ECONOMIC DYNAMICS OF HOMOGENEOUS MIDDLE-AGED SUBURBS

Technical report for Task 1 of the SACOG Economic Research Project

A report prepared for SACOG

By Martin P. Buchert and Arthur C. Nelson,
Metropolitan Research Center, University of Utah

Contents

List of Abbreviations

AF: Areal Fraction

CECI: Community Economic Condition Index. The index is hierarchically composed of three sub-indices, each of which is composed of several indicator variables.

GFN: Green-Field Neighbor

HMS: Homogeneous Middle-aged Suburb

IV: Indicator variable; one of thirteen component indicators which, aggregated together, constitute sub-indices of the CECI index.

SI: Sub-index. One of the three immediate sub-components of the CECI index; each is itself composed as the mean of several indicator variables.

SACOG: Sacramento Area Council of Governments

Introduction

SACOG, with its vital role in regional land use, transportation, and air quality planning, serves a region with remarkable needs and opportunities in these areas. As a booming metropolitan area with a growing regional population, land use and development issues ranging from affordable housing, urban sprawl, and community economic vitality are urgent matters for many of SACOG's local governments. And in an era of growing resolution to act on the threat of global climate change, the roughly two-million residents of SACOG's six county and 22 municipal member communities face significant challenges as they seek to coordinate planning and development activities to reduce greenhouse gas emissions and become a sustainable society. On all of these fronts, SACOG works with Federal and State government divisions while coordinating county and municipal government needs and actions and reaching out to the individuals who desire and deserve a voice in determining the shape and behavior of Greater Sacramento going forward.

SACOG has responded to California's SB 375 mandate to integrate transportation, land use, and air quality planning and has completed its Sustainable Communities Strategy. In its multivalent role as COG, MPO, and RTPA, SACOG is now turning to the vital work of *actualizing* the potential sustainability benefits envisioned in the SCS, by assisting its member communities as they *implement* the Strategy on the ground. By analyzing the factors associated with the economic dynamics of aging, troubled suburbs, SACOG can empower local communities who seek to craft policies and act to improve their citizen's well being. And by steering future business development to locations that accord with the SCS, SACOG can help revitalize struggling, aging suburbs while facilitating increased transit mode share for mobility, reducing traffic congestion, and decreasing rates of carbon emissions.

Findings

Against this backdrop, the University of Utah's Metropolitan Research Center was commissioned to investigate the economic dynamics of aging suburbs and to produce a longitudinal tract-level database of community economic indicators. Some key findings from this study are:

- The proportion of homogenous middle-aged suburbs in the SACOG region is nearly twice that of the nation as a whole (17.5% vs. 9.7%). Between 1980 and 2010, these suburbs have been more challenged (across 9 of 13 metrics) in the Sacramento area than equivalent suburbs in the rest of the country.
- Over the past 30 years, homogenous middle-aged suburbs in the Sacramento region have experienced higher home values coupled with much lower rates of home ownership than similar communities in the rest of the U.S. And while Sacramento's homogenous middle-aged suburbs have outpaced the nation's in fostering attainment of high-school diplomas among its workforce, they missed a decade of strong growth in a college-educated work force seen during over this time in other U.S. aging communities.
- Relative to the rest of the country, the entire Sacramento area has performed poorly on measures of households receiving public assistance, home ownership, rate of mortgage-

free homeowners, housing costs for mortgaged homeowners and for renters. The area has outperformed the nation on measures of household income, housing unit occupancy, home value, labor-force participation, and high-school diploma and college degree attainment.

- At the national scale, between 1980 and 2010 homogenous middle-aged suburbs consistently *out-performed* the rest of the nation on 8 out of 13 indicators measured by this study. However, this low observed performance for the rest of the country unquestionably masks economic diversity among aging suburbs that has already been commented on by other researchers.

- There is some evidence that sprawling growth in young, green-field neighbors of older suburbs has a negative impact over time on those older communities, but the signal is only weakly distinguishable given the limitations of this study, particularly the operational definition of sprawl necessary to generate a three-decade-long, tract-level, national database. More research is merited on this question, as any connection between *degree* of sprawl in neighbors and negative consequences to older existing communities is currently not well understood.

Background

Aging suburbs around the country are undergoing the decline that downtowns experienced in previous decades. Economic challenges to these communities stemming from aging infrastructure and housing stock are exacerbated by the migration of employment to exurban centers, and in combination these factors are leaving more inner-ring suburbs (IRS's) in worse shape (Lucy and Phillips 2000). Sacramento has some of these early-generation suburbs, and they are struggling. While recognizing the transformative opportunity that implementation of SACOG's Sustainable Communities Strategy represents, local and regional planners grapple with questions of how to help them. The University of Utah's Metropolitan Research Center was commissioned to investigate the economic dynamics of aging suburbs from their growth phase and through the past three decades, in the context of a) the remainder of the SACOG region, b) the rest of the U.S. as comparisons to help understand regional changes, and c) possible impacts related to the growth of newer suburbs.

City of Rancho Cordova Mayor David Sander, a key SACOG stakeholder, observed that older communities that have seen competing new development on their borders have fared poorer than those with no substantial competition on their borders. Accordingly we focused our research to determine if there is a relationship between economic stress in inner-ring suburbs and contiguous (new suburban) development in their younger neighbors.

Methods

Our basic approach was to develop a multi-dimensional Community Economic Condition Index (CECI) applicable to all areas, then to classify census tracts as Homogeneous Middle-aged Suburbs (HMS's), and finally to compare them to the remainder of the metropolitan region, both within SACOG and across the country.

We explain our methods in brief in this section; full details are given in Appendix 1.

In creating CECI, we developed a national database of 13 different variables, each available at the tract level for every decadal year from 1980 to 2010, and well recognized for their association with community economic condition. These indicator variables (IV's) were grouped into measures of households (HH), housing units (HU), and workforce (WF), spatially allocated to constant 2010 tract boundaries, adjusted for inflation to real (2011) dollars, and normalized to consistent 0-to-1 scales, where 0 represents the less and 1 represents the more desirable economic condition. Within each group, we took the mean across indicators to generate HH, HU, and WF sub-indices (SI's) and then computed CECI as the mean of the three SI's. Each census tract was scored with this method for each census (1980, 1990, 2000, 2010).

Extending from earlier research by Hanlon (2009), we defined HMS's as tracts where the current housing stock is dominated by units constructed during a single decade. We defined "domination" in this sense using a 40% threshold that we selected after sensitivity analysis and in consultation with key SACOG staff.

This study focused on tracts dominated by housing stock built during the '50's, '60's, and '70's, HMS's. These post-war decades represent the initial wave of auto-driven suburban expansion and were selected in consultation with key SACOG staff as characteristic of the aging Sacramento-region communities of concern. As Table 1 indicates, the tracts within SACOG's planning region are disproportionately rich in HMS's when compared with the nation as a whole. A partial explanation for the larger share of census tracts in the HMS category is that the Sacramento region is relatively young, especially compared to the eastern half of the country. See Figure 9 in the appendix for a count of tracts with pre-WWII housing stock. Given the nature and magnitude of the issues facing these communities, this constitutes a significant challenge for planners and stakeholders in the region.

Table 1 Count of HMS and non-HMS tracts, in SACOG and the nation

	Is HMS	Not HMS	HMS Fraction:
SAGOG	88	415	**17.5%**
USA	7134	66535	**9.7%**
SACOG Fraction:	1.22%	0.62%	

A second group of census tracts was defined for newer suburban areas, Green-Field Neighbors (GFNs). We defined as GFN's tracts where the mean age of the 2010 housing stock post-dates the HMS's dominant decade. It is noted that much of the study compares HMSs to non-HMSs. The Green-Field Neighbor tracts are just a subset of non-HMSs, which also includes

tracts that may be dominated by housing from older decades, rural areas, or may be composed of more mixed housing stock, making it conceptually less homogenous than the HMS group.

Implicit in Mayor Sander's informal hypothesis is that HMS's lose attractiveness in comparison with GFN's because of uniformly aging housing stock and other infrastructure as they age compared to newer communities freshly built. A regression analysis was conducted to test the hypothesis. We expect to see a time lag before GFN growth manifests a negative impact on HMS CECI, and would expect the largest impact to occur with the longest lag. The report's final section describes the analysis.

HMS's and GFN's

Figure 1 shows the distribution of HMS's and GFN's by decade in the Sacramento area.

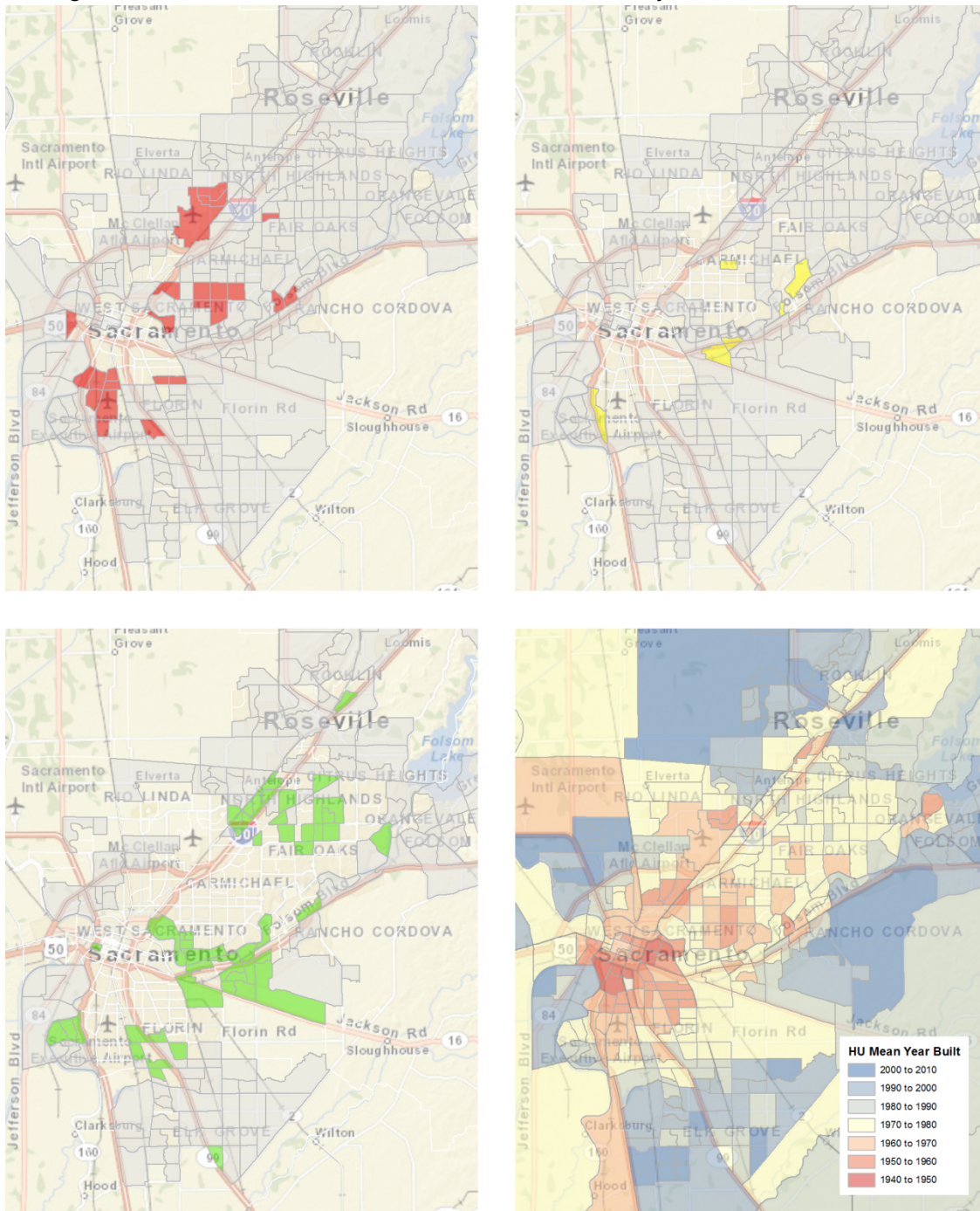**Figure 1 HMS's and GFN's in the Sacramento region. Top left: HMS's dominated by housing stock from the 50's (red). Top right: HMS's dominated by housing stock from the 60's (yellow). Bottom left: HMS's dominated by housing stock from the 70's (green). In each of these figures, GFN's for the respective decade are shown in gray. Bottom right: tracts colored by mean age of housing unit stock, regardless of dominance.**

Results

CECI

CECI and its component variables represent a significant informational resource, particularly because it allows us to compare tracts across the country and over three decades (1980 to 2010), in consistent measures. We anticipate that SACOG planners will utilize the CECI database in future work and will continue to discover planning applications for the data.

In this section we discuss some initial insights provided by this data set. We start with observations about CECI, sub-indices (SI), and indicator variables (IV) behavior generally, including the geographic distribution of CECI and its sub-indices in the Sacramento region. See Table 3 in the appendix for variable definitions. Next, we discuss the specific situation of HMS's vs. non-HMS's in the Sacramento region, a story that is more complicated than it may first appear. Finally, we examine differences and similarities between SACOG and the rest of the nation, and differences nationally between HMS tracts (dominated by housing built in the 1950s, 1960s, and 1970s) and non-HMS tracts.

National Trends

The national story that CECI tells is one of weakly improving economic conditions over the 30 year span from 1980 to 2010, subject to significantly divergent behavior between the various IV's. Overall slowing from 2000 to 2010 provides a clear signal of the Great Recession of 2008. Worsening conditions in almost all of the CECI HU components were a brake on CECI over the entire span, but strong growth in several CECI WF components (particularly high-school and college attainment) and moderate growth in CECI HH (driven by decreased dependence on public assistance) were enough to push CECI higher until the decade of the 2000's, when the index flat-lines. Figure 2 shows the dynamics of CECI, its sub-indices, and the 13 indicator variables, considering all tracts in the nation collectively.
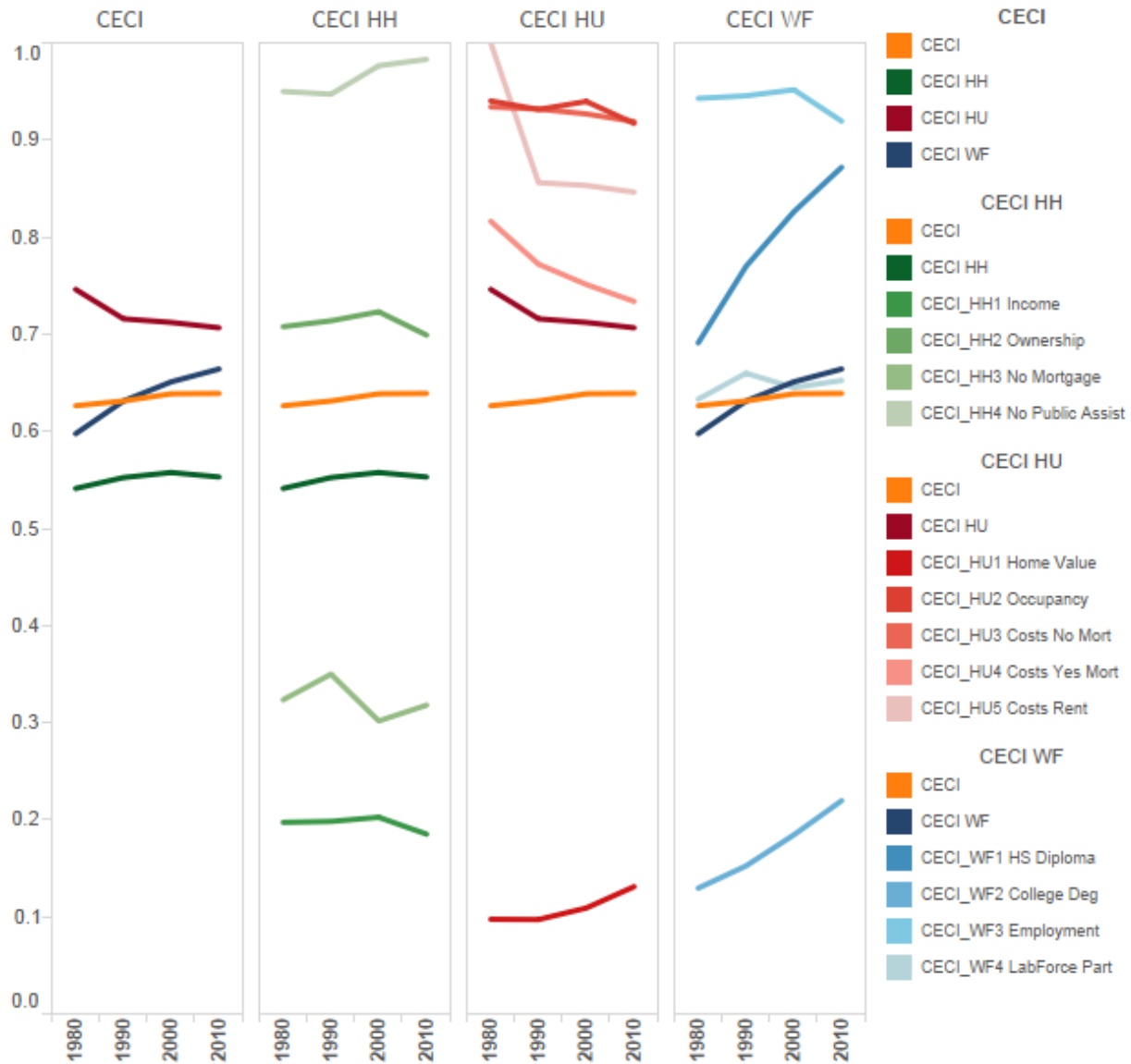
Figure 2 Median values of CECI, SI's, and IV's over time: all tracts nationally.

CECI HH is composed of four IV's, and three different patterns are noticeable. First, median household income and median home ownership rates both grew slowly over the '80's and '90's, and then fell distinctly between 2000 and 2010, although home ownership grew noticeably more strongly than did income. These IV's show the fingerprints of the mortgage lending tragedy of the early and mid-2000's, and the epidemic of job and home loss following the financial crisis of 2007/2008. Second, the public assistance IV reflects two major waves of structural change to public assistance in the nation. The first, noticeable in the jump from 1990 to 2000, was the passage of major Federal welfare reform in 1996. The other is seen in the continued rise of this IV from 2000 to 2010 in spite of falling household incomes during this period, and likely reflects the wave of cuts to Federal, State, and local assistance programs instituted in response to the Great Recession. Finally, the energetic movement of the mortgage-free homeowner IV is curious. The median fraction of homeowners who are mortgage-free increased over the '80's,

12

dropped sharply over the '90's, and rose again in the 2000's although not enough to prevent an overall net decline. The significant net decline over the '90's and 2000's could reflect the growth and popularity of equity lines of credit that flourished while rising home values and loose financing prevailed. The 2000's also coincide with the rising wave of Baby Boomers in their maximum purchasing power and the tail end of a roughly generational 30-year mortgage time frame, which may help explain the rise in this IV from 1990 to 2000.

CECI HU consists of five IV's: home value, housing unit occupancy rate, and three different measures of housing unit costs (for owners with and without mortgages, and for renters). The home value index increases across all three decades of the time series. The CECI decadal time series does not sample at a high-enough frequency to directly show the housing bubble that played out over this decade, but other indicators such as the S&P/Case-Shiller U.S. National Home Price Index corroborate that at the national scale, net growth over the decades of the 2000's was positive[1]. The home occupancy IV drops noticeably between 2000 and 2010, reflecting the surge in foreclosures following the sub-prime lending crisis of that decade. All three of the housing cost IV's decline over the span of the time series, but the precipitous drop in the rental cost IV from 1980 to 1990 is an artifact of limitations in the source Census data[2] due to data suppression in that census. This same artifact is evident when SACOG-region tracts and HMS's are considered separately. The artifact's impact is to *overstate* CECI HU and therefore CECI performance in 1980, and therefore to *understate* CECI growth between 1980 and 1990. Note that declines in these IV's reflect rising costs for housing, which is consistent with the rising home value IV.

CECI WF is made up of four IV's. Two of them reflect the educational attainment of the workforce (fraction of 25+ year olds with high-school diplomas and undergraduate college degrees), and these together are the largest contributors to the net gain in CECI nationally over the time series. The high-school diploma IV alone climbs more than 18%, lifting CECI by 1.5% between 1980 and 2010 on its own. This tremendous climb in the median fraction of 25+ year-olds with high-school diplomas and undergraduate degrees, while a positive development in itself, does not directly reveal other key education outcomes that have and will have an impact on U.S. economies from the national down to the neighborhood scale. For instance, while an increasing fraction of the nation's youth are graduating from high-school, national test scores show that U.S. twelfth-graders in 2013 are poorer readers on average than they were in 1992[3], and these indicators say nothing about burdensome student debt. The other two workforce IV's measure employment rate and labor force participation rate. The former has declined over the 1980-2010 span, with an abrupt fall as expected in the 2000's, after moderate growth during the '90's. The labor force participation IV is one of *only three* IV's at the national level that show appreciable acceleration from the decade of the 90's into the 2000's (the others are the no-mortgage and home value indicators). This uptick in workforce participation from 2000 to 2010 could be a consequence of retirees going back to work since the great recession.

A final note about the national picture is in order. Figure 1 gives an aggregate national picture by plotting the median value across all tracts in the country for each variable. At the coarsest level, irrespective of inter-decadal dynamics, the relative magnitude of the different IV's

---

[1] http://us.spindices.com/indices/real-estate/sp-case-shiller-us-national-home-price-index
[2] See Appendix 1: Methods for a discussion of data suppression issues in the 1980 Census.
[3] http://nationsreportcard.gov/reading_math_g12_2013/#/student-progress

is controlled by the highly skewed distribution of most variables. For highly right-skewed IV's (the majority of tracts with lower values and a few tracts with extremely high values) median values will predictably tend to be low (household income, home value, and fractions of mortgage-free homeowners and college graduates). For strongly left-skewed IV's (the majority of tracts have high values and a few have low values) median values will likewise tend to be high (fraction of public-assistance-free households, housing unit occupancy, housing costs for mortgage-free homeowners and for renters, employment, and high-school attainment). The remaining IV's (home ownership rate, housing costs for mortgage holders, and labor force participation rate) have distributions that are not as extremely skewed.

It is the balance of these very high and very low IV's that controls the relative magnitude of the SI's. For example, CECI HH, the household sub-index, is composed of two very right-skewed IV (median household income and fraction of homeowners who are mortgage free), and one very left-skewed IV (fraction of households who receive no public assistance). Consequently, this sub-index is generally low across all years. Figure 3 shows the distribution of CECI and its three sub-indices in the Sacramento area. These figures focus on the more densely urbanized area around Sacramento for the sake of clarity. The CECI HH sub-index is generally lighter shaded than the other two SI's because of the right-skewed variables described above.
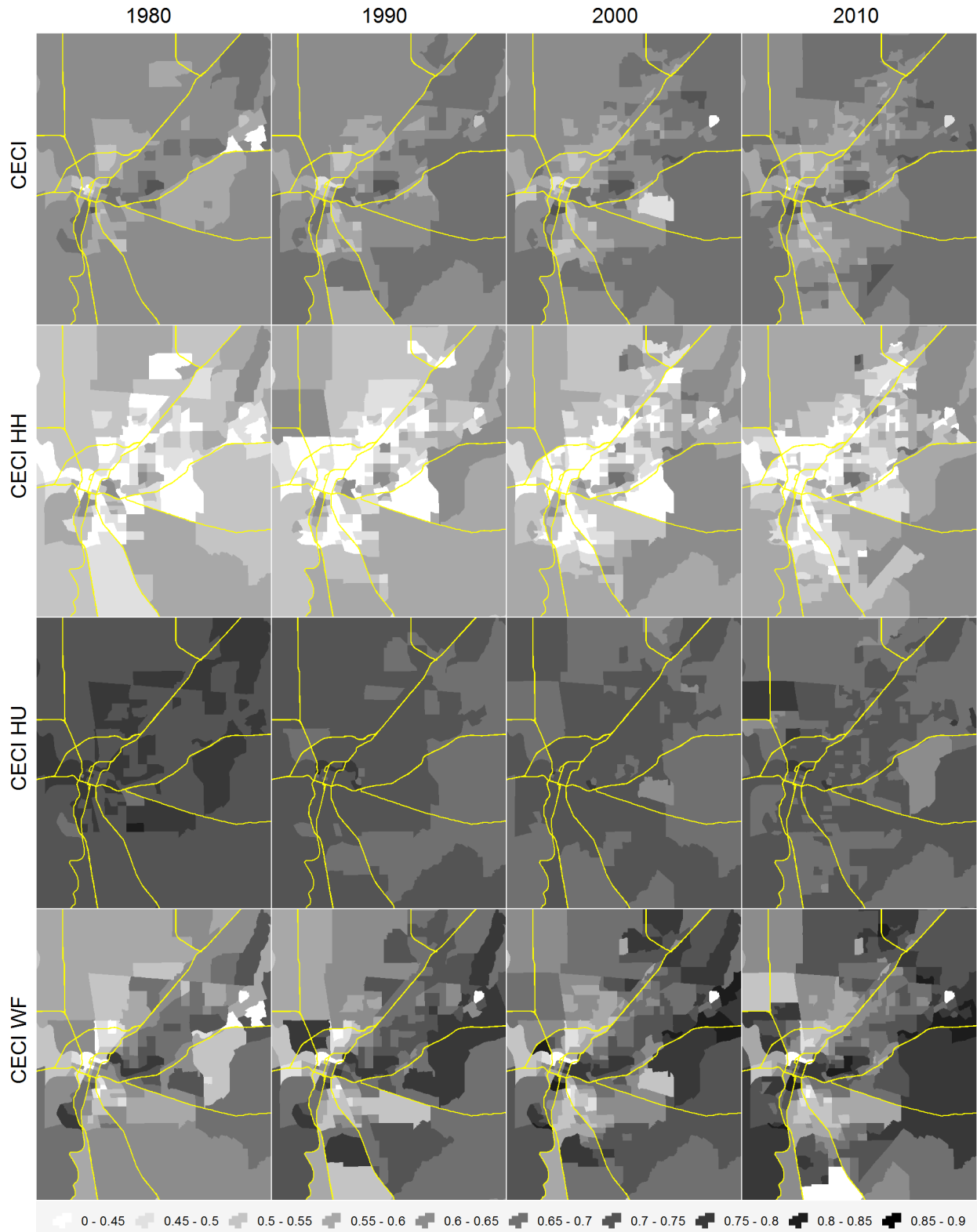
Figure 3 CECI and CECI SI's by decade for the urban SACOG region. Major regional highways in yellow.

HMS's vs. non-HMS's in the SACOG region

Of greatest interest to planners and stakeholders in the Sacramento area is how Sacramento-region HMS's perform relative to the rest of the area. This is shown in Figure 4 below.
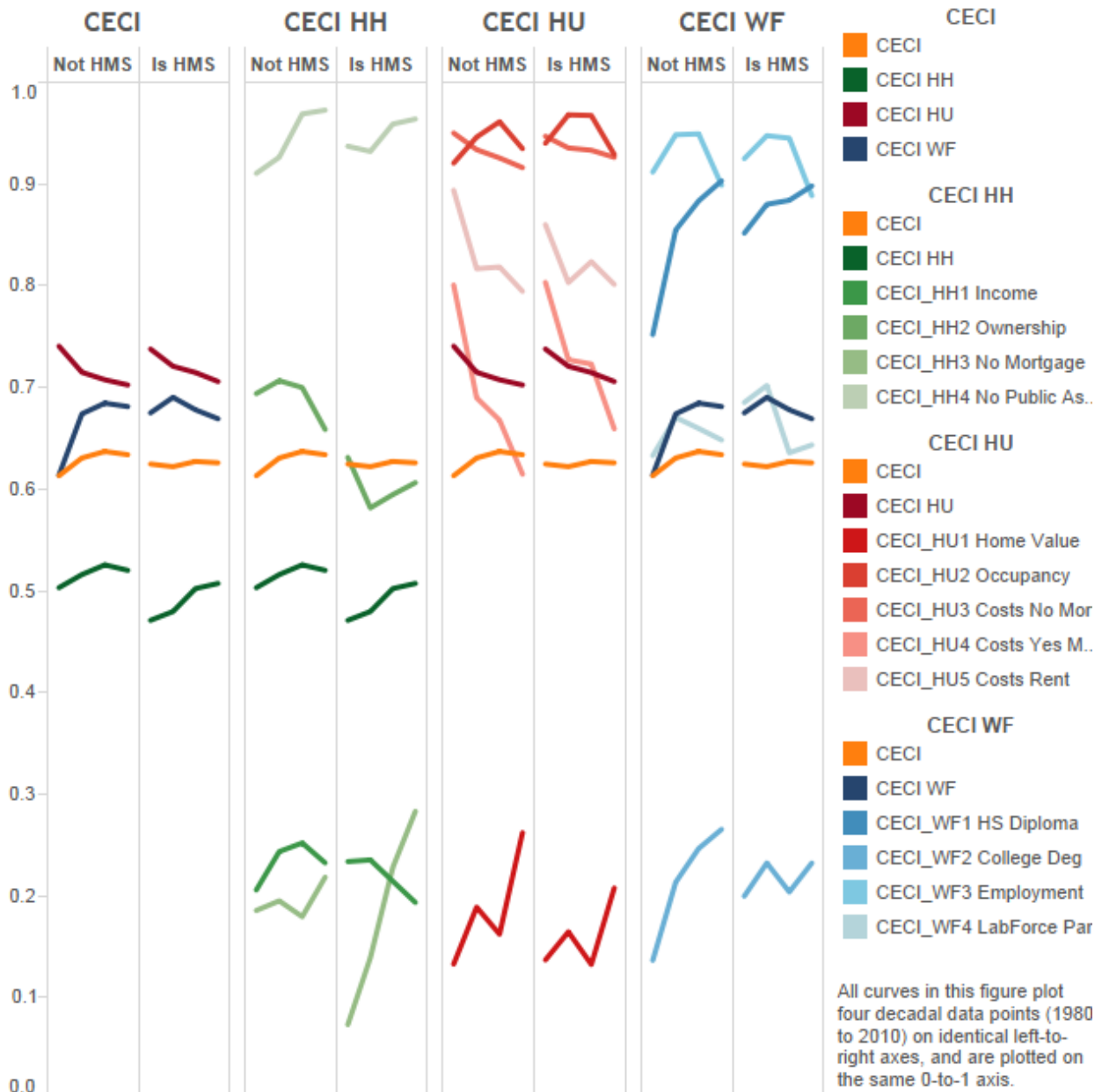
Figure 4 Median values of CECI, SI's, and IV's over time: SACOG region only, non-HMS vs. HMS

Here the picture is quite clear: the region's HMS's show definite signs of struggling relative to non-HMS's, in ways observed anecdotally by local planners. Starting with the CECI HH indicators, Sacramento's HMS's did not see as much of a bounce upward over the '90's in the public assistance IV and continue to perform more poorly than non-HMS's here. Home ownership is substantially lower in these communities (although as of 2010 on an increasing trend, in distinction to the rest of the region, where home ownership rates are falling). HMS

16

household income, while starting higher in 1980, didn't see the lift that the rest of the region did during the '80's and '90's and has fallen substantially over the 30 year period where non-HMS's experienced a net gain in spite of the losses of the 2000's. The dramatic exception to the pattern of stronger performance in non-HMS's than HMS's is in the rate of mortgage-freedom among homeowners, where Sacramento's HMS's have shot past the rest of the region in spite of starting from a baseline that was less than half as high as the non-HMS's, quadrupling this measure from .07 to .28.

Within the CECI HU indicators, occupancy has generally been higher in HMS's than non-HMS's, although in 2010 this pattern reversed itself, indicating that the sub-prime mortgage crisis and resulting wave of foreclosures during the Great Recession have hit Sacramento-area HMS's particularly hard. Costs for mortgage-free homeowners have been comparable (marginally lower) across all tracts over this period, but housing costs for renters and mortgage-holding homeowners, while matching the regions dramatic net increases between 1980 and 2010, remain lower, especially for mortgaged owners. Home values within HMS's have responded to the same regional signals in the 1980's as non-HMS's, but immediately in the '90's lost their initial slight lead over the rest of the region and have failed to capture as much of the region's dramatic gains here as non-HMS's.

The CECI WF indicators tell a mixed story of HMS performance. As with many of the indicators already discussed, Sacramento's HMS's began the '80's in a state of strength relative to the rest of the region. These older suburbs had strongly higher rates of high-school and college attainment, marginally higher employment, and noticeably higher workforce participation than the rest of the region. On both the educational attainment and the labor indicators, HMS's grew ahead of non-HMS's into 1990, albeit at significantly lower rates than their non-HMS counterparts. But HMS high-school attainment flattened in the '90's and college attainment lost almost all of its gains from the '80's, even while these indicators continued to climb in the rest of the region. And the flattening and fall that non-HMS's experienced in employment and labor-force participation respectively were stronger in HMS's, particularly in labor-force participation; however, HMS's reversed this decline in the 2000's and were climbing upward in 2010 where non-HMS's continued to slide.

National vs. SACOG

Next we examine the dynamics of CECI, its sub-indices, and the 13 IV's, considering SACOG-region tracts separately from the rest of the nation. In most years the rest of the country as a whole out-performs the SACOG region.[4] At the level of the individual IV's however, the situation is more mixed, as can be seen in Figure 5.

---

[4] The lone exception is 1990, when median CEHI for SACOG-region HMS tracts was barely higher than that of nation-wide HMS tracts. The difference for this subset of tracts in this year is largely driven by the dramatic gap in this year in the High-school diploma and College degree IV's.

Figure 5 Median values of CECI, SI's, and IV's over time; USA vs. SACOG.

CECI HH: Median IV values for SACOG are generally lower than but otherwise typically show similar dynamic behavior as the rest of the nation. SACOG's public assistance IV, while finishing lower than that of the country, started lower and rose further over the time series. SACOG home ownership fell much further over this span than it did nationally, largely because this indicator fell ~2.5% over the '80's while climbing 0.6% during the same period nationally. This is likely a reflection of California's significantly higher home prices and monthly housing costs, relative to the rest of the country. However, household income was generally stronger in SACOG than the nation as a whole, even managing a net gain from 1980 to 2010 in spite of the widespread losses of the 2000's; in contrast, the nation started lower and fell over the time series. This is the flip-side of the Golden State's general high cost of living. The strongest

difference between SACOG and the nation among these IV's is in the fraction of mortgage-free homeowners. Both appear to have responded to the same (presumably national) signals by decelerating during the '90's and accelerating again in the 2000's, but while this IV started in 1980 nearly twice as high in the nation as in SACOG, the rest of the country saw a marginal net loss during the time that SACOG climbed ~7.5%. One potential explanation for this growth in SACOG tracts is California's Proposition 13, which incentivizes homeowners to stay in homes they have lived in for many years.

CECI HU: SACOG performs better than the nation, across all years of our time series, on three of the five housing unit IV's: occupancy, monthly housing costs for mortgage-free homeowners, and most dramatically, home values (which posted an even more dramatic gain through the decade of the 2000's housing crash than did the nation). SACOG's renters and mortgage-holders however have had a harder time than their counterparts in the rest of the nation; these IV's are substantially lower (indicating higher costs) within the SACOG region than in the rest of the country, and the precipitous decline in the mortgage-holders housing cost IV in particular is the largest 1980-2010 drop (17.6%) made by any IV. The Sacramento area's heated housing market has taken a toll on mortgage-holders, placing them under a dramatically heavier burden than is the norm for the nation.

CECI WF: collectively, the employment-related IV's show that the Great Recession hit the Sacramento region harder it did the nation as a whole. Encouragingly, the region outperforms the nation as a whole in the fraction of 25+ year-olds attaining both high-school diplomas and college degrees.

HMS's vs. non-HMS's Nationally

Figure 6 shows how HMS's performed over time compared with non-HMS's. Interestingly, our analysis shows that when pooled nationally, HMS tracts as defined in this study—i.e. dominated by housing built during the 1950's, 1960's, and 1970's—have fared *better* than non-HMS tracts on 9 of the 13 CECI IV's. The IV's that consistently performed better in HMS's than in non-HMS's are: public assistance, home ownership, and household income (3 of 4 CECI HH IV's); occupancy and home value (2 of 5 CECI HU IV's, including none of the housing cost indicators); employment[5], labor force participation, and high-school diploma and college degree attainment (4 of 4 CECI WF IV's). As well as in their absolute performance HMS's also typically exhibit similar inter-decadal dynamics as non-HMS's.

---

[5] Nationally, the median value of the employment IV for HMS's is greater than that for non-HMS's, for all years except 2010.

Figure 6 Median values of CECI, SI's, and IV's over time: non-HMS nationally vs. HMS nationally.

The exceptions to these nation-wide patterns are the housing cost IV's, all of which are lower across all years for HMS's than for non-HMS's, and the mortgage-free homeowners IV. This last was dramatically lower in HMS's than non-HMS's in 1980, but nearly doubled its median value across HMS tracts over the following 30 years (from .15 to .29) while the median value for non-HMS tracts declined over the same period (from .34 to .32).

It is critical to note that this nation-wide observation does not always hold for SACOG's HMS's.[6] For instance, for non-HMS tracts the fraction of homeowners who are mortgage-free has consistently been dramatically lower (.13 - .15) in the SACOG region than in the rest of the

---

[6] Appendix 2 contains an combined version of Figures 5 and 6, showing cross-tabulations of the USA/SACOG and non-HMS/HMS categories.

nation (.30 - .36). The story is different for HMS tracts, where home ownership itself in the SACOG-region lags the nation by about .13 (.58 - .63 in the SACOG regions vs. .71 - .76 nationally). Although further analysis was outside the scope of this project, it is likely that other regional variation nuances this narrative further; see our further comments in the discussion section.

Limitations of a HMS's relative to non-HMS's paradigm

Both the definition of HMS and the HMS/non-HMS used in this research were tailored for the Sacramento area and our particular research question. They are effective at identifying the kinds of distressed suburbs of concern within the region, but as is evident in the national comparison of HMS vs. non-HMS index behavior, it appears that they fail to explain variability among HMS's in other parts of the country, as well as under-specifying important distinctions between non-HMS tracts. Previous research has highlighted a trend of economic polarization among old and middle-aged suburbs, with proportionate growth in the numbers of both high- and low-income suburbs of these ages (Lucy & Phillips, 2006; p.142). Also, "non-HMS" in the context of this study means tracts composed of everything from new tract housing to distressed inner city; the broad variability non-HMS tracts by our definition likely explains the lower performance of non-HMS tracts nationally as observed here, and should be a high priority for further analysis outside of the Sacramento region.

Relationship between growth in GFN's and CECI in HMS's

The hypothesis in this analysis is that growth in Green-Field Neighbors (GFNs) in later time periods can contribute to (negative) changes in CECI as HMSs age. The GFN growth metric is the change in housing units in a tract. The regression model used dwelling units per square mile as the independent variable, since the tract size does not change. A summary of pertinent results from the regression models is reported in Table 2. When not controlling for region (i.e. regional membership dummy covariates) we found no significant correlation between growth in GFNs and change in HMS CECI. Once regional effects are controlled, we did find correlation between these two key variables, but the strength and significance of the correlation was mixed and, importantly, was mediated by the degree of lag. Models explained more of the variation (model R- squared) in CECI for older HMS's than for younger ones, and ANOVA Regression F statistics show stronger support for these correlations in the oldest HMS's (dominated by housing stock built during the 1950's) as well. The model coefficients for the sprawl variable show mixed support for the hypothesis. The sprawl coefficient is only both in the expected sign and significant at the p>.01 level in one instance, for the model with the strongest contrast (HMS's from the '50's) and at the longest lag (sprawl in the '70's and CECI in the 2000's). In general then, we found the strongest association in 1950's-dominated HMS's, and at the longest lags.

Table 2: Regression Model Results. Statistically significant coefficients with anticipated sign are shown in italics.

HMS	Sprawl decade	CECI decade	lag (years)	Model R-squared	ANOVA Regression F	Sprawl variable model coefficient	Sprawl variable Sig
1950's	1970's		30	0.217	80.756	*-0.060*	*0.001*
	1980's		20	0.214	79.344	0.014	0.419
	1990's		10	0.214	79.479	0.022	0.205
1960's	1970's		30	0.168	29.926	-0.021	0.434
	1980's	2000's	20	0.168	29.915	-0.019	0.468
	1990's		10	0.171	30.398	0.052	0.042
1970's	1970's		30	0.088	31.176	0.034	0.073
	1980's		20	0.090	31.823	0.053	0.004
	1990's		10	0.087	30.891	0.017	0.351

Impact of GFN sprawl relative to other CECI internal factors

Because the CECI IV's are endogenous to CECI and not independent predictors, it is specious to directly test their strength relative to the strength of the green-field growth-impact we modeled. However, the effect size noted in the significant model case suggests that any impact is minor relative to the sorts of variability we observe in the CECI IV's. It is impossible from this analysis to draw conclusions about the spatial distribution of this effect across the SACOG region.

Conclusions and recommendations

In the course of this project, we developed a national, tract-scale 30-year decadal time series of 13 community economic condition indicators, and synthesized into Housing, Households, and Workforce sub-indices and an overall Community Economic Condition Index. We also developed an analogous tract-level longitudinal database of the spatiotemporal progress of housing development. These datasets should be highly useful to SACOG in regional planning work to support economic development.

Strength of CECI factors

There is significant variability in the strength of the various IV's contributions to sub-index and overall CECI dynamics, both across years and between SACOG HMS's/non-HMS's. A full breakout is given in Appendix 3 but the key drivers are also visually evident in Figure 6 above. We give a summary for SACOG's HMS's and non-HMS's by decade here. Entries are rank ordered by absolute magnitude of their contribution to CECI and colored blue where change was positive and red where change was negative. The reader should remember that declines in housing cost IV's indicate rising, not falling, costs.

SACOG HMS's:
- 1980's: mortgage-free homeowners, mortgaged-owner housing costs, home ownership, renter housing cost.
- 1990's: mortgage-free homeowners, labor force participation, college degree attainment, home value.
- 2000's: home value, employment, mortgage-free homeowners, mortgaged-owner housing costs

SACOG non-HMS's:
- 1980's: high-school diploma attainment, mortgaged-owner housing costs, college degree attainment, renter housing cost.
- 1990's: public assistance, college degree attainment, high-school diploma attainment, home value.
- 2000's: home value, employment, mortgaged-owner housing costs, home ownership.

While this research did not directly investigate the nature or strength of factors driving these CECI IV's, we have made observations about likely drivers in the results section above. We recommend that local planners and policy makers focus on these variables, as this analysis indicates that they are the key drivers of dynamics. In particular, many of the key drivers seem related in a complex around affordability of housing. SACOG planners should be able to use the longitudinal index of housing affordability that was produced in conjunction with this study to guide planning efforts in this vein.

GFN growth vs. HMS CECI: major or minor effect?

This research sought to quantify the relationship between the growth rates in neighboring green-field communities and the economic trajectories of homogeneous middle-aged suburbs. We find a weak association between these two factors, after controlling for regional differences around the country. While the association is somewhat weak in absolute terms (r^2 values ≈ 0.2; model coefficient = -0.06) it is notable given the complex system of social and economic factors that influence the economic state of a community.

Limits of the HMS paradigm in the study

Given the points we make above about regional and other unexamined variables differentiating HMS's from one another, we recommend against drawing further conclusions about HMS's outside of the Sacramento region without examining the influence of other factors aside from age homogeneity of housing stock.

While earlier work has suggested that more recent development competes with and degrades the economic fortunes of older post-war suburbs, we are aware of no previous research suggesting a further connection between the *degree* of neighboring urban growth rates and declining economic fortunes of older suburbs. However, this work detects only statistical association and as such suggests possible causation but does not demonstrate it. What is it about the green-field competitors itself, independent of the simple fact of newer development, which hurts the older suburb? This question is left for further research designed to identify causal mechanisms by which accentuated sprawl in neighboring development damages the economic condition of older suburbs.

Literature Cited

Bruegmann, R. (2006). Sprawl : a compact history. Chicago, IL: University of Chicago Press.

Burchell, R. W., Shad, N. A., Listokin, D., Phillips, H., Downs, A., Seskin, S., Davis, J., Moore, T., Helton, D., Gall, M. (1998). TCRP Report 39 The Costs of Sprawl — Revisited. Washington, D.C.

Ewing, R., Pendall, R., & Chen, D. (2003). Measuring Sprawl and Its Transportation Impacts. Transportation Research Record, 1831(1), 175–183.

Ewing, R., & Hamidi, S. (2014). Measuring urban sprawl and validating sprawl measures (p. 189).

Fotheringham, A. Stewart, and David WS Wong. "The modifiable areal unit problem in multivariate statistical analysis." Environment and planning A 23.7 (1991): 1025-1044.

Galster, G., Hanson, R., Ratcliffe, M. R., Wolman, H., Coleman, S., & Freihage, J. (2001). Wrestling Sprawl to the Ground: Defining and measuring an elusive concept. Housing Policy Debate, 12(4), 681–717. doi:10.1080/10511482.2001.9521426

Hanlon, B. (2009). Once the American Dream (p. 224). Temple University Press.

Johnson, M. P. (2001). Environmental impacts of urban sprawl: a survey of the literature and proposed research agenda. Environment and Planning A, 33(4), 717–735. doi:10.1068/a3327

Johnstone, I. M. (2001). Energy and mass flows of housing: estimating mortality. Building and Environment, 36(1), 43–51. doi:10.1016/S0360-1323(99)00066-9

Kwan, M. (2012). The Uncertain Geographic Context Problem. Annals of the Association of American Geographers, 102(5), 958–968. Retrieved from http://www.kgeography.or.kr/admin/file_data/20130319-1.pdf

Lucy, W. H., & Phillips, D. L. (2000). Confronting Suburban Decline (p. 321). Washington, D.C.: Island Press.

Lucy, W. H., & Phillips, D. L. (2006). Tomorrow's Cities, Tomorrow's Suburbs (p. 354). Chicago: American Planning Association.

Minnesota Population Center. (2011). National Historical Geographic Information System: Version 2.0. Retrieved from www.nhgis.org

Minnesota Population Center. (2014). Data Documentation: GIS Boundary Files. Retrieved June 05, 2014, from https://www.nhgis.org/documentation/gis-data

Siedentop, S., & Fina, S. (2012). Who sprawls most? Exploring the patterns of urban growth across 26 European countries. Environment and Planning A, 44(11), 2765–2784. doi:10.1068/a4580

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2009). Thematic Cartography and Geographic Visualization (3rd Edition., p. 591). Upper Saddle River, N.J.: Pearson Prentice Hall. Retrieved from http://books.google.com/books?id=P_URAQAAIAAJ

Sobel, J. (1991). Principal components of the Census Bureau's TIGER File. In D. J. Peuquet & D. F. Marble (Eds.), Introductory Readings In Geographic Information Systems (1st Ed., p. 371). CRC Press.

US Census Bureau. (2012). 2010 Geographic Terms and Concepts - Census Tract. Retrieved June 05, 2014, from http://www.census.gov/geo/reference/gtc/gtc_ct.html

Wombold, L. (2007). Changes and Challenges: Understanding American Community Survey Data. ArcUser. Retrieved from http://www.esri.com/news/arcuser/1207/census.html

Wu, S., Qiu, X., & Wang, L. (2005). Population Estimation Methods in GIS and Remote Sensing : A Review, (1), 80–96.

Appendix 1: Methods

Development of research question

The specific research question targeted by this project changed significantly part-way into this project. The initial focus was to use regression modelling and investigate correlates of annual changes in community economic condition. While the research design was correlative and would not test explicitly for causation, such studies can suggest reasons for differences in economic outcomes; these in turn would suggest possible policy levers for local decision makers. We designed this research using Census Places as units of analysis and proxies for "communities", and began building an annual longitudinal index of community economic condition.

An initial review of the place-based analysis identified concerns on the part of SACOG personnel, most significantly that the spatial grain of places was too coarse for us to observe the sort of dynamics observed in the SACOG region, thus limiting the usefulness to regional planning application. Further, in the interim SACOG's research interest had intensified on the specific question of possible interactions between recent green-fields (newer suburbs) growth and the economic fortunes of older HMS's. For instance, City of Rancho Cordova Mayor David Sander, a key SACOG stakeholder, observed that older communities that have seen new development on their borders have fared poorer than those with limited competition on their borders.

This led to SACOG refocusing the research question. Accordingly we shifted our research: specifically, we refocused the examined predictors of Community Economic Health Index (CECI) dynamics from the broad suite of predictors originally proposed to the specific predictor of neighboring green-field growth.

Data Sources

Our rescoped research question required that all data be available (1) at the census tract level, (2) for all of the contiguous 48 states, and (3) for four successive decadal Census years (1980 - 2010; referred to hereafter as "epochs"). The data sources listed in Table 2 were initially proposed for inclusion in our analysis of community economic condition but were eventually excluded from the analysis for not meeting the three data availability criteria. Note that while these data sets vary in their spatial grain (i.e. unit of analysis) and spatial extent (i.e. total geographic footprint of data coverage), none are available at decadal frequency and going back to 1980. This key limitation to available data sources eventually constrained us to variables available through the Census (and more recently, the American Community Survey).

Variable	Data Source	Tract level	48 States	All epochs
Household income	MRC's modified Infogroup Household database	Y	Y	-
Household combined Housing and Transportation ("H+T") costs	Center for Neighborhood Technology (CNT)	Y	-	-
Total payroll	Census County Business Pattern	-	Y	-
Wage-housing balance	Previous MRC research	Y	Y	-
Sectoral diversity of employment base	Census LED	Y	-	-
Employment by 2- or 3-digit NAICS sector	Census LED; MRC's modified Infogroup Business database	Y	-	-
Employment wages by NAICS sector	Census LED	Y	-	-
Real estate market changes	Zillow, REIS	-	-	-
Sales tax and other local government revenue	Census of Local Government Finance	-	Y	-
Urban Sprawl metric	Census LED	Y	Y	-

Defining Community Economic Condition

Ultimately, our research examines the economic condition of communities and seeks to determine if there is a relationship between this and the degree of sprawling development in the community's neighborhood. Central to this question is the definition of community economic condition. For this purpose, we created the Community Economic Condition Index (CECI).

We developed CECI with several principles in mind:

- To answer our longitudinal research question, all variables used in the analysis need to be available at the tract level for all four decennial census epochs (1980 - 2010). This allowed us to calculate measures of trend as well as create snapshot values. We ignored measures of volatility given the relatively sparse (ten-year) frequency of the time series[7].

- Because both household income and expenses vary widely across the country, CECI should measure wealth-generating factors (e.g. income) relative to wealth-consuming factors (e.g. housing costs); this would allow for comparisons between, for instance, high-income but economically distressed communities in large metro areas and low-cost, healthy communities in smaller metros. Wealth-positive factors should be positively correlated with CECI, while wealth-negative factors should be negatively correlated with

---

[7] In simple terms, sampling theory indicates that sampling frequency must be at least twice the signal frequency in order to reconstruct the signal from the sample. In this case, this means that the only signals of actual volatility that could be reconstructed from variability in our time series would be on time scales of >= 20 year cycles. See (Weisstein, 2014)

CECI.
- Since a full cost accounting of major household income and expense streams is unavailable for the temporal scope of our research, CECI should include other variables with an accepted theoretical connection to community economic circumstances.[8]
- Since CECI is a dependent aggregation of a variety of indicator variables (IV's), it is not logically sound to analyze these measures as predictors per se of CECI.

Following the principles outlined above, we evaluated 13 indicator variables of community economic condition, grouped in three categories (measures of households, housing, and workforce) that we used as sub-indices. These indicator variables are described in Table 3.

---

[8] In its original scope, this work proposed to develop a positive (i.e. descriptive) index of community economic condition *as well as* a database of additional possible explanatory correlates, and then to perform regression analysis using these data to investigate multiple possible drivers of community economic condition.

Table 3 Indicator variables (IV's) in CECI

Sub-Index	Variable	Notes
Households	Income	Median household income, in real 2011 dollars
	Ownership	Percentage of households who own the housing unit they occupy
	No Mortgage	Percentage of home-owning households who are mortgage-free
	No Public Assistance	Percentage of households who do NOT receive public assistance
Housing Units	Home Value	Median single-family home value, in real 2011 dollars
	Occupancy	Percentage of housing units that are occupied
	Housing Costs, No Mortgage	Median monthly housing costs for home-owning households who are mortgage-free, in real 2011 dollars
	Housing Costs, Yes Mortgage	Median monthly housing costs for home-owning households who have a mortgage, in real 2011 dollars
	Housing Costs, Rent	Median gross monthly rent for renting households, in real 2011 dollars
Workforce	High-school Diploma	Percentage of 25+ year-old adults with a high-school diploma or higher
	College Degree	Percentage of 25+ year-old adults with a 4-year college degree or higher
	Employment	Percentage of the civilian labor force who are employed
	Labor-force participation	Percentage of the 16+ year-old population who are either employed or who are actively seeking employment

Data were downloaded from the National Historical Geographic Information System (NHGIS) portal (Minnesota Population Center, 2011). Data for years 1980, 1990, and 2000 were from both the short (population counts) and long (other variables) forms used by the U.S. Census Bureau for those censuses. Analogous data for the 2010 epoch were obtained through NHGIS from the 2010 decadal census (population) and 2007 - 2011 five-year American Community Survey (ACS). ACS five-year data characteristically represents a period estimate rather than the single-day point estimate given by the census (Wombold, 2007). We accept this as an unavoidable limitation of working with thematically specialized census data in the present day; there are no alternative sources for these data that do not require assembling and processing data from fragmented, raw sources.

CECI was computed for all 2010 tracts in the country, subject to data availability as detailed above. In producing CECI, we reallocated all measures from their respective epoch tract boundaries to 2010 Census tract geographies, as described previously. For all variables denominated in dollars, which included housing cost variables, home value, and household income, we adjusted for inflation using the Bureau of Labor Statistics Consumer Price Index

(CPI) inflation calculator to 2011 (the year for which the 2010-epoch 5-year ACS data for 2006-2011 were denominated).

We then normalized each IV to a 0-to-1 scale (IV's such as occupancy or tenure were natively on this scale, being calculated as fractions of a tract-wide total), and reversed the raw order of the variable where necessary so that 1 consistently represents the preferred economic circumstance, all else being equal. Where the raw data did not have a natural limit of 1 we normalized using the maximum value observed for that IV. We normalized each dollar-denominated IV using the maximum value observed across all decades so that inter-decade normalized values are truly comparable. For three IV's measuring housing costs, we normalized all three measures using the same single maximum observed value, rendering these normalized values comparable with one another as well as across years.

For each tract, we computed the mean within each group of indicators to generate HH, HU, and WF sub-indices (SI's) and then computed the Community Economic Condition Index as the mean of the three SI's.

Reallocating historic data to 2010 tract boundaries

Our study uses time-series census data reported at the tract level to analyze community economic and development dynamics over time. However, census tract boundaries are not constant over time. The Census Bureau endeavors to make tract boundaries as stable as possible[9] but subdivides or merges tracts over time as necessary to maintain tract populations in the range of 1200 to 8000 people, with a preferred size of about 4000 people. Nonetheless, boundary changes are expected over time. This conceptual fluidity of tract boundaries is compounded in practice by the Census Bureau's evolving technological capacity, reflected in geodata accuracy improvements over the 40+ years since the 1980 census. The ironic upshot is that while the Census Bureau's pioneering national data set (TIGER/Line: Topologically Integrated Geographic Encoding and Referencing; Sobel, 1991) has had topologically integrated (i.e. clean, non-overlapping) tract boundaries within a given census epoch for decades, tracts (and other units) are not as yet topologically integrated between census epochs. These boundary changes required that we reallocate reported values to consistent geographies so that any longitudinal analysis compares truly commensurate entities.

We performed a spatial reallocation of values based on the assumption that the quantity reported for any variable for a given tract was uniformly distributed over the area within the tract boundaries. Given this assumption, stock variables are proportionate to the area of the spatial reporting unit and a portion of the source tract that encompasses 50% of the original tract area, for example, would represent 50% of the stock represented by the variable. This simplifying assumption of internal uniform distribution is demonstrably false for many measures of human geography, given that human use of space is non-uniformly distributed in space at every scale.[10] We retained the assumption because other, more accurate methods of reallocation such as dasymetric mapping were beyond the scope of our national, four-decade research design. We discuss likely implications for our research results in a subsequent section.

---

[9] It uses a strategy to follow "visible and identifiable features" over time. (Census Bureau, 2012.)

[10] There is a rich literature in Geography on methods estimating the intra-unit distribution of a variable, typically referencing remotely sensed or other ancillary proxies of population density (see Wu et al, 2005 and Slocum et al 2009 Ch. 15 for reviews).

We obtained tract boundary geometry data for census years 1980, 1990, 2000, and 2010 from the National Historical GIS in ESRI shapefile format (NHGIS, 2014). Upon importing these data into a geodatabase, we computed the geometric union of each earlier decade's tract polygons with the 2010 tract polygons. The output in each case was a new polygon feature class of "daughter" polygon features (we referred to these as "fragments"), each recording both the unique identifier (UID) and the original area of both its "parent" feature tracts.

Using the areas of the daughter fragment and its earlier census-year parent, we calculated the areal fraction (AF) of the pre-2010 tract retained in the daughter fragment. We then reallocated stock or count variables for pre-2010 epochs as follows: we joined the source data table to the fragments attribute table, using the UID of the earlier-epoch tracts as the secondary and primary keys, respectively. Next we weighted the variable of interest by the areal fraction of the pre-2010 tract. We then aggregated the table to 2010 tract UID's, summing values of all AF-weighted fragments.

Some variables do not represent stocks themselves but are conceptually rates. For example, inasmuch as the median is representative of a distribution, median single family home value can be understood as a representative dollar amount *per house* in the tract. All such variables were transformed for each fragment by de-normalizing (e.g. multiplying median home value by count of owner-occupied single family homes) prior to retabulating, then were retabulated to 2010 tract UID's, and finally were renormalized by the new reallocated denominator value.

Exhaustive partitioning of the national territory into census tracts was not completed until the 1990 census, although since this tracting effort began with urban areas the 1980 tracts cover the urbanized and high-population-density regions of the country. Year 2010 tracts with data for the 1980 epoch account for 90.4% of the 2010 national population (see Figure 7). The index production effort resulted in a longitudinal, four-decade time series of CECI for 61,330 Census 2010 tracts (from a 2010 nationwide total of 74,003 tracts). We note that CECI and its sub-indices and constituent indicator variables have considerable planning and research value of their own, independent of their direct instrumental utility in this project.
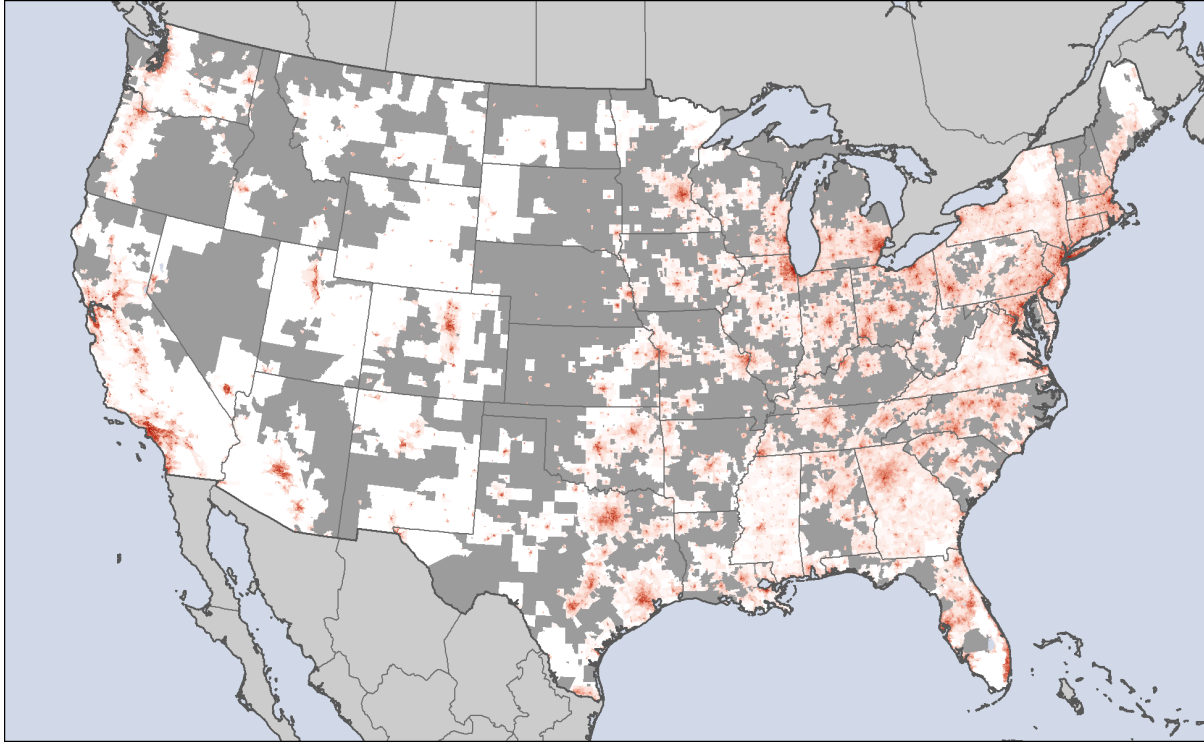
Figure 4: Population Density and 1980 Tract coverage. Population density is on a light-to-dark red scale; dark gray areas are 2010 tracts with no corresponding tract coverage during the 1980 Census.

Bad Data

A specific challenge to our longitudinal, spatially reallocated database was the issue of bad and missing data. In collating and republishing historical census data, the NHGIS does not include original Census bad data flag codes, leaving the data analyst with the job of judging (potentially via reference to multiple other source Census tables) whether records with a value of zero represent actual zeroes, or rather represent tracts with suppressed data.[11]

From the raw data, we flagged as having suppressed data those tracts where both the total count of HH's was reported as greater than zero *and* any of the following conditions were true:

- Median HH income = 0
- # owning = 0 AND # renting = 0
- # owning >0 AND #owners with mortgage =0 AND #owners with no mortgage = 0
- # receiving public assistance = 0 AND # receiving no public assistance = 0
- # occupied = 0 AND # unoccupied = 0
- Median home val = 0
- # renting = 0 AND median monthly rental expenses = 0
- # owning >0 AND median monthly owner housing expenses = 0 (for both mortgage-holders AND non-mortgage-holders)
- # of working age > 0 AND #HSG = 0 AND #CollegeGrad = 0
- #employed = 0 AND #unemployed = 0

---

[11] The Census Bureau suppressed records in the 1980 census when the number of enumerated units within a tract was low enough that a tabulated count of a sensitive value would be too revealing.

- # in labor force = 0 AND # not in labor force = 0

For each epoch, here is the total number of tracts flagged. Note that a tract was flagged on the basis of *any* of the above criteria, but would not necessarily be suppressed in all of those categories:

- 1980: 29,674 out of 46,728 tracts (63.5%)
- 1990: 1,709 out of 61,258 tracts (2.8%)
- 2000: 1,246 out of 65,443 tracts (1.9%)
- 2010: 4,200 out of 74,002 tracts (5.7%)

Of the 1980 tracts flagged, it is likely that the majority represent actual data suppressions by the Census Bureau. As these tracts are definitionally low-population tracts they would tend to be away from populated centers focused on in our research. However, there is the chance that they could be on the margins of growth and represent locations that were "grown into" in following epochs.

For the 1990 and 2000 epochs, the Census Bureau's suppression strategy was replaced with more effective techniques that eliminated the large-scale voids of the 1980 data. For these years flagged tracts likely represent unusual circumstances that were not anticipated by our error-flagging logic but which could not be further resolved without reference to ancillary data sources. We feel that the small volume of remaining flagged errors represents an acceptable level of effort for these years.

We removed from our analysis all 2010 tracts that:

- were untracted during the 1980 epoch (i.e. had no data across the board for the 1980 epoch)
- had counts of zero total, owner, and renter households
- had no match between boundary files and indicator variables.

Defining Homogeneous Middle-aged Suburbs

We operationalized our definition of Homogeneous Middle-aged Suburbs (HMS's) starting with the definition of Inner-Ring Suburbs (IRS's) from Hanlon's "Once the American Dream" (2009). Hanlon's work examines aging "first tier" suburbs - communities built during the first wave of auto-driven suburban growth (roughly 1945 - 1970) that occurred during the post-WWII-boom. For purposes of our research, we call them "inner-ring suburbs" (or IRS's). During this period new development was characterized by growth outward from historic city centers, especially in older cities with dense urban cores. Hanlon defines IRS's as census tracts where at least 50% of the housing stock was built prior to 1970, which are contiguous with a central city (or are in a contiguous block with another tract that meets the definition of IRS). This definition includes both a compositional and a locational element (see Figure 8).

Using counts of Housing Units (HU's) by decade from the 2007-2011 5-year ACS, we mapped tracts that meet this definition. However, SACOG staff concluded, from visual inspection, that the results failed to highlight the communities of concern in the SACOG planning area where the aging and troubled suburbs are not the result of spatially contiguous growth patterns (see Figure 1). Consequently, we dropped the locational selection criterion.
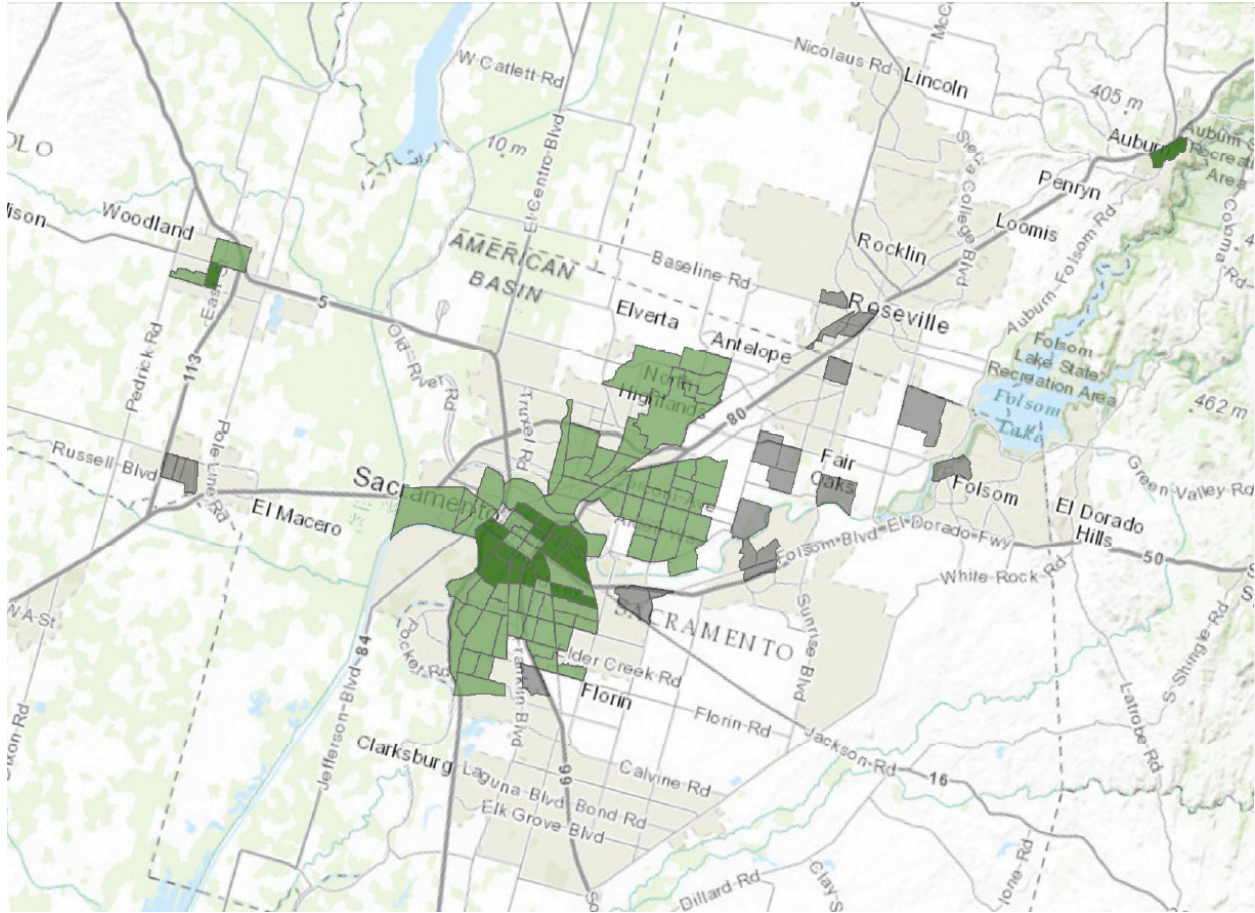
Figure 5: IRS classification as per Hanlon 2009. Dark green tracts are "central cities"; light green tracts are "inner ring suburbs" (>50% of housing stock predates 1970 AND are in a contiguous block with a central city); gray tracts meet the compositional but not the locational criterion for IRS's.

Because the research question focuses on looking explicitly at relationships between variables over time, we concluded that the simple pre/post-1970 housing composition implied by Hanlon's definition was inappropriate. Our study focuses on homogeneous, aging suburbs, regardless of the spatiotemporal pattern of development that played out in the metro region in which they are found. We can operationalize "homogenous" and "middle-aged" by looking for tracts where housing stock is dominated by units built during the same narrow period. Figure 9 shows, from 74,003 tracts nationwide, the number of tracts that meet this definition of HMS, by decade of dominance.
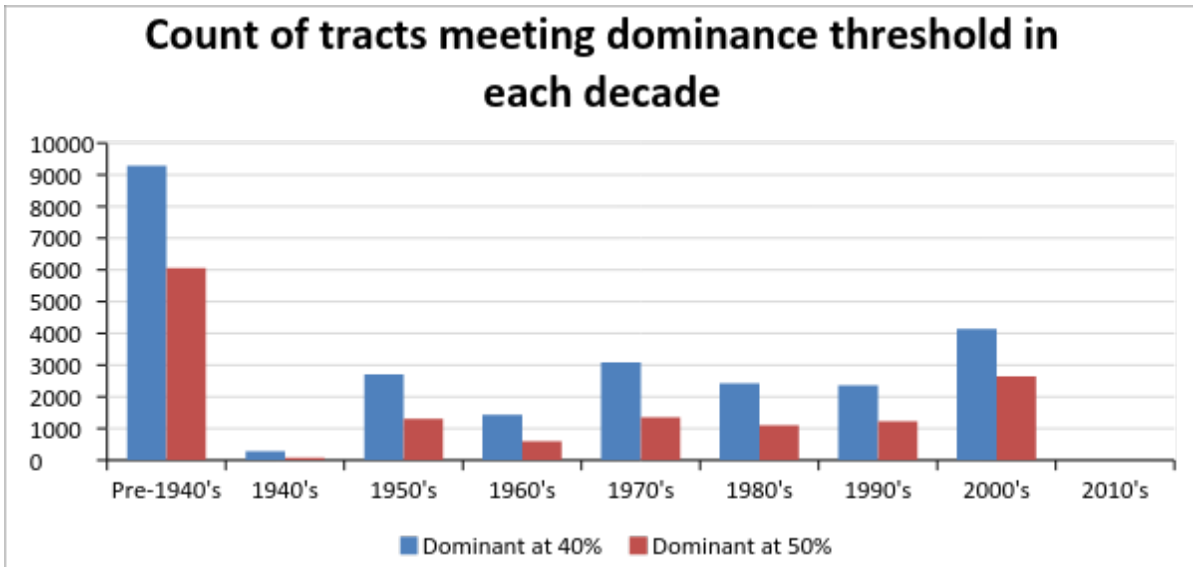
Figure 6: Frequency of housing units count, by tract (all decades)

We tested the impact of varying the dominance threshold, and in consultation with SACOG staff decided that a value of 40% more accurately captured the neighborhoods of local concern. Our final operationalization defines HMS's as 2010 census tracts where 40% or more of existing housing units were built in a particular decade. We identified HMS's from the 50's, 60's, and 70's and used these in our subsequent analysis (see Figure 10 below).
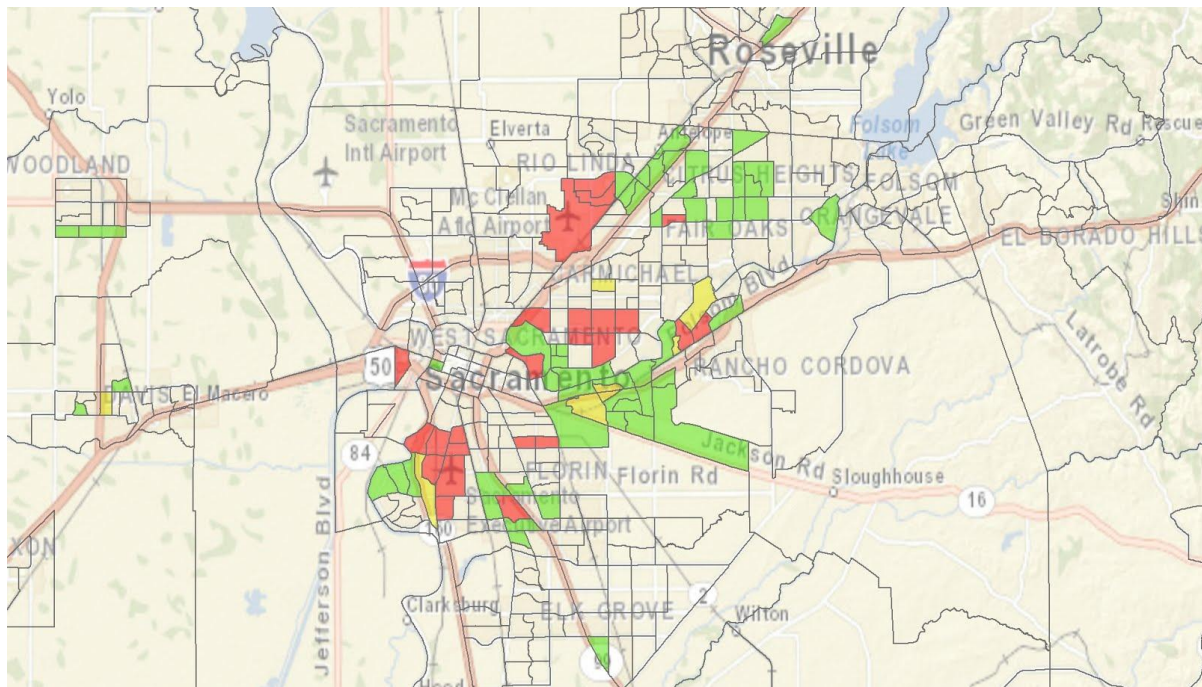

Figure 7: HMS's in the Sacramento region. HMS's dominated by housing stock from the '50's, '60's, and '70's are shown in red, yellow, and green respectively.

34

Defining Greenfield Neighborhoods

Conceptually, Greenfields Neighbors (GFN's) are communities near HMS's that were developed on largely undeveloped lands in the decades after HMS's. Where the HMS definition looks specifically for communities with aging and homogeneously aged housing stock, GFN's conceptually are (1) younger than the HMS they are neighbors too, and (2) represent the prevalent late-20th century pattern of housing development forces creating new communities from scratch. We defined GFN's as tracts within 3 miles of an HMS and in which the majority of housing units post-date 1980 in construction (i.e. median year of construction is after 1980). This criterion filters for tracts that have seen the bulk of their development in recent decades, and in decades following the development of the HMS's we examine in this research.

Defining Urban Sprawl

Urban sprawl is a complex, multi-dimensional urban form phenomenon. In its original scope, this project proposed to use nation-wide state-of-the-art sprawl metrics reflecting several dimensions of urban form, computed by the Metropolitan Research Center. However, these sprawl metrics are only available in consistent formulation for 2010, and our rescoped research question mandated a longitudinal research approach. We therefore adopted a simplified definition of sprawl that reflects the "density" variable only.

We thus operationalized density (and by extension sprawl) as the number of Housing Units (HU) per square mile of the census tract, using HU counts by 2010 census tract, reported in 2007-2011 5-year ACS data. Because population density varies over time and space not only as a function of the built environment, but also as a function of sociocultural factors that play out as variation in mean household size, we reasoned that HU density should be a more valid (i.e. direct) measure of this dimension of sprawl than population density. We used data from the 2007-2011 5-year ACS because these were the same data with which we identified HMS's and GFN's, as well as that used for most 2010-epoch CECI indicator variables.

Measuring Changes in Sprawl and CECI

The research question asks whether a *change* in a neighbor's sprawl (actually a neighbor's growth since only the numerator of the sprawl metric changes over time) correlated with a *change* in an HMS's economic condition. Our 2010 HU age data report (estimates of) Housing Unit (HU) counts per tract, by decade of construction, with all years prior to the 1940's binned together and with the "2010's" estimates representing only construction for 2010 and 2011.[12] Accordingly, the by-decade-of-construction estimates of HU counts in this dataset comprise our time series of development and, by extension from our operational definition, sprawl. It is important to recognize the limitations inherent in interpreting these data this way. The density calculation is a gross density definition since housing units are measured against the entire census tract, regardless of the mix of uses within the tract. We assumed that the present by-decade counts estimated in this dataset represent the full count of HU's built in each decade. In truth some of the HU's built in each decade are lost to demolition through accident or active redevelopment with every successive decade. There are methods for estimating "mortality" rates of HU cohorts (c.f. Johnstone 2001) but we concluded that this was outside the scope of

---

[12] i.e. The two years of the five-year sampling frame falling within this decade.

this study. Our study therefore likely undercounts the stock built in earlier decades somewhat and therefore overstates the rate of growth in HU density between time points.

For both sprawl and CECI, we computed inter-epoch change as the ratio $epoch_{later}/epoch_{earlier}$. The resulting value ranges theoretically from zero (large decline between dates) to one (no change between dates) to infinity (large increase between dates). We then log-transformed the data so that negative values indicate decline over time, zero indicates no change, and positive values indicate growth. We note that there is an advantage in formulating change in sprawl this way; i.e. as a ratio rather than as a difference. The ratio approach magnifies the resulting measure when sprawl (HU density) increases from a low baseline, in accordance with this study's focus on greenfields sprawl.

Associating Sprawl in GFN's with CECI of HMS's

Using ArcMap 10.2, we performed a spatial join of GFN attributes (specifically, inter-epoch sprawl measures) to HMS's. We used a search radius of 3 miles and because more than one GFN could fall within this radius we recorded the mean value of all qualifying GFN's. We performed this analysis for three different sets of HFN's: HFN50, HFN60, and HFN70 (dominated by development from the 1950's, 1960's, and 1970's respectively).

Regression Modeling

With four point-in-time measures for both sprawl and CECI it is possible to compute changes for 3 between-decade periods, as well as for 1980 - 2000, 1990 - 2010, and 1980 - 2010 periods. We modeled HMS's inter-epoch changes in CECI as a function of GFN's mean inter-epoch changes in sprawl, using multiple linear regression models in IBM SPSS Statistics v. 19.0.0 (IBM, 2010). We used binary regional membership variables as dummy covariates to control for regional effects, and also included MSA-wide inter-epoch population growth as a predictor variable in order to control for effects of region-wide growth. Implicit in Mayor Sander's informal hypothesis is that HMS's lose attractiveness in comparison with GFN's because of uniformly aging housing stock and other infrastructure as they age - in other words, we expect to see a time lag before GFN sprawl manifests a negative impact on HMS CECI, and would expect the largest impact to occur with the longest lag. We tested for evidence of impacts (i.e. significant correlation between GFN sprawl and HMS CECI) at three different lags: 90's sprawl v. 2000's CECI; 80's sprawl v. 2000's CECI, and 70's sprawl v. 2000's CECI.

Likely sources of error

We identified several likely sources of error that should be addressed in future work. First, because our method of spatially reallocating values between tracts of different census epochs assumes that variables are uniformly distributed within the tract, in places where development grows radially or linearly we likely overestimate measures outside but on the periphery of development, and underestimate them at the outer edges of previously developed areas. While there are existing data products that use dasymetric mapping techniques for normalizing older measures to constant 2010 geographies, it will be a challenge to acquire or produce data in time series and at national extent that have sufficient thematic resolution to replicate this study. For example, Geolytics sells census data normalized to consistent 2010 geographies using dasymetric techniques. Critically, these data were unavailable when we began this research, as Geolytics' 2010 renormalization effort extended more than a year and a half beyond their

published production schedule.

Second, our use of by-decade existing HU stock data from 2010 as proxy for housing starts data undercounts older stock that has been removed from the inventory of housing stock. We believe that our methodology minimizes this by restricting the analysis to tracts with relatively homogeneous housing stocks.

Third, for logistical reasons we were restricted in this analysis to a simplistic definition of sprawl, contrary to the consensus in the literature that sprawl represents more than simply low density development. As evidenced by the rich literature on measures of sprawl and the conflicting assessments that come out of this literature, it is likely that adding measures of land use diversity, urban design, and degree of centeredness would alter and further enlighten these results. The difficulty of course is in developing national extent, fine- (i.e. tract) grained, time series data from which to develop these measures, but the Metropolitan Research Center has investigated and begun to develop databases from which these could be computed. Given the attention with which Ewing and Hamidi's (2014) recent single-decade update of a national tract-scale multi-dimensional sprawl measure was received, there is certainly an appetite among planners for data of this sort.

Finally, this analysis was subject to what has been defined as the Uncertain Geographic Context Problem (UGCoP). This is "the problem that findings about the effects of area-based attributes could be affected by how contextual units or neighborhoods are geographically delineated and the extent to which these areal units deviate from the "true causally relevant" geographic context" (Kwan 2012). The problem is separate from the better-known Modifiable Areal Unit Problem (MAUP; Fotheringham and Wong, 1991). In this case, the relevant geographic context is the area within which sprawl in GFN's causes (or at least is hypothesized to cause) declining CECI in HMS's. We used a three mile radius as the geographic context defining GFN's in this research, but further work should seek to examine this explicitly.

Directions for Further Work

Our work involved a large complicated research effort involving hundreds of data sets. As with any such project, many potential options diverge along the path of the analysis workflow and not all can be followed. These results, while confirmatory of the research question and local expert's observations, suggest several avenues of further research.

First, CECI represents a unique resource and should be further explored and characterized. Such work could include inward-oriented examination, such as correlation analysis of CECI indicator variables, or outward examination such as the relationship between sprawl and the various submetrics or specific indicator variables. In these respects, do some submetrics show a greater responsivity to competing, sprawling development than others? Second, the database of development dynamics should be extended by modeling cohort mortality in HU's. And finally, this study's correlation results are interesting but not definitive. It remains to be determined whether the correlation noted is reflective of a causal relationship, and assuming that this is the case, what is the mechanism of effect for sprawling development on economic health? These and other questions stand to further inform the work of SACOG and other regional planners around the country. We anticipate that SACOG's planning work will benefit from the findings of the current research.