

## Methodological Appendix

### How PRIM Works and Interpreting PRIM Results

Each uncertainty parameterization is referred to as a *scenario* though they don't resemble the fully fleshed-out stories of exploratory scenarios. Each scenario is run through the modeling suite without policy and with each policy intervention (Gross, 2018). For each completed run, or *future*, I calculate seven indicators: core household growth, inner suburban household growth, outer suburban household growth, household growth beyond the region, household growth in high accessibility areas, growth in targeted ecological areas, and median housing price. The indicators of each future are translated into *regret*, the difference between the highest-performing future for each scenario and the performance of the selected future. Regret allows for comparison across futures in unlike scenarios.

With regret values calculated, scenario discovery utilizes data mining to explore the broader uncertainty space for regions in which a policy performs particularly well or poorly. I employ the patient rule induction method (PRIM), an algorithm that searches for lower dimensional boxes of concentration within higher dimensional space (Friedman & Fisher, 1998). Bryant et al (2010) promote this method for scenario discovery because it provides easier interpretations of results than comparable methodologies, such as Classification and Regression Tree, which produce similar results, but without orthogonal boxes. To operationalize PRIM, I convert regret into binary values of success and failure utilizing a semi-arbitrary threshold – such as all policy scenarios with regret above the median (Bryant & Lempert, 2010; Gross, 2018). Binary values allow for calculating the density of failure within a subspace. The PRIM algorithm sorts through each dimension of uncertainty, slicing it into two parts that exhibit the highest difference in failure densities. The lower

density region is discarded and the algorithm moves on to the next dimension of uncertainty until all have been examined. The remaining subspace is declared a region of high regret and can be interpreted as futures that generally produce undesirable results. PRIM then iterates over the remaining space to determine secondary regions of concern. The process can similarly be used to determine regions of relative success. A key limitation of utilizing PRIM is that results are highly sensitive to the order in which the dimensions are analyzed.

PRIM is designed such that whether a point is within or outside of the box in one dimension is independent of all other dimensions. Essentially this is a high dimensional rectangular box. The results for any dimension of the PRIM box can thus be interpreted directly in terms of which side of the threshold points fall on. For instance, when measuring the core households indicator in the baseline scenario, the PRIM box for distance to work is -.09 to .45. This indicates that failures are concentrated for distance to work LHS values below .45. While one could interpret this as a hard threshold, visual examination of the PRIM boxes in each dimension usually indicate tendencies. Thus, I prefer to interpret the PRIM box range as indicating that lower values of distance to work are associated with greater regret in choosing the baseline policy.

## Density of Scenarios

Given this investigative framework, how many scenarios are sufficient? This is of particular concern given the runtime limitations of the SILO model and other models researchers and agencies would use for similar exercises. While there is no way to develop an absolute target, Swartz and Zegras (2013) developed a measure of the density of an LHS sample:

$$J = n_s^{1/k},$$

where  $n_s^{\square}$  is the number of runs performed in any sample and  $k$  is the dimensions of uncertainty explored. Previous research has ranged in density from figures near 1.1 to 3.16. As such, I have selected to generate 100 scenarios, which corresponds with a density of 2.15, which compares favorably to the other experiments. For instance, Augusdinata and Dittmar (2009) generated 75,000 scenarios – however, in 18 dimensions, their scenario density was 1.87 despite the far larger number of runs.

## SILO Model Details

The following section presents the baseline weights in the SILO model for calculating regional and dwelling unit utilities. The logit model weights are determined heuristically from the literature. For details on the equations in SILO, see Moeckel (2016). Most of the input data – that for people, households, and dwellings - comes from the Census PUMS (Public Use Microdata Sample). In this application, we augmented our data with each state’s publicly available 2015 school quality data, 2015 Location Inc. crime data, and development capacity data built by the National Center for Smart Growth from county zoning maps and other sources.

### Replaceable Location Factors

Replaceable location factors are qualities of a region or dwelling that a household values, but which they are they can trade off against other qualities. Together their combined utility constitutes one factor in the overall utility of a region or unit.

HH Type	Location Factors						
HH size	Income Quartile	Dwelling Size	Dwelling quality	School quality	Auto accessibility	Transit accessibility	Crime Index
1	1	0.120	0.330	0.000	0.250	0.300	0.000
2	1	0.140	0.285	0.025	0.250	0.300	0.000

3	1	0.160	0.240	0.050	0.250	0.300	0.000
4+	1	0.180	0.170	0.100	0.250	0.300	0.000
1	2	0.170	0.330	0.000	0.200	0.250	0.050
2	2	0.190	0.210	0.100	0.200	0.250	0.050
3	2	0.210	0.140	0.150	0.200	0.250	0.050
4+	2	0.230	0.070	0.200	0.200	0.250	0.050
1	3	0.220	0.400	0.000	0.150	0.150	0.080
2	3	0.240	0.230	0.150	0.150	0.150	0.080
3	3	0.260	0.110	0.250	0.150	0.150	0.080
4+	3	0.280	0.040	0.300	0.150	0.150	0.080
1	4	0.270	0.430	0.000	0.120	0.080	0.100
2	4	0.290	0.110	0.300	0.120	0.080	0.100
3	4	0.310	0.090	0.300	0.120	0.080	0.100
4+	4	0.330	0.070	0.300	0.120	0.080	0.100

## Essential Location Factors

Essential location factors must be fulfilled at a minimal level or the household will not consider the unit.

Below are the weights for determining the final utility of a region or unit.

HH Type		Location Factors			
Household size	Income Quartile	Replaceable Location Factors	Dwelling price	Distance to work	Transportation costs
1	1	0.200	0.350	0.175	0.275
2	1	0.200	0.350	0.175	0.275
3	1	0.200	0.350	0.175	0.275
4+	1	0.200	0.350	0.175	0.275
1	2	0.350	0.300	0.200	0.150
2	2	0.350	0.300	0.200	0.150
3	2	0.350	0.300	0.200	0.150
4+	2	0.350	0.300	0.200	0.150
1	3	0.500	0.100	0.225	0.175
2	3	0.500	0.100	0.225	0.175
3	3	0.500	0.100	0.225	0.175
4+	3	0.500	0.100	0.225	0.175
1	4	0.600	0.050	0.250	0.100

2	4	0.600	0.050	0.250	0.100
3	4	0.600	0.050	0.250	0.100
4+	4	0.600	0.050	0.250	0.100