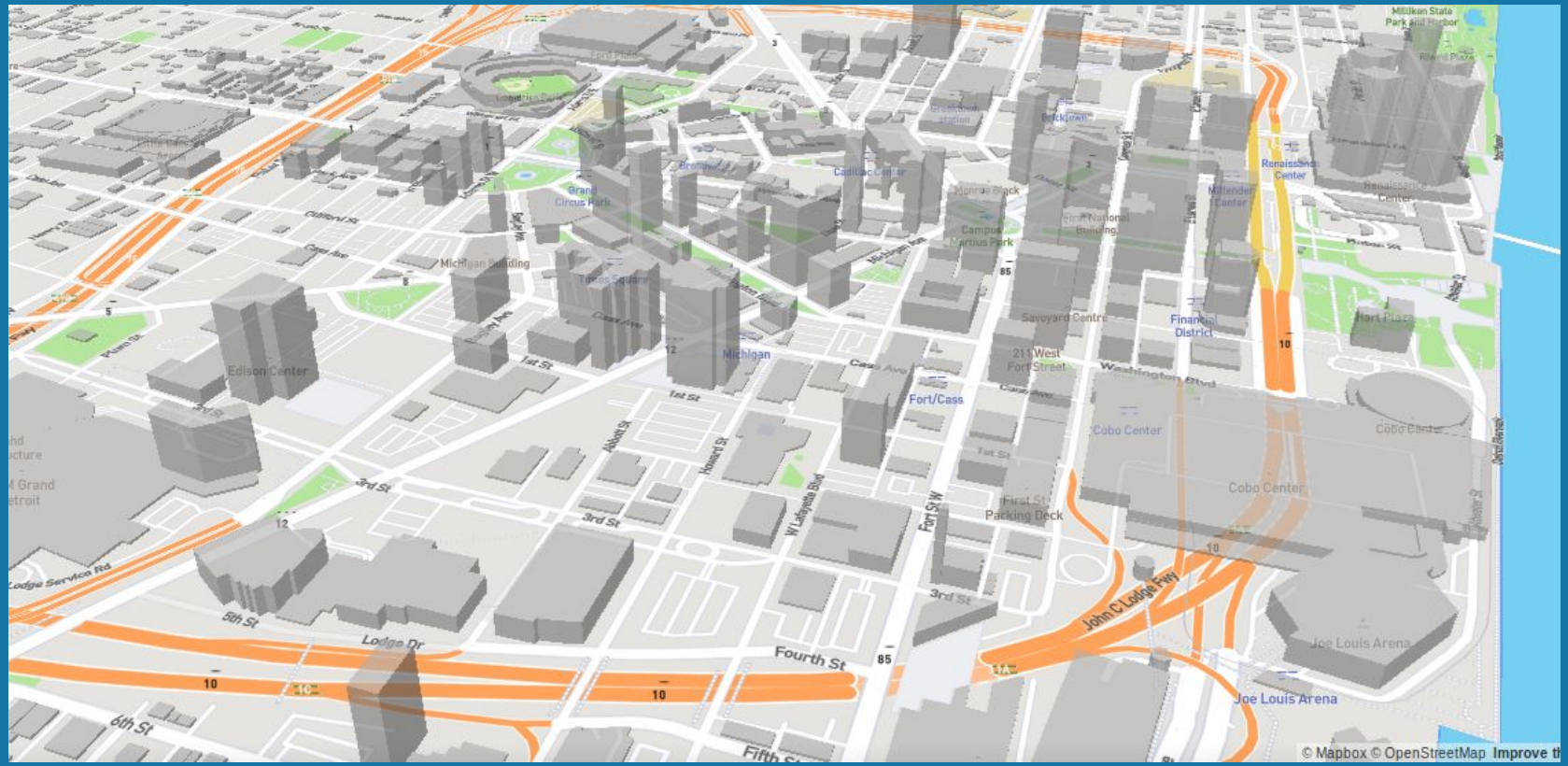
SEMCOG-UrbanSim Bi-weekly



May 30, 2019

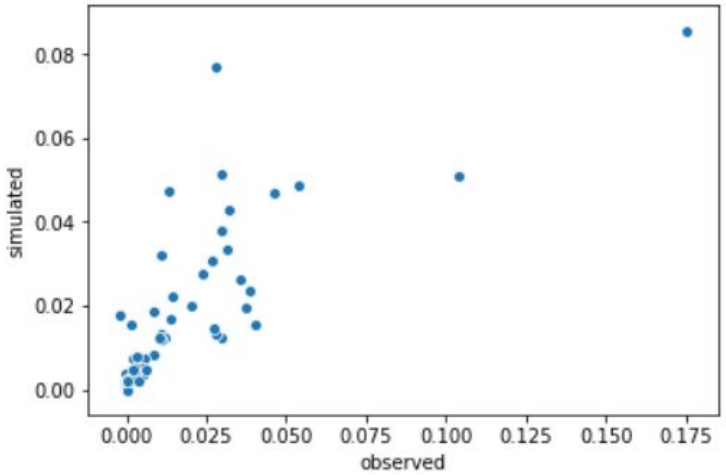
Effect of calibration- HLCM1 in Oakland County

`income_quartile ==1 & persons <=2 & age_of_head >= 65`

pre-calibration

```
Validation metrics for HLCM1 growth, 2016 - 2025
  MSE is 0.000296343783352
  RMSE is 0.0172146386355
  Prediction R2 is 0.603724680394
  Correlation is 0.782805462924
```

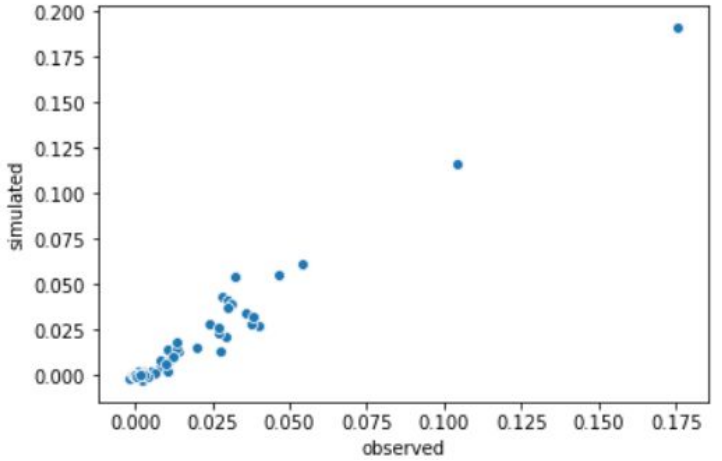
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6cf3c1fdd0>
```



post-calibration

```
Validation metrics for HLCM1 growth, 2016 - 2025
  MSE is 3.89766105225e-05
  RMSE is 0.00624312506062
  Prediction R2 is 0.947879896054
  Correlation is 0.98313355284
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f326f03ad90>
```



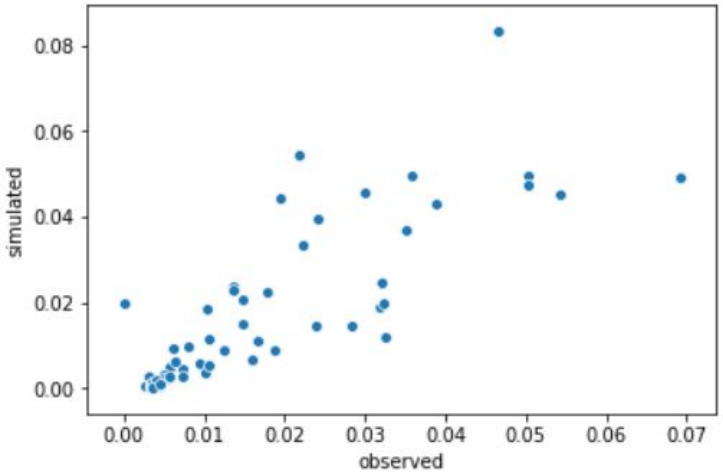
Effect of calibration- HLCM7 in Oakland County

`income_quartile in [2,3] & persons <=2 & age_of_head >= 65`

pre-calibration

```
Validation metrics for HLCM7 growth, 2016 - 2025
  MSE is 0.000108252547977
  RMSE is 0.0104044484706
  Prediction R2 is 0.538121651989
  Correlation is 0.827181100854
```

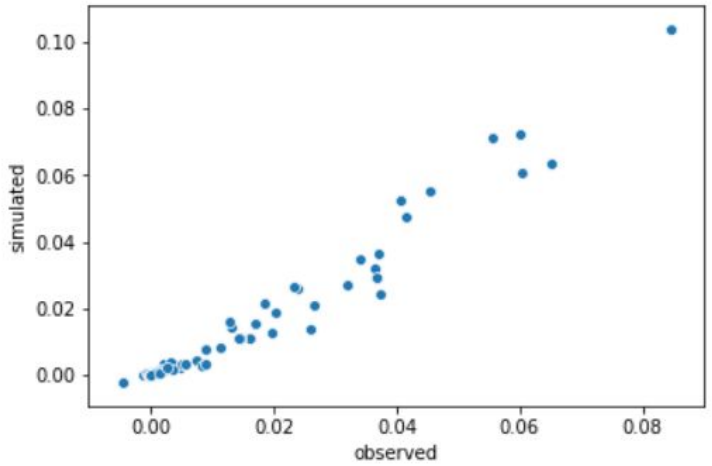
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6cf19c7e50>
```



post-calibration

```
Validation metrics for HLCM7 growth, 2016 - 2025
  MSE is 2.79974853339e-05
  RMSE is 0.00529126500318
  Prediction R2 is 0.928159600977
  Correlation is 0.975864977478
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f326f0e4a10>
```



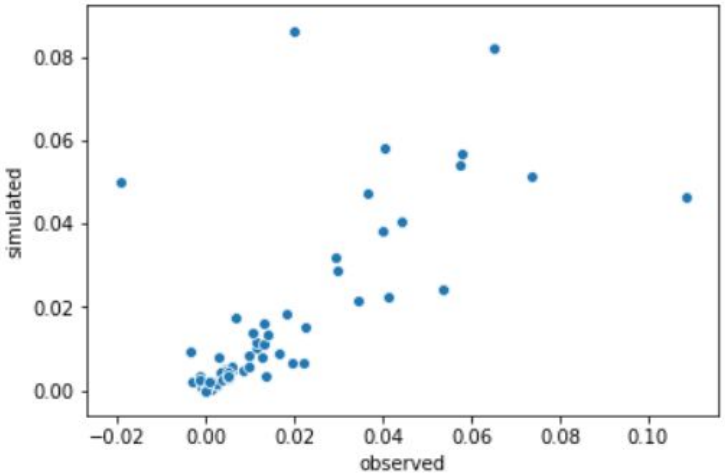
Effect of calibration- HLCM13 in Oakland County

`income_quartile ==4 & persons <=2 & age_of_head >= 65`

pre-calibration

```
Validation metrics for HLCM13 growth, 2016 - 2025
  MSE is 0.000269301390916
  RMSE is 0.0164104049589
  Prediction R2 is 0.452686843286
  Correlation is 0.707322604764
```

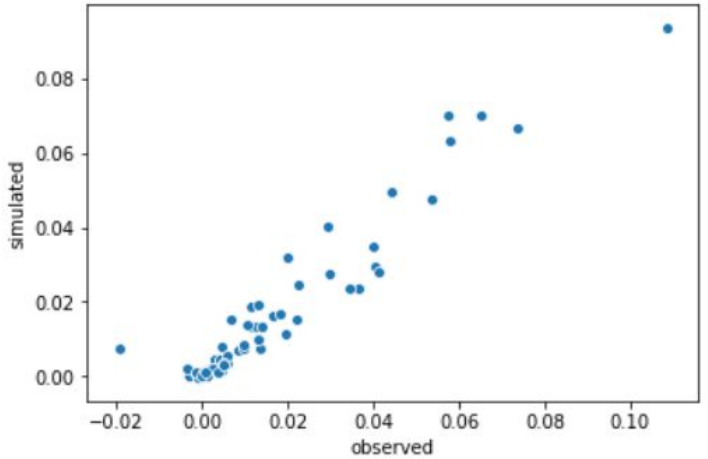
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6cf3d3cd50>
```



post-calibration

```
Validation metrics for HLCM13 growth, 2016 - 2025
  MSE is 4.19986669712e-05
  RMSE is 0.00648063785219
  Prediction R2 is 0.914644247029
  Correlation is 0.956560034508
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f326ea649d0>
```



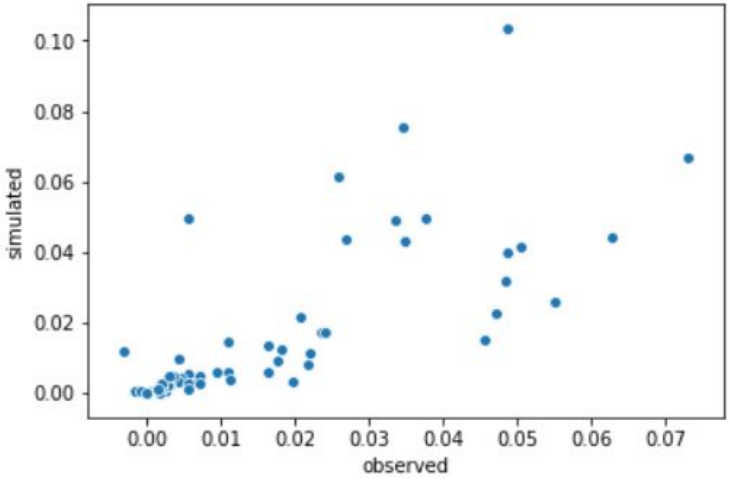
Effect of calibration- HLCM16 in Oakland County

`income_quartile ==4 & persons > 2 & age_of_head >= 65`

pre-calibration

```
Validation metrics for HLCM16 growth, 2016 - 2025
  MSE is 0.000211545986026
  RMSE is 0.0145446205184
  Prediction R2 is 0.387157592517
  Correlation is 0.753201952982
```

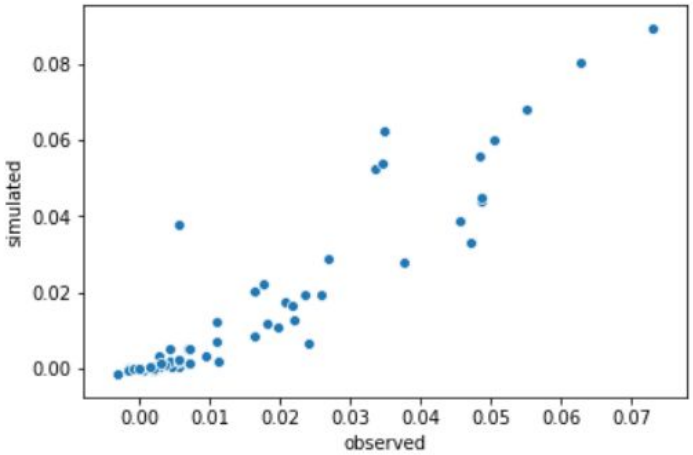
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6cf3174cd0>
```



post-calibration

```
Validation metrics for HLCM16 growth, 2016 - 2025
  MSE is 7.71586219941e-05
  RMSE is 0.00878399806433
  Prediction R2 is 0.776473775044
  Correlation is 0.927769553947
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f326e824950>
```



Benefits of a calibration approach that utilizes auto-differentiation + gradient descent

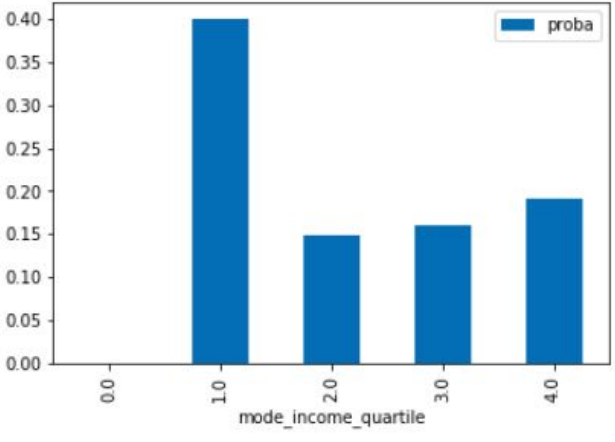
1. Don’t need to run an actual simulation, so calibration can be faster and less influenced by randomness
2. Multi-geographic levels possible- e.g. some iterations with zone-level loss function, and then some iterations with city-level loss function. Or a loss function that accounts for both.
3. Can fine tune estimated parameter values with small step sizes, using estimated parameters as the starting values, to preserve some of the information from the cross-sectional data
4. Tune behavioral variables jointly along with spatial dummies

Building vs unit alternatives in estimation

BUILDING-LEVEL ESTIMATION

Segment: household type = 1
Null Log-liklihood: -13815.511
Log-liklihood at convergence: -11289.880
Log-liklihood Ratio: 0.183

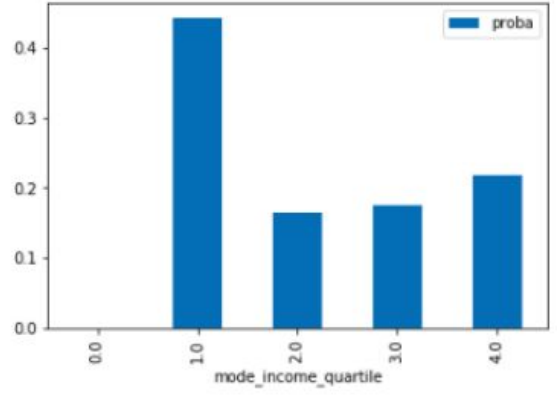
Component	Coefficient	Std. Error	T-Score
st_b_ln_residential_units	0.448	0.006	80.731
st_b_ln_sqft_per_unit	-0.468	0.086	-5.441
st_zones_population	0.046	0.027	1.695
st_zones_prop_income_1	0.367	0.032	11.454
st_zones_mean_age_of_head	0.345	0.026	13.424
st_zones_mean_hh_size	-0.111	0.026	-4.261
st_zones_prop_race_1	0.237	0.031	7.593



UNIT-LEVEL ESTIMATION

Segment: household type = 1
Null Log-liklihood: -13815.511
Log-liklihood at convergence: -13377.959
Log-liklihood Ratio: 0.032

Component	Coefficient	Std. Error	T-Score
st_b_ln_residential_units	0.075	0.005	15.481
st_b_ln_sqft_per_unit	-0.592	0.091	-6.510
st_zones_population	0.014	0.026	0.526
st_zones_prop_income_1	0.325	0.031	10.426
st_zones_mean_age_of_head	0.404	0.026	15.723
st_zones_mean_hh_size	-0.024	0.025	-0.946
st_zones_prop_race_1	0.173	0.030	5.796



Building vs unit alternatives in estimation

3- RUN SIMPLIFIED WITH REGIONAL SPECS CLUSTERING VARS

	Overall MSE
zones	3,522,233
large_areas	36,552
cities	495,160
counties	40,603

3 B- RUN SIMPLIFIED WITH REGIONAL SPECS UNIT LEVEL HLCMS - CLUSTERING VARS

	Mean MSE
zones	1,317,909
large_areas	32,943
cities	495,681
counties	36,500

Moving beyond test cases- next steps:

1. Job-space level estimation of the ELCM (job spaces as the alts in estimation)
2. Unit-level estimation of location choice models, simulated by large area
3. Calibrate proforma
4. Tune estimated parameter values (plus city dummies) with the calibrator- for each large area and each submodel
5. Run full simulation with large area controls + calibrated coefficients

Potentially:

- Test joint calibration of all components together, over the forecast horizon