (2005) estimated a model with data from 33 large US cities for the time period 1950–1990. Wassmer (2006) and Spivey (2008) (separately) estimate cross-section models using the size of census-defined "urban areas" in the US using data from 2000. The empirical estimation of monocentric models has also recently been applied to cities in China by Deng et al. (2008), following Brueckner, Fansler and McGrath's methodology.

Because of the great interest in many disciplines in measuring and modeling the causes of metropolitan spatial expansion (sometimes called "sprawl"), this paper contributes to the empirical literature on the determinants of urban expansion. It estimates the monocentric city model with a unique data set covering nearly all US metropolitan areas for the time periods 1980, 1990, and 2000. The extension of data on urban expansion to cover over 300 metropolitan areas and multiple time periods allows estimations in a panel framework. Until recently, the lack of temporally and spatially-consistently interpreted data on urban extent for US metropolitan areas has limited the ability to estimate panel models of land use change. These rich data and methods also allows us to confirm whether or not previously estimated relationships hold over all metropolitan areas, and to test whether the underlying determinants of city-size expansion vary across city size categories and regional polycentricity. For purposes of robustness checks with previous research, this paper also presents estimates utilizing a consistently-interpreted satellite-derived land cover product not previously used in most economic research estimating city-size models.

* Tel.: +1 608 262 8990.
E-mail address: kpaulsen@wisc.edu.

0166-0462/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.regsciurbeco.2012.02.002

This paper is organized as follows. In the first section, I briefly review the now standard urban economic theory which explains city size and review the empirical literature. I then outline the construction of different measures of the spatial extent of urban land area. The next three sections present empirical estimates and elasticities of the determinants of urbanized land area. Section 4 presents simple cross-section estimates utilizing two different measures of urban land area, one produced by Census-based population density definitions of "urban" and one from a satellite product measuring "developed" land area. Section 5 presents estimates from a fixed-effect panel model. In Section 6, I present tests for structural breaks in the model of land development based on city size categories and based on a measure of polycentricity. Conclusions and discussion follow.

2. The theory and empirics of city size

Although there have been many contributors to the development of the traditional bid-rent model of urban spatial structure, the model is commonly called the Alonso–Mills–Muth or the Mills–Muth model after its early developers (Mills, 2000). Although the purpose of the model was to explain the equilibrium distribution of land uses within a city as a function of land rents (which decline with distance to the CBD), the model makes strong and empirically estimable predictions about the size of the developed land area in a city by means of predictions about $\bar{x}$, the radius of the urban area. As cited in Brueckner and Fansler (1983), McGrath (2005) and Spivey (2008), Wheaton (1974) developed the comparative statics for Alonso's model for the spatial extent of urbanized area within a city.

Letting $\bar{x}$ represent the radius of the urbanized area, the developed land area of a city, A, is given by $A = \pi\ \bar{x}^2$. The population of the area is given by P, y represents income, r_a represent agricultural land rent, and t represents per-mile commuting costs. Wheaton (1974) showed that, under the assumptions of the monocentric model with monetary commuting costs

$$\frac{\partial A}{\partial P} > 0, \frac{\partial A}{\partial y} > 0, \frac{\partial A}{\partial r_a} < 0, \frac{\partial A}{\partial t} < 0$$

That is, developed land area is increasing in population, because additional households require additional land. Land area is increasing in income as land demand is income elastic. Likewise, increases in agricultural land prices should shift $\bar{x}$ closer to the CBD as land on the fringe becomes more costly to develop, reducing the size of the urban area. Increases in per-unit transportation costs are likewise expected to reduce the size of the city as households make equilibrium tradeoffs between lot-sizes and commuting times (Glaeser and Kahn, 2004).

Empirical estimation of the predictions of the monocentric city model can be undertaken in either one of two ways (McMillen, 2006): estimating changes within one city over time, or estimation across a sample of cities. Although the theory originally predicts what should happen within a city as a function of changes within that city (in population, income, agricultural land prices), it has most often been estimated using cross-section data on a sample of cities. Measured at any one point in time and fitting the model with cross-section data should demonstrate that, *ceteris paribus*, cities with higher income and higher population would be expected to have a larger urbanized land area, while cities with higher land costs and per-unit transportation costs should show smaller urbanized land areas. Indeed, cross-section studies such as Brueckner and Fansler (1983), Wassmer (2006) and Spivey (2008) do find that the monocentric model fits the data quite well. Population, income and land price variables in these studies almost always have the expected sign and explain a substantial portion (at least 80%) of the variation in urbanized land area across a sample of cities. Wassmer (2006), however, finds agricultural land values to be statistically insignificant and Spivey (2008) finds a negative income elasticity in some specifications. In most specifications across the research, proxies for per-unit transportation costs were not found to be significant predictors in cross-city variation in spatial size.

The alternative empirical approach is to examine changes within a city over time. This approach could involve acquiring data for a city or cities over a long series of time points, as shown by McMillen (2006) who presents data for Chicago from 1860 to 2000. Or, this could involve combining cross-sectional and time-series data to estimate a fixed-effects panel model on a sample of cities over time. McGrath (2005) does this, fitting a pooled cross-section, time series model with 33 cities over 5 decadal observations (1950–1990). In this paper, I present a panel model for over 300 cities with 3 decadal observations (1980–2000). The interpretation of coefficient estimates in a city fixed-effects panel structure of a monocentric model would be the extent to which changes in a city's population, income or land prices would be expected to result in changes in urbanized land area.

3. Data construction

Within the general urban model described above, land is converted into "urban" or "developed" uses at the urban fringe as a function of distance to the city center, land demand (income, population, travel costs), and agricultural rents/land values. However, the appropriate measure of what land areas count as "urbanized" or "developed" across cities has been the subject of much debate and different approaches from different research traditions. The methodology most commonly seen approach within the economics literature (based on Census Bureau definitions and data) is to define "urban" based on a population density threshold. A population-density-derived measure of "urban" or land represents the idea that cities are agglomerations of people. The advantage of such a definition is not only its conceptual ability to measure urbanization, but its practical implementability. For years, the only data broadly available were from the Census and its computation of "urban areas."

One of the possible biases introduced into spatial-extent estimates from using population-density-derived threshold measures of urbanization is that lower-density development on the exurban fringe or in traditionally rural areas is not measured as "urban," even when this constitutes a significant amount of physical land conversion. Rural and exurban housing construction can bring significant changes to physical landscapes and development patterns in ways that some might consider to be part of the process of urbanization (Esparza and Carruthers, 2000; Newburn and Berck, 2006; Radeloff et al., 2005; Wu and Cho, 2007). Much of the more recent research within bio-physical sciences and ecology points out that population density measures significantly understate the amount of land development (Radeloff et al., 2005; Schneider et al., 2010)

However, with more recent advances in geospatial and remote sensing (satellite) technologies, many fields in the geographical and environmental sciences measure "urban" or "developed" lands based on the criteria of physical alteration of land cover wherein vegetative cover is removed and/or impervious surfaces are created. Using high-resolution satellite imagery, researchers with specialized classification algorithms examine changes in landscape patterns. Land which is not "developed" can be identified by the spectral signature of the vegetation. Land which has been converted from its natural state, or "developed" can also be detected by different spectral signatures. Researchers have developed a wide variety of decision-tree classification algorithms to classify land cover as developed/not developed, which also involves using "training sites" (of known urban land) and post-estimation quality control and accuracy assessments. With the increasing availability of satellite data and the processing power of computers, measurement and detection of changes in urbanized areas has greatly improved (Schneider and Woodcock,

2008; Schneider et al., 2010). Satellite-interpreted data thus offers the potential for more fine-grained measures of changes in land use and also allow for temporally-consistent measures of change.

Utilization of satellite data for measurement of urban land use change within the economics literature has been limited but is increasing (Vesterby and Heimlich, 1991; Burchfield, et al., 2006). One caution for urban social science researchers in using satellite data is the comparability of data from different time periods when the underlying satellite technology (resolution, mapping unit size) and/or interpretation methodology has changed. Burchfield, et al. (2006), for example, utilize satellite data from different time periods at different and incompatible resolutions.

One of the contributions of this paper is its presentation of estimates of urbanized land area (spatial extent of city growth) using both population-density and satellite-land-cover approaches. Comparison of estimated elasticities of land consumption across different data sources would therefore suggest the extent to which measurement errors might affect urban models. Conversely, if estimated elasticities are similar across different data products, this would be evidence to suggest the robustness or validity of previous studies which used only population-density derived data. For the satellite data, I utilize the National Land Cover Database (NLCD) 1992–2001 Land Cover Change Retrofit Product. NLCD data is produced by the Multi-Resolution Land Characteristics Consortium (federal agencies), is derived from Landsat imagery and contains 30-meter resolution consistently-interpreted land-cover data for the continental US. (Consistently-interpreted means that the resolution of the data, and the classification system are the same such that measured changes are actual and not a result of different methodologies.) The NLCD's "change retrofit product" has been specifically constructed to provide consistent estimates of land cover for both 1992 and 2001 for each pixel (30 m resolution). This enables consistent measurement of changes in land cover from 1992 to 2001.[1] Classification of land cover within the NLCD is a modified Anderson classification system (Anderson et al., 1976). Using the higher level of classification, all land surface of the continental US is classified as being in water, "developed", barren, forested, shrubland, herbaceous, planted/cultivated, and wetlands. In GIS, I overlay consistent metropolitan area boundaries (from the year 2000 census) on the 1992 and 2001 NLCD data to calculate the total amount of "developed land" in each of the metropolitan areas (MSAs and PMSAs) in the continental US for both 1992 and 2001.

For the population-density derived measure of urbanized land area using Census data, I develop a new method of overcoming the changing definition of "urban" by the Census in different decades. Although the Census has reported "urban" land estimates for Census geographies each decade, the classification and definition of what counts as "urban" is different each decade.[2] Thus, reported "urban" area estimates from the census are not comparable across years because of changing methodologies. As well, the underlying boundaries of census block-groups on which the data is derived also change from year to yet. The data set developed in this paper overcomes this inconsistency by applying the year 2000 definition of "urban" and applying that same definition to Census data for 1990 and 1980. The definition of "urban" in the year 2000 involves core block-group areas of at least 1000 persons per square mile, and all adjacent or nearby block-groups of at least 500 persons per square mile.

In order for estimates of "urban" to be consistent across years, the underlying data must be in consistent boundaries, because most census data products are produced with boundaries which change from decade to decade. I first acquire census block data from 1980, 1990 and 2000, all in year-2000 consistent boundaries, available commercially from Geolytics. Only areas located within metropolitan areas (MSAs/PMSAs) are utilized. Following the Census 2000 methodology, the first stage is to identify census block groups with population densities exceeding 1000 persons-per-square mile (PPSM). These areas form the core of the urban area. Next, any block groups "contiguous" to these urban cores and which contain at least 500 PPSM are identified. Because the Census Bureau allows for a maximum "jump" of 0.5 miles between blocks to still fall within the definition of "contiguous" (because of industrial areas, parks, airports, rivers, etc.) this method also uses a 0.5 mile buffer. All block groups that meet the contiguity and density thresholds are then added to the core area. This process is then repeated, each time out by 0.5 mile buffers, adding only those block groups with a minimum 500 PPSM, until no new block groups remain to be added. This algorithm is applied to the 1980, 1990 and 2000 data sets, providing consistent estimates of the extent of "urbanized" area for each metropolitan area for each year.

Therefore, the measures of "developed" land area (from the satellite-derived data) and of "urbanized" land area (from the Census-derived data set reported here) are summarized at the metropolitan area level, and reported in square miles. These variables on developed/urban land area become the dependent variable measure of urban spatial area extent in the regressions reported.

The independent variables (income, population, and agricultural land values) are constructed as follows. The income variable is the median household income in the metropolitan area, for each time period, expressed in constant 1999 dollars. Although median household income data is available for each of the block-groups within the metropolitan area, we cannot compute an income variable unique to these "urban" areas because the actual distribution of income among households is not known at the block group level. So, although urban-area median income is likely higher than metropolitan-area median income, the metropolitan figures are the only ones able to be used. The value of agricultural land comes from Census of Agriculture, which reports average agricultural land values for counties, per acre, going back to the 1950s. I calculate averages using data on both sides of any particular decennial census (e.g. 1978 and 1982 for 1980 census) to arrive at 1980, 1990 and 2000 estimates of land values. I then compute metropolitan specific weighted-averages of per-acre land prices, weighting by a county's share of the metropolitan land area, and converted to 1999 dollars.

Because the underlying data set (on urbanized area) is built up from census block groups, population in each census block is known and therefore we can estimate the exact population residing in areas now classified as "urban." Previous studies have used either the entire population of the county or the metropolitan area as the measure of population. However, because the satellite-derived data product on developed land area covers the whole metropolitan area (including areas outside of "urban areas") and because the satellite data are not spatially connected to census blocks, the entire metropolitan population is used as the population data when using the satellite "developed" area estimates.

All 4 of the major monocentric empirical papers for the US (Brueckner and Fansler, McGrath, Wassmer, and Spivey) have had great difficulty utilizing consistent, available, and appropriate measures of transportation costs. Wassmer (2006) does not include transportation costs, while Brueckner and Fansler (1983) find that proxy measures of transportation costs (percentage of households owning cars and percent using transit) are statistically insignificant. McGrath finds that a regional transportation expenditure CPI measure is significant and negative only when included with a time-trend variable.

[1] For technical details, see Fry, J.A., Coan, M.J., Homer, C.G., Meyer, D.K., and Wickham, J.D., 2009, Completion of the National Land Cover Database (NLCD) 1992–2001 Land Cover Change Retrofit product: U.S. Geological Survey Open-File Report 2008-1379, 18 p. Available at: http://pubs.usgs.gov/of/2008/1379/. Last accessed Sept. 19, 2009.

[2] See "Urban area criteria for Census 2000," Federal Register 67:51 (March 15, 2002), p. 11663. Available at: http://www.census.gov/geo/www/ua/uafedreg031502.pdf, last accessed Sept. 19, 2009.

Spivey (2008), when restricting her sample to the 85 areas covered by the Texas Transportation Institute's "time travel index" and "congestion cost index," does find the expected negative coefficient (increased congestion or longer travel times reduce the size of the urban area). However, inclusion of the travel time or congestion indices results in land prices no longer being statistically significant. When she uses the census variable "average travel time to work" (which, unfortunately, also includes transit use which has higher per-trip fixed time-costs), the coefficient is positive, the same result I find in results not presented here. However, all of these approaches are admittedly poor proxies for the variable "per-mile (monetary) travel costs" which would be predicted in the model. Monetary per-mile costs, however, should not be expected to differ all that much across metropolitan areas, except for differences in gasoline prices. We are unable to find consistent or available metropolitan-area gasoline price data going back to the 1980 time period of the data.

Many have argued that the true cost to households of living further out from the CBD is not the pecuniary costs of travel, but the value of time spent commuting (Bajari and Kahn, 2005; Glaeser and Kahn, 2004). The best correlate of the value of time is the opportunity cost as measured by wages or income, which is already included the model. However, travel time for a household will be a function not only of the distance traveled but also the efficiency and density of the transportation network. As such, neither distances traveled nor average travel time measure the household tradeoff between housing consumption and commute time. Like Spivey (2008), I try to use the census data on average travel time to work, which is the only transportation related data consistently available across all metropolitan areas over the time period of this study, and is only available at the metropolitan (not urbanized area) level. This variable, however, presents a number of problems. First, introducing it into the model introduces severe multicollinearity, causing the signs and significance of the other variables to change. Second, like Spivey, the estimated sign is positive, opposite of that predicted by the monocentric model: increased travel times are associated with increased land area of the urbanized area. Third, the variable "average travel time" is almost tautologically the same as the size of the urbanized areas, because larger urban areas have longer average travel times. At the risk of underspecifying the model, and based on the variable's confusing and inconsistent results in other research, the models presented here do not include measures of travel time.[3]

In this study, the unit of analysis is the metropolitan area, using the year-2000 census boundaries. The sample consists of all MSAs (metropolitan statistical areas) and PMSAs (primary metropolitan statistical areas) in the continental US (excluding Hawaii and Alaska). Descriptive statistics of all the variables used in this study are found in Table 1.

Table 1
Descriptive statistics.

Variable	Obs	Mean	SD	Min	Max
Urbanized land area (in sq. miles)					
1980	329	163.43	240.10	15.98	2011.03
1990	329	186.66	270.89	18.73	2122.32
2000	329	223.23	321.57	19.05	2321.41
Developed land area (in sq. miles), from NLCD					
1992	329	235.75	258.88	29.51	1853.50
2001	329	250.36	278.13	31.99	2028.50
Urbanized population (population in "urbanized areas")					
1980	329	432,188	878,402	20,395	8,062,729
1990	329	485,771	960,231	35,074	8,583,784
2000	329	557,170	1,076,177	37,821	9,257,317
Average metropolitan agricultural land value (per acre, in 1999$)					
1980	329	$3016	$2715	$236	$28,162
1990	329	$2766	$3194	$146	$25,943
2000	329	$3497	$3799	$189	$29,260
Metropolitan median household income (in 1999$)					
1980	329	$34,975	$5454	$23,222	$64,038
1990	329	$39,245	$8224	$22,529	$88,482
2000	329	$41,172	$7827	$24,863	$76,554

4. Cross section estimates of urbanized and developed land area

As described above, one commonly used approach to fitting the monocentric model is to take a cross-section sample of cities. Underlying this approach is the assumption that sampled cities share the same determinants of growth and that estimated parameters represent some natural or equilibrium trajectory of urban growth. Nearly all research in this area (see, especially Malpezzi and Guo, 2001; McMillen, 2006; and Spivey, 2008) interprets the predictive power of the cross-section tests as an indirect test of the applicability of the stylized monocentric model to city size.

Mills (2000), in a retrospective, famously seems surprised that this simple cross-sectional model, with employment concentrated at the CBD, continues to fit the data quite well: "…the remarkable fact is not that the chimp types so badly, but that it types at all." (Mills, 2000, p 18). Urban spatial structure is, however, remarkably durable (Glaeser and Gyourko, 2005). As Spivey writes, "City structure, while it has grown increasingly complex, is still governed by market forces and not uncontrolled sprawl. It also suggests, with regard to the Mills–Muth model, that the chimp still types, if at an increasingly slower rate." (Spivey, 2008, p. 306). I offer, in partial support of Spivey's "still-typing-but-slower" assertion, the R^2 metrics in Table 2. While still explaining 75% of the variation in cross-city urban areas, this predictive ability declines ever so slightly each decade, from 78.6% in 1980. In relative terms, a 3% reduction in predictive ability over 20 years suggests the old "chimp" might still have many good years of typing left!

I first present cross-section results, both to present comparisons with previous research and to examine the robustness of estimates across population-density and satellite-derived measures of urbanization. As well, the "developed" land area data from the satellite are only available for 2 time periods, making the panel estimations (as below) unavailable for this data product. Table 2 presents cross section regressions for each of the 3 years in the "urbanized area" sample (1980, 1990, and 2000) and the 1992 and 2001 "developed land area" NLCD data. The dependent variable is the area of land urbanized or developed (in square miles) for the 329 metropolitan areas in the sample. The measure of population in the "urbanized area" regressions is urban population, while for NLCD "developed" estimates it is the total metropolitan area population. Regressions are estimated with standard errors robust to heteroskedasticity.

The models, as expected, perform well, explaining about 75–78% of the variation in the size of urban areas across all 329 metropolitan regions. The signs are consistent with theory and previous work: population is significant and positive. Median household income is significant and positive in the urbanized land area regressions, and significant at the 10% level in the developed land area regressions. Agricultural land values are significant and negative. Similar signs and coefficient magnitudes between the census-derived "urbanized" measurement and the satellite-derived "developed" measure suggests that the monocentric model is reasonably robust to different concepts of measuring urbanization. This consistency could imply that "more accurate" satellite data might not be necessary, at least at the geographical level of a metropolitan area.

Because the marginal estimates across different data sets and methods can become critical for future landscape change models,

[3] Regressions with travel time variables included are available from the author upon request. In none of the estimates, does the magnitude of the estimated elasticities of land area with respect to population or income change significantly.

Table 2
Cross section estimates of monocentric model.

	"Urbanized land area"			"Developed land area"	
Independent Variable	1980	1990	2000	1992	2001
Population	0.0002382	0.0002397	0.0002543	0.0002193	0.0002154
	(0.000)*	(0.000)*	(0.000)*	(0.000)*	(0.000)*
Median household income	0.0073305	0.0071663	0.0080227	0.0049405	0.0050457
	(0.023)**	(0.012)**	(0.023)**	(0.064)	(0.089)
Value of agricultural land (per acre)	−0.0119932	−0.0174123	−0.0156261	−0.0217993	−0.0184129
	(0.056)	(0.038)**	(0.032)**	(0.008)*	(0.006)*
Constant	−159.7107806	−162.8494492	−194.1409121	−29.3785974	−40.189578
	(0.063)	(0.046)**	(0.076)	(0.69)	(0.653)
Observations	329	329	329	329	329
R^2	0.786	0.768	0.751	0.741	0.76

Robust *p* values in parentheses.
* Significant at 1%.
** Significant at 5%.

and to compare the different estimates between census and satellite products, it is worthwhile to look very specifically at the cross-sectional coefficients in Table 1, and recalling that the unit of measurement is square-miles. The population coefficients indicate that, all else being equal, one more "urban" resident would be associated with between 6640 and 7089 ft^2 of additional urbanized land area. Within the "urbanized" model, the difference between the 1980 and 2000 coefficients suggests that the marginal land consumption per new urban resident has increased 448 ft^2 over 20 years or about 0.01 acres. Within the "developed" model, the coefficient indicates that one additional metropolitan area resident is associated with 6005 to 6113 ft^2 of additional developed land area. When we compare the developed to the urban coefficients, for example for the year 2000/2001, the urbanized land area population coefficient is about 1000 ft^2 larger per person. I would argue that this difference is small enough to conclude that census-derived estimates are "good enough" for use in a wide range of models, without necessarily requiring the additional precision and expense of satellite-models. As well, the fact that these estimates are very close to each other is a confirmation of the robustness of previous estimations of the model using only census data. Because panel data techniques cannot be performed on the satellite-data (only 2 time periods), the cross-sectional correspondence gives us confidence that the panel estimates on the census-data (3 time periods) are more likely to be accurate.

Cross-section elasticities of urbanized or developed land area, evaluated at the sample means, are presented in Table 3. For comparison purposes, Table 3 also presents those elasticities for the similar variables reported in previous studies. In terms of the cross-section elasticity of developed or urbanized land area with respect to population, both the urban and developed elasticities presented here are significantly lower than Brueckner and Fansler's estimates of 1.086, and Spivey's estimate of 0.91, and modestly lower than McGrath's of 0.76.

The income elasticity of urbanized land with respect to changes in median household income, [1.48 to 1.56] is consistent with Brueckner and Fansler's estimate of 1.497, but significantly different from McGrath's of 0.33. The elasticity of developed land area with respect to median household income is lower than the elasticity with regard to urbanized land area (1.48 to 0.83), likely because developed land area includes exurban and rural lands detectable as developed by satellite imagery. My estimates of the elasticity of land area with respect to agricultural land prices also confirms Brueckner's estimates, with estimated elasticities about −0.25. McGrath's estimates of agricultural land value elasticities are likely different because of the use of only state-wide average land prices.

The difference between income elasticities between developed and urban land in Table 3 is substantial enough to warrant further discussion. Comparing the 2000 (urban) and 2001 (developed) estimates, a 1% increase in metropolitan median household income would increase demand for urbanized land 1.48%, but increase demand for overall developed land area by 0.83%. Recall that developed land includes land development which is outside of "urban" areas such as rural or exurban development. Under the usual definitions within economic theory, developed land area, then, would be a "normal" good and urbanized land area would actually be a "luxury" good. However, it is important in considering these comparison to remember what cross-sectionally-estimated elasticities actually mean. As I argue below, panel-estimated elasticities which provide better insight, and the robust panel elasticities of land with respect to income (over all metros) is substantially less than unity (0.399).

5. Panel estimation results

As above, most empirical estimates of the monocentric model estimates a sample of cities, but less common, because of data availability, has been to estimate these models in a fixed-effects panel framework. McGrath (2005) is the only study to use multiple time periods in the US, 33 cities from 1950 to 1990 in a pooled-cross section with a time-trend variable. Panel models would indicate how the spatial extent of land area in a city (rather than across cities) is expected to change with changes in income, prices, and population.

Using the data on urbanized land area from 1980, 1990 and 2000, a metropolitan fixed-effects panel model is estimated. Breusch-Pagan and Hausman tests strongly indicated the appropriateness of a fixed-effects estimation strategy. Panels are estimated with Driscoll and Kraay (1998) standard errors, robust to both heteroskedasticity and

Table 3
Estimated elasticities of urbanized/developed land area with respect to model exogenous variables, evaluated at sample means.

	"Urbanized land area"			"Developed land area"		Brueckner and Fansler (1983)	McGrath (2005)	Spivey (2008)
	1980	1990	2000	1992	2001	Data = 1970	1950–1990	Data = 2000
Population	0.630	0.624	0.635	0.558	0.588	1.086	0.76	0.91
Median household income	1.568	1.510	1.480	0.822	0.830	1.497	0.33	−0.39
Value of agricultural land (per acre)	−0.220	−0.258	−0.245	−0.256	−0.257	−0.231	−0.1	−0.03

Table 4
Panel estimates of monocentric model.

Independent variable	All metros	Metros by size categories		By number of principal cities	
		Big (over 500 k)	Small (under 500 k)	Only 1	More than 1
Urban population	0.0003136	0.0002992	0.0005121	0.0003360	0.0002847
	(0.000)*	(0.000)*	(0.000)*	(0.000)*	(0.000)*
Median household income	0.0019802	0.0034366	0.0003879	0.0013022	0.0030683
	(0.000)*	(0.000)*	(0.000)*	(0.000)*	(0.000)*
Value of agricultural land (per acre)	−0.0003917	−0.0007769	0.0006676	0.0008671	−0.0018068
	(0.739)	(0.704)	(0.002)*	(0.367)	(0.340)
Constant	−38.0609769	−64.1323264	−14.9899106	−31.96366	−46.92159
	(0.026)**	(0.122)	(0.000)*	(0.006)*	(0.074)*
Within R^2	0.694	0.700	0.731	0.704	0.692
Observations	908	306	681	687	300
Number of groups	329	102	227	229	100

Robust *p* values in parentheses.
* Significant at 1%.
** Significant at 5%.

autocorrelation.[4] Results for the full sample of 329 metro areas are presented in the first column of Table 4.

The urban population and median household income variables are significant and positive as predicted by theory, indicating that the amount of urbanized land should be expected to increase with increases in a city's population or median household income. Panel estimated coefficient on population is higher than the cross-sectional estimates, suggesting that adding one additional urban resident in a decade to a city would increase the size of the urban area by 8742 ft^2. Panel estimates measure marginal land consumption while cross-section estimates compared across time measure changes in average land consumption rates.

The agricultural land value variable, however, is insignificant across the sample of all metros, suggesting that changes in a metropolitan area's land prices do not significantly impact changes in urbanized land area. Although the variation in the price of land on the urban fringe explains variations of city size across cities, changes in land prices do not significantly impact urban development rates. The most likely explanation for this insignificant panel coefficient is shown in Table 1, that land values (adjusted for inflation) move in opposite directions across the study period: declining in real terms from 1980 to 1990 because of the agricultural credit crises, and then increasing from 1990 to 2000.

Panel estimated elasticities means are presented in the first column of Table 5. The estimated elasticity of urbanized land area with respect to change in a metropolitan region's urban population is 0.807 and the elasticity for median household income is 0.399, both of which are close to McGrath's estimations ("Model 2" with time-trend variables) of 0.76 and 0.33, respectively. Models of future land use change should consider that panel estimated elasticities provide more precise estimates of changes in urbanized land area expected for changes in income and population. The estimates presented here, confirming McGrath's, do show that both income and population elasticities are significantly less than unity when estimated in a panel framework. Many older land use models had utilized elasticity assumptions of unity or greater.

6. Do growth patterns vary by region size and structure?

The results presented in Section 5 cover all metropolitan areas in the continental US, based on the assumption that a similar growth trajectory characterizes all areas. One advantage of the extension of the data set on urban land area reported here to a larger number of metro regions is that it allows testing the structural stability of the model of growth determinants across different regimes. Two different regimes of city structure are tested here: size and polycentricity. The first test involves comparing small metropolitan areas with large metropolitan areas, and the second tests whether the model differs between monocentric and polycentric regions. For each of these two structural breaks, the test for model stability involves estimating the panel model across the two sub-samples and testing whether estimated coefficients are different.

The first test of structural stability in the model is across size categories of cities. To test the difference between small and large cities, I hypothesize a break around the year 2000 metro population of 500,000. While there is not necessarily any theoretical or *a priori* reason to chose 500,000 as a the cutoff, a large number of researchers limit their analysis to roughly the 100 largest metro areas. In this data, there are 102 "large" metropolitan regions (over 500 k) and 227 "small" metros. These large metros contain over 77% of the metropolitan-area population of the continental US and 75% of the urbanized land area.

The test for structural stability proceeds as the "Chow" test. I create a dummy variable for large regions with a year 2000 population of greater than 500,000 persons and estimate a panel model of all metros with a dummy variable for region size, and interaction terms on all the coefficients. The test is that the dummy variable and interaction terms are all jointly non-zero. Because dummy variables for region size are time invariant, this test cannot be estimated in a fixed-effects panel framework, so is instead implemented in a linear regression with panel-corrected standard errors. The test is significant at the 1% level, leading to the conclusion that the difference in estimated coefficients across the two sub-samples are not the result of sampling error.

The estimated coefficients for panel regressions by size categories are then shown in columns 2 and 3 in Table 4, and the estimated elasticities for each group are shown in Table 5. The panel-estimated elasticity of urbanized land area with respect to population is significantly higher in smaller metropolitan regions, and near unity while it

Table 5
Panel estimated elasticities of urbanized land area, evaluated at sample means.

	All metros	Metros by size categories		By number of principal cities	
		Big Metros	Small Metros	Only 1	More than 1
Population	0.807	0.833	0.986	0.887	0.698
Median household income	0.399	0.312	0.209	0.28	0.542
Value of agricultural land (per acre)	n.s.	n.s.	0.024	n.s.	n.s.

Note: n.s.—estimated coefficient is not statistically significantly different from zero.

[4] Panels are estimated in Stata using the "xtscc, fe" command. Diagnostic statistics available from author upon request.

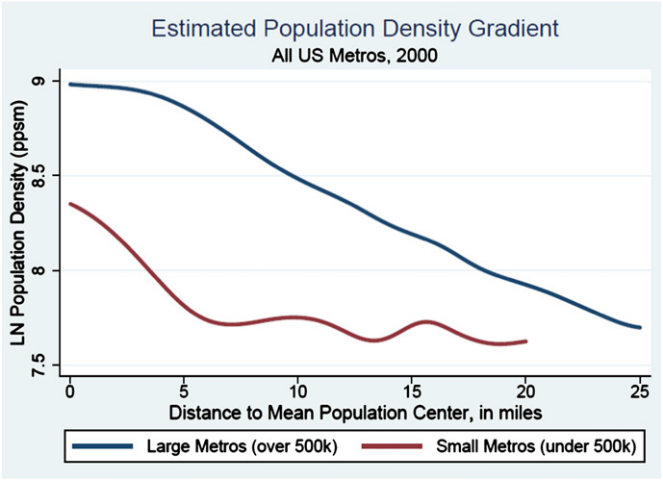

Fig. 1. Estimated population density gradient, large and small metros.

remains less than 1 in the large regions. The income elasticity in smaller regions is lower than in larger regions.

No previous research, either theoretically or empirically, has examined the difference in growth patterns across large/small regions so our discussions here are tentative. Perhaps the diagram in Fig. 1 below can motivate this understanding. For all of the metro areas in the sample, I compute the geographic "mean population center" and then compute the distance for each block group from the center in order. Regressing distance to center on the log of population density is a standard technique to produce a "density gradient." The data from all metro regions is pooled for both large and small metros, and Fig. 1 represents a non-parametric kernel density estimate of regressing distance on the natural log of density. Density gradients for both look as predicted in theory. The smaller metros' gradient is estimated out to only 20 miles because of a very low number of observations out past 20 miles. Out past 6 miles, the function exhibits waviness, likely picking up other smaller towns in the region. The gradient for larger areas is always higher and flatter than the one for small metros.

The results in Fig. 1 are consistent with the comparative statics of the AMM model. As population increases, the density gradient shifts upwards, and as income increases the gradient rotates towards the origin (becomes flatter). The large metros have a statistically significant higher median household income (year 2000) than small metros, $45,280 to $39,326 ($t=6.81$, $p<0.000$).

A second test for structural stability of the monocentric model is testing its applicability in polycentric regions. Although theoretical concern regarding the effect of polycentricity on housing and labor markets has been around for years, only recently have McMillen and Smith (2003) estimated the number of sub-centers in urbanized areas for the largest 60 urban places. Spivey (2008) uses these data for those 60 urbanized areas and finds a statistically significant and negative coefficient in her model of urban area, indicating that places with more employment sub-centers had lower levels of urbanized land area, all else being equal.

Instead of limiting the sample to only 60 places with counts of employment subcenters, in this paper I implement a simpler measure of the polycentric nature of urban areas, measuring whether a metropolitan region has more than one principal city or not. Obviously, there are many different concepts and approaches which could represent the idea of polycentricity, but the Census Bureau defines principal city or cities of a metropolitan area relative to its share of population and employment. I hypothesize that regions with more than 1 principal city would show a reduced urban land area, all else being equal. This is equivalent to hypothesizing that the estimated elasticity of urbanized land area with respect to urban population in regions with more than 1 principal city should be lower than in regions with only 1 principal city. Two hundred twenty-nine metropolitan regions in the sample have only 1 principal city, while 100 have more than 1.

The test proceeds as with the test on the size dummy variable above. A dummy variable for more than 1 principal city along with interaction terms of the dummy variable with each independent variable is estimated in the model, and the test is that all these coefficients are jointly zero. At the 1% level, we conclude that there is a statistically significant difference between the estimated coefficients in the model across groups (only 1 principal city, or more than 1 principal city). Estimation results are in the last columns of Table 4, and the estimated elasticities by groups are in the last columns of Table 5. As predicted, the coefficients and estimated elasticities of urbanized land area with respect to population are significantly lower in polycentric regions than in monocentric regions. Income elasticities of land demand are nearly twice as high in more polycentric cities, even though there is no statistically significant difference in median incomes between polycentric and monocentric cites ($t=1.61$, $p=0.11$).

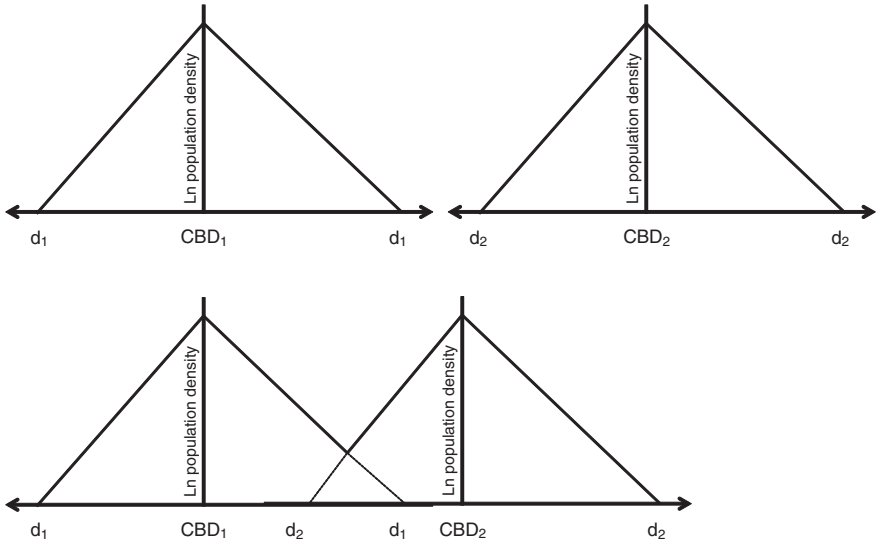

Fig. 2. Density gradients in polycentric and monocentric cities.

Fig. 2 presents an argument as to why we might expect the urban land area of polycentric regions to be smaller.[5] In Fig. 2, there are two principal cities, and the standard monocentric population density gradient is shown with the natural log of population density on the y axis. The figure shows the density gradient going out in both directions, such as in a city without natural constraints such as lakes or mountains. The radius of urban extent would be d_1 for city 1 and d_2 for city 2. In the top panel of Fig. 2, the two cities are far enough apart such that the spatial extent of city 1 does not intersect with the spatial extent of city 2. The bottom panel, however, illustrates the situation in a metropolitan region with more than 1 principal city. The spatial extent of urbanized land development in each city overlaps with the spatial extent of the other city, and the land use patterns shown in the dashed lines no longer exist. To accommodate the same population within a polycentric region, then the overall size of the urbanized area is smaller and the population elasticity of urbanized land demand is smaller.

7. Discussion and conclusions

The spatial expansion of the urbanized land area of cities – commonly, but perhaps misleadingly labeled "sprawl" – has been a perennial concern of academics, policy makers, planners and environmentalists for years (Fischel, 1999; Brueckner, 2001; Malpezzi and Guo, 2001). The literature on "sprawl" has itself grown to cover nearly every conceivable variable, cause, method, or data source, without an overall consensus as to either the measurement of or the cause(s) of urban expansion. In contrast, the standard monocentric city model of urban economics continues to provide a seemingly robust explanation for and prediction of city size. Of course the model is heavily stylized and simplified and ignores government policy; yet, it remains empirically robust and analytically tractable. In this paper, data for more areas, and including more time periods, combined with panel methods continue to demonstrate the empirical relevance of this reasonably simple model.

In addition to the historically-consistent data on more metropolitan areas to enable panel estimation, this research makes two important contributions to the debate about metropolitan spatial expansion. First, panel estimates of population and income elasticities, under all specifications, are substantially less than unity. This has implications for the wide variety of transportation and/or land use models which scientists and policy makers use. To the extent that conventional wisdom or practice had utilized larger population elasticities in modeling future land use scenarios, these models would have over-predicted the amount of land expected/needed to be developed. Secondly, this paper has identified differences in growth patterns by the structural characteristics of regional size and polycentricity. Large metropolitan regions have lower elasticities of urbanized land area with respect to population, but higher income elasticities of urbanized land area. Polycentric regions (more than 1 principal city) have lower population elasticities of urbanized land area, but higher income elasticities.

Although the standard monocentric model is still empirically tractable and robust in the presence of these differences, understanding the nature and meaning of these differences will continue to be important for urban modeling and policy analysis. Future research with this data (covering more metropolitan areas) could test a range of structural and policy variables on urban growth patterns.

In this paper, I have presented robust evidence that the monocentric standard model continues to offer a robust empirical account of the patterns of urban growth in the United States. This highly-stylized, monocentric, 3-variable model of the world continues to explain 75% of the variation in city sizes is therefore useful and offers "stylized facts" about fundamental drivers of urban land markets. Of course, cities and regions exhibit substantial heterogeneity of urban form, physical geography and regional policies which need to be more fully understood and incorporated to make these models even richer and more policy-relevant.

Acknowledgements

Yan Zhang, a doctoral student, provided expert GIS assistance in developing the urbanized area algorithm. Arthur Nelson, George Galster, Annemarie Schneider and Steve Malpezzi provided helpful comments and feedback on this paper, and Prof. Galster suggested the explanation in Fig. 2. I especially want to thank my (tor)mentor, Jack Huddleston. All mistakes are, of course, my own.

References

Anderson, J., Hardy, E., Roach, J., Witmer, R., 1976. A Land Use and Land Cover Classification System for Use with Remote Sensor Data. United States Government Printing Office, Washington, D.C.
Bajari, P., Kahn, M., 2005. Estimating housing demand with an application to explaining racial segregation in cities. Journal of Business and Economic Statistics 23 (1), 20–33.
Brueckner, J., 2001. Urban sprawl: lessons from urban economics. Brookings-Wharton Papers on Urban Affairs, pp. 65–97.
Brueckner, J., Fansler, D., 1983. The economics of urban sprawl: theory and evidence on the spatial size of cities. The Review of Economics and Statistics 55, 479–482.
Burchfield, M., Overman, H., Puga, D., Turner, M., 2006. Causes of sprawl: a portrait from space. Quarterly Journal of Economics 121 (2), 587–633.
Deng, X., Huang, J., Rozelle, S., Uchida, E., 2008. Growth, population and industrialization, and urban land expansion in China. Journal of Urban Economics 63, 96–115.
Driscoll, J.C., Kraay, A.C., 1998. Consistent covariance matrix estimation with spatially dependent panel data. The Review of Economics and Statistics 80, 549–560.
Esparza, A., Carruthers, J., 2000. Land use planning and exurbanization in the rural mountain west. Journal of Planning Education and Research 20, 23–36.
Fischel, W., 1999. Does the American way of zoning cause the suburbs of metropolitan areas to be too spread out? Governance and Opportunity in Metropolitan America. National Academy Press, Washington, D.C.
Glaeser, E., Gyourko, J., 2005. Urban decline and durable housing. Journal of Political Economy 113 (2), 345–375.
Glaeser, E., Kahn, M., 2004. Sprawl and urban growth. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, 4. Elsevier, pp. 2481–2527.
Kalnay, E., Cal, M., 2003. Impact of urbanization and land-use change on climate. Nature 423, 528–531 (29 May 2003).
Malpezzi, S., Guo, W.-K., 2001. Measuring "sprawl": alternative measures of urban form in U.S. metropolitan areas, center for urban land economics research. University of Wisconsin.
McGrath, D., 2005. More evidence on the spatial scale of cities. Journal of Urban Economics 58, 1–10.
McMillen, D., 2006. Testing for monocentricity. In: Arnott, R.J., McMillen, D.P. (Eds.), A Companion to Urban Economics. Blackwell Publishing, Malden, MA, pp. 128–140.
McMillen, D., Smith, S., 2003. The number of subcenters in large urban areas. Journal of Urban Economics 53, 321–338.
Mills, Edwin S., 2000. A thematic history of urban economic analysis. Brookings-Wharton Papers on Urban Affairs, pp. 1–38Brookings-Wharton Papers on Urban Affairs.
National Research Council, Committee for the Study on the Relationships Among Development Patterns, Vehicle Miles Traveled, and Energy Consumption, 2009. Driving and the built environment : the effects of compact development on motorized travel, energy use, and CO_2 emissions. Transportation Research Board National Research Council Special Report, 298.
Newburn, D., Berck, P., 2006. Modeling suburban and rural—residential development beyond the urban fringe. Land Economics 82 (4), 481–499.
Radeloff, V., Hammer, R., Stewart, S., 2005. Rural and suburban sprawl in the U.S. Midwest from 1940 to 2000 and its relation to forest fragmentation. Conservation Biology 19 (3), 793–805.
Schneider, A., Woodcock, C., 2008. Compact, dispersed, fragmented, extensive? A comparison of urban growth in 25 global cities using remotely sensed data, pattern metrics and Census information. Urban Studies 45 (3), 659–692.
Schneider, A., Friedl, M., Potere, D., 2010. Monitoring urban areas globally using MODIS 500m data: new methods and datasets based on 'urban ecoregions'. Remote Sensing of Environment 114, 1733–1746.
Spivey, C., 2008. The Mills–Muth model of urban spatial structure: surviving the test of time? Urban Studies 45 (2), 295–312.
Vesterby, M., Heimlich, R., 1991. Land use and demographic change: results from fast growth counties. Land Economics 67 (3), 279–291.
Wassmer, R., 2006. The influence of local urban containment policies and statewide growth management on the size of United States urban areas. Journal of Regional Science 46 (1), 25–65.
Wheaton, W., 1974. A Comparative Static Analysis of Urban Spatial Structure. Journal of Economic Theory 43, 258–272.
Wu, J., Cho, S.-H., 2007. The effect of local land use regulations on urban development in the Western United States. Regional Science and Urban Economics 37, 69–86.

[5] I thank George Galster for suggesting this explanation to me.