Determinants of Urban Sprawl: A Panel Data Approach

Joseph S. DeSalvo

Department of Economics

University of South Florida

4202 E. Fowler Ave. CMC 342

Tampa

FL 33620-5700, USA

Qing Su

Department of Economics and Finance

Northern Kentucky University

Nunn Drive, BC 338

Highland Heights

KY 41099, USA

Abstract

This paper applies fixed effects (within-groups) and between-groups estimations to panel data to test hypotheses of the monocentric urban model with urbanized area data for the period 1990–2010. The paper examines the impact of population, household income, transportation cost, and land rent at the urban fringe on urbanized area spatial size. The fixed effects regression finds that a 1- percent increase in population and a 1- percent decrease in travel cost causes an urbanized area to expand by 1.087 percent and 0.127 percent, respectively. The impact of household income is non-linear. The regression results from the between-groups estimation indicate that geographic and political factors help explain the spatial size differences across urbanized areas. The spatial size of an urbanized area is larger with a higher percentage of the urban fringe overlying aquifers, a higher percentage of local revenues from intergovernmental transfers, a higher percentage of urban fringe incorporated in 1980, and a lower elevation range in the urban fringe.

1. Introduction

Urban sprawl is a topic that has generated much debate and has become an important policy issue in the United States in the past twenty years. (Note 1) Because discussion of the costs and benefits of urban expansion is an ongoing process and no consensus has been reached regarding the sign of net benefits, it becomes more and more important for us to understand the underlying driving forces of urban sprawl and their impact over time. This is of great interest given the inherent impact of urban spatial expansion on public services and public finance. For example, transportation policies in most states in the U.S. involve drawing up long-term plans for road capacity expansion and construction. For these plans, traffic forecasts must be made and updated regularly depending on estimated urban spatial expansion and many other variables. If urban spatial growth is underestimated in an area, roads and highways could become more and more congested. If overestimated, scarce tax dollars would be wasted on unnecessary expansion of roads and highways. In addition, measuring and understanding contributors to urban spatial expansion may also benefit many other fields, including modeling of water quality and greenhouse gas emissions (Kalnay and Cal, 2003; Paulsen, 2012).

Much of our understanding of the determinants of urban spatial structure stems from the monocentric urban model. Development of this model was begun by Alonso (1964), Mills (1967), and Muth (1969). Wheaton (1974) provided an important advance by developing a general equilibrium version of the model and obtaining its unambiguous comparative static results. Wheaton's model did not include a housing-production sector, which was added by Brueckner (1987) in a synthesis of the work of Mills and Muth. Brueckner's paper is the starting point for most empirical estimations, applications, and extensions of the urban monocentric model, some of which we review below.

Most empirical studies of the monocentric urban model have used single-year cross-section data (Brueckner and Fansler, 1983; Song and Zenou, 2006; Spivey, 2008; Su and DeSalvo, 2008; Geshkov and DeSalvo, 2012). Two have used pooled multi-year cross-section data (McGrath, 2005; Deng, et al., 2008). Only two papers we know of have used panel data (Paulsen, 2012; Oueslati, Alvanides, and Garrod, 2015).

Although we appreciate the contributions of empirical studies based on single-year and multi-year cross-section data, it is important to note that urban spatial expansion is a dynamic process. Certainly studies based on data from a single year cannot capture these possibly time-varying effects, and even multi-year cross-section data do not track changes over time for the same set of urban areas. Capturing such time-varying changes is of crucial importance for models relying on forecasts of urban spatial growth. Empirical evidence based on panel data can better serve this purpose.

The preceding discussion dealt only with the monocentric model, but real urban areas have become more polycentric over time. To reflect this reality, polycentric urban models were

developed, beginning with White (1976). We believe that the polycentric urban theory of Anas and Kim (1996) and Anas and Xu (1999) addresses most of the criticisms leveled at the monocentric model. Although polycentric urban models incorporate more realistic factors, Anas (2007) leaves the empirical application of his theory for future work.

The monocentric model, however, lends itself easily to empirical estimation. The empirical evidence based on samples of relatively small U.S. urbanized areas (Brueckner and Fansler, 1983; Su and DeSalvo, 2008; Spivey, 2008; Geshkov and DeSalvo, 2012), almost all U.S. urbanized areas (Song and Zenou, 2006), and multi-year datasets (McGrath, 2005; Deng, et al., 2008; Paulsen, 2012; Oueslati, Alvanides, and Garrod, 2015) have demonstrated the robustness of the monocentric urban model.

This paper examines the determinants of urban sprawl using a panel dataset of urbanized areas for the period 1990–2010. We begin with a brief discussion of those papers that have used multi-year datasets and then turn to our empirical work.

2. Brief Discussion of Empirical Articles Using Multi-Year Datasets

McGrath (2005) used a pooled dataset for 1950 through 1990 to create a sample of 153 urbanized areas contained within the thirty-three largest U.S. metropolitan areas. He estimated two OLS regressions with and without a time-trend variable and found that population, income, and agricultural land values were statistically significant and had signs consistent with expectations. The coefficient on the transportation cost index was statistically significant in the model with the time-trend variable but not in the other model.

Paulsen (2012) used both census and satellite data, consistently defined for urbanized areas and developed areas in 329 MSAs in the year 2000, giving him data for 1980, 1990, and 2000. He provided cross-section estimates for urbanized areas and developed land areas separately for each year, as well as panel estimates for large and small urban areas in terms of population and number of principal cities. For the panel estimation, Paulsen used a fixed effects model and found statistically significant results and expected signs for population and income and the expected sign but insignificance for agricultural land value. Paulsen omitted a transportation-cost variable because, he says, of the poor performance of previously used transportation-cost proxies. In fact, of the nine articles cited earlier, two-thirds found this variable statistically significant and with the correct sign (Deng, et al., 2008; McGrath, 2005; Oueslati, Alvanides, and Garrod, 2015; Song and Zenou, 2006; Spivey, 2008; Su and DeSalvo, 2008).

Oueslati, Alvanides, and Garrod (2015) used two indices to measure urban sprawl based on a panel dataset of European cities: the total artificial area in square kilometers as a proxy for the size of the urban area and a scattering index measuring urban fragmentation. Their findings were considered robust and consistent with the prediction of the monocentric city model when the total artificial area was used as the dependent variable. Contrary to the positive impact of population and income (measured by GDP per capita), these two variables were found to have a negative impact when the scattering index was used as the measure of urban sprawl. Their finding suggests that larger population and higher income levels increase

both the spatial size and compactness of urban areas.

The monocentric urban model has also been tested using data on China. Deng, et al. (2008), noted that 70 percent of land allocation in China was state controlled, the rest being market controlled. This casts doubt on the applicability of the monocentric model, which assumes perfect competition in the land and housing markets. Nevertheless, the empirical results were good. The authors used satellite data covering all of China, which resulted in pooled data for 1987–89, 1995–96, and 1999–2000. Their sample size was 4,482 counties, which consisted of the county urban core and other contiguous built-up areas. The estimated equation that conforms most closely to the monocentric model found all exogenous variables highly statistically significant and consistent with theoretical expectations. Additional geographical variables also performed well, leading the authors to suggest that urban economists use geographical variables along with the usual economic variables in the empirical analysis of urban area spatial size.

Although not working strictly within the framework of the monocentric model, some economists, using various datasets for the U.S. (Baum-Snow, 2007; Duranton and Turner, 2012) and China (Baum-Snow et al., 2013; Faber, 2013), found that building and extending highway and railroad networks also contributed to urbanization and decentralization.

We conclude this section with some comments on the articles reviewed above. Of the papers using multi-year data sets, neither McGrath (2005) nor Deng, et al. (2008), used panel data. Thus, the papers most relevant to our study are Paulsen (2012) and Oueslati, Alvanides, and Garrod (2015), so it may be useful to note how what we do differs from what they did. Differing from the fixed effects model used by Paulsen (2012) and the Hausman-Taylor model used by Oueslati, Alvanides, and Garrod (2015), we employ the between-groups estimator to make use of the cross-sectional information from the panel data. In addition, we use a transportation-cost variable, as did Oueslati, Alvanides, and Garrod (2015), as well as geographic and political variables in the between-groups regression. Finally, we add another year's worth of data (2010) for U.S. urbanized areas. By using additional variables and addition data, we obtain additional results.

3. Data and Variables

Variables used in this paper are from two major sources: U.S. Bureau of the Census, Census (1990, 2000, 2010), and Burchfield, et al. (2006). Using 2010 Census data is important in part because, in contrast to the period 1980-2000 when real household incomes grew consistently, the median household income in many areas actually declined from 2000 to 2010 while fuel prices continued to increase. Data from this period may, therefore, help shed more light on the determinants of urban sprawl. Burchfield, et al. (2006), provides us with four geographic and political variables, which are interesting in their own right but which may also help reduce possible omitted-variable bias.

In this paper, urban sprawl is measured by the spatial size of an urbanized area. The use of spatial size is consistent with the monocentric urban model and has been used many times over nearly thirty years in empirical studies based on that model (see Brueckner and Fansler,

1987; McGrath, 2005; Song and Zenou, 2006; Spivey, 2008; Su & DeSalvo, 2008; Geshkov & DeSalvo, 2012; Paulsen, 2012; Oueslati, Alvanides, & Garrod, 2015). Among the several definitions of "urban" used by the census, economists consider the "urbanized area" to be the best because it includes closely settled areas outside the central city that are economically related to the central city (see, for example, Mills and Hamilton, 1994, p. 6). To avoid the potential problems of measurement errors due to changes in methodologies of identifying urban boundaries over time by the Census, we apply the same method proposed by Paulsen (2012) to make data on the spatial size of urbanized areas comparable over time based on the criteria adopted in 2000 by the Census: a population density of at least 1000 per square mile for the urban core and, to be delineated as "contiguous," a population density of at least 500 for nearby areas with a maximum of a half-mile "jump." Our variable of urban spatial size is measured in square miles for 1990, 2000, and 2010. (Note 2)

We divide explanatory variables into two groups: (1) standard variables from the monocentric urban model, and (2) geographic and political characteristics from Burchfield, et al. (2006).

3.1 Standard Variables from the Monocentric Urban Mode

The following four variables from the unextended monocentric urban model have been used in empirical work (Note 3): population or number of households, per-capita or per-household income, commuting cost per round-trip mile, and land rent at the urban-rural boundary. In the monocentric model, population and income are directly related to urban spatial size, while commuting cost and rural land rent are inversely related. We obtain population as well as mean and median household income for urbanized areas from the Census for the period 1990–2010. We use real income in 1990 dollars.

Data on transportation cost are not available by urbanized area. In providing the justification for excluding a measure of transportation cost in his panel-data regressions, Paulsen (2012) discussed in detail problems associated with a few potential proxies (including average travel time and average gasoline prices). While we encounter the same problems in obtaining a meaningful proxy for transportation, our focus turns to those performing relatively well in the existing literature.

Of the proxies used for transportation cost in studies of urban sprawl, government expenditure on transportation per commuter (Zenou and Song, 2006) and a transport-cost index (McGrath, 2005) were statistically significant. Since government expenditure on transportation per commuter was drawn from 1997 county-level data and the transport cost index was only available for the largest 33 MSAs, we cannot borrow directly from these two studies. We therefore construct a proxy similar to government transport expenditure per commuter, specifically the fraction of a state's highway expenditures spent in an urbanized area per 1,000 urbanized area residents, which enables us to increase the variation of this measure. This proxy is calculated as follows: state highway expenditure × total road miles in an urbanized area ÷ by total road miles in the state ÷ by the urbanized area's population (in thousands).

Since government highway expenditures are a burden on users in the form of motor vehicle

taxes, motor carrier taxes, road and crossing tolls, and intergovernmental transfers on transportation, we expect that a higher value of our proxy variable reflects a higher transport cost. This expectation is based on the fact that, since the Interstate Surface Transportation Efficiency Act of 1991 was implemented, US highway expenditures have shifted from primarily maintenance to primarily expansion of the highway network (Federal Highway Administration, 2012). In the 1993–1995 period, around 20 percent of expenditures were on added capacity, while 57 percent were for existing system maintenance and improvement in efficiency (Federal Highway Administration, 1997). Based on the newest data release in 2011, only 6.8 percent of expenditures were on new construction of highways and bridges, while 52.8 percent were on highway maintenance and improvement in 2010. The same trend was followed by state and local governments (American Roads and Transportation Builders Association, 2014).

Using data for 1990 and 2000, we find that our measure is positively correlated (0.3995) with average gasoline prices at the state level (a measure of variable monetary travel cost, identified by Paulsen as a viable measure, but not available in his dataset) and positively correlated with average travel time to work (0.109). (Note 4) Therefore, we think it is reasonable to believe our measure is positively correlated with transport cost at the urbanized area level.

We realize that Geshkov and DeSalvo (2012) used state highway expenditure per user, assuming it to be negatively correlated with transport costs. This could be true for their dataset since it included only small-urbanized areas, each located in a single county with one central city. For small-urbanized areas with little or no congestion, higher expenditures on highways could reduce travel time, an important component of travel cost. (Note 5)

Data on land rent at the urban-rural boundary are also unavailable by urbanized area. Following many scholars (Brueckner & Fansler, 1987; McGrath, 2005; Song & Zenou, 2006; Spivey, 2008; Su & DeSalvo, 2008; Geshkov & DeSalvo, 2012, Paulsen 2012), we use mean estimated market value of farmland per acre for the county (or counties) in which the urbanized area is located.

3.2 Geographic and Political Characteristics of an Urbanized Area

Four variables describing an urbanized area's geographic and political characteristics are also used in this paper (for detailed description, see Burchfield, et al., 2006, pp. 609-615). (Note 6) Burchfield, et al., justified the use of these variables and their expected relation to urban sprawl on the basis of the urban monocentric model as well as polycentric and city-system models. Burchfield, et al., expected that the presence of aquifers might have allowed households to sink their own water wells, thereby avoiding the necessity of connecting to a public water system, which may induce sprawl. Their proxy for this variable was the percentage of urban fringe area that overlaid aquifers. They also paid attention to the impact of physical barriers, hypothesizing "rugged terrain to naturally encourage scattered development, while "high mountains...are likely to make development more compact" (Burchfield, et al., pp. 612–613). They also expected that sprawl would be more prevalent where local taxpayers pay a smaller share of local government expenses to provide public

services. Their proxy for this variable was the amount of intergovernmental transfers as a percentage of local revenues in 1967. When there were unincorporated areas on the urban fringe, they expected that developers might have avoided municipal regulations by building outside municipal boundaries, which led to urban expansion. Their proxy for this variable was the percentage of urban fringe land that was incorporated in 1980.

Our sample of urbanized areas does not include all the urbanized areas delineated by the Census in 1990, 2000, and 2010. This is so because of inconsistencies in the names of the central city (or cities) in certain urbanized areas. For example, Albany-Schenectady-Troy, NY, was listed as the name of the urbanized area in 1990, but in 2000 we have Albany, NY, and in 2010 we have Albany-Schenectady, NY. To avoid potential measurement errors, our sample includes those urbanized areas that have the same central city or cities for 1990, 2000, and 2010. Table 1 provides descriptive statistics for our balanced panel dataset with 450 observations.

Table 1. Descriptive Statistics, Panel Data Set, 1990–2010

Variable	Mean	St. Dev	Minimum	Maximum
Spatial Size (sq. mi.)	347.62	463.37	20.29	2789.26
Population (000s)	999.36	19,90.86	101.36	18,400
Median Household Income ($ in 1990)	30065.5	4714.8	16,703	50998.8
Highway Expenditure Per 1000 Residents ($)	165,846	125,975	13,521	785,672
Rural Land Value ($ per acre)	1647.53	1148.04	146.70	8151.67

4. Model Specifications and Regression Results

A baseline panel-data model starts with the following expression with two-way error components disturbances:

$$y_{i,t} = \beta x_{i,t} + \mu_i + \lambda_t + v_{i,t}, \; i = 1, 2, ...N, \; t = 1, 2, ..., T \tag{1}$$

where μ_i denotes an individual effect, λ_t denotes a time effect, and $v_{i,t}$ is the remainder stochastic disturbance term. If μ_i and λ_t are assumed to be fixed parameters to be estimated and $v_{i,t} \sim \text{IID}(0, \sigma_v^2)$, (1) is a two-way fixed effects error-components model where $x_{i,t}$ is assumed to be independent of the disturbances for all i and t. On the other hand, if

$\mu_i \sim \text{IID}(0, \sigma_\mu{}^2)$, $\lambda_t \sim \text{IID}(0, \sigma_\lambda{}^2)$, and $\nu_{i,t} \sim \text{IID}(0, \sigma_\nu{}^2)$, independent of each other, then

(1) represents a two-way random-effects model where $x_{i,t}$ is independent of the μ_i, λ_t, and $\nu_{i,t}$ for all i and t (Baltagi, 2005, pp. 33–35).

Our search for the model specification that best fits our data starts with the fixed effects model since our dataset includes the same urbanized areas for the periods 1990, 2000, and 2010. (Note 7) We also apply the random-effects model. Based on the preliminary results, we conduct the Hausman test, which rejects the hypothesis that the coefficients from the fixed effects and random effects models are not systematically different at the level of 0.01 or better. The results from the Hausman test suggest that the random-effects estimator is not consistent.

4.1. *Model Specification*

Given the results of the Hausman test, model specification tests are based on within-groups and between-groups estimators. In this context, by *within-groups* we mean the same urbanized areas over time, and by *between-groups* we mean different urban areas in the same time period. We apply graphic analysis to help identify whether there are potential non-linear relationships between urban spatial size and the four standard monocentric variables. We find that the relationship between urban area size and median household income may not be linear. Consequently, we add a quadratic term for income. We also create dummy variables indicating the time periods 1990 and 2000 to capture time effects.

4.1.1 Fixed Effects (Within-Groups) Estimation

The fixed effects model examines the relationship between urbanized area spatial size and the explanatory variables that vary over time within an urbanized area, i.e., OLS of $(y_{it} - \bar{y}_i)$ on $(x_{it} - \bar{x}_i)$. Since the variables used by Burchfield, et al. (2006) are time-invariant, the impact of those variables cannot be captured by the fixed effects model. Since our observations are obtained from the census at 10-year intervals, time fixed effects may exist, which is verified by our test results. Thus, our fixed effects model includes the four standard monocentric variables and two dummy variables capturing the time fixed effects of 1990 and 2000 as follows:

$$\begin{aligned}
\ln(SpatialSize)_{i,t} = {} & \alpha + \beta_1 \ln(\text{Pop})_{i,t} + \beta_2 \ln(\text{HighwayExpPer1000Residents})_{i,t} \\
& + \beta_3 \ln(RHouseholdIncome)_{i,t} + \beta_4 (\ln(RHouseholdIncome)_{i,t})^2 \quad (2) \\
& + \beta_5 \ln(RuralLand\,\text{Re}\,nt)_{i,t} + \beta_6 Year2000 + \beta_7 Year2010
\end{aligned}$$

4.1.2 Between-Groups Estimation

In contrast to our fixed effects model, which uses the time-series or within-urban-area information, our between-groups estimator uses the cross-sectional information reflected in

the changes between urbanized areas. We also include those time-invariant variables used by Burchfield, et al. (2006). Joint significance tests are then applied to remove those variables that are jointly insignificant. The between-groups estimating equation is as follows:

$$\ln(SpatialSize)_{i,t} = \alpha + \beta_1 \ln(\text{Pop})_{i,t} + \beta_2 \ln(\text{HighwayExpPer1000Residents})_{i,t}$$
$$+ \beta_3 \ln(RHouseholdIncome)_{i,t} + \beta_4 (\ln(RHouseholdIncome)_{i,t})^2$$
$$+ \beta_5 \ln(RuralLand\,\text{Re}nt)$$
$$+ \beta_6 (\%UrbanFringeOverlyingAquifers)$$
$$+ \beta_7 (ElavationRangeinUrbanFringe)$$
$$+ \beta_8 (IntergovernmentalTransferasPercentageofLocalrevenues)$$
$$+ \beta_9 (\text{ % Urban Fringe Land incorporated})$$

(3)

4.1.3 Tests of Unit Roots and Endogeneity

There may be some concern that the panels used in this paper may contain unit roots, which makes the usual panel data approach, such as the fixed effect model, inappropriate. Since the panels include observations in only three time periods, we applied the Harris-Tzavalis unit-root test (Stata command: xtunitroot ht). The test results indicate that we can reject the hypotheses that the panels contain unit roots at the level of 0.001 or better. These results support the hypothesis that the four standard monocentric variables used in this paper are stationary. Thus the panel-data approach used is appropriate.

We adopt the method proposed by Woodridge (2012, p. 24) to test whether the urbanized area highway expenditure per 1000 residents is endogenous. Our first step is to estimate the reduced form, fixed effects regression of this variable on the instrumental variable urbanized area freeway lane miles and all other exogenous explanatory variables and to obtain the fixed effects residuals. The second step is to run a fixed effect regression with the potential endogenous variable and the saved fixed effects residual from the first step, together with other exogenous explanatory variables. The test results indicate that the coefficients on the residuals are statistically insignificant for the within-groups estimation ($P > |t|$: 0.746) and the between-groups estimation ($P > |t|$: 0.423). The results indicate that we cannot reject the hypothesis that this measure is exogenous at any reasonable level. We added another instrumental variable (average gas prices by region) to the instrumental variable fixed effects regression and conducted an over-identification test, which also indicated that we cannot reject the hypothesis that the instrumental variables used are exogenous.

4.2 Regression Results

Regression results from the fixed effects (within-groups estimator) and the between-groups estimator for the full sample are reported in Table 2.

The regression results from the fixed effects model indicate that an area's population has a positive and statistically significant impact on urban spatial size during the period 1990–2010, which is consistent with the theoretical prediction of the monocentric urban model. Since the fixed effects model examines the relationship between the spatial size of an area and the

explanatory variables that vary over time within an urbanized area, the regression results indicate that when an area's population increases by 1 percent, the spatial size of the area will increase by 1.087 percent, other things equal. The coefficient on population is higher than those obtained by Paulsen (2012) and Oueslati, Alvanides, and Garrod (2015). A possible explanation is that our sample includes many small urbanized areas while Paulsen's empirical analysis was based on a much larger sample within a different time framework and lacked a measure of transportation cost. On the other hand, Oueslati, Alvanides, and Garrod used a sample of European cities, many of which experienced slow population growth in the study period.

The coefficient on transportation cost, proxied by annual highway expenditure per 1000 residents, is negative and statistically significant at the 0.01 level. This result is also consistent with the prediction of the monocentric urban model. When annual expenditure per 1000 residents falls by 10 percent, the spatial size of the urbanized area increases by 1.27 percent. The coefficient on rural land rent is negative and but not statistically significant at the 0.1 level.

The impact of household income is not linear, but increases at a decreasing rate. This finding is similar to that obtained by Su and DeSalvo (2008). For those areas with median household income less than $33,088 in 1990 dollars, its impact is positive. A possible explanation for this non-linearity could be that the time-cost of transportation, which rises with income, is absent from both the monocentric model and our explanatory variables. When the time-cost of transportation is below a certain threshold, people are willing to consume larger and newer housing in the suburb at the expenses of longer commute. When the time-cost of transportation becomes sufficiently large, however, households may want to move closer to the employment center. This may reverse the monocentric model's theoretical result, resulting in a more compact area. Our result may also be capturing the impact of urban development in moderate and high density areas made by many local governments since 2000. It may also reflect the impact of lower median incomes due to the most recent recession in the US.

Table 2. Panel Data Regression Results

Variable	Within-Groups	Between-Groups
ln(Population)	1.087*** (17.52)	0.8814*** (36.21)
ln(Real median household income)	15.183* (1.87)	29.0861*** (2.63)
$(\text{ln(Real median household income)})^2$	–0.729*	–1.1412**

	(1.93)	(2.64)
ln(Highway expenditure per 1000 residents)	−0.1271***	−0.1084**
	(3.48)	(2.06)
ln (Real rural land rent)	−0.0239	−0.0372
	(0.47)	(0.75)
Year 2000	0.138***	
	(4.71)	
Year 2010	0.194***	
	(3.77)	
Percent urban fringe overlying aquifers		0.0018***
		(3.37)
Elevation range in urban fringe		−0.0001***
		(5.10)
Percent of urban fringe incorporated		0.0088**
		(2.18)
Intergovernmental transfers (percent of local revenues 1967)		0.0076***
		(3.31)
Constant	−86.259**	−157.238***
	(1.97)	(2.76)
R^2/fitness	Within: 0.79 Between: 0.90 Overall: 0.89	Within: 0.73 Between: 0.94 Overall: 0.93
N	450	450

We turn now to our between-groups regression. This differs from the fixed effects model,

which uses time-series information from the panel dataset, while the between-groups estimator uses the cross-section information reflected in changes between urbanized areas. In Table 3, the results are labeled "Between-Groups."

The between-groups regression confirms the monocentric model's theoretical predictions for population and transportation cost. When the difference in population between urbanized areas is 1 percent, on average, the urbanized area with the larger population is 0.88-percent larger spatially. When the difference in travel cost between urbanized areas is 10 percent, the area with the lower transport cost (in terms of a lower transportation burden on travelers) is 1.1 percent larger spatially.

The effect of income on an urbanized area's spatial size is not linear. For the urbanized areas included in the sample, the overall impact of household income is positive, which is consistent with all but one of the published articles (Brueckner & Fansler, 1983; McGrath, 2005; Song & Zenou, 2006; Deng, et al., 2008; Spivey, 2008; Geshkov & DeSalvo, 2012; Paulson, 2012; Oueslati, Alvanides, & Garrod, 2015), the exception being Su and DeSalvo (2008) who also used a quadratic term. The coefficient on rural land rent, in the between-groups model is negative but not statistically significant at the 0.1-level or better.

We include the four time-invariant variables from Burchfield, et al. (2006), discussed earlier, in the between-groups estimation. These variables provide evidence that some geographic and political factors have an impact on an urbanized area's spatial size. The regression results indicate that the spatial size of an urbanized area is larger when (1) a higher percentage of the urban fringe overlies an aquifer, (2) there is a lower elevation range in the urban fringe, (3) a higher percentage of the urban fringe was incorporated in 1980, and (4) a higher percentage of local revenues is from intergovernmental transfers, all of which are in accord with the expectation of Burchfield, et al. (2006).

4.3 Robustness Check

Since our sample does not include all urbanized areas for the 1990-2010 period, we conduct a robustness check to test the sensitivity of the results.

The first robustness check concerns measures of household income. As mentioned earlier, the monocentric urban model assumes that all households have the same income, for which we use median income as a proxy. In the robustness check, we use mean income. The second robustness check tests whether or not results from (3) are sensitive to exclusion of the variables that are omitted due to joint insignificance based on the pooled OLS results. This robustness check is conducted using the method of Barslund, et al. (2007). The core variables are those used and reported in Table 2. The testing variables are the geographic and political variables previously excluded. For the robustness check, we run regressions that include all the core variables and different combinations of the testing variables.

While the robustness of the results of this paper is encouraging, a major caveat remains, stemming from potential inaccuracy of or measurement errors in the explanatory variables used. Since accurate measures of land rent at the urban fringe and transportation cost are not available, potential bias cannot be ruled out.

5. Conclusion

We apply fixed effects (within-groups) and between-groups estimations to panel data to test hypotheses of the monocentric urban model with urbanized area data for the period 1990–2010. The paper examines the impact of population, household income, transportation cost, and land rent at the urban fringe on urbanized area spatial size. This is done not only within, but also across, urbanized areas over time.

The fixed effects regression finds that a 1- percent increase in population and a 1- percent decrease in travel cost causes an urbanized area to expand by 1.087 percent and 0.127 percent, respectively. The impact of household income is non-linear. In the between-groups estimation, we obtain results similar to those of the fixed effects model. The regression results from the between-groups estimation indicate that geographic and political factors help explain the spatial size differences across urbanized areas. The spatial size of an urbanized area is larger with a higher percentage of the urban fringe overlying aquifers, a higher percentage of local revenues from intergovernmental transfers, a higher percentage of urban fringe incorporated in 1980, and a lower elevation range in the urban fringe.

Urban spatial expansion is a dynamic process jointly determined by many factors. Nevertheless, our empirical analysis generally supports the theoretical results of the monocentric urban model. Our research has policy relevance as well. Most of the urbanized areas in our sample have populations of 500,000 or less. Our estimate of the population elasticity of urban spatial size suggests that travel-demand modelers and land-use forecasters should assume a unit-elastic effect of population on urban expansion for small and medium-sized urbanized areas. Finally, our findings on the impact of geographic and political characteristics on urban spatial size suggest that planners and policy makers should adjust their model forecasts accordingly when those factors apply.

References

Alonso, W. (1964). *Location and Land Use.* Cambridge: Harvard University Press.

American Roads and Transportation Builders Association. (2014). *Do we spend more money each year building new roads and bridges or repairing and maintaining our existing roads and bridges?* Retrieved from http://www.artba.org/about/transportation-faqs/#8

Anas, A., & Kim, I. (1996). General Equilibrium Models of Polycentric Urban Land Use with Endogenous Congestion and Job Agglomeration. *Journal or Urban Economics*, *40*, 232-256. https://doi.org/10.1006/juec.1996.0031

Anas, A. (2007). A Unified Theory of Consumption, Travel, and Trip Chaining. *Journal of Urban Economics, 62*, 162-186. https://doi.org/10.1016/j.jue.2006.05.002

Anas, A., & Xu, R. (1999). Congestion, Land Use, and Job Dispersion: A General Equilibrium Model. *Journal of Urban Economics, 45*, 451-473. https://doi.org/10.1006/juec.1998.2104

Barslund, M., Rand, J., & Tarp, F. (2007). Understanding Victimization: The Case of

Mozambique. *World Development, 35*, 1237-1258. https://doi.org/10.1016/j.worlddev.2006.09.018

Anas, A., & Ikki, K. (1993). Income Distribution and the Residential Density Gradient. *Journal of Urban Economics, 31*, 164-180. https://doi.org/10.1016/0094-1190(92)90051-L

Baum-Snow, N. (2007). Did Highways Cause Suburbanization? *Quarterly Journal of Economics, 122*(2), 775-805. https://doi.org/10.1162/qjec.122.2.775

Baum-Snow, N., Brandt, L., Henderson, J., Turner, M., & Zhang, Q. (2013). Roads, Railways and Decentralization of Chinese Cities. Retrieved from http://www.econ.brown.edu/fac/nathaniel_baum-snow/china_transport_all.pdf

Brueckner, J. K. (1987). The Structure of Urban Equilibria: A Unified Treatment of the Muth-Mills Model. In Edwin S. Mills (Eds.), *Handbook of Regional and Urban Economics, Vol. II, Urban Economics* (pp. 821-845). Amsterdam: Elsevier.

Brueckner, J. K., & David, A. F. (1983). The Economics of Urban Sprawl: Theory and Evidence on the Spatial Size of Cities. *Review of Economics and Statistics, 65*, 479-482. https://dx.doi.org/10.2307/1924193

Burchfield, M., Henry, G. O., Diego, P., & Matthew, A. T. (2006). Causes of Sprawl: A Portrait from Space. *Quarterly Journal of Economics, 121*, 587-633. https://doi.org/10.1162/qjec.2006.121.2.587

Davis, M., & Palumbo, M. (2007). The Price of Residential Land in Large US Cities. *Journal of Urban Economics, 63*(1), 352-384. https://doi.org/10.1016/j.jue.2007.02.003

Deng, X. Z., Jikun, H., Scott, R., & Emi, U. (2008). Growth, Population and Industrialization, and Urban Land Expansion of China. *Journal of Urban Economics, 63*, 96-115. https://doi.org/10.1016/j.jue.2006.12.006

Duranton, G., & Turner, M. (2012). Urban Growth and Transportation. *Review of Economic Studies, 79*, 1407-1440. https://doi.org/10.1093/restud/rds010

Faber, B. (2013). Trade Integration, Market Size, and Industrialization: Evidence FromChina's National Trunk Highway System. Working Paper.

Geshkov, M. V., & Joseph, S. D. (2012). The Effects of Land-Use Controls on the Spatial Size of U.S. Urbanized Areas. *Journal of Regional Science, 52*(4), 648-675. http://dx.doi.org/10.1111/j.1467-9787.2012.00763.x

McGrath, D. (2005). More Evidence on the Spatial Size of Cities. *Journal of Urban Economics, 58*, 1-10. https://doi.org/10.1016/j.jue.2005.01.003

Mills, E. S. (1992). The Measurement and Determinants of Suburbanization. *Journal of Urban Economics, 32*, 377-387. https://doi.org/10.1016/0094-1190(92)90025-G

Mills, E. S., & Bruce, W. H. (1994). *Urban Economics* (5th ed.). Upper Saddle River, NJ: Harper Collins College Publishers.

Oueslati, W., Seraphim, A., & Guy, G. (2015). Determinants of Urban Sprawl in European Cities. *Urban Studies, 52*(9), 1594-1614. https://dx.doi.org/10.1177%2F0042098015577773

Paulsen, K. (2012). Yet Even More Evidence on the Spatial Size of Cities: Urban Spatial Expansion in the U.S., 1980–2000. *Regional Science and Urban Economics, 42*, 561-568. https://doi.org/10.1016/j.regsciurbeco.2012.02.002

Song, Y., & Yves, Z. (2006). Property Tax and Urban Sprawl: Theory and Implications for U.S. Cities. *Journal of Urban Economics, 60*, 519-534. https://doi.org/10.1016/j.jue.2006.05.001

Spivey, C. (2008). The Mills-Muth Model of Urban Spatial Structure: Surviving the Test of Time? *Urban Studies, 45*, 295-312. http://dx.doi.org/10.1177/0042098007085964

Su, Q., & Joseph, S. D. (2008). The Effect of Transportation Subsidies on Urban Sprawl. *Journal of Regional Science, 48*, 567-594. http://dx.doi.org/10.1111/j.1467-9787.2008.00564.x

U.S. Department of Agriculture. (2014). 2012 Agriculture Census. Retrieved from http://www.agcensus.usda.gov/Publications/2012/

U.S. Bureau of the Census. (2013). Retrieved from http://www.census.gov

U.S. Energy Information Administration. (2013). "Motor Gasoline Sales through Retail Outlets Prices. Retrieved from http://www.eia.gov/dnav/pet/pet_pri_allmg_a_epm0_ptc_dpgal_m.htm

U.S. Federal Highway Administration. (2012). Legacy of a Landmark: ISTEA after 10 years. Retrieved from https://www.fhwa.dot.gov/publications/publicroads/01novdec/legacy.cfm

U.S. Federal Highway Administration. (1997).Trends and Forecasts of Highway Costs. Retrieved from https://www.fhwa.dot.gov/policy/hcas/final/three.htm

Wheaton, W. C. (1974). A Comparative Static Analysis of Urban Spatial Structure. *Journal of Economic Theory, 9*, 223-237. https://doi.org/10.1016/0022-0531(74)90068-4

White, M. (1976). Firm Suburbanization and Urban Subcenters. *Journal of Urban Economics, 3*, 323-343. https://doi.org/10.1016/0094-1190(76)90033-4

Woodridge, J. (2012). Panel Data Models with Heterogeneity and Endogeneity. Retrieved from http://www.ifs.org.uk/docs/wooldridge%20session%204.pdf

Notes

Note 1. In this paper, we define *urban sprawl* non-normatively as the increase in urban spatial size. We recognize that there are other definitions of urban sprawl.

Note 2. We thank Liren Zhou for performing the GIS work necessary for this data adjustment.
Note 3. Two other theoretical variables (the radians of land available for residential use and

the price of the composite structural input to the production of housing) have not been used in empirical analysis of the model.

Note 4. We do not have data on this measure for 2010.

Note 5. Using the Geshkov-DeSalvo dataset of 152 observations graciously provided by the authors, we find state highway expenditure per user to be negatively correlated with urbanized area average travel-to-work time (–0.022).

Note 6. Burchfield, et al. (2006), used other variables in their regressions.   We used a pooled OLS regression to determine which of these to include. Variables not jointly statistically significant at the 0.1 level or better were excluded.

Note 7. We tested pool ability across time, and both the Chow and Roy-Zellner tests rejected the hypothesis at the 5-percent level.

Copyright Disclaimer