Uncertain Economic Growth and Sprawl: Evidence from a Stochastic Growth Approach*

Belal Fallah[1]
Mark Partridge[2]
Rose Olfert[3]

June 26, 2011

Acknowledgement: We thank Infrastructure Canada for their support in funding part of this research under a grant titled. "Mapping the Rural-Urban Interface: Partnerships for Sustainable Infrastructure Development." We also thank the Canada Rural Revitalization Foundation and the Federation of Canadian Municipalities for their support of this project, in particular Robert Greenwood. An earlier version of this paper was presented at the Canadian Agriculture Economics Society Meeting. We thank Elena Irwin for her helpful comments and suggestions.

[1] College of Administrative Sciences and Informatics, Palestine Polytechnic University. bfallah2000@yahoo.com.
[2] Swank Professor of Rural-Urban Policy and Professor, Department of Agricultural, Environmental, and Development Economics, Ohio State University. Partridge.27@osu.edu (contact).
[3] Johnson-Shoyama Graduate School of Public Policy, University of Saskatchewan. rose.olfert@usask.ca

* The final publication is available at springerlink.com

Uncertain Economic Growth and Sprawl: Evidence from a Stochastic Growth Approach

Abstract: This paper examines how the volatility of local economies, represented by uncertainty over future land rents, affects urban sprawl. We develop a theoretical model that links sprawl to shocks to expected rent from future land development, among other factors. The econometric analysis draws upon panel data from U.S. metropolitan areas. To measure urban sprawl, we construct a distinctive measure that captures the distribution of population density within metropolitan areas. Using a proxy for uncertainty over future land rents, we confirm the theoretical prediction that across U.S. metropolitan areas, higher levels of uncertainty are associated with lower levels of sprawl.

Keywords: Economic Growth; Sprawl

1. Introduction

The emergence of urban sprawl in North America in the mid-20th Century has spawned a literature investigating its forms and effects, as well as aimed at understanding its underlying causes. Understanding the causes of urban sprawl is important because it represents an important shift in population distribution and land-use as it reshapes the rural-urban interface, pushing development well out into what were historically rural communities, changing the distribution of economic activities across urban areas. These in turn potentially affect productivity (Fallah et al., forthcoming) and potentially foster environmental degradation (Carruthers and Vias 2005; Glaeser and Kahn 2004). Sprawl can also affect the quality of life through impacts on the cost of infrastructure provision and the work-commute relationship (Carruthers and Ulfarsson 2008; Wu 2010). Because attempts to mitigate greenhouse gas emissions will likely entail more planning and zoning to reduce energy usage, the need to understand the causes of sprawl will likely intensify.

Urban economists have traditionally used the theory of the static monocentric city to explain the forces underlying urban sprawl and the implied decrease in population density farther from the central business district (CBD). In the absence of a single precise definition of urban sprawl, we use the term objectively to mean low density development with the expansion of urban development into (previously) undeveloped rural areas and interior urban areas (e.g., Nechyba and Walsh, 2004). Thus, we assess the key feature of most discussion of sprawl, namely the relatively low land-use intensity of development, leaving the impacts on efficiency to further empirical investigation.

The expansion of the urban structure hinges on the tradeoff between land rent and commuting costs. In equilibrium, this requires lower land rent at the urban fringe to offset the higher commuting cost to the CBD. The declining rent gradient leads to a decline in the density gradient (often modeled as a constant decline) as distance increases from CBD to the urban fringe (Alonso 1964; Muth 1969; Mills 1967). Research on sprawl has typically used this model as a starting point to propose a variety of causes of sprawl.[1]

Despite its success in reproducing general aspects of the urban spatial structure, the static monocentric model has been criticized by many urban economists. For example, Anas (1978) suggests that the monotonic decline of density away from the CBD occurs only under special conditions, such as rising income levels. Moreover, in dynamic models with durable housing and myopic landowners, Anas (1978), Harrison and Kain (1974), and McFarlane (1999) argue that residential development around the employment center is an incremental process and that density depends on the economic conditions at the time of development.

Dynamic models have conventionally assumed that a risk-neutral developer will choose to invest when the present value of expected cash flows exceeds the cost of development. However, if the future cash flow is uncertain and the investment is durable and irreversible, as in land development, this principle must be modified. For example, as in our model, consider an open urban area experiencing shocks in *labor demand.*[2] The resulting shocks in population growth and housing demand lead to an uncertain future land/housing rent to be earned from developing new lots. Titman (1985), McDonald and Siegel (1985, 1986), and others suggest that under

[1]For example, past research has proposed many causes of sprawl such as rising income (Margo 1992), declining transportation costs (Glaeser and Kahn 2004), fragmented local governance (Carruthers 2003), and improved infrastructure such as improved highways (Baum-Snow 2007).
[2]The shock could be to households—changes in labor supply—perhaps due to amenities. However, it is more likely due to be a labor demand shock.

uncertainty and illiquid durable investment, developers may choose to delay investment. The ability to delay investment decisions has economic value, i.e., a real option (Pindyck 1991).

Cappoza and Helsley (1990) were among the first to explore the relationship between urban spatial structure and uncertainty. Assuming that future rent follows a stochastic pattern, they found that in cities with a higher level of uncertainty, developers would delay land development until the expected future rent (usually referred to as reservation or hurdle rent) compensates not only for the agricultural land value and conversion cost, as in the case of a certain future, but also for the real option value. In their model, the real option value reflects the ability to delay land development while awaiting more information on future land prices. As a result of the development delay, Cappoza and Helsley (1990) suggest that the expected city size decreases, all else constant.[3]

In examining the effect of uncertainty on city size, Capozza and Helsley (1990) assume that the lot size is fixed across urban areas. Capozza and Li (1994) extend this work by investigating the interaction between uncertainty and the ability to vary residential capital intensity. Assuming that developers choose the optimum capital intensity and time of development, they find that uncertainty raises the reservation rent and delays the development decision.

To date, urban economists have done little to empirically test the implications of models that relate uncertainty about future land rent to urban spatial structure.[4] The objective of this paper is to address this gap. Specifically, we investigate how variations in uncertainty over future rent explain differences in the extent of U.S. metropolitan sprawl. The importance of this quest derives from policymakers' and urban planners' need to understand how an expanding and volatile

[3]Bar-Ilan and Strange (1996) extend the Capozza and Helsley (1990) model to allow for lags between the decision to develop a project and its completion. In this scenario, development lags may lead to a leap frog pattern where more distant land is developed prior to that closer to the city center. Mills (1981) also modeled scenarios in which a leap frog development pattern occurs under uncertainty and perfect foresight assumption.

[4]Most of the related literature on uncertainty (real options) has focused on capital investment (e.g, Caballero and Pindyck 1996; Paddock et al. 1988; Moel and Tufano 2002). Fewer studies have explored linkages between uncertainty and real estate. For example, Holland et al. (2000) examined the effect of price uncertainty on changes in property size.

urban economy can influence urban sprawl so that they can better address its impacts on the livability of the urban environment, residence-work commute relationships, and the provision of physical infrastructure (roads, schools, sewers, and other public facilities).

We depart from the work of Capozza and Helsley (1990) and Capozza and Li (1994) as follows: Capozza and Li's (1994) model is site-specific, predicting the irregular intra-metropolitan density profile rather than the overall density distribution. The latter is the emphasis of this paper. In addition, Capozza and Helsley (1990), focusing on the relationship between uncertainty and urban size, assume that each household consumes one unit of a *fixed-sized* lot, leading to constant density across urban areas. We relax the fixed-size assumption by allowing household preferences to determine lot size. Note that 'the lot' remains the unit of new land development with implications for how uncertainty in population growth is represented, namely in terms of absolute growth and how much land is developed.

In this paper, we also seek a theoretically consistent measure of sprawl that links metropolitan population density to uncertainty, differing from previous work in two ways. First, our model does not include structural capital. Instead, the variation in the density (sprawl) across urban areas arises from variations in lot size. Second, unlike Capozza and Li (1994), we assume that the representative household chooses the optimal lot size through utility maximization behavior (see Wu 2010), while a representative developer takes the optimal lot size as given and chooses the optimal *time* to convert agricultural land into residential use. In practice, this is akin to assuming the construction industry is relatively competitive, allowing household preferences to determine lot size (perhaps through the political process and zoning), which is more realistic than a fixed lot size across the entire urban area. To account for varying density across an urban area, we assume there is a composition effect where new lot development in the interior of the city is denser than development at the urban periphery, while the elasticity of lot size also plays a differential role between the periphery and interior.

Assuming that the future rent of developed land evolves in a stochastic fashion and solving for optimal lot size and optimal conversion time, we derive an expression that links population density to a reservation land rent. As in Capozza and Helsley (1990), the reservation rent is a function of shocks to expected future developed-land rent (among other factors), reflecting the real option value. A higher level of uncertainty raises the reservation rent in order to cover the real option value of delaying development.

Our econometric analysis uses panel data from U.S. metropolitan areas over the 1980-2000 period. Consistent with the theoretical model, we construct a distinct sprawl measure that captures both overall density and population distribution. The sprawl index measures the share of the population that lives in low density block groups within a metropolitan area (MA). The panel data allows us to capture unmeasured aspects of each MA. Then using variations in population change as a proxy for how labor demand shocks affect future land rent, we provide robust evidence that greater uncertainty is negatively related to urban sprawl.

The only empirical study that has tackled the issue of uncertainty and sprawl, that we are aware of, is Burchfield et al. (2006). Unlike our approach, their sprawl definition is specific to leap frog-type sprawl, measured as the amount of undeveloped land surrounding an average urban dwelling. Their measure of sprawl is different from ours in that we use a population density based measure of sprawl. They find that uncertainty is positively associated with delaying the development decision.[5]

In what follows, Section 2 presents the theoretical model. Section 3 describes the sprawl measure, followed by the empirical model and empirical results in section 4. Section 5 presents sensitivity analysis. Finally, section 6 concludes with some policy implications.

[5]A related study is Cunningham's (2006) assessment of how real-option behavior influences vacant land prices. His findings are consistent with real option theory—a higher level of uncertainty delays development and increases land prices.

2. Theoretical model

2.1 Household Problem

Our theoretical model is consistent with an open city model, where economic activities are concentrated in its central business district (CBD), a point to which households commute daily. Residential locations are indexed by their distance, z, from the CBD. The cost of commuting is normalized to \$1 per kilometer. Households are identical in terms of tastes and income, y, which is assumed, for now, to be exogenous to the size of the city population. Households in each period derive their utility, $u(m,q)$, from consuming land (lot) denoted by q, and numéraire non-land goods, denoted by m. The price of m is normalized to \$1. The budget constraint of households is given by: $m + Rq = y - z$, where R is land rent.

The (residential) mobility of urban households across urban areas is assumed costless. Therefore, utility is constant across space. Also, over time, we assume that each urban area attains a competitive short-run equilibrium at every point in time (t) (see Anas 1978), such that:

$$u(m,q)_{t1} = u(m,q)_{t2} = u(m,q)_{t3} = \ldots u(m,q)_{tm} = v \quad , \tag{1}$$

where v is constant utility. The bid rent function, obtained from the budget constraint, is the maximum rent per lot size that a household can pay for residing at distance z while enjoying a fixed utility level (v). The utility function takes the following Cobb Douglas form: $u = m^a q^b$, where $a > 0$, $b > 0$, and $a+b=1$. The bid rent function at time t is given by:

$$R(z,t) = \max_{q,m}\left[\frac{y(t) - z - m(t,z)}{q(t,z)}\right] , \tag{2}$$

subject to: $$m^a q^b = v \; .$$

(3)

The objective of the representative household is to choose the optimal bundle of q and m, which can be represented by differentiating equation (2) subject to the utility constraint (equation (3)), such that:

$$\tilde{q}(t,z) = \phi[y(t) - z]^{\frac{-a}{b}} \quad , \tag{4}$$

where the constant $\phi = v^{\frac{1}{b}} a^{\frac{-a}{b}}$, and

$$\tilde{m}(t,z) = a[y(t) - z] \quad . \tag{5}$$

Substituting $\tilde{q}$ and $\tilde{m}$ into the bid rent function (equation (2)) yields:

$$\tilde{R}(t,z) = \theta[y(t) - z]^{\frac{1}{b}} \quad , \tag{6}$$

where the constant θ=(1-a)/ø. As explained below, equation (6) is crucial in determining the expected city boundary.

2.2 Developer problem

We assume that land development occurs at the urban periphery (i.e., undeveloped, agricultural, exterior land) and within the urban area (i.e., in currently undeveloped interior land). The undeveloped interior land is modeled exogenously, which can take the form of leap-frog development. To insure that areas closer to CBD earn higher rent, we assume that returns to undeveloped interior land (R_i^A) is higher than that at the periphery (R_j^A). The cost of development is c, which by assumption does not depreciate over time. The optimal time of development is t*, which is specified as $t^*=t + s$ (also known as the first striking time), where s is the stopping time. Land development is irreversible due to prohibitive cost (at least in the time period under consideration by the developer), and once developed, the land earns an urban rent $R(s,z)$, which decreases as we move farther from the CBD. Therefore, the following is the price of undeveloped peripheral lot at location z_j, conditional on information available at time t:

$$P^A(t,s,z) = E\left\{ \int_t^{t^*} R_j{}^A(s,z)e^{-r(s-t)}ds + \int_{t^*}^{\infty} R_j(s,z)e^{-r(s-t)}ds - (R_j{}^A - C)e^{-r(t^*-t)} \middle| R(t,z) \right\}$$

(7)

The first term in equation (7) is the net return to undeveloped peripheral land up to the date of development (t*). The second term is the net return to the developed land after the date of development. We assume that the representative developer is risk neutral and the discount rate (r)

is constant across urban areas. With lots being the unit of development, equation (7) assumes that the lot size is exogenous to the developer. The equation of undeveloped interior land price is the same as equation 7, except that returns to undeveloped interior is $R^A{}_i$ once developed, earning a return of R_i.

We assume that the developed land rent R in either peripheral or interior areas has an exogenous stochastic pattern. We follow the literature (Capozza and Helsley, 1990; Capozza and Li, 1994; Plantinga et al., 2002) and employ a Brownian motion process with a drift g > 0 and variance σ^2, such that at time $t+s$, a developed land parcel's rent is:

$$R(t+s,z) = R(t,z) + gs + \sigma B(s) \tag{8}$$

Equation (8) implies that the distribution of rent after s periods is equivalent to current time period (t) rent plus a drift and a random component evaluated after s periods. Capozza and Helsley (1990) assume that income is stochastic. Since rent is linear in income, due to their assumption of fixed lot size, it follows that rent is also stochastic. In our case, income does not linearly enter the bid-rent function (equation (6)), making the model mathematically less tractable. Rather, we assume that the stochastic nature of rent is a consequence of exogenous shocks such as labor demand (including productivity shocks and shocks to local firms) and labor supply shocks (e.g., Nieuwerburgh and Weill 2006; Partridge and Rickman 2003, 2006).

Substituting equation (8) into (7) and integrating by parts, the expected value of an undeveloped peripheral lot is:

$$P_t^A(t,z) = \frac{R_j{}^A}{r} + E\left\{\left[\frac{R_j\ (t,z)}{r} + \frac{g}{r^2} - \frac{R_j{}^A}{r} - C\right] e^{-r(t^*-t)} \middle| R(t,z)\right\} \tag{9}$$

A representative developer chooses the optimal time t^* for converting the agricultural land into residential use. This occurs when the land development rent (R_j) reaches the optimal reservation rent R'_j. The conversion time t^* is defined as:

$$t^* = \min_S \left[t+s \geq t \middle| R_j(t+s,z) \geq R'_j \right] \tag{10}$$

From Karlin and Taylor (1975, pp. 361-362), the expected value of the Laplace transformation of t^*, conditional on the initial value of development rent (R) and R', is given by:

$$E\left[e^{-r(t^*-t)}\middle| R_j(t,z), R'_j\right] = e^{-\alpha[R'_j - R_j(t,z)]} \tag{11}$$

where $\alpha = [(g^2 + 2\sigma^2 r)^{\frac{1}{2}} - g]/\sigma^2$. Substituting equation (11) into (9) yields:

$$P^A(t,z) = \frac{R_j{}^A}{r} + \left[\frac{R'_j}{r} + \frac{g}{r^2} - \frac{R_j{}^A}{r} - C\right] e^{-\alpha(R'_j - R_j(t,z))} \Big| R_j(t,z) \tag{12}$$

The developer chooses R'_j that maximizes the land value. Differentiating equation (12) with respect to R'_j yields:

$$R'_j = R_j{}^A + rC + (r - \alpha g)/\alpha r$$

(13)

where (r-αg) ≥ 0. Equation (13) reveals that the optimal reservation rent is a function of returns to undeveloped land land ($R^A{}_j$), cost of conversion (C), rate of change in future land development rent (g), and shocks to future land development rent (σ), which reflects the uncertainty effect. The latter is subsumed in the term (r-αg)/αr, which as discussed earlier, reflects the option value arising from delaying the land development due to future rent uncertainty. The derivation of the interior reservation rent (R'_i) is the same is in equation (8-13), except that returns to interior undeveloped land is ($R^A{}_i$) and returns to developed land is $R^A{}_i$.

2.3 Density and land value

The expected overall density at time t^* can be written as the weighted average of density in the urban peripheral and interior areas:

$$D(t^*,z^*) = \underbrace{\delta \int_{z=0}^{z} \frac{1}{\tilde{q}(t^*,z)^{(\gamma_i - 1/\gamma_i)}}}_{Interior\ area} + \underbrace{(1-\delta) \int_{z}^{z=z^*} \frac{1}{\tilde{q}(t^*,z^*)^{(\gamma_j - 1/\gamma_j)}}}_{Periphery}$$

(14)

Where δ is the share of developed lots in urban interior. The first term of equation 14 refers to interior area density, while the second term corresponds to density at the periphery. The term (γ_i-

1)/γ_i captures the lot-size elasticity (γ_i >0) effect.[6] Assuming different lot-size elasticity captures some of the differential effects between interior and peripheral areas. Lot size elasticity is a function of many factors such as prior zoning and available land and it is likely that at least in the case of single unit homes, the elasticity is greater at the periphery than in the interior. This pattern follows because of past land-use patterns and zoning often place greater limitations on interior development and there is more undeveloped land at the periphery (we describe the case of multiunit housing in footnote 8).

The only unknown in the density function (equation (14), is the expected city boundary (z^*) . From equation (6) and (13), the expected urban boundary can be written as:

$$z^*(t^*) = y(t^*) - \left(\frac{R'_j}{\theta}\right)^b \tag{15}$$

Substituting Equations (15) and (4) into equation (14) and solving for the integral, the expected population density (equation 14) becomes:

$$D(t^*) = \delta \underbrace{\left(\phi^{a-1}\left(\frac{1}{1-a}\right)^a (R'_i)^a\right)^{(\gamma_i - 1/\gamma_i)}}_{Interior\, area} + (1-\delta) \underbrace{\left(\phi^{a-1}\left(\frac{1}{1-a}\right)^a (R'_j)^a\right)^{(\gamma_j - 1/\gamma_j)}}_{Exterior\ Area} \tag{16}$$

Substituting the reservation rent function (equation (13)) into equation (16) yields:

$$D(t^*) = \delta \underbrace{\left(\phi^{a-1}\left(\frac{1}{1-a}\right)^a \left(R_i{}^A + rC + (r-\alpha g)/\alpha r\right)^a\right)^{(\gamma_i - 1/\gamma_i)}}_{Interior\quad area} + (1-\delta) \underbrace{\left(\phi^{a-1}\left(\frac{1}{1-a}\right)^a \left(R_j{}^A + rC + (r-\alpha g)/\alpha r\right)^a\right)^{(\gamma_j - 1/\gamma_j)}}_{Exterior\quad area}$$

(17)

Equation (17) shows that the expected population density is a function of the equilibrium reservation rent, which as described above, is a function of returns to agricultural land (R^A), the share of land in the interior (δ), cost of conversion (c), rate of change in future land development rent (g), and shocks to future land development rent (σ). There are two key effects in equation

[6] Setting γ=1 produces a constant lot size across urban areas (Capoza and Helsley 1990).

(17) that we consider: (1) a composition effect and (2) a lot-size elasticity effect. The composition effect refers to how overall sprawl is affected by having relatively more or less development in the interior or periphery.

Using algebraic analysis to examine the effect of future rent shock (σ) on density (sprawl) is cumbersome. Alternatively, we used a simulation approach that showed that the findings depend on lot-size elasticity.[7] Specifically, an increase in the value of σ delays lot development, leading to a higher option value and a higher reservation rent (all else equal).

A higher reservation rent has a differential effect on density across interior and peripheral urban components. First, if delay is more likely to take place on the periphery, then greater uncertainty implies relatively more contemporaneous development in the interior, which reduces sprawl. One factor that may push development towards the interior is that greater uncertainty may lead developers to delay development in more remote locations. Overall, the share of new development going into the interior versus the periphery is a function of δ and the amount of available land in the interior. Second, the elasticity effect suggests a greater decrease in lot-size in the periphery relative to the interior because of our assumption of greater periphery lot-size elasticity.[8] Both the elasticity and composition effects suggest a pattern in which greater uncertainty increases density and reduces sprawl, all else equal, which will be the focus of our empirical assessment.

As for the effect of expected rent growth (g), the simulation exercise shows that it cannot be determined *a priori*, as it depends on both lot-size elasticity and the composition effects. For example, an increase in g reduces (increases) lot size if lot-size is price inelastic (elastic). Due to

---

[7] The simulation results are available on request.

[8] Our assumption of a lower elasticity in the interior than the periphery applies more to single-unit housing. However, we can also envision a scenario where uncertainty causes reservation land prices to rise to the extent that it encourages significant building of high-rise apartments (or condominiums). In this case, the land-use elasticity would be relatively greater in the interior, which would *reinforce* the negative relationship between sprawl and uncertainty. Yet, Mills (2006) points out the political difficulties of building high-rise apartments in urban settings even when it is allowed by zoning.

the interior/periphery composition effect, high growth may increase sprawl if it development is more prone to occur on the periphery, or it may make infill development more profitable—which reduces sprawl. Thus, it is also an empirical question as to the net effect of growth on sprawl.

3. Empirical Model and Data

3.1 Measuring Urban Sprawl

Urban sprawl can take different forms: low density development, clustering of population and economic activities at the urban fringe (edge cities), and fragmentation of land use (also known as leapfrog development) (Nechyba and Walsh 2004). Although there have been several attempts to develop alternative measures of sprawl, researchers have mainly focused on population density.[9] Density has been widely used because of its intuitive appeal and the difficulty and cost of obtaining alterative measures that require geographical information system (GIS) or other technologies (Lopez and Hynes 2003).[10] Moreover, higher density development is seen by many sprawl critics as an antidote to unwanted aspects of U.S. urban structure that accompany sprawl, such as infill land, loss of open space, and agricultural land.

Typically, metropolitan population density has been defined as the total metropolitan population divided by total metropolitan land. One major drawback of this measure is that large areas of counties within MAs are rural, leading to an upward bias (Lang 2003). Alternatively, other researchers (e.g. Fulton et al. 2001) measure sprawl over smaller geographic bases, namely, census urbanized areas.[11] However, the latter measure excludes relatively large areas of 'developed' land at the urban fringe leading to a downward bias in the sprawl measure (Cutsinger et al. 2005).

[9]Galster et al. (2000) devise eight measures that capture many dimensions of sprawl: density; continuity; concentration; compactness; centrality; nuclearity; diversity; and proximity.
[10]Burchfield et al. (2006) and Irwin and Bockstael (2007) are examples of studies that use GIS technology.
[11]A census urbanized area is a densely settled territory that consists of core census block groups or blocks that have a population density of at least 1,000 people per square mile, and surrounding census blocks that have an overall density of at least 500 people per square mile.

Another important concern of using aggregate measures of sprawl is that they typically do not consider the distribution of population within an urban area. There is no distinction between MAs where population is evenly distributed and those with highly concentrated population. Indeed, two otherwise equal MSAs in terms of overall population density may have very different types of land use and density patterns—e.g., a handful of very high density neighborhoods with low density elsewhere compared to a more equally distributed density pattern in the MA. Hence, we construct a sprawl index that incorporates both density and population distribution aspects of sprawl using a very small-scale geographic area—the census block group. A census block group is smaller than a tract (or neighborhood) and comprises a group of co-located census blocks that contain between 600 and 3,000 people, with a typical size of 1,500 people. Using block groups ensure that we capture fine differences in the intra-metropolitan distribution of population; these would be missed in an aggregate MSA measure such as average population density.

We measure metropolitan sprawl as:

$$Sprawl = ((L\% - H\%) + 1))*0.5$$

(18)

where *L%* is the share of metropolitan population living in block groups with population density *below* the overall U.S. metropolitan median block group. *H%* is the share of metropolitan population living in block groups with density *above* that of the overall U.S. metropolitan median block group. Our *Sprawl* index follows the definition of Lopez and Hynes (2003), which we keep so as to not add further confusion in the literature, though this measure simplifies to *L%*.

The sprawl measure in equation (18) is an index that ranges between 0 and 1. Values closer to 1 represent more sprawl. The population density of each census block is calculated by dividing its population by its land area (square miles). For each MA, block groups are sorted and aggregated into high density (above the U.S. metropolitan median block group) and low density (below the U.S. metropolitan median). A MA with a high percentage of its population living in block groups with density below the median has a greater share of its population living in a low density environment (more sprawl).[12] Unlike a simple average population density measure, our measure captures elements of density, differences in land use, and population gradient from the central business district. Though, we use the median density in our base analysis, sensitivity analysis (described below) shows that our sprawl findings are not sensitive to dividing *H%* and *L%* at places other than the median.

To account for rural clusters in MAs, we exclude all block groups with density below 200 persons per square mile. This cut-off corresponds to one residential unit per eight acres—about the lowest residential density of recent-vintage exurban housing developments. The source of census block group data is Geolytics Census data base (www.geolytics.com), which provides consistent census block group data for 1980, 1990, and 2000.[13] In sensitivity analysis, we consider all block groups in the MA, including those with less than 200 people per square mile.

Lopez and Hynes (2003) use a similar measure of sprawl, but use fixed density cut offs (high density corresponds to greater than 3,500 persons per square mile, while low density lies between 200 and 3,500). Our measure extends theirs because their choice of high and low density cut offs is determined somewhat arbitrarily while ours is based on the national median density and could be used for international comparisons.

Using the sprawl index in equation (18), we now provide a descriptive analysis of sprawl across U.S. MAs. Col. 1 in Table 1 reports the median sprawl index across all MAs in the sample. The 1980 median metropolitan sprawl index is 0.665, indicating that two thirds of the median MA population lived in a low density environment (sprawl).[14] Consistent with previous research (e.g.

[12] In measuring the degree of sprawl, Glaeser and Kahn (2004) consider the density at which the average city citizen lives. Their measure is specified as: $\sum(N_i/N)/(N_i/A_i)$, where N_i and A_i are the total population and area of a metropolitan block group. N is total metropolitan population. The correlation between their measure and our sprawl index (equation 18) is -0.60, -0.61, and -0.60 for the 1980, 1990, and 2000, respectively. The respective correlations between the sprawl index and the aggregate population density are -0.68, -0.71, and -0.71.

[13] Due to data limitations, we use 2000 census boundary definition instead of 1990 or 1980 definitions. An advantage of doing so is that we are able to consider the entire urbanized settlement pattern over the period.

[14] For the 1980, 1990, and 2000 censuses, the median U.S. density for metropolitan block group is 4,161, 4,358, and 4,367, respectively, showing little variation over time.

Nechyba and Walsh 2004), sprawl increases over time. That is between 1980 and 2000, sprawl expanded by 8 percent for a median MA using our measure.

The spatial distribution of sprawl is not uniform. As shown in cols 2 to 5 of Table 1, southern MAs exhibit the greatest sprawl, with the least in the west. Over time, sprawl has increased for all regions except in the west, which experienced a decrease over the 1980s but stabilized in the 1990s. The size of MAs is strongly associated with the degree of sprawl. Cols 2 to 4 in Table 2 divide MAs into: small (population less than 350,000), medium (350,000 - 1 million), and large (greater than 1 million). The results show that sprawl decreases as we move up the urban hierarchy. In sum, while sprawl in the median MA is increasing over time, there is significant variation across urban size and region. For example, Table 3 lists the 2000 sprawl index for 42 large U.S. cities and contrasts it with two widely used measures of sprawl. The first is the aggregate population density measure, constructed as the total metropolitan population divided by the metropolitan area's square miles after excluding rural clusters. The second measure uses Glaeser and Kahn's (2004) weighted average density (see footnote 12), which in our case measures the block group density at which an average person lives.

3.2 Empirical Model

Our primary question revolves around how uncertain future economic activity (σ) affects metropolitan sprawl. We use equation (19) to inform our econometric model. Yet, before discussing the estimation, we address several empirical issues.

First, the main estimation techniques we employ are fixed effects (FE) panel estimation and first difference (FD) approaches. These control for time-invariant metropolitan characteristics that are likely to affect sprawl, but assumed fixed in the theoretical model. Examples are soil quality, topographical and other geographical characteristics (Burchfield et al. 2006), historical zoning policies (Fischel 1985), and natural amenities (Wu 2001; Brueckner et al. 1999).[15] Therefore, any bias resulting from the correlation of those unmeasured urban characteristics and urban sprawl will be removed. A key implication is that correlated *omitted* MA factors that can cause endogeneity bias in cross-sectional models are accounted for in our approaches (Angrist and Pischke, 2009). The second estimation issue concerns measuring conversion cost. We are not aware of any data that estimate the cost of converting agricultural land into residential use at a metropolitan level. However, since conversion cost is associated with topography and other city specific factors, such as soil quality (Irwin and Bockstael 2004), inter-metropolitan differences in such conversion costs are controlled for.

Third, we control for other time variant economic and social variables that are not addressed in the theoretical model but that other research has shown to be influential in explaining sprawl. Doing so ensures that the uncertainty measure is not picking up the effects of other time varying omitted effects. Yet, this might come at the expense of introducing multicollinearity. Therefore, the sensitivity analysis section undertakes robustness checks in which estimates of more parsimonious models are reported (Partridge 2005). Finally, to mitigate any direct endogeneity, the explanatory variables are measured at a period prior to that of the dependent variable.[16] The empirical model is specified as:

$$\begin{aligned} sprawl_{it} = {} & \beta_1 avepop_{it-1} + \beta_2 stdpop_{it-1} + \beta_3 \log Income_{it-1} + \beta_4 Gini_{it-1} \\ & + \beta_5 Povertycentral_{it-1} + B_6 undevelopedland_{it-1} + \varphi_i + \mu_{t+} e_{it} \end{aligned} \tag{19}$$

The unit of observation is the Metropolitan Statistical Area (MSA) or Primary Metropolitan Statistical Area (PMSA).[17,18] The dependent variable is from equation (18) is from 1980, 1990,

[15]Burchfield et al. (2006) provide evidence that physical geography (natural barriers, ground water, terrain ruggedness) and climate explain 23.5 percent of the variation in sparse development across U.S. metropolitan areas. Wu (2006) finds that amenities provide an explanation for the leap-frog type of sprawl.
[16]Unless specified, the explanatory variables are measured at the initial periods (1970, 1980, and 1990).
[17]MSAs that cross state boundaries are assigned to the state in which the majority of their population resides.
[18]The general concept underlying MSAs is that of a large population nucleus combined with adjacent counties with close economic and social ties. MSAs are not closely associated with other metropolitan areas because they are typically surrounded by rural areas. PMSAs are individual metropolitan areas

and 2000.

To assign suitable proxies for the rates of change and shocks to expected future land development rent (g_i and σ_i), we follow a long literature that uses past changes to forecast future expectations (Burchfield et al. 2006; Plantinga et al. 2002; Cunningham's 2006; Holland et al. 2000). Land developers form their expectations regarding future rent in an adaptive manner based on *past* patterns of labor demand and supply (migration) shocks. The differences in the expected development rent across MAs are largely accounted for using expected MA population growth because lots can be developed anywhere in the MA (Burchfield et al. 2006; Plantinga et al. 2002).[19] This is consistent with recent urban literature that typically models shocks in housing demand and supply (rent shocks) as a function of changes in urban population (Moretti 2010). An adaptive model is reasonable in this setting because population growth across MAs is quite persistent over decades, i.e., past changes are highly correlated with current changes (Blanchard and Katz, 1992; Partridge and Rickman, 2003). Therefore, the *average annual change of MA population*[20] ($ave\Delta pop_{it\text{-}1}$), measured over 1969-75, 1969-85 and 1969-95, proxies for urban growth (g_i). For similar reasons, the *standard deviation of annual change of MA population* ($std\Delta pop_{it\text{-}1}$) is our proxy for uncertainty (σ_i) for the same periods.[21]

Using deep past annual population change is advantageous as $std\Delta pop_{it\text{-}1}$ and $ave\Delta pop_{it\text{-}1}$ are less likely to be endogenous. Yet, one probable drawback of these longer-term proxies is that they might be less informative for forming expectations than measures derived in the more recent past.

---

centered on a large central city or several closely related central cities. For more information, see: http://www.census.gov/geo/www/cob/ma_metadata.html.

[19] We also considered the standard deviation of per-capita income as another measure of the variation in demand shocks, but this measure did not perform nearly as well as population change. This result should not come as a surprise because population shocks more directly affect new housing demand for lots. Conversely, shocks to per-capita income would disproportionately affect existing households in the MA who already have housing.

[20] The data on population change are constructed from the U.S. Department of Commerce's Regional Economic Information System (REIS). Since REIS is data at the county level, all data are aggregated over the counties that belong to the same MA using the 2000 census definition. The same applies for the other explanatory variables.

[21] In constructing the proxies for g_i and σ_i, Plantinga et al. (2002) use changes in population density as a control variable. We cannot use their proxies due to possible simultaneity.

We experimented with using shorter periods such as 10 years to form our average population growth and standard deviation measures, but our conclusions were unaffected.

The rationale for using population change instead of percentage population change is that it is the level of population change, and the standard deviation of this change, that is most closely related to housing units or lots, which are the unit of development for the developer. Population change also more accurately reflects the amount of undeveloped land-supply that is needed for housing. The implied percentage change is largely irrelevant in our model because the actual number of housing units would differ greatly across small and large MAs for the same percentage changes. Another practical gain is that FE and FD estimation require sufficient within-MA variation in all explanatory variables to draw inference on metropolitan sprawl (e.g., Partridge 2005). Unlike most of the control variables, which are measured across censuses, constructing the growth and uncertainty proxies using annual percentage population change would be problematic as the time-series variation within a MA is typically very small. We discuss the consequences of using population change further below. For example, to check the validity of uncertainty-sprawl negative relationship, we construct uncertainty measures based on quarterly rent data available for large MAs (Davis and Palumbo 2008).

In this fixed effects setting, both urban growth and uncertainty are measured with five-year lags to help mitigate direct simultaneity (causality) problem.[22] While fixed effects estimation accounts for unobserved metropolitan factors, the identification requires a strict exogeneity of the explanatory variables and the error term. The endogeneity concern is further investigated in the sensitivity section. Previewing those results, potential endogeneity does not appear to affect the results. In the sensitivity analysis section, we further tackle the endogeneity concern using FD estimation, which, unlike in FE model, is not susceptible to the strict exogeneity restriction. The OLS and IV estimates of the FD model show that uncertainty-sprawl relationship is robust to

[22]The earliest data on estimated annual population is 1969. In order to utilize as much historical data as possible, $ave\Delta pop_{it-1}$ and $std\Delta pop_{it-1}$ are lagged for only 5 years.

using different estimation techniques.[23]

Regarding the time-varying control variables, $logincome_{it\text{-}1}$ is added to control for the income effect. Previous studies have emphasized the effect of income levels on the extent of sprawl and urban structure (Margo 1992; Brueckner and Fansler 1983; Anas 1978; Ottensmann 1977). Higher income makes it affordable for people to buy more space, and thus higher-income MAs are expected to have more sprawl. However, the effect of income is not clear since higher income is positively correlated with land value (Van Nieuwerburgh and Weill 2006), leading to a higher density (less sprawl). Furthermore, the Gini coefficient ($Gini_{it\text{-}1}$) is included to control for factors associated with income inequality effects including related social factors. However, many urban economists suggest that urban sprawl is affected by socioeconomic factors including ethnic segregation (Mills and Price 1984), crime rate (Sigelman and Heing 2001) and local public finance considerations (Tiebout 1956). Whereas most of these factors are accounted for by the metropolitan fixed effects, the central city poverty rate ($centralpoverty_{it\text{-}1}$) is added to control for time-variant inner-city socioeconomic conditions (Partridge and Rickman, 2008).[24] A higher central city poverty rate may overburden city governments, leading to higher taxes and/or deterioration of public services, which might encourage flight to low-density suburbs (Jordan et al. 1998). The source for the poverty and income variable is various issues of the U.S. Census Bureau, *County Data Book*.

Panel data on agricultural rent is not readily available at the MA level. As a proxy, we use the ratio of undeveloped to developed land, $undevelopedland_{it\text{-}1}$.[25] All else equal, a MA with a higher supply of undeveloped land is expected to have lower agricultural land rents, which is likely to increase sprawl. Data on developed/undeveloped land, measured in 1982, 1987, and 1997, is

---

[23]The FD technique will allow us to use deeper lags (10 years) for growth and uncertainty proxies. This would further mitigate the simultaneity problem.

[24]The effect of racial tension is discussed below (see footnote 24). However, consistent time series data on crime rate is not available for all three census periods (1970, 1980, and 1990), though crime's persistent effects are accounted for in the metropolitan fixed effect.

[25]Descriptive statistics for all variables are shown in Appendix Table 1.

derived from National Resource Inventory (NRI).[26] Finally, φ_i is MA dummies to capture metropolitan fixed effects, such as soil quality, topographical and other geographical characteristics (Burchfield et al. 2006). Such factors, as explained above, capture metropolitan differences in conversion cost of development. Other fixed effects include historical zoning policies (Fischel 1985), and natural amenities (Wu 2001; Brueckner et al. 1999). μ_t includes the time dummies, which control for common national time-varying effects that are not captured by the explanatory variables. Examples of such time-varying effects are changes in mortgage rates and national trends in public infrastructure expenditures.[27] As for the MA fixed effects, failing to control for time dummies might induce omitted variable bias.

A potential estimation concern is that the residuals could be spatially correlated, which would negatively bias the standard errors. To correct for this, the empirical model is estimated assuming that the residuals are correlated within a particular geographical cluster, but uncorrelated across clusters.[28] The advantage of using the clustering approach is that it does not impose restrictions on the spatial correlation of the residuals within clusters, unlike other spatial econometric models that use more restrictive assumptions, such as, distance or adjacency weight matrix.

4. Base Empirical Results

Most of the empirical discussion will be focused on the sprawl-uncertainty relationship, though other notable results are indicated. Starting with control variable results, as shown in col. 1 in Table 4, the coefficient of $logpercapita_{it-1}$ is positive and highly significant at the 5% level. This suggests richer MAs are likely to sprawl more (Brueckner and Fansler 1983; Margo 1992), which

[26]NRI defines developed land is "A combination of land cover/use categories, large urban and built-up areas, Small built-up areas, *and* Rural transportation land." Undeveloped land includes: cultivated cropland, non-cultivated cropland, pasture land, rangeland, forest, and rural transportation (roads and railways), http://www.nrcs.usda.gov/TECHNICAL/NRI/2002/glossary.html. Like REIS, NRI data is at the county level.
[27]Baum-Snow (2007) provides evidence that construction of highways contribute to suburbanization.
[28]The clusters used in the empirical model are BEA economic areas, which consist of one or more economic nodes that reflect regional centers of economic activities. The sample consists of 157 BEA economic areas. The Stata cluster command is used for the estimation. For more information on the definition of BEA economic area, see http://www.bea.gov/regional/docs/econlist.cfm.

is not surprising as homeowners can afford bigger housing lots. In addition, the Gini coefficient is positive and significant at around 5% level, indicating that as income dispersion increases, the rich are more likely to segregate in wealthy low-density areas, leading to more sprawl. Nonetheless, the results also show that central city poverty rates are not statistically associated with greater sprawl.[29]

Our base model's results show that uncertainty is associated with less sprawl, with the $std\Delta pop_{it-1}$ coefficient being negative and highly significant at the 1% level. This finding suggests that a one standard deviation increase in the $std\Delta pop_{it-1}$ is associated with a 2 percentage point decrease in sprawl, which compares to the sprawl index having a standard deviation of about 20 percentage points. The $ave\Delta pop_{it-1}$ coefficient is negative and significant at the 5% level. A one-standard deviation increase in the $ave\Delta pop_{it-1}$ is associated with a 1.4 percentage point decrease in sprawl. Also, the share of undeveloped land is positively associated with urban sprawl, suggesting that land availability allows developers to build bigger lots.

A possible criticism of using long series of annual population changes dating back to 1969 is that it might be less useful than more recent changes to form future predictions. To assess this possibility, we also used only recent changes (1969-74, 1975-84, and 1985-95) to construct the respective proxies for the uncertainty and growth variables, but these results were qualitatively similar.[30]

As indicated above, the main advantage of using FE estimation is to capture the unobserved metropolitan factors that are correlated with sprawl and the uncertainty proxy, which would otherwise produce biased estimates if the omitted factors are correlated with the other explanatory variables. Still, nonrandom state and time unobserved shocks that are not captured by the FE

[29]In an unreported regression, we added the ratio of black to white population in central cities to the base model to reflect racial segregation (data source *County Data Book*). These results suggest that the effect of uncertainty remains robust.
[30]The estimates and t-statistics of the uncertainty and growth variables are -1.7e-06 (-3.78) and -9.5e-07 (-4.75).

model might create an endogeneity concern. This endogeneity bias would occur if there are time trends that are specific to each state, perhaps due to changes in zoning law. Thus, in an unreported model, we include a vector of dummy variables that interact the time dummies with state dummies, where the reference year is 1970. The results show that the uncertainty-sprawl relationship is robust.

5. Sensitivity Analysis

This section further assesses whether the negative relationship between uncertainty and urban sprawl reported in the previous section is robust. To save space, the sensitivity analysis discussion will mostly be limited to the key uncertainty ($std\Delta pop_{it\text{-}1}$) results. The sensitivity analysis addresses the following: 1) collinearity; 2) endogeneity; 3) metropolitan size effect; 4) alternative sprawl measures; 5) model specification; 6) alternative estimation technique to the FE approach, and 7) alternative uncertainty measure.

5.1 Collinearity. To check if collinearity is a driving factor, the results of three parsimonious models are reported in cols. 2 to 4 of Table 4. Because some of the explanatory variables could be correlated with the residuals in other periods, omitting these variables and treating the model as a parsimonious reduced form may also help meet the assumptions of the fixed effects model. First, the ‘super’ parsimonious model (col. 2) includes only $ave\Delta pop_{it\text{-}1}$, $std\Delta pop_{it\text{-}1}$, and the time dummies. Col. 3 has the results of the base-model just omitting $centralpoverty_{it\text{-}1}$ and $undevelopedland_{it\text{-}1}$. Col. 4 reports the base model omitting the other economic variables, $logpercapi_{it\text{-}1}$ and $Gini_{it\text{-}1}$. In all three cases, the findings are robust, i.e., the $std\Delta pop_{it\text{-}1}$ coefficient is negative and significant at the 1% level. Also, the size of the $std\Delta pop_{it\text{-}1}$ coefficient in the parsimonious models is similar to the base models.

Another robustness check is related to possible collinearity between $ave\Delta pop_{it\text{-}1}$ and $std\Delta pop_{it\text{-}1}$. MAs with a higher population growth rate may have greater population growth variation. Yet, the correlation between $ave\Delta pop_{it\text{-}1}$ and $std\Delta pop_{it\text{-}1}$ is only 0.5. To investigate if this is a concern, we re-estimate the same models reported in Table 4 excluding first $ave\Delta pop_{it\text{-}1}$ and then $std\Delta pop_{it\text{-}1}$. These unreported results show that the negative uncertainty-sprawl relationship is robust.

5.2 Endogeneity. In the base model, we mitigate direct endogeneity between the uncertainty proxy ($std\Delta pop_{it\text{-}1}$) and the extent of sprawl by lagging the former 5 years. In the panel FE estimation, the lagging approach would hold under the assumption of strict exogeneity, i.e., the error terms have to be uncorrelated with $std\Delta pop_{it\text{-}1}$ at every date. To further assess endogeneity, we use 2SLS estimation in which $std\Delta pop_{it\text{-}1}$ is instrumented using a shift-share type prediction of the metropolitan-level standard deviation of population growth. This instrument directly assumes that the standard deviation of population growth is a function of underlying MA industry demand shocks that affect job growth.[31]

We perform a weak instrument test based on critical values provided by Stock and Yogo (2005). Reported at the bottom of Table 4, the Cragg-Donald Wald F-statistics exceeds the critical value for 10% maximal IV size distortion, indicating that the IVs are strong, confirming the uncertainty's negative impact on the extent of sprawl.[32] As reported in col. 5, Table 4, the estimate of $std\Delta pop_{it\text{-}1}$ is negative and significant, while the magnitude of standard deviation coefficient is even larger when employing IV.[33] The urban growth variable ($ave\Delta pop_{it\text{-}1}$) is

[31] The shift-share approach for instrumental variables became popular with the work of Bartik (1991) and Blanchard and Katz (1992). We use this approach as it captures the property that it only predicts sprawl *indirectly* through how it predicts stdΔpop (for more on instrumental variables, see Angrist and Pischke, 2009). Specifically, the shift share first calculates the standard deviation of annual employment growth for each one-digit industry *j* for the relevant period (1969-75, 1969-85, and 1969-95). Then, the standard deviation in industry *j's* growth rate (SD_j) is multiplied by the initial-period MA employment share (1969, 1975, 1985) for industry *j* (ES_j). These products are summed across all one-digit industries, not including government: $MASD=\sum_i(ES_i\times SD_j)$. The resulting sum *MASD* is approximately the predicted standard deviation of employment growth if the MA's private industries experienced annual employment shocks the size of the corresponding national average. The standard deviation of national industry employment growth should be exogenous to each of the given 313 MAs, meaning the identifying information comes from cross-sectional differences in the initial-period MA industry composition. We then multiply *MASD* by the initial MA population to obtain our final instrument. Thus, the shift-share type instrument captures our underlying notion of examining the impacts of labor demand shocks to population growth—i.e., having a different industry composition places the MA at a different risk of experiencing population/employment shocks due to exogenous national variation.

[32] In separate unreported regression, we estimated an IV version of the parsimonious models such as those depicted in Table 3, but the results continue to produce robust estimates.

[33] In a separate regression, we treat both urban growth and uncertainty variable as endogenous variables. In

statistically insignificant**.** The Hausman test statistic is statistically significant, suggesting that potential endogeneity may be biasing the OLS results, and thus we put more weight on the IV results (though they are similar to the OLS results).

5.3 Metropolitan size effect. One concern of using population change instead of percentage population change to measure $std\Delta pop_{it\text{-}1}$ is the underlying estimate might be heavily influenced by larger annual inflow (outflow) in big urban areas relative to small areas, magnifying the uncertainty measure of large urban areas. However, note that the MA fixed effect already controls for the average metropolitan area size effect, thus we are considering an additional robustness test for time-series variation in MA population. We first assess this issue by restricting our sample to MAs with population below 2 million in any census. The results (not reported) show that the magnitude of $std\Delta pop_{it\text{-}1}$ estimate did not significantly change.

Second, we carefully control for the MA size effect using the OLS and IV version of the FE model. Specifically, we now add metropolitan population ($\mathrm{MA\_pop}_{i,t\text{-}1}$) and interact stdΔpop and $ave\Delta pop$ with $\mathrm{MA\_pop}_{i,t\text{-}1}$. The results in Table 5 show that the stdΔpop estimate remains negative and highly significant, while the coefficient of (stdΔpop* $\mathrm{MA\_pop}_{1980}$) is positive and significant. Using the IV results, this pattern suggests that expected shocks to population growth have a negative influence on sprawl until the MA population reaches just over 5.7 million people (1990); that is, for all but three of the most populous MAs. In addition, the main 1990 MA population variable is also negative and statistically significant, consistent with the notion that larger MAs have less sprawl. Nonetheless, our general finding remains that uncertainty is associated with less sprawl in the vast majority of MAs

5.4 Alternative sprawl measures. A possible concern of measuring sprawl using our index is that it might not be comparable over time as the median cut-off density can vary over time. Yet,

---

this case the urban growth variable is instrumented using a similar instrument to that of the uncertainty variables, except that we use the average annual national employment growth of industry *i* instead of the standard deviation of the annual national employment growth. Unreported, the results indicate that base model's estimate of $std\Delta pop_{it\text{-}1}$ are robust.

as reported above (footnote 12), the median density changes little over the three censuses (1980, 1990, and 2000). We thus reconstruct the sprawl measure using 1980 national median population as a benchmark. The re-estimated base model (not shown) reveals that the uncertainty-sprawl relationship is robust. Another inter-temporal comparison concern is that in restricting the measurement of sprawl index to block groups with density above 200 people, the metropolitan block groups that qualify may vary across the three censuses as densities fall below or rise above the 200 threshold. To account for this, we reconstruct the sprawl index including all metropolitan block groups regardless of their densities. Still, the OLS and IV estimates of uncertainty are robust with the coefficient of $std\Delta pop_{it-1}$ being negative and highly significant as respectively reported in cols. 1 and 2 of Table 6.

We argue that our specification of the sprawl index, as shown in equation (18), is superior to other measures based on aggregate population density because it accounts for *both* overall density and intra-metropolitan distribution of population. Still, one concern about the sprawl index is it does not distinguish between block groups with extreme density and those with density just above the median. However, to check whether the negative uncertainty/sprawl relationship is an artifact of the specification of the sprawl index, we undertake several sensitivity checks. First, we use the 25th percentile of the overall metropolitan block group density as a more conservative cut-off than the median.[34] We then use the 75th percentile as a second alternative. The OLS estimates (cols. 3 and 5) and IV estimates (cols. 4 and 6) reported in Table 4 suggest that the uncertainty estimates are robust to using the alternative cutoffs. Second, we next measure urban sprawl using aggregate population density, with rural clusters excluded. In this case, urban density is expected to be *higher* in MAs with greater uncertainty, which is supported by the OLS and IV results (see column 7 and 8, Table 4).[35,36]

[34]This implies that a metropolitan area with a high percentage of its population living in block groups with a density below the 25th percentile would have more low-density sprawl compared to those with block group densities above. For 1980, 1990, and 2000, the 25th percentile population densities of overall metropolitan block groups are respectively 1641, 1835, and 1931 people per square mile.

[35]The Hausman test statistics for alternative sprawl models are statistically significant suggesting that

We repeat the same exercise using Gleaser and Kahn (2004) weighted average density as the dependent variable. This measure gives more weight for high density block groups. The greater the value of this measure, the denser (less sprawl) is the urban area. Using a log-linear FE regression, the uncertainty OLS and IV estimates are positive and highly significant, supporting the aforementioned findings (not shown).

5.5 Model specification. The above results are obtained when the model is estimated at the level (linear) form. To check whether the results are specific to the model specification, we re-estimate the base model using log-linear specifications. Unreported, the estimates show that the negative effect of uncertainty is robust. The same specification is applied for sprawl measures using several alternative cut-off densities relative to the national median in addition to measuring sprawl using aggregate density, as specified above. Here again, the estimates are robust to the model specifications.

5.6 Alternative estimation techniques (First Difference Estimation). A key assumption of the FE model is that the explanatory variables are uncorrelated with the residuals at all dates, i.e., the strict exogeneity assumption. However, the strict exogeneity assumption does not apply to the first difference model (FD). Another advantage of using the FD estimation is that we are able to measure urban growth and uncertainty with deeper lags (10 years), which further mitigate any simultaneity. FD models account for metropolitan fixed effects by differencing them out, using the following specification: $Y_{(2000\text{-}1990)} = XB_{(1990\text{-}1980)}$. In this case, $std\Delta pop_{it\text{-}1}$ and $ave\Delta pop_{it\text{-}1}$ are constructed as innovations between consecutive dates, i.e., 1990 $std\Delta pop_{it\text{-}1}$ and 1990 $ave\Delta pop_{it\text{-}1}$ are measured between 1980 and 1990; 1980 $std\Delta pop_{it\text{-}1}$ and 1980 $ave\Delta pop_{it\text{-}1}$ are measured between 1969 and 1980. Nonetheless, the resulting FD estimates reported in col. 1 of Table 7 show that the estimate of $std\Delta pop_{it\text{-}1}$ is negative and significant at the 5 percent level.

---

potential endogeneity may be affecting the OLS results.

[36]In an unreported model, we also estimate the density models with $ave\Delta pop_{it\text{-}1}$ excluded, though the coefficient of s$td\Delta pop_{it\text{-}1}$ remains positive and highly significant.

Even though we lag the uncertainty proxy by 10 years, an endogeneity problem might arise if population change is persistent over time. We consider this issue with IV estimation using two instrumental variables that are similar to the FE instrumental variable. The first is specified as the standard deviation of the annual national employment growth of that industry, 1969-1980, multiplied by the MA's industry employment share in 1969. The sum of the industry shift shares is then multiplied by the metropolitan population of 1969 to form the first instrument (see footnote 31). The second IV is the same except that the standard deviation is measured between 1980 and 1990 and the sum of the industry shift shares is multiplied by the metropolitan population of 1980. As shown in col. 2, Table 7, the two instruments are strong predictors of the first difference change in stdΔpop. The resulting regression results suggest that the *stdΔpop* IV estimate is negative and significant at the 5% level.[37] The magnitude of the IV standard deviation coefficient is larger than for the OLS results, again suggesting that our conclusions are unaltered by using IV.

5.7 Alternative measure of uncertainty. The negative uncertainty-sprawl findings we reported above are based on constructing the uncertainty proxy using historical population data—i.e., we consider the underlying labor demand and labor supply shocks. To further check the validity of our findings, we construct the uncertainty measure using quarterly rent data available for 42 large MAs reported by Davis and Palumbo (2008).[38] The estimation technique is similar to the above FD model, such that the uncertainty ($\text{std}\Delta\text{rent}_{2000\text{-}1989}$) and growth ($\text{ave}\Delta\text{rent}_{2000\text{-}1989}$) variables are constructed as $\text{std}\Delta\text{rent}_{2000}$ - $\text{std}\Delta\text{rent}_{1989}$ and $\text{ave}\Delta\text{rent}_{2000}$ - $\text{ave}\Delta\text{rent}_{1989}$, respectively. The $\text{std}\Delta\text{rent}_{2000}$ and $\text{std}\Delta\text{rent}_{1989}$ are measured as the standard deviation of quarterly rent change between 2000 and 1990, and 1989 and 1984 respectively. The average quarterly rent change

[37]When treating both the uncertainty and growth variable as endogenous, the uncertainty estimate is still robust. In this case, the instruments for the growth variable are the same as those for the uncertainty variables, except that the industry shift share is now calculated using the average annual industry employment growth.

[38]The rent uncertainty and growth variables are constructed as above but for different periods. For uncertainty, we employ the standard deviation of quarterly rent change. For the average growth variable, we use the average quarterly rent change. A Hausman test suggested that we could not reject the null hypothesis that the rent variables were exogenous, and thus we employ OLS.

(aveΔrent) of the same periods is used to construct the aveΔrent$_{2000}$ and aveΔrent$_{1989}$. To control for the size effect, we include MA population (MA_pop$_{1990\text{-}1980}$) and interact stdΔrent$_{2000\text{-}1989}$ and aveΔrent$_{2000\text{-}1989}$ with MA_pop$_{1990}$. The other control variables are specified as the above. The results show that the uncertainty effect is robust. The coefficient estimates (t-statistics) for stdΔrent$_{2000\text{-}1989}$ and stdΔrent$_{2000\text{-}1989}$* MA_pop$_{1990}$ are -1.0E-05 (-6.39) and 5.4E-12 (4.81), respectively. Hence, these results are very similar to the corresponding results in Table 5.

6. Conclusion

This paper investigates how urban sprawl responds to an expanding and volatile urban economy. We modify the theoretical work of Capozza and Helsley (1990) and Capozza and Li (1994) by assuming that changes in population density across urban areas are due to changes in lot size. Second, we assume that households chose the optimal lot size, while developers take lot size as given and choose the optimal time for land development. We derive a theoretical expression that links population density (sprawl) to uncertainty over future land development rent, among other factors. The major prediction of the theoretical model is that an urban area with higher levels of uncertainty is expected to have higher density (less sprawl).

The empirical analysis draws upon panel data from U.S. MAs over 1980-2000 censuses. Based on the theoretical model, we construct a distinctive sprawl index that captures both overall density and population distribution. In constructing the index, we use a disaggregated scale of data to permit identification of intra-metropolitan differences in the distribution of population density and land use.

The theoretical model shows that urban sprawl can be affected by factors other than those derived from static and dynamic models with perfect future foresight (Carruthers 2003). Consistent with the theoretical prediction, we provide evidence that sprawl is negatively influenced by uncertainty regarding future land development rent. As a proxy for uncertainty, we use the standard deviation of past annual metropolitan population change.

A key implication of the findings is that it is inappropriate to design urban planning policies in isolation from considering the underlying economic volatility. For two otherwise identical urban areas, the results suggest that one with greater uncertainty will manifest less sprawl and therefore may require a different urban planning package. For urban planners and policymakers, appropriate recognition of the role of uncertainty regarding land use will improve their ability to deliver efficient growth management of urban areas and reduce costs of servicing residential development.

References

Alonso W (1964) *Location and land use*. Cambridge: Harvard University Press.

Anas A (1978) Dynamics of urban residential growth. *Journal of Urban Economics* 5:66-87.

Angrist, J. D. and Pischke, J. (2009). *Mostly Harmless Economics, An Empiricist's Companion.* Princeton, NJ: Princeton University Press.

Bar-Ilan A, Strange WC (1996) Urban development with lags. *Journal of Urban Economics* 39:87–113.

Bartik, Timothy J., 1991. *Who Benefits From State and Local Economic Development Policies?* Kalamazoo, MI: Upjohn Institute.

Baum-Snow N (2007) Did highways cause suburbanization? *Quarterly Journal of Economics*, 122: 775-805.

Blanchard O. and Katz L. F. (1992) Regional Evolutions. *Brookings Papers on Economic Activity* 1-75.

Brueckner JK, Fansler DA (1983) The Economics of urban sprawl: Theory and evidence on the spatial size of cities. *The Review of Economics and Statistics,* 65: 479-482.

Brueckner JK, Thisse JF, Zenou Y (1999) Why is central Paris rich and downtown Detroit poor?: An amenity-based theory. *European Economic Review*, Elsevier, 43: 91-107.

Burchfield M, Overman H, Puga, D, Turner M. (2006) Causes of sprawl: A portrait from space. *Quarterly Journal of Economics*, 121: 587-633.

Caballero RJ, Pindyck RS (1996) Uncertainty, investment, and industry evolution, *International Economic Review*, 37: 641–662.

Capozza DR and Helsley RW (1990) The stochastic city. *Journal of Urban Economics*, 28: 187-203.

Capozza DR, Yuming Li. (1994) The intensity and timing of investment: The case of land. *American Economic Review*, 84: 889-904.

Carruthers John I. 2003. Growth at the Fringe: The Influence of Political Fragmentation in United States Metropolitan Areas. *Papers in Regional Science* 82(4): 475-99.

Carruthers, John I., and Gudmundur F. Ulfarsson, 2008. Does 'Smart Growth' Matter to Public Finance? *Urban Studies* 45(9): 1791-1823.

Carruthers John I., and Alexander C. Vias. 2005. Urban, Suburban, and Exurban Sprawl in the Rocky Mountain West: Evidence from Regional Adjustment Models. *Journal of Regional Science* 45: 21–48.

Cunningham CR (2006) House price uncertainty, timing of development, and vacant land prices: Evidence for real options in Seattle. *Journal of Urban Economics,* 59: 1-31

Cutsinger J, George G, Wolman H, Hanson R, Towns D (2005) Verifying the multi-dimensional nature of metropolitan land use: Advancing the understanding and measurement of sprawl.

Journal of Urban Affairs, 27: 235–259.

Davis, *A.* and Palumbo, M. G *(*2008) “The Price of Residential Land in Large U.S. Cities*" Journal of Urban Economics, 63: 352-384.*

Fallah, Belal, Mark D. Partridge, M. Rose Olfert. (forthcoming) “Urban Sprawl and Productivity: Evidence from U.S. Metropolitan Areas.” *Papers in Regional Science*. DOI: 10.1111/j.1435-5957.2010.00330.x.

Fischel WA (1985) The economics of zoning laws: A property rights approach to American land use controls. Baltimore, md: Johns Hopkins University Press.

Fulton WR, Pendall MN, Harrison A (2001) Who sprawls most? How growth patterns differ across the U.S. Washington, DC: Brookings Institution.

Galster GR, Hanson HW, Coleman S, Friebage J (2000). Wrestling sprawl to the ground: Defining and measuring an elusive concept. Washington, DC: Fannie Mae.

Glaeser EL, Kahn ME (2004) Sprawl and urban growth. *Handbook of Regional and Urban Economics,* 4:2481-2527.

Harrison D, Kain J (1974) Cumulative urban growth and urban density functions. *Journal of Urban Economics* 1: 61-98.

Holland AS, Ott SH, Riddiough TJ (2000) The role of uncertainty in investment: An examination of competing investment models using commercial real estate data. *Real Estate Economics*, 28: 33–64.

Irwin EG, Bockstael NE (2004) Land use externalities, open space preservation, and urban sprawl. *Regional Science and Urban Economics*, 34: 705-725

Irwin EG, Bockstael NE (2007) The evolution of urban sprawl: Evidence of spatial heterogeneity and increasing land fragmentation.” *Proceedings of the National Academy of Sciences* (PNAS), 104: 20672-20677.

Jordan S, Ross JP, Usowski KG (1998) U.S. suburbanization in the 1980s. *Regional Science and Urban Economics*, 28: 611-627(17).

Karlin H, Taylor M (1975) A first course in stochastic processes, Academic Press, New York.

Lang R (2003) Open spaces, bounded places: Does the American west’s arid landscape yield dense metropolitan growth? *Housing Policy Debate,* 13, 755–778.

Lopez R and Hynes H (2003) Sprawl in the 1990s: measurement, distribution, and trends. *Urban Affairs Review,* 38: 325–355.

Margo RA (1992) Explaining the postwar suburbanization of population in the United States: The role of income. *Journal of Urban Economics,* 31: 301-310.

McDonald R, Siegel D (1985) Investment and the valuation of firms when there is an option to shut down. *International Economic Review*, 26: 331–349.

McDonald R, Siegel D (1986) The value of waiting to invest, *Quarterly Journal of Economics*, 101: 707–727.

McFarlane A (1999) Taxes, fees, and urban development. *Journal of Urban Economics*, 46: 416-436.

Mills ES (1967) An aggregative model of resource allocation in metropolitan area. *American Economic Review*, 197: 197-210.

Mills DE (1981) Growth, speculation and sprawl in a monocentric city. *Journal of Urban Economics* 10:201–226.

Mills, ES, (2006) Sprawl and Jurisdictional Fragmentation. *Brookings-Wharton Papers on Urban Affairs,* 231-256. DOI: 10.1353/urb.2006.0023.

Mills, ES, Price R (1984) Metropolitan suburbanization and central city problems. *Journal of Urban Economics*, 15: 1-17.

Moel A, Tufano P (2002) When are real options exercised? An empirical study of mine closings. *Review of Financial Studies*, 15: 35–64.

Moretti, Enrico (2010) Local Labor Markets. IZA Discussion Papers: 4905.

Muth R F (1969) Cities and housing: the spatial pattern of urban residential land use. Chicago, IL: University of Chicago Press.

Nechyba T, Walsh R (2004) Urban sprawl, *Journal of Economic Perspectives*, 18:177-200.

Ottensmann JR (1977) Urban sprawl, land values and the density of development. *Land Economics*, 53: 389-400.

Paddock JL, Siegel D, Smith JL (1988) Option valuation of claims on real assets: The case of offshore petroleum leases, *Quarterly Journal of Economics*, 103: 479–508.

Partridge Mark D. (2005) Does income distribution affect U.S. state economic growth? *Journal of Regional Science*, 45: 363-394.

Partridge, Mark D., D.S. Rickman. (2003) The Waxing and Waning of U.S. Regional Economies: The Chicken-Egg of Jobs Versus People. *Journal of Urban Economics* 53: 76-97.

Partridge, Mark D., Dan S. Rickman. (2006) Fluctuations in Aggregate U.S. Migration Flows and Regional Labor Market Flexibility. *Southern Economic Journal* 72: 958-980.

Partridge, Mark D., D.S. Rickman. (2008) Does a Rising Tide Lift All Metropolitan Boats? Assessing Poverty Dynamics by Metropolitan Size and County Type" *Growth and Change* 39: 283-312.

Pindyck R (1991) Irreversibility, uncertainty, and investment. *Journal of Economic Literature*, 29: 1110-48.

Plantinga AJ, Lubowski RN, Stavins RN (2002) The effects of potential land development on agricultural land prices. *Journal of Urban Economics*, 52: 561-581.

Sigelman L, Heing J (2001) Crossing the great divide – race and preferences for living in the city versus the suburbs. *Urban Affairs Review*, 37: 3–18.

Stock, J. H., and M. Yogo (2005) Testing for Weak Instruments in Linear IV Regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, (eds. D.W. Andrews and J. H. Stock) Cambridge: Cambridge University Press. pp. 80-108.

Tiebout CM (1956) A pure theory of local expenditure. *Journal of Political Economy*, 64: 416-24.

Titman S (1985) Urban land prices under uncertainty. *American Economic Review*, 75: 505–514.

Van Nieuwerburgh, Stijn and Pierre-Olivier Weill (2006) Why Has House Price Dispersion Gone Up? NBER Working Paper #12538.

Wu, JunJie (2001) Environmental amenities and the spatial pattern of urban sprawl. *American Journal of Agricultural Economics*, 83: 691–69.

Wu, JunJie (2006) Environmental amenities, urban sprawl, and community characteristics. *Journal of Environmental Economics and Management*, 52: 527-547.

Wu, JunJie (2010) Economic Fundamentals and Urban-Suburban Disparities. *Journal of Regional Science*, 50: 570-591.

Table 1: Extent of Sprawl in U.S. Metropolitan Areas (MAs), by Region, 1980-2000

Year	All MSAs (1)	Northeast (2)	Midwest (3)	South (4)	West (5)
1980	0.665	0.653	0.611	0.785	0.595
1990	0.684	0.677	0.644	0.826	0.532
2000	0.720	0.705	0.694	0.861	0.539
# of MAs	328	57	80	126	65

Note: Extent of Sprawl indicated by the median sprawl index of the respective regions.

Table 2: Extent of Sprawl in U.S. Metropolitan Areas, by Urban Size, 1980-2000

Year	All MSAs (1)	Small (2)	Medium (3)	Large (4)
1980	0.665	0.716	0.578	0.437
1990	0.684	0.738	0.615	0.429
2000	0.720	0.777	0.663	0.467

Note: Extent of Sprawl indicated by the median sprawl index of the respective metropolitan sizes. Small MSAs are defined as < 350,000 people; medium MSAs, 350,000 - 1 million people; large MSAs > 1 million.

Table (3) 2000 Sprawl Measures of large U.S. Metropolitan Areas.

Metropolitan Areas	sprawl index	Aggregate density[a]	Weighted average Density[b]
New York, NY PMSA	0.074	10437	64214
Los Angeles--Long Beach, CA PMSA	0.124	6388	14883
San Jose, CA PMSA	0.162	4489	9555
Miami, FL PMSA	0.177	5164	10605
San Francisco, CA PMSA	0.185	4825	20784
Oakland, CA PMSA	0.261	3261	9415
San Diego, CA MSA	0.311	3093	8600
Chicago, IL PMSA	0.329	3187	12030
Denver, CO PMSA	0.332	2808	6159
New Orleans, LA MSA	0.348	2442	7186
Phoenix--Mesa, AZ MSA	0.384	2787	6004
Sacramento, CA PMSA	0.391	2155	5640
Salt Lake City--Ogden, UT MSA	0.443	2425	4978
Philadelphia, PA--NJ PMSA	0.460	2063	10344
Seattle--Bellevue--Everett, WA PMSA	0.466	1985	6061
Detroit, MI PMSA	0.467	2355	5334
Washington, DC--MD--VA--WV PMSA	0.475	2004	7573
Houston, TX PMSA	0.482	2171	5690
Dallas, TX PMSA	0.486	2075	6195
Riverside--San Bernardino, CA PMSA	0.489	1927	5104
San Antonio, TX MSA	0.490	2113	4807
Baltimore, MD PMSA	0.494	1723	7623
Buffalo--Niagara Falls, NY MSA	0.497	1677	5837
Milwaukee--Waukesha, WI PMSA	0.498	1715	6393
Cleveland--Lorain--Elyria, OH PMSA	0.535	1729	5387
Norfolk--Virginia Beach--Newport News, VA--NC MSA	0.561	1971	4519
Fort Worth--Arlington, TX PMSA	0.587	1691	4216
Tampa--St. Petersburg--Clearwater, FL MSA	0.607	1779	3858
Minneapolis--St. Paul, MN--WI MSA	0.623	1697	4602
St. Louis, MO--IL MSA	0.641	1635	4087
Memphis, TN--AR--MS MSA	0.642	1645	3972
Pittsburgh, PA MSA	0.648	1168	4462
Cincinnati, OH--KY--IN PMSA	0.655	1422	3938
Rochester, NY MSA	0.662	1357	4312
Oklahoma City, OK MSA	0.669	1546	3573
Kansas City, MO--KS MSA	0.725	1595	3406
Indianapolis, IN MSA	0.733	1423	3263
Hartford, CT MSA	0.755	678	2409
Columbia, MO MSA	0.841	1091	3178
Birmingham, AL MSA	0.842	981	2410
Atlanta, GA MSA	0.848	1161	2664
Portland, ME MSA	0.904	567	1487
Charlotte--Gastonia--Rock Hill, NC--SC MSA	0.915	840	1962
Providence--Fall River--Warwick, RI--MA MSA	0.977	535	1146

[a]Measured as total metropolitan population divided by total metropolitan area square miles, excluding rural clusters (block groups with density below 200 persons per square mile.

[b]Measured as: $\sum(N_i/N)/(N_i/A_i)$, where N_i /N is the share of a block group population to total metropolitan population. A_i is the area of a metropolitan block g.

Table 4: The Effect of Uncertainty on Sprawl, Fixed Effects Models

Variable	OLS Base Model -1-	OLS Parsimonious Models -2-	-3-	-4-	IV Base model -5-
Std∆pop	-2.80E-06	-3.12E-06	-2.97E-06	-2.86E-06	-1.09E-05
	(-3.69)	(-3.81)	(-3.58)	(-3.38)	(-2.75)
Ave∆pop	-1.29E-06	-1.02E-06	-1.13E-06	-1.19E-06	1.38e-06
	(-2.25)	(-2.02)	(-2.12)	(-2.21)	(0.78)
Log per capita income	0.08		0.018		0.073
	(2.46)		(1.25)		(2.09)
Gini	0.172		0.164		0.171
	(1.96)		(2.12)		(1.81)
Poverty rate in central cities	0.143			0.0194	0.134
	(1.39)			(0.23)	(1.27)
Share of undeveloped land	0.0001			0.0001	0.0001
	(3.71)			(3.64)	(3.58)
Period dummies	Y	Y	Y	Y	Y
N[a]	939	957	950	939	939
Overall R^2	0.06	0.19	0.13	0.14	0.26
Weak Inst. Wald F[b]					251
Hausman test: p-val[c]					0.00

Note: Robust (spatially clustered) t-statistics are in parenthesis. In calculating the robust t-statistics, the clusters are formed based on BEA economic areas, which are defined as the relevant regional markets surrounding metropolitan or micropolitan statistical areas. See: http://www.bea.doc.gov/bea/regional/docs/econlist.cfm.

[a]The number of observations varies across models due to missing data for some explanatory variables.

[b]Exceeds Stock and Yogo (2005) critical values for 10% maximal IV size distortion indicating that the IVs are strong.

[c]Hausman test for the null hypothesis that the OLS results do not suffer from (statistical) endogeneity bias. P-values greater than 0.05 suggest that the null hypothesis cannot be rejected, suggesting that OLS is the preferred estimator.

Table (5) The effect of uncertainty on Sprawl: controlling for Metropolitan size effect

Variable	OLS	IV
StdΔpop	-2.49E-06	-2.00E-05
	(-2.53)	(-2.66)
AveΔpop	-1.91E-06	4.49E-06
	(-2.00)	(1.38)
MA population 1990	-1.15E-07	-9.90E-08
	(-6.49)	(-3.12)
StdΔpop * MA population 1990	9.47E-13	3.50E-12
	(4.64)	(2.45)
AveΔpop * MA population 1990	1.54E-13	-6.57E-13
	(1.29)	(-1.05)
Log per capita income	0.078	0.085
	(2.45)	(2.25)
Gini	0.108	0.163
	(1.20)	(1.59)
Poverty rate in central cities	0.087	0.164
	(0.9)	(1.45)
Share of undeveloped land	0.0001	0.0001
	(3.96)	(3.67)
Period dummies	Y	Y
N	939	939
Overall R^2	0.22	0.17
Weak Inst. Wald F[b]		13.2
Hausman test: p-val[c]		0.0002

Note: Robust (spatially clustered) t-statistics are in parenthesis. In calculating the robust t-statistics, the clusters are formed based on BEA economic areas, which are defined as the relevant regional markets surrounding metropolitan or micropolitan statistical areas. See: http://www.bea.doc.gov/bea/regional/docs/econlist.cfm.

[a]The number of observations varies across models due to missing data for some explanatory variables.

[b]Exceeds Stock and Yogo (2005) critical values for 15% maximal IV size distortion, i.e., the IVs are strong.

[c]Hausman test for the null hypothesis that the OLS results do not suffer from (statistical) endogeneity bias. P-values
greater than 0.05 suggest that the null hypothesis cannot be rejected, suggesting that OLS is the preferred estimator.

Table 6: The Effect of Uncertainty on Sprawl, FE Models, Alternative Sprawl Measures.

Variable	Median cut-off (Including Rural Clusters)		75th Percentile cut-off		25th Percentile cut-off		Population density	
	-1- OLS	-2- IV	-3- OLS	-4- IV	-5- OLS	-6- IV	-7- OLS	-8- IV
StdΔpop	-1.82E-06	-7.46E-06	-2.60E-06	-1.2E-05	-1.28E-06	-4.97E-06	0.006	0.03187
	(-3.13)	(-2.49)	(-3.27)	(-2.58)	(-3.55)	(-2.75)	(2.21)	(2.53)
AveΔpop	-1.43E-06	4.19E-07	-1.16E-06	1.89E-06	-1.11E-06	9.48E-08	0.006	-0.00194
	(-3.00)	(0.29)	(-2.67)	(0.99)	(-2.84)	(0.1)	(3.68)	(-0.30)
Log per capita inc.	0.075	0.071	0.086	0.078	0.039	0.035	-121.93	-101.027
	(2.36)	(2.14)	(3.1)	(2.62)	(1.27)	(1.18)	(-0.68)	(-0.53)
Gini	0.22	0.22	0.026	0.024	0.218	0.217	-1278.	-1274.
	(3.07)	(2.92)	(0.31)	(0.25)	(2.87)	(2.89)	(-2.59)	(-2.57)
Poverty rate in central cities	0.184	0.18	0.2261	0.22	-0.022	-0.024	-199	-182
	(2.07)	(1.95)	(2.67)	(2.31)	(-0.22)	(-0.24)	(-0.4)	(-0.37)
Share of undeveloped land	0.00008	7.01E-05	9.55E-05	8.08E-05	6.30E-05	5.713E-05	-0.21	-0.17
	(3.19)	(3.06)	(3.35)	(3.04)	(3.83)	(3.25)	(-1.69)	(-1.19)
Period dummies	Y	Y	Y	Y	Y	Y	Y	Y
N[a]	939	939	939	939	939	939	939	939
Overall R^2	0.06	0.16	0.09	0.25	0.05	0.13	0.03	0.20
Weak Inst. Wald F[b]		251		251		251		251
Hausman test: p-val[c]		0.00		0.00		0.00		0.00

Note: Robust (spatially clustered) t-statistics are in parenthesis. In calculating the robust t-statistics, the clusters are formed based on BEA economic areas, which are defined as the relevant regional markets surrounding metropolitan or micropolitan statistical areas. See: http://www.bea.doc.gov/bea/regional/docs/econlist.cfm.

[a]The number of observations varies across models due to missing data for some explanatory variables.

[b]Exceeds Stock and Yogo (2005) critical values for 10% maximal IV size distortion indicating that the IVs are strong.

[c]Hausman test for the null hypothesis that the OLS results do not suffer from (statistical) endogeneity bias. P-values greater than 0.05 suggest that the null hypothesis cannot be rejected, suggesting that OLS is the preferred estimator.

Table 7: The Effect of Uncertainty on Sprawl: First Difference Models

Variable	FD[†](OLS)	FD[†] (IV)
StdΔpop	-6.92E-07	-3.72e-06
	(-2.17)	(-2.18)
AveΔpop	-2.89E-07	-6.57e-07
	(-2.21)	(-1.85)
log percapita income	-0.003	-0.003
	(-3.38)	(-3.30)
Gini	0.28	0.23
	(3.43)	(2.49)
poverty rate in central cities	0.088	0.18
	(2.66)	(1.74)
share of undeveloped land	-9.33E-05	-7.48E-04
	(-2.36)	(-2.18)
Constant	0.038	0.035
	-10.49	(8.22)
N[a]	320	319
R^2	0.07	0.05
Weak Inst. Wald F[b]		59.0
Hausman test: p-val[c]		0.00

Note: Robust (spatially clustered) t-statistics are in parenthesis. In calculating the robust t-statistics, the clusters are formed based on BEA economic areas, which are defined as the relevant regional markets surrounding metropolitan or micropolitan statistical areas. See: http://www.bea.doc.gov/bea/regional/docs/econlist.cfm.

[†]The sprawl index is constructed using median cut-off density of the overall metropolitan block group density.

[a]The number of observations varies across models due to missing data for some explanatory variables.

[b]Exceeds Stock and Yogo (2005) critical values for 10% maximal IV size distortion indicating that the IVs are strong.

[c]Hausman test for the null hypothesis that the OLS results do not suffer from (statistical) endogeneity bias. P-values greater than 0.05 suggest that the null hypothesis cannot be rejected, suggesting that OLS is the preferred estimator.

Appendix Table 1: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
2000					
Sprawl index (median cut of density)	328	0.703	0.196	0.028	1
Sprawl index (25th percentile cut of density)	328	0.419	0.184	0.017	1
Overall density (population/area)	328	1473.7	1219.7	276.8	16156.1
AveΔpop	320	6484.8	11967.1	-10664	80734.8
StdΔpop	319	4792.5	7704.5	298.4	70922.6
log per capita income	319	9.79	0.175	9.14	10.34
Gini	318	0.41	0.029	0.336	0.508
share of undeveloped land	316	16.05	46.94	0.415	778.34
poverty rate in central cities	316	0.131	0.049	0.000	0.411
1990					
Sprawl index (median cut of density)	328	0.676	0.194	0.030	1
Sprawl index (25th percentile cut of density)	328	0.380	0.174	0.013	1
Overall density (population/area)	328	1513.37	1128.88	280.23	14657.04
AveΔpop	320	5696.56	11307.40	-33287.10	83314
StdΔpop	319	4417.35	6961.41	313.56	73959.45
log per capita income	317	9.15	0.240	6.95	9.70
Gini	318	0.44	0.039	0.370	0.637
share of undeveloped land	316	19.67	65.71	0.483	1120.05
poverty rate in central cities	316	0.118453	0.041621	0.000	0.346
1980					
Sprawl index (median cut of density)	328	0.669	0.181	0.031	1
Sprawl index (25th percentile cut off density)	328	0.348	0.168	0.011	1
Overall density (population/area)	328	1517.00	1175.59	340.21	16410.04
AveΔpop	320	5586.01	10243.11	-56457.2	65251.67
StdΔpop	319	4461.50	6801.22	215.58	76681.16
log per capita income	319	8.23	0.170	7.57	8.73
Gini	317	0.324	0.030	0.257	0.459
share of undeveloped land	316	28.277	108.052	0.5	1798.21
poverty rate in central cities	316	0.135	0.061	0.000	0.486