August 29-30, 2019

Real Estate Market Model Peer Review Report

WASATCH FRONT REGIONAL COUNCIL

MAG

Table of Contents

Goals of Peer Review 1
REMM Overview 2
Geographic Extent 2
History 3
Use in Transportation Planning 4
Primary Data Inputs 5
Model Structure 6
Gathering Local Knowledge 8
Panel Discussion: REMM Overview 8
Model Inputs 10
Panel Discussion: Streamlining Process of Updating Base Year Input Data 11
Panel Discussion: Creating and Managing Policy Layer Inputs 13
Panel Discussion: QA/QC of Base Year Data and Policy Layers 14
Accessibility Sensitivity 16
Panel Discussion: Incorporating Appropriate Accessibility Sensitivity into REMM 18
Developer Model 20
Panel Discussion: How to Model Development and Redevelopment 20
Panel Discussion: Mixed Land Use Development 21
Panel Discussion: Incremental Development 23
Panel Discussion: “Leapfrog” Development 24
Panel Discussion: Over-Building by Developer Model 25
Panel Discussion: Monte Carlo Process 26
Panel Discussion: Spatial Unit for the Model 28
Model Calibration and Validation 30
Panel Discussion: Price and Location Choice Estimation Variables 30
Panel Discussion: Price and Location Choice Model Estimation 31
Panel Discussion: REMM Sub-Model Parameter Calibration 32
Demographic Transition 33
Panel Discussion: Household Synthesis for Every Year 33
Outputs and Visualization 34
Panel Discussion: Summary Statistics and Visuals Used for Model Output 34

Panel Discussion: External Review 39
Panel Recommendations 42
Observations 42
Recommendations 42
Attendees A-1
Panelists A-1
Attendees A-1
Panelist Biographies A-3
Agenda A-4
Variables and Specifications Handouts A-6
Handout 1 – Variables A-6
Handout 2 – Modeling Tables A-8
REMM Documentation A-16

Real Estate Market Model Peer Review Report

Goals of Peer Review

Introduction of the panelists:

- **Guangyu Li**, Southeast Michigan Council of Governments (SEMCOG)
- **Jesse Ayers**, Maricopa Association of Governments (MAG-Phoenix)
- **James Kolberg** (by phone), Mid-Region Council of Governments (MRCOG)

Wasatch Front Regional Council (WFRC) and Mountainland Association of Governments (MAG-Provo) organized the 2019 REMM peer review to:

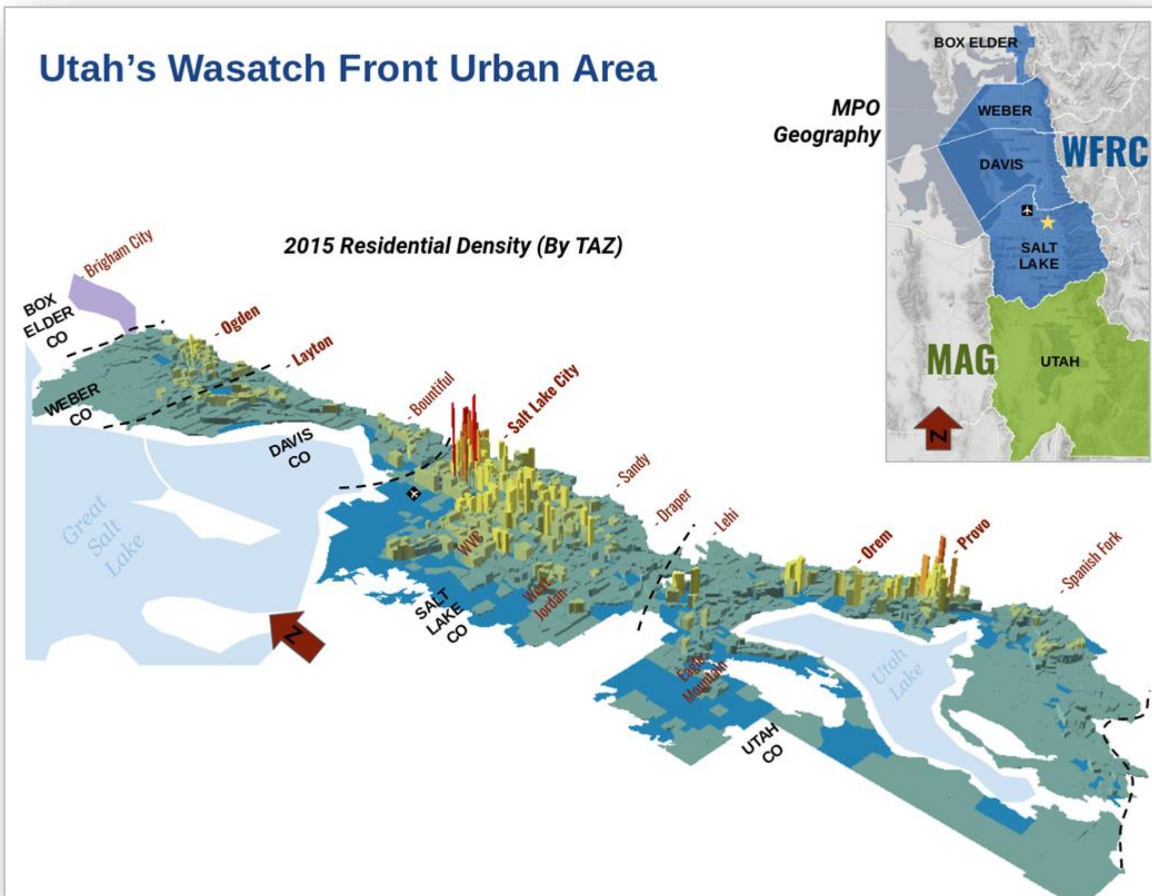
- Evaluate the Real Estate Market Model (REMM) logic, input data preparation, results, and processes relative to best practices employed by technical experts working toward similar requirements and objectives elsewhere in the nation
- Identify and assess opportunities for future REMM enhancements and begin prioritization for future REMM work
- Present a technical overview and discussion of REMM to further the knowledge of staff at partnering stakeholder agencies
- Strengthen relationship and future collaboration potential between land use model professionals at metropolitan planning organizations

After an introduction to REMM, the review was organized around six specific topic areas with discussion questions presented around challenges perceived by WFRC/MAG-Provo technical staff. Each 'challenge' discussion question was presented during the review together with additional context, including considerations, current approach, and strength/weaknesses observations. This format was chosen in order to maximize the exposure to, and constructive feedback received on perceived key challenges.

REMM Overview

Geographic Extent

WFRC and MAG-Provo jointly maintain and operate REMM. REMM covers the urbanized portions of Weber, Davis, Salt Lake and Utah Counties (the area covered by the Wasatch Front Travel Demand Model). The travel model also includes a portion of Box Elder County from Brigham City southward. However, this area is not currently in REMM. Utah County is part of the MAG-Provo planning area. The remaining counties belong to the WFRC planning area. The following graphic depicts the geographic extents covered by the REMM model.

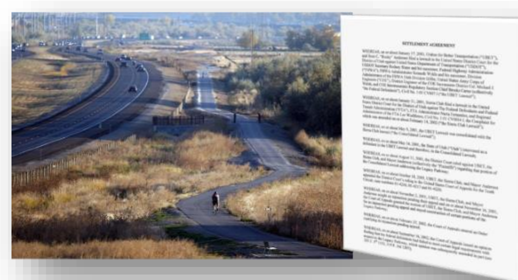


History

The history of land use modeling in Utah dates back to the early 1990s. In 1995, Governor Mike Leavitt convened Utah's Growth Summit, bringing together planning for transportation, land use, water, and air quality. Government agencies and private sector organizations participated. One result of the summit was the establishment of the Quality Growth Commission, and the Quality Growth Efficient Tools (QGET). As part of QGET, the state and MPOs jointly evaluated the possibility of developing an integrated land use/transportation model. Many of the leading researchers in the field of land use modeling were brought to Utah to present their models. In the end, the committee selected Paul Waddell and his UrbanSim econometric model.

The data compilation and computing requirements of the model necessitated an "all hands on deck" approach from several state and local agencies. In the end, the maintenance requirements of the model were too great for a single agency to shoulder, and the original UrbanSim model was shelved.

In 2003, the Sierra Club and other environmental groups sued the Utah Department of Transportation (UDOT) to halt construction of the Legacy Parkway in western Davis County. As part of a settlement agreement to that lawsuit, WFRC agreed to, once again, evaluate the usefulness of a land use model. Picking up where the first effort left off, a basic UrbanSim implementation was developed and used in the development of the socioeconomic projections for the 2007 Long Range Plan. Lack of staff expertise and resources again shelved the land use model.



In 2012, WFRC made the institutional decision to put sufficient resources towards developing a living land use/transportation model. Once again, the UrbanSim platform was selected, and this third implementation effort was named the Real Estate Market Model (REMM) in a rebranding effort. The model that came from this latest effort is the model that was reviewed during this panel.

In general, the land use modeling effort to date has not been without challenges. These include:

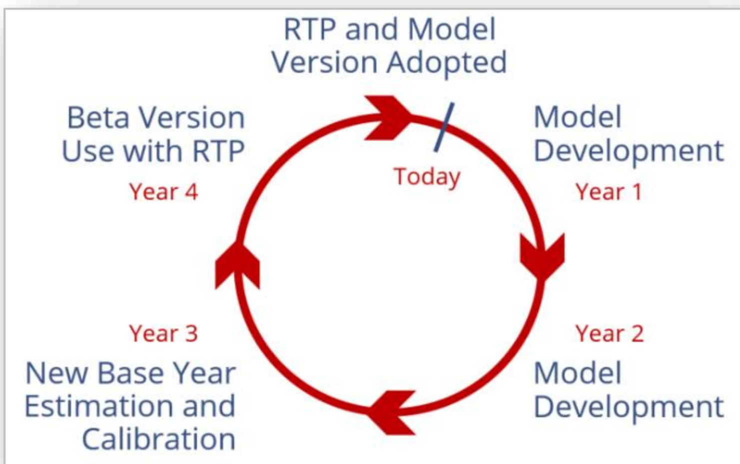
- Resources, or lack thereof, at the MPO level
- Buy-in, or lack thereof, from management and policy levels
- Distrust of a "black box" model
- Availability of input data

MAG-Provo supported the previous UrbanSim modeling efforts, primarily on the data side. However, it wasn't until the most recent Regional Transportation Plan adopted in 2019 that MAG-Provo formally made REMM a part of its regional planning process.

Use in Transportation Planning

MAG-Provo and WFRC are on a coordinated, 4-year model development cycle aligning with the Regional Transportation Plan (RTP) adoption cycle. The first two years of the planning cycle are devoted to improving the land use and travel demand models and input datasets. Year 3 entails calibrating the models to a new base year, followed by running the models to support the preparation of regional transportation plans. In year 4, the new model input data layers (e.g. land uses policy layers, demographic forecast, and transportation projects) are finalized and supporting analyses (performance measures and air quality analyses) are done.

Model Adoption Cycle



In the 2019 plan development, MAG-Provo utilized REMM to assess the impact on land use of a highway and transit "build" versus "no-build" transportation scenario. WFRC utilized REMM to produce two different socioeconomic forecasts that were released with the travel demand model:

- **Fiscally Constrained** – the transportation supply was constrained to the assumed funding available for the plan horizon and each phase
- **Fiscally Unconstrained** – the transportation supply represented the "needed" projects by phase and was not

constrained based on assumed funding availability

In addition, REMM was employed to test the sensitivity of socioeconomic allocation and transportation supply for two regional transportation planning efforts, the West Davis Connector Environmental Impact Study and the Wasatch Front Central Corridor Study.

Primary Data Inputs

The following lists the primary REMM inputs that were presented as an overview for the peer review panel discussion:

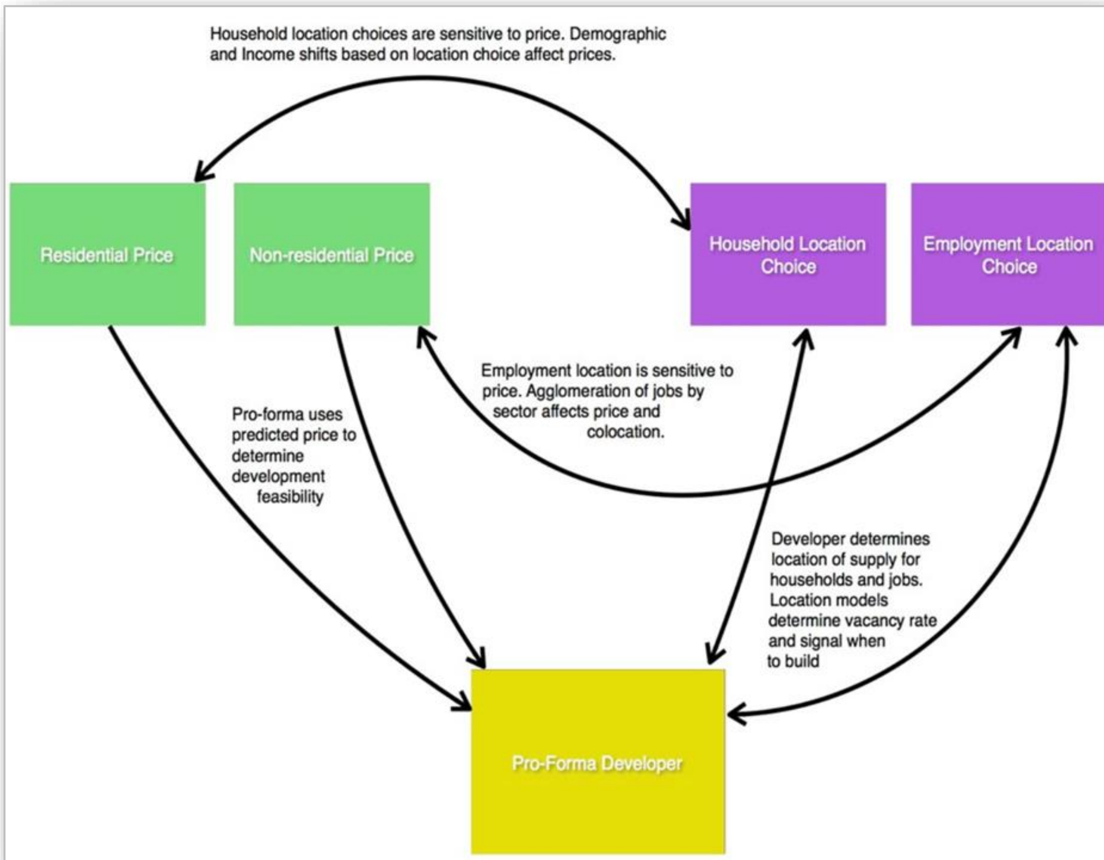
- Base Year Buildings Database
    - Location, size, value, age, lease values, sales values, type
    - Sources: county assessor, Utah Department of Workforce Services, and Coldwell Banker
    - Lease values and sales values are from MLS samples
- Base Year Households and Population
    - Synthesized household characteristics
    - Source: Census 2010 and Public Use Microdata Areas (PUMA) data
- Base Year Employment
    - Job type, quantity, and location
    - Source: State Department of Workforce Services (DWS)
    - A gap exists between the employment county control totals and the observed employment data from DWS; the gap was closed by adding missing jobs, by sector, and distributing them within the region using REMM's Location Choice Model
- Policy Layer
    - Zoning and General Plan data is collected from each jurisdiction, then compiled and transformed into a land use policy/capacity layer, taking a common, generalized format
    - Base year and future year changes are made by location and year and compiled in a policy/capacity change table
    - This change table allows for shifts in density and use over time
    - "Pipeline" projects may be asserted into the developer model for identified post base-year construction projects or to evaluate different development proposals

- Accessibility Variables
    - Distance to freeway exits, traffic volumes, distance to rail stations, commute times, population and jobs within "x" minutes
    - These accessibility inputs are derived from the travel demand model
- County Level Control Totals

- Households, population, employment by sector
- Source: Utah Population Committee and University of Utah’s Kem C. Gardner Policy Institute

Model Structure

REMM runs on the standard UrbanSim 2 platform. Of the many sub-modules contained in UrbanSim, the agencies chose a pared down set with the primary focus to get a solid tool that produced plausible forecasts. The basic UrbanSim structure and the sub-modules included in the REMM model are presented in the following graphics (note the first graphic is from Paul Waddell, Urbansim Inc).



The agencies shared the following details of the REMM model structure at the peer review meeting:

- Real Estate Price Model
    - Uses Ordinary Least Square (OLS) Regression
    - Product type: 2 Residential, 3 Non-residential
    - Relevant input variables include: size and age of structure, vacancy rate of the area, accessibility, availability of transit service
- Real Estate (Residential and Non-residential) Development Project Location Choice Model
    - Pro Forma based, uses probability weighted by profit (price minus cost) to inform the building construction model
- Redevelopment Module
    - Uses economic logic to determine when a land parcel is prime for redevelopment
    - Takes into account: current price, zoning/capacity, potential redevelopment, opportunities and estimated values/prices for each, price of demolition, age of existing structures
- Household and Employment Location Choice Models
    - Multinomial Logit Models
        - Segregated by choosing group (income class or job type)
        - Uses Random Utility Theory to determine probabilities of location choice by chooser to determine where they want to locate within the available building stock
    - Examples of relevant variables: age of structure, income of surrounding residents, traffic volume of adjacent roads, distance from freeways
- Sub-Model Residential Specification
    - Market Segmentation (demand) – 4 categories of households based on income
        - $0 - $35k
        - $35k - $59k
        - $60k - $90k
        - $91k +
    - Market Segmentation (supply)
        - Single Family
        - Multi-Family
- Sub-Model Non-residential Specification
    - Market segmentation (demand) – 8 employment categories forecasted by REMM
        - Retail
        - Food and Accommodations
        - Manufacturing
        - Wholesale and Transportation
        - Office
        - Health
        - Government and Education
        - Other

        - Note: 4 additional employment categories are calculated via post-processing and not forecast by REMM:
            - Agricultural
            - Mining
            - Construction
            - Home-based Jobs
    - Market Segmentation (supply) – 3 categories of non-residential buildings
        - Retail
        - Industrial
        - Office
- Sub-model specification informed by:
    - Expert panel events with invited realtors, home builders, commercial lenders, commercial leasing agents, developers
    - Extensive GIS visual analysis by small area by the REMM model development team
    - Statistical analysis of data, including approximately 175 independent variables

Gathering Local Knowledge

Prior to the REMM model development, the agencies met with multi and single-family home developers, commercial developers, realtors, and finance and leasing agents to obtain an "on the ground" knowledge from local professionals. During the calibration and validation process, the agencies met again with local professionals, including developers, real estate professionals, planning officials, and demographer professionals, to obtain feedback on the plausibility of REMM's forecasts.

Panel Discussion: REMM Overview

The panelists asked if the agencies are collecting pipeline projects from member agencies (e.g. local governments). The agencies are not yet doing this currently but would like to do it.

One panelist commented that the UrbanSim sub-modules selected by the agencies to use in REMM is not too far off from how they have structured their UrbanSim application.

The panelists asked if the employment categories are aggregation of NAICS. The REMM model does aggregate NAICS to their employment categories.

The panelists inquired how much local developers checked the pro forma profit calculations. The agencies reported that in the meetings with the developers, the developers shared their complete costs, though they were generalized to protect the developers' trade secrets. There were some elements that the agencies could not account for in the pro forma costs, such as quality of schools or crime rate, since these items could not be forecast. The panelists agreed that there are elements that cannot be included in the pro forma costs and that assumptions would need to be used when no forecasted variable data is available. One panelist remarked that in a slow growth area patterns don't change very often. Also, checking with local communities is a good way to validate the results of the model. If the locals feel comfortable with the future growth pattern, the agency can feel comfortable.

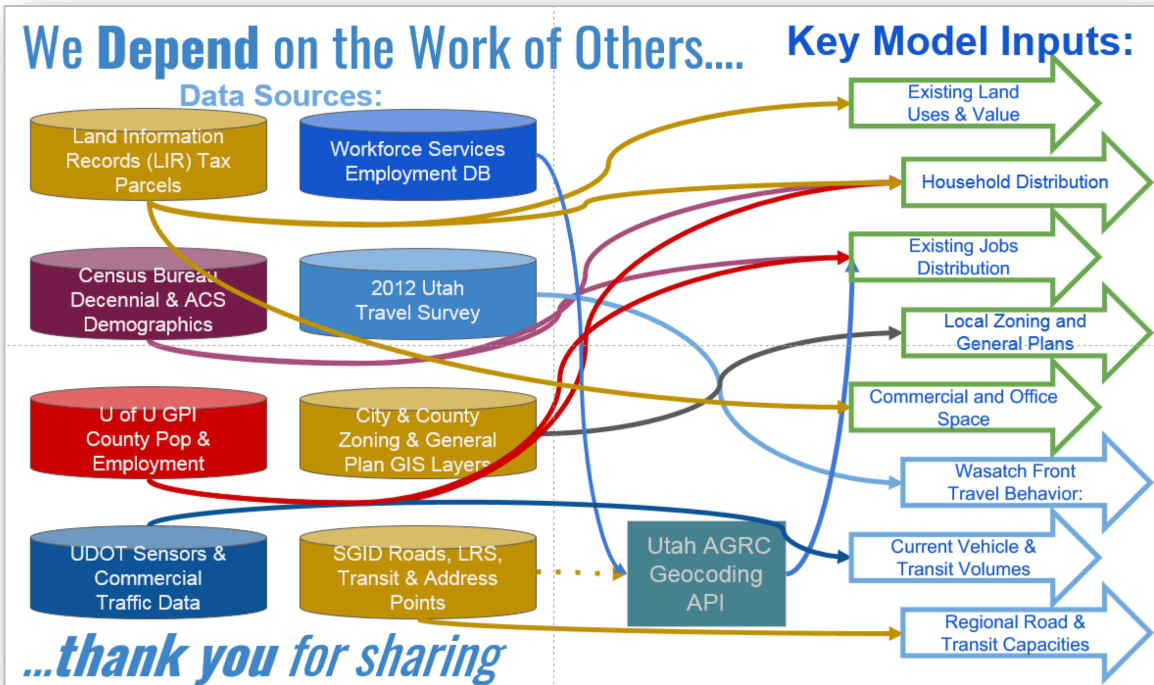
The panelists asked if the agencies have considered changing from TAZ to Census block groups for reporting output. The agencies have not had that conversation.

The panelists asked if the agencies see a trend of city general plans becoming more general (e.g. 20 categories down to 6) and that their plans have much less specificity (e.g. large areas labeled mixed use) making it harder to define densities. The agencies have not seen this trend along the Wasatch Front.

The panelists asked if REMM has different plan types and building types. One of the panelists dealt with this particular issue and commented that using a sample by weight could be used to force new construction to be reasonable. This was done by the agency using a formula because of the number of parcels. Another panelist commented that they went to a one-to-one relationship between plan type and building type, but this required the plan type to have detailed building types to match the policy or general plan layer. The agencies reported they use a mix of building types, stratified by a list of rules. Some rules maintain a one-to-one relationship and others maintain a one-to-many. REMM does not currently use a formula to associate building type with plan type. REMM's solution was to keep both very simple and more general.

Model Inputs

The agencies reviewed how the data needed to support REMM is highly dependent on the work of other organizations and the data is being developed for purposes other than land use modeling. Much effort is required to get the data into a REMM model-ready format.



A primary example of this is the use of the county assessor data which provides assessed value, property characteristics, and other information at the parcel level. A few years ago, Utah began to provide assessor data via an online statewide GIS database. Included in the database are land value and a semblance of a land use code. However, the assessor data varies by county requiring work to generalize/standardize the data and to perform other cleanup work.

The agencies sought specific feedback from the panelists on the following topics:

- Streamlining process of updating base year input data
- Creating and managing policy layer inputs
- QA/QC of base year data and policy layers

Panel Discussion: Streamlining Process of Updating Base Year Input Data

The agencies shared the following information:

> **Key Considerations:** Data availability, formatting, cleaning, standardization.
>
> **Our Current Solution:** Base year data inputs are created from Census data, assessor and recorder datasets, and QCEW employment data. Cleaning and standardizing data and formatting for use in the land use model take an extraordinary amount of time.
>
> **Strengths, Weaknesses, Opportunities, Threats Overview:** We are considering a process for updating the base data layer annually with the annual release of assessor data. Lag time with the employment data is an issue. Land value calculations vary from county to county. Lack of current building footprint data.
>
> **Discussion Question:** What are some best practice considerations for streamlining the data development issues mentioned?

The panelists acknowledged that getting and cleaning data for use in UrbanSim is a huge lift and they feel they are in the same boat as the local agencies.

SEMCOG uses parcels originally obtained from the county assessor. Maintaining the parcel shape is not always easy since parcels are always changing (e.g. being split or merged). The assessor data also did not provide building footprint information. SEMCOG contracted with a firm, Sanborn, to build a very basic footprint map. This process used LIDAR data and aerial imagery to get building footprints and heights. Once SEMCOG received the building footprints, additional work was required to verify the building footprint data and to identify building type. County assessor data helped identify building type, with the acknowledgement that assessor data is somewhat limited (e.g. not that accurate for certain variables or some counties have their own standards and do not report the same level of data). Employment data was also used to identify building type. Aerial imagery was used in checking the building footprint data. Once the building footprint data was created, SEMCOG incrementally updated the database using building permit data and aerial imagery. The online aerial image source they used had a time slider for the most recent 5 years which helped them accurately identify when a structure was been built.

Creating the initial building footprint database took SEMCOG a long time, about 2 ½ years. Incremental updates are done to the database constantly as staffing resources allow and as data becomes available. The constant updating of the data poses a challenge. Small changes to the input data can amplify the random seeding from the Monte Carlo process in UrbanSim. This presents the dilemma of wanting to use the latest data but not wanting results to be constantly shifting, which may bring negative light on the model. SEMCOG freezes the land use model for a period of time so there is a stable release of the model and results. Future releases include all significant changes and updates since the last release. Knowing when to release the model has been a challenge. SEMCOG sometimes holds constant the random seed and base year data in order to have a stable baseline to do model calibration.

The update of the SEMCOG base year database takes about 1 to 1 ½ years to complete. Most of this time is in processing and reviewing messy input data. SEMCOG estimates that the full process of rebuilding the entire base year database would take 1 ½ to 2 years.

SEMCOG does most of their data processing in-house. They try to use third party data when available. They recognized that some agencies do this differently, such as Puget Sound Regional Council which contracts out the majority of their data development then does some additional work in-house. SEMCOG has automated some of the process and more automation could be done. However even with automation, the data still requires a lot of manual checking and clean up. SEMCOG estimated that about 80-85% of their staff time is devoted to dealing with data, particularly in the beginning stages when the effort is almost entirely devoted to data. Model calibration and validation is also dependent on data being accurate and ready. Having experienced staff who are familiar with the region and the data and who know where to look when there are issues, can greatly reduce the amount of time needed to update the data (i.e. by half). SEMCOG also employs interns when additional help is needed. The SEMCOG land use forecasting group includes five to six data analysts and two modelers. It does not include any traffic/travel demand forecasting staff who work in a separate group. The SEMCOG land use forecasting group provides data to many parts of the agency. For instance, the building database is used for other planning purposes besides UrbanSim. SEMCOG also provides certain datasets to outside groups when the data is ready and not subject to confidentiality concerns.

MAG-Phoenix is required by mandate to provide a socioeconomic forecast three times a decade, for every year ending in 2, 5 and 8. MAG-Phoenix updates the UrbanSim base year for each of these forecast years. They update the entire base year from scratch each time but have the advantage of working with fewer counties.

It takes MAG-Phoenix about six months or less to update the base year database. MAG-Phoenix uses lots of Python code automation to do this, specifically utilizing the ArcPy and Pandas libraries. MAG-Phoenix has found that if they do the data processing in ArcMap in the beginning, it is difficult to remember the actions performed on the data. Using Python helps with organization and in keeping track of the edits. The Python code is organized in a series of four to five Jupyter notebooks that do partial data assembly and fill in certain data. MAG-Phoenix staff is primarily self-taught in using Python.

The MAG-Phoenix employment database is updated every year. One staff person is responsible to maintain and update the employment database as one of their primary tasks. The parcel data is being worked on constantly. Some key problematic databases are maintained by other MAG-Phoenix groups which develop and use the data for other purposes, thus reducing the data development burden on the land use forecasting staff. MAG-Phoenix also mentioned that their assessor data is generally pretty good. The data validation includes an extensive data checking and review process with local governments. MAG-Phoenix is under a constant data review process with its member agencies. The MAG-Phoenix land use forecasting group employs 14 analytic staff, including data analysts, and three modelers.

MRCOG uses land use data from the county assessors with a good amount of updating from third-party contractors including Costar, Metrostudy, ESRI, Infogroup as well as manual observation. The Costar data also helps MRCOG inform vacancy rates. MRCOG allocates base year employment to the zonal level based on Infogroup's data. Employment is then allocated back down to buildings based on building type and employment sector. MAG-Phoenix works within parcel based on building square feet.

SEMCOG, MAG, and MRCOG only model jobs not firms. MAG-Phoenix indicated that modeling firms would be a next step for them but they are not yet ready to take on this effort.

Panel Discussion: Creating and Managing Policy Layer Inputs

The agencies shared the following information:

Key Considerations: We maintain multiple versions of policy layers: general plans, vision land uses, and various scenarios that combine the two. Challenges lie with converting general plans into usable and consistent capacity values.

Our Current Solution: Our current capacities were created using brute force, i.e. lots of intern time, compiling and standardizing general plan data and converting to capacities. Vision land use is based off the general plans and augmented in areas cities identify for densification.

Strengths, Weaknesses, Opportunities, Threats Overview: Creating and maintaining multiple policy layers and keeping official versions separate from scenario layers is a challenge. Best data management practices. Incorporating diverse zoning data and converting it to capacity layers.

Discussion Question: Do you maintain multiple policy layers, official vs. scenario? How is your official policy layer created and is it officially adopted? How often is your policy layer updated? Best practices for streamlining data development?

SEMCOG doesn't use zoning because it changes quickly and doesn't necessarily represent future or envisioned land use. Instead, SEMCOG uses general plans. But general plans have their own set of challenges, specifically, over generalization and a lack of details (such as lack of nonresidential density values). Sometimes general plans are missing capacities altogether, meaning if these were read directly into UrbanSim the model could build to any height or have unrealistic densities. In SEMCOG's experience, UrbanSim has the tendency to go to maximum capacity allowed.

SEMCOG found that as they reviewed their forecasted data, there were instances that the land use "type" was correctly identified in the general plans but the densities were either too high or too low resulting in socioeconomic forecasts that were unrealistic. SEMCOG used a process to impute density constraints based on nearby densities based on the same building type. This was done as a preprocessing step. In the case of a new subdivision with nothing nearby, SEMCOG set the densities based on what they felt was reasonable for that area. This method doesn't account for future densification, however.

SEMCOG has built the policy layer for UrbanSim using data collected from local communities. SEMCOG has a staff person tasked with collecting the local governments' general plans. As part of their process, SEMCOG solicits general plan updates from the local governments during their forecast preparation.

MAG-Phoenix's process is similar to SEMCOG's. MAG-Phoenix doesn't use zoning and instead uses general plans. MAG-Phoenix has a redevelopment layer for known redevelopment events to supplement the policy layer. MAG-Phoenix has been collecting known redevelopment data from the communities for 20 years. This is done on an ongoing basis. The collection of the development data is added to the general plans layer to become the new the policy layer. The amount of known development or redevelopment is significant, representing approximately fifteen years of growth for the area. MAG-Phoenix hasn't yet used UrbanSim to test policy scenarios.

MAG-Phoenix has an annual review with the local governments to vet the inputs and outputs. Before MAG-Phoenix runs the model, they bring the input data to the local governments for verification. Later, MAG-Phoenix meets again with the local governments to review the outputs. MAG-Phoenix performs lots of up/down adjustments to the policy layer based on feedback received.

MAG-Phoenix has a staff person dedicated roughly 30% of their time to working with the local governments. This person is responsible, in part, for coding the known redevelopment events into the policy layer. MAG-Phoenix receives known redevelopment information in every form. PDF's are most common but sometimes information comes in as just a description on paper. MAG-Phoenix digitizes the new development into GIS.

MRCOG uses zoning for the policy layer development. They also collect known or scheduled development events from agencies and developers. They receive a large list and some have very large, unrealistic densities. If all the known development events were added, MRCOG would have more than enough capacity to accommodate 2040 household growth. MRCOG chose to put in known development events that have gone through enough process and which they are confident will happen.

MRCOG receives updates to the policy layer on an ongoing basis. MRCOG has a dedicated staff person responsible for this work. The staff person also has other responsibilities, though mostly related to this task. MRCOG receives data in every format, though PDF's are the most common, and will digitize the information into GIS. Sometimes the development event causes a modification to the parcel geometry.

MRCOG uses a trend scenario to determine "upzoning" in specific locations, such as around transit stations. Local governments provide feedback to help ascertain reasonableness. MRCOG also does a cost shift routine in certain geographies (such as around transit-oriented zones) to allow for higher profit margins in the pro forma model calculations. This is done to mimic incentives in those areas.

Panel Discussion: QA/QC of Base Year Data and Policy Layers

The agencies shared the following information:

> **Key Considerations:** Non-policy layer inputs from various sources, all with varying degrees of data quality.
>
> **Our Current Solution:** Ad hoc review of individual datasets. No structural review of data with other datasets.
>
> **Strengths, Weaknesses, Opportunities, Threats Overview:** Without a robust process of data review, we run the risk of missing issues on the various datasets.
>
> **Discussion Question:** How big is your QA/QC effort for data development?

Much of MAG-Phoenix's quality control work is done by others. Since MAG-Phoenix receives much of their data from other departments, these departments perform the primary quality checking of the data. Local governments provide a lot of additional quality checking. However, MAG-Phoenix shared that once the base year database is built, they immediately begin modeling. MAG-Phoenix acknowledges that they could benefit from a more robust quality assurance/quality control process.

SEMCOG performed passive checks of their data in previous forecasts. SEMCOG used a self-developed python program to check for data integrity. The program checked an SQL database when triggered by table updates and reported data violations. SEMCOG also performs quick statistical checks to see if things look correct (e.g. check if building square feet in the downtown area seems plausible). These checks look for outliers or to see if the data seems plausible. Often an error in an input dataset is discovered from analyzing the model's output after the model is run. In the latest 2045 forecast, SEMCOG tested UrbanSim data checker program but didn't end up using it due to an issue with the performance of the checker tool. SEMCOG currently has a python-based input program to help import model inputs from a GIS/SQL database as well as perform simple data verification tasks.

MRCOG feels that they are often lacking in the quality assurance/quality control area. Part of this is because they often don't create the data themselves but get it from third party vendors. MRCOG feels that they would benefit from more manual checks of the data.

The agencies asked this follow up question: if the panelists had another FTE to do data what would you put them on?

MRCOG responded that they would put another FTE on developing small level employment data. Employment changes so quickly, they could have a much more up to date database with another resource. One of MRCOG's biggest areas for potential inaccuracy in their data deals with deriving building square foot/job based on matching the geocoded infoUSA employment data with the buildings. MRCOG spends a lot of time trying to fix this.

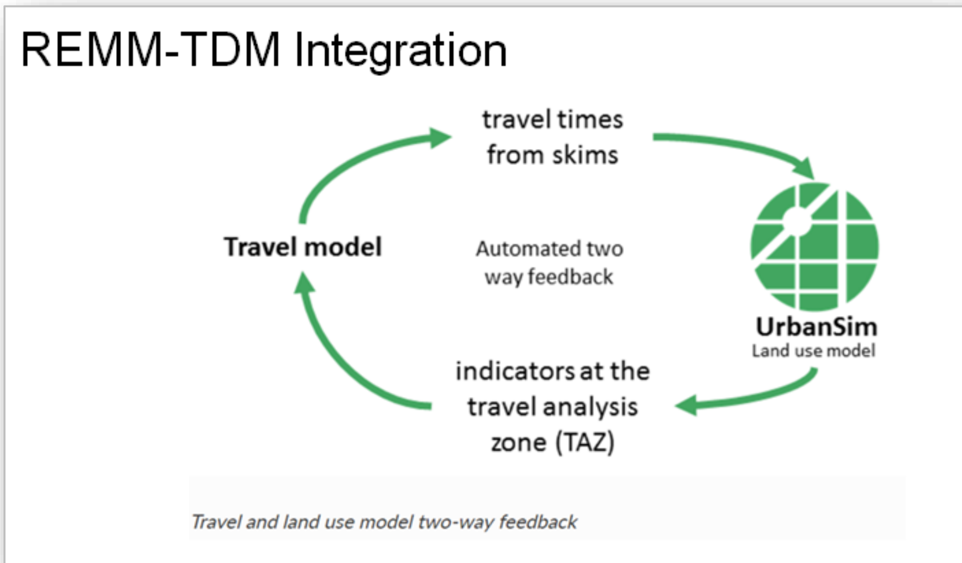
MAG-Phoenix echoed MRCOG's recommendation. The task of distributing base year employment data at the building level as an UrbanSim input is an onerous task. For this reason, MAG-Phoenix decided to farm it out to not burn out their staff. MAG-Phoenix also commented that the building square foot/job has biggest effect on the model than any other table.

SEMCOG would put another FTE resource over quality assurance/quality control and review of data. SEMCOG uses a lot of data developed by different people. When data is combined, there is always missing links or bad data. It can be very time consuming going back and forth running the model and looking at input data. SEMCOG focuses on improving overall data quality rather than maximizing the currency of the data.

SEMCOG also noted that working with multi-use buildings is complicated. SEMCOG has been working with the UrbanSim team to address this but recognizes that this is not an easy issue to solve. Coming up with a universal formula that works everywhere is hard.

Accessibility Sensitivity

The agencies reviewed REMM's integration with the travel demand model. REMM provides the travel demand model with updated TAZ-level socioeconomic data. These data are then used to produce trips which are then assigned to the transportation networks. Congested travel times and other measures are then supplied back to REMM. Accessibility variables are used in REMM's location choice and price modules.



The following accessibility variables from the travel demand model are used by REMM:

- **Commute_time** - Average Commute Time for the Commuters living in the TAZ
- **fwy_exit_dist** - Distance to the Nearest Freeway Exit Parcel
- **Rail_stn_dist** - Parcel's Distance to Nearest Train Station
- **parcel_volume** - Parcel Two-Way Traffic Volume Calculated from Arterial Volume Shapefile
- **train_stn_dist** - Distance to Nearest Transit Station Parcel
- **logsum_jobs** - Logsum Value to Jobs
- **logsum_hhincx** - Logsum Value to Households (x=Income Quartile)
- **logsumpopulation** - Logsum Value to Population
- **population_within_20_min** - Population Within 20 Minutes by Automobile
- **jobs_within_20_min_transit** - Jobs within 20 Minutes by Transit
- **Jobs_5_within_10_min** - Manufacturing Jobs within 10 Minutes by Automobile

The observed data showed a non-linear relationship of a household's desire to locate near a freeway exit. Households of all income groups like to be close but not too close to interchanges. The agencies plan to use this information to refine estimation.

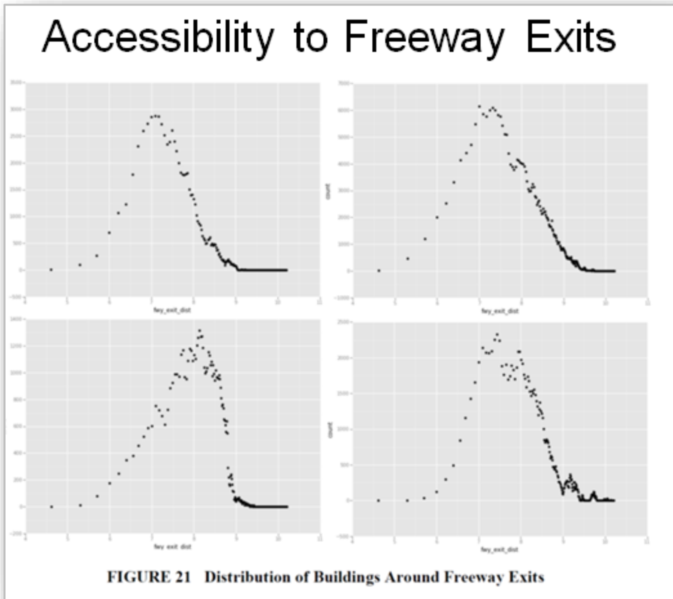
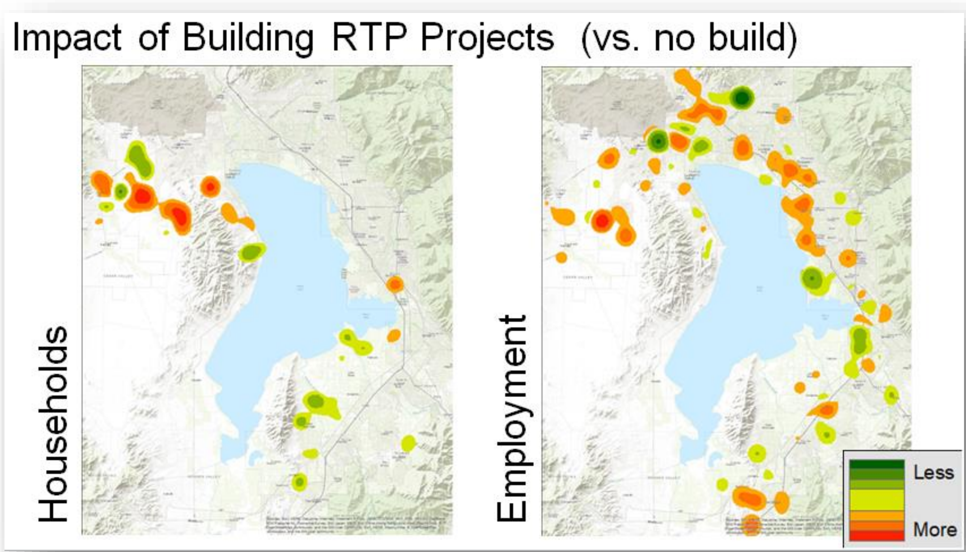


FIGURE 21 Distribution of Buildings Around Freeway Exits

To test REMM's sensitivity to transportation supply, REMM was given two sets of congested travel time skims: one based on building the Regional Transportation Plan projects and one representing a no-build condition. The no-build scenario was much more congested and REMM responded by increasing the distribution of households and employment to locations with more accessibility.



Panel Discussion: Incorporating Appropriate Accessibility Sensitivity into REMM

The agencies shared the following information:

> **Key Considerations:** The purpose of the REMM exercise is to facilitate transportation planning and infrastructure selection. REMM is designed to integrate with the Travel Demand Model. REMM needs to be appropriately sensitive to the travel model.
>
> **Our Current Solution:** The integrated modeling system produces many transportation related variables. OLR (for price) and multinomial logit were used to determine the statistical significance of the price and location choice variables, respectively. If no accessibility variable was significant, the commute time was asserted.
>
> **Strengths, Weaknesses, Opportunities, Threats Overview:** REMM sensitivity to accessibility is based on sound statistical practice. However, for some sub-modules, no accessibility variable produced a significant relationship with price and/or location choice.
>
> **Discussion Question:** Have you encountered this issue? How are your travel demand model and your UrbanSim integrated? How much sensitivity of land use to transportation accessibility is appropriate? Any other variables that we should be testing?

MAG-Phoenix found they got better results with the travel time variables when dummy variables were used rather than straight line variables. MRCOG and MAG-Phoenix found that jobs within 20 minutes produced good results and is correlated to other things. MRCOG likes to start with jobs within 20 minutes and then go from there.

In testing UrbanSim's response to travel accessibility, MRCOG found that the model's allocation tended to stay around the core with a more congested travel time skim and spread out from the core with a less congested skim. The allocation was pretty sensitive to transportation. MRCOG also found that using peak congested times produced good results. MAG-Phoenix and SEMCOG have not yet done this type of analysis.

MRCOG asked how often the travel model is run. After the 2015 base year which may be run before REMM begins, REMM runs a modified version of the travel model four times:

- 2019 – base year of the transportation plan
- 2027 – midway between the TIP (2024) and phase 1 (2030)
- 2035 – midway between phase 1 (2030) and phase 2 (2040)
- 2045 – midway between phase 2 (2040) and phase 3 (2050)

In each case the later-year network is used to allow REMM sufficient time to react to the upcoming transportation investment. REMM's interaction with the travel model is automated so there is no manual moving files or kicking off a new travel model run. The travel model also introduces the greatest cost in model runtime.

SEMCOG does not use future travel times in their land use forecast but plans to add this function later. The SEMCOG transportation network is pretty well established and new projects are mainly widening projects. Because the region is growing slowly, the future build scenarios are usually not too different from the current in terms of travel times.

MRCOG's UrbanSim model receives feedback from the travel model twice: once at the halfway point, and then right before last year. They are still trying to figure out how often to run the travel model but estimate about every 5-10 years would be good. MRCOG would also like to investigate how to automate changing out the transportation networks as the current process is labor intensive.

MAG-Phoenix indicated they also run the travel model several times (every 3-5 years). They plan to employ Paul Waddell and MAG-Phoenix's transportation group to figure out how to make this better.

SEMCOG and MRCOG have found that regional and local accessibility matters depending on which UrbanSim module is being run. SEMCOG and MRCOG use a local street network to determine local accessibility or the number of different types of jobs or activities within a certain proximity to the parcel.

SEMCOG uses Pandana network analysis software and Michigan Geographic Framework (MGF) roads to determine the walk accessibility variable. They use a range of ¼ to ½ mile for this analysis. This calculation is being done inside UrbanSim at the parcel level. The process uses a Pandana network graph converted from state GIS data, which includes trails and roads beyond the SEMCOG region boundaries. SEMCOG also uses UrbanSim's 'Urban Access' module to calculate transit accessibility, which utilizes Pandana and General Transit Feed Specification (GTFS) data. SEMCOG acknowledged that many of the variables needed to create local accessibility may not exist for a future build scenario, such as the local network for a brand-new subdivision. SEMCOG bases the future calculations for these variables off of the base network.

MRCOG uses Pandana and local street data to create walk accessibilities. For example, MRCOG is using walk accessibility to food service jobs as a variable in some of the location choice models.

MAG-Phoenix does not use Pandana but would like to. They ran into time constraints to get this functionality into their current model.

Developer Model

The agencies reported that the developer model dominates the modeling process in REMM. The agencies sought specific feedback from the panelists on the following topics:

- How to model development and redevelopment
- Mixed land use development
- Incremental development
- "Leapfrog" development
- Over-building by developer model
- Monte Carlo process
- Spatial unit for the model

Panel Discussion: How to Model Development and Redevelopment

The agencies reviewed the primary developer tasks:

1. Macroeconomic analysis
2. Evaluate capital markets
3. Local supply and demand analysis
4. Evaluate zoning/planning
5. Evaluate local politics
6. Estimate rents
7. Estimate costs
8. Identify land opportunities
9. Evaluate land control options
10. Back of the envelop pro formas
11. Evaluate investment thresholds
12. Evaluate organization strategy
13. Estimate project timeline
14. Estimate project scope
15. Evaluate programmatic options
16. Financial underwriting – 1st draft

Pro-Forma Summary :
Hypothetical Residential Subdivision

Project Revenues	
Number of Units	50
Average Sale Per Unit	$ 400,000
Gross Sales	$ 20,000,000
Less Commissions, Fees	- $ 800,000
Net Project Revenues	$ 19,200,000
Project Costs	
Land Acquisition	$ 2,575,000
Planning, Design & Approvals	$ 600,000
Sitework & Building Construction	$ 12,175,000
Amenities, Off-Site Costs	$ 100,000
Management & Overhead	$ 1,760,500
Total Project Costs	$ 17,210,500
Net Cash Flow Before Financing	$ 1,989,500
Financing Interest	$ 1,102,400
Net Cash Flow to Developer	$ 887,100
Cash Investment	$ 1,020,600
Total Cash-On-Cash Return	86.9 %
Annualized Cash-On-Cash Return	19.9 %
Internal Rate of Return	22.4 %

The agencies shared the following information:

Key Considerations: The developer model has a strong impact on REMM outputs. Different sub-areas have different development and redevelopment patterns due to existing land use, land use plans, political resistance, etc. How should these be addressed?

Our Current Solution: The developer model evaluates expected prices by real estate development type and costs to construct each type in each location using a pro forma computation. This process includes some constraints (e.g. redevelopment friction factors) to simulate political resistance, building condition and other development realities.

Strengths, Weaknesses, Opportunities, Threats Overview: Current method is based on historical data and is guided by a local general plan policy layer that constrains development type and quantity.

Discussion Question: How are you guiding growth in your agency's UrbanSim model? What calibration tools are you using to encourage or discourage development/redevelopment? Which level of friction values have you found to be reasonable?

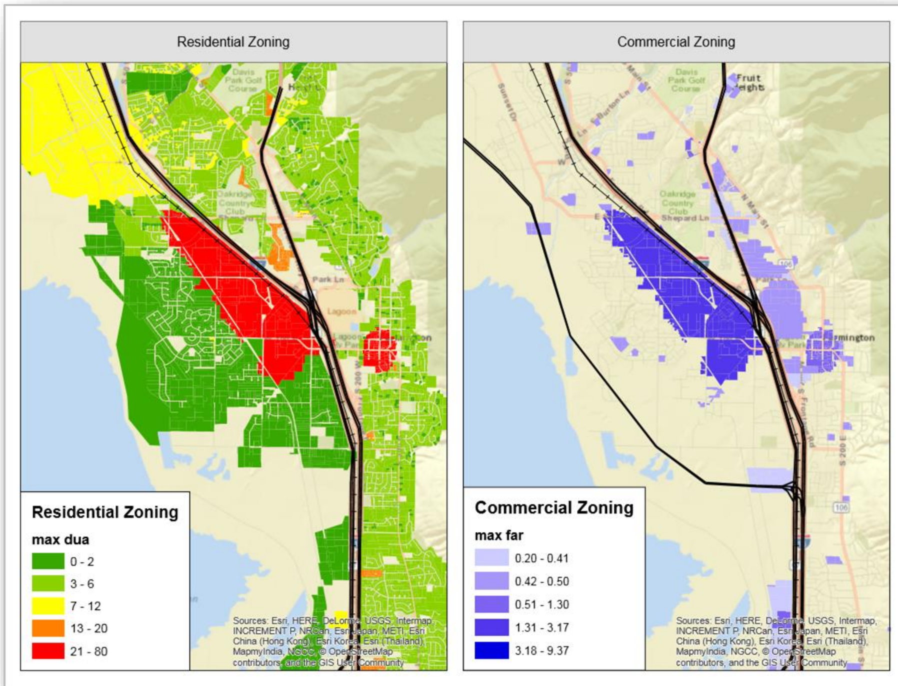
MRCOG suggested that using zoning on the policy layer is a good way to discourage development. MRCOG shared the example where they dealt with a subdivision that had covenants preventing any more density. They created a new zoning category that had restricted development for this subdivision and then documented why the zoning was changed. MRCOG also shared an example of how they use the YAML file plus a geographic cost shifter to guide development. In their model, the YAML file has a general minimum parking requirement. Since their downtown has no parking requirement, MRCOG applies a cost shifter in the downtown. The result is the downtown gets a bit of a discount for future development since they don't have to build parking.

MAG-Phoenix uses an in-house developed real estate model and doesn't use UrbanSim's pro forma model. They also shared the experience of Puget Sound Regional Council (PSRC) who struggled with their pro forma model. PSRC ended up limiting outputs to the top 20 most profitable projects the development model would select from. MRCOG also uses the policy layer to guide development. For example, they reset the capacities of historic districts to zero to prevent development. However, they try not to change capacities too much. MRCOG noted that transitioning between the base and first modeled year has been problematic. To smooth this out, MAG-Phoenix uses building permits to seed development to get the trajectory right. MAG-Phoenix focuses on getting first 5-10 years right, then transitions to a longer-term trajectory model.

SEMCOG used to have a land-value-to-building-improvement-value ratio as a switch to determine redevelopment. However, SEMCOG was unable to get reliable data in their latest round of modeling and was not able to use this ratio. SEMCOG also saw high value single family housing turn to multifamily housing. To overcome these issues, SEMCOG applied automatic and manual filters to remove buildings from redevelopment consideration. Each community in the SEMCOG model gets a certain amount of development share based on a historical trend. SEMCOG reported they had problems using the current price in the pro forma model, so they tried to increase future real estate prices based on REMI forecasted income growth to make certain properties become profitable.

Panel Discussion: Mixed Land Use Development

The agencies shared their challenges working with land designated as mixed use. The agencies code both residential and employment capacities in same location for mixed use land designations in the policy layer. The pro forma model determines which to develop first. In reality, residential and non-residential development could be built simultaneously.



The agencies shared the following information:

Key Considerations: Mixed land use is a complicated land use type to accommodate in UrbanSim. Mixed use is very important for CBD and TOD areas, among others. Local agencies have expressed interest in this land use designation.

Our Current Solution: The non-residential and residential capacities were established for mixed use areas using floor area ratio (FAR) and dwelling units per acres respectively. The pro forma model decides which type of building type is going to be built and how much to build.

Strengths, Weaknesses, Opportunities, Threats Overview: The sequential development conforms with the model process, but in reality, mixed land use is generally built simultaneously. Currently most land zoned “mixed use” in our region develops as multifamily housing since that is in higher demand than commercial or retail.

Discussion Question: Do you include mixed use as one of your land use types (why or why not)? How does your model handle the mixed land use type? Have you found mixed use to be functional/helpful in your model? What guidance do you have for including it and/or improving our practice?

SEMCOG indicated that modeling mixed use became an issue in UrbanSim version 2 because the development choice is sequential (this problem did not exist in UrbanSim version 1). SEMCOG put a request into Paul Waddell to address this issue in the next round of UrbanSim development. They also suggested working with other MPOs to try to get the software developer to fix this issue.

MRCOG set up their UrbanSim model by tacking on a small amount of retail where there are mixed-use areas. Usually retail rents are high so developer model adds a little retail to mixed use developments. MRCOG doesn't have a lot of mixed-use opportunities in their land use policy layer, so this hasn't really been an issue for them.

Panel Discussion: Incremental Development

There are many big parcels in the REMM model. Some of these big parcels are located in areas where intense development is expected, such as the northern part of Utah County west of Utah Lake.

The agencies shared the following information:

Key Considerations: Some of the developable land in our region is currently very big agriculture parcels whose future land use is zoned for residential. In reality, these big parcels are subdivided into smaller parcels and built gradually across time. UrbanSim wants to build each large agricultural parcel as a single parcel in one year.

Our Current Solution: We used ArcGIS to split the big parcels into gridded smaller ones (approximately 10 acres). Then the developer module develops the larger parcel one grid cell at a time.

Strengths, Weaknesses, Opportunities, Threats Overview: The strength of the current method is easy implementation and seems to produce reasonable results. The weakness is trying to understand if this simplification is sufficient to encapsulate the complexities of development in these areas.

Discussion Question: How are you dealing with large parcels in your models? Is there a better way to deal with these big parcels than our current method?

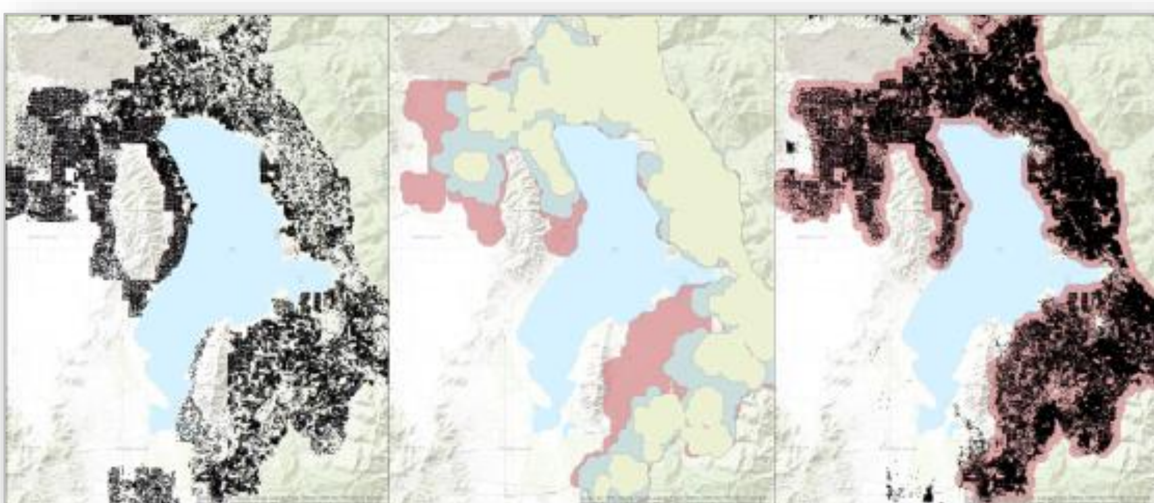
MAG-Phoenix has dealt with the same problem by cutting big parcels into smaller ones. MAG-Phoenix divides their big parcels along section lines which are approximately a square mile (640 acres) in area. MAG-Phoenix acknowledges that the divided parcels are still pretty large and having a large parcel develop in a single year is unrealistic. MAG-Phoenix uses velocity curves applied to single family subdivisions to reduce the amount of development that can be built in any given year. The velocity curve is based on the size of the development. The largest developments can take up to six years or so to build out. MAG-Phoenix would like to automate the parcel splitting, but this is a long-term goal and the current parcels serve pretty well.

SEMCOG doesn't really have a lot of large developable parcels. Larger vacant parcels have been filtered out from development or are parks.

MRCOG has had to deal with this issue a lot in their area. MRCOG divides large parcels first based on master plan land uses. If no master plans exist, MRCOG uses a 10-acre grid. Water is also a constraint in developing these larger parcels, which is accounted for in the zoning layer. MRCOG has also had to deal with the opposite issue where antiquated platting required aggregating some parcels. MRCOG doesn't currently get too worried over annual, parcel-level results. They compare 2040 model results against the base year at the TAZ level, which is also what they give to the public. The travel model uses TAZ-level inputs for 5-year increments.

Panel Discussion: "Leapfrog" Development

The agencies observed that UrbanSim, as a non-spatial model, builds miles away from current development. They wanted to constrain the model to prevent leap frog developments miles from existing areas. The agencies apply a ½ mile GIS buffer every three years around existing development to restrict growth.



The agencies shared the following information:

> **Key Considerations:** Leapfrog development (e.g. in south Utah County), over eager redevelopment (everywhere), and non-redevelopable land uses (primarily in downtown SLC).
>
> **Our Current Solution:** 1) Temporally staged land releases for development in south Utah County; 2) 20-year-old requirement for redevelopment consideration and calibrated redevelopment friction factor; 3) created a friction factor to disallow all future redevelopment for some land uses/areas.
>
> **Strengths, Weaknesses, Opportunities, Threats Overview:** This method is working but external validation would be valuable for confirmation/confidence in this approach.
>
> **Discussion Question:** Do these approaches seem reasonable? Have you encountered these same issues? If so, how did your agency address them? What other issues have you encountered that have required similar adjustments?

SEMCOG experienced a similar issue where they saw leap frog development in the remote areas of the region in the later years of the model run. SEMCOG also buffers around utilities (sewer/water). The buffer is extended in future time periods.

MAG-Phoenix commented that the agencies' approach seemed reasonable. MAG-Phoenix is not seeing leap frog development occurring in their modeled area. This might be due to MAG-Phoenix's accessibility variables that keep it from going too far too quickly. MAG-Phoenix has a lot of accessibility variables in each model and are unsure what else might be keeping the model in check.

In a previous plan, MRCOG dealt with this issue by making a boundary layer from water/sewer line data received from utility companies. They also used a forecast of where utilities would go in the future. This effort was similar to the agencies' approach but less automated. For the current planning effort, they did not see the leap frog issue and feel the accessibility variables they were using kept development pointed toward downtown.

Panel Discussion: Over-Building by Developer Model

The agencies shared the following information:

> **Key Considerations:** In the base year, nearly all parcels are profitable for development for nearly all development types. The unconstrained developer model over-builds all development types in all four counties. How should we rationally constrain the developer model?
>
> **Our Current Solution:** Currently we are using a sub-county vacancy rate to constrain development behavior. The developer will not currently build beyond a 7% (residential) and 10% (non-residential) expected vacancy rate in any area.
>
> **Strengths, Weaknesses, Opportunities, Threats Overview:** This method is currently working fairly well and constraining developer behavior to produce reasonable build quantities and sequencing. Perhaps this doesn't allow overbuild/underbuild cycling in the real estate market. We are looking for feedback on the reasonableness of this approach. By potentially over

constraining development in a path dependent model like REMM, we may be artificially affecting model outcomes.

Discussion Question: Does this approach seem reasonable? How are your agencies constraining the developer model in UrbanSim?

SEMCOG commented that UrbanSim version 2 runs the location choice model first then the developer model. UrbanSim version 1 was the opposite.

MRCOG also uses vacancy rate to curb the model from over developing. MRCOG uses a 6% vacancy rate for residential and a 9% vacancy rate for non-residential. Vacancy rates are applied at the TAZ level. MRCOG also includes the area's vacancy rate in the price model. In the price model, as vacancy rates go up prices go down.

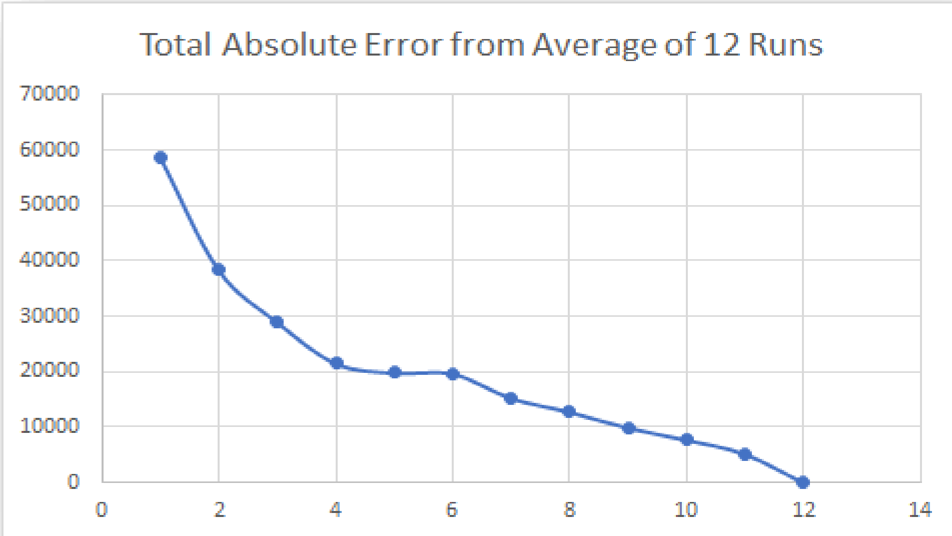
MAG-Phoenix commented that they felt the agencies were using a good approach. MAG-Phoenix uses a process that looks like the agencies' process but in a looser way.

Panel Discussion: Monte Carlo Process

The agencies indicated that they average the results of six REMM runs to minimize the impact of the randomness from the Monte Carlo process.

;TAZID	RUN1	RUN2	RUN3	RUN4	RUN5	RUN6	RUN7	RUN8	RUN9	RUN10	RUN11	RUN12
141	461	491	466	478	488	478	505	420	462	446	449	439
142	185	218	186	199	209	199	222	206	198	198	184	243
143	143	120	140	132	135	132	136	138	148	147	155	154
144	66	70	56	73	65	73	58	67	56	67	62	67
145	537	499	417	538	581	538	481	507	496	508	487	508
146	287	255	276	318	298	318	277	304	228	314	291	291
147	408	519	508	409	412	409	459	490	497	429	445	423
148	492	537	458	496	513	496	494	502	519	543	574	572
149	679	661	615	637	627	637	640	607	703	591	684	602
150	594	593	615	623	592	623	575	599	598	597	599	607
151	381	325	313	332	337	332	342	337	378	329	360	358
152	680	729	716	802	779	802	831	769	787	769	809	734
153	48	50	53	47	36	47	42	48	59	48	47	55
154	334	315	292	318	276	318	319	327	340	303	301	283
155	233	219	192	259	272	259	243	218	210	217	226	216
156	140	149	151	150	160	150	143	144	136	152	156	154
157	94	97	94	98	93	98	81	95	100	98	99	88
158	478	451	510	439	501	439	474	479	450	500	492	445
159	881	862	846	879	874	879	802	817	852	870	903	850
160	257	306	327	313	314	313	331	255	261	281	317	250
161	443	554	512	485	483	485	487	445	465	534	493	566
162	814	810	816	813	798	813	845	833	803	829	819	813

The agencies found that between four to six runs the total absolute error of the results starts to level off. Averaging additional runs beyond six continues to reduce the error but by a more marginal amount.



The agencies shared the following information:

> **Key Considerations:** Random Utility Theory is the basis of UrbanSim. To facilitate this, Monte Carlo simulation with a random seed is used to generate a discrete choice of location within the model. As a result, every model run may produce a different outcome.
>
> **Our Current Solution:** Our current practice is to average six model runs for every scenario or build alternative. Even with this approach, REMM could, theoretically, produce conflicting outcomes in two unique six model run averages.
>
> **Strengths, Weaknesses, Opportunities, Threats Overview:** The strength of this approach is that it accommodates the randomness in actual urban futures. It is also robust for scenario and infrastructure alternatives analysis. The reality is not determined. However, this process can be time consuming and challenging when a single official forecast is desired.
>
> **Discussion Question:** How do you simulate the choice model and how do you present the results? Is a six-run average sufficient? How have your agencies dealt with this issue?

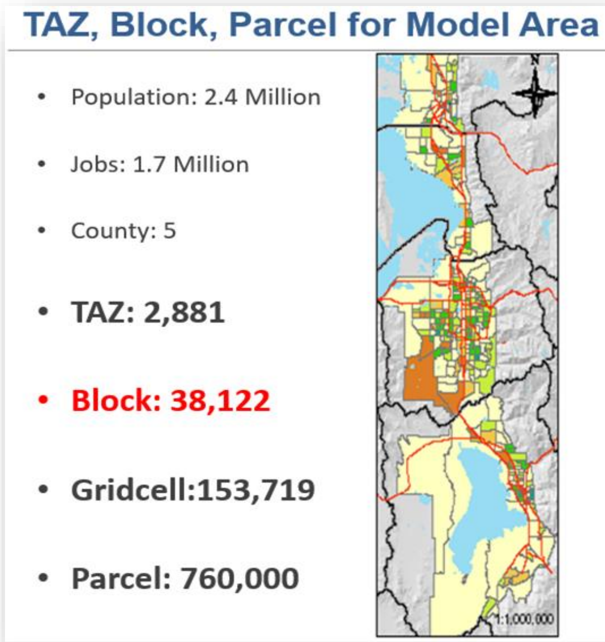
SEMCOG asked how the agencies perform the averaging and how they maintain cross-tabulated data and control totals. The agencies do not average the cross-tabulated results. Averaging is done at the TAZ level. No averaging is performed as REMM runs interactively with the travel model; only total results are averaged after the REMM run is complete.

MAG-Phoenix agreed that the agencies' approach was a reasonable way to address uncertainty, if they are not averaging the micro data. Having the sub-level datasets no longer match would be a concern if the agencies wanted to use that level of data.

SEMCOG said they sometimes fix the random seed to avoid randomness issues in calibration.

Panel Discussion: Spatial Unit for the Model

The initial implementation of UrbanSim in Utah used grid cells. REMM currently uses parcels as the primary input structure.



The agencies shared the following information:

Key Considerations: UrbanSim can be run at parcel level, block level, or zone level. Understanding the optimum level of spatial resolution for runtime (not critical), input preparation, and output interpretation are important considerations.

Our Current Solution: We did a survey of many MPO's land use models. MPOs often process data at the parcel level for inputs. Outputs are typically aggregated to larger geographies for interpretation. This is our practice with REMM--run at the parcel level, outputs presented at the TAZ level.

Strengths, Weaknesses, Opportunities, Threats Overview: REMM is estimated on parcel level data and the base year data are at the parcel level as well. The strength of this approach is that the outputs are easily aggregated to multiple levels of geography for presentation to various audiences. Sometimes we have requests for parcel level outputs but parcel level outputs do not always make sense.

Discussion Question: Which level model do you use and why? What do you see are the advantages to using different levels of geography as inputs and outputs? What weaknesses do you see in our current approach? What recommendations do you have for us to improve this process?

In 2013 MAG-Phoenix experimented with “super parcels” which was their own micro geography. For example, a subdivision would be a super parcel. Super parcels worked well, but MAG-Phoenix went back to parcels for their current modeling efforts. MAG-Phoenix felt taking the extra step to create super parcels was getting further from parcel-level decisions and was a step they didn’t need to take. They also felt that stakeholders have an easier time understanding parcels. MAG-Phoenix also uses different geographies for displaying outputs, such as grids or hexagons.

SEMCOG thought parcels were a more natural way to simulate the real world. The logic of project development happens at parcel level (e.g. pro forma calculation). When data is aggregated, some details are lost. However, credibility challenges could arise when showing parcel level data. That said, SEMCOG agreed with MAG-Phoenix that parcels were probably the way to go. SEMCOG acknowledged that processing data at the parcel level is a big challenge, although super parcels would be a good way to keep some land from developing.

MRCOG uses parcel level data, but they don’t release anything below the TAZ level. The decision to use parcels was an inherited decision. However, they have found that parcels allow for informal conversations. For example, in reviewing pro forma data with local governments, one can point to certain parcels and talk about why these are not developing.

Model Calibration and Validation

The agencies sought specific feedback from the panelists on the following topics:

- Price and location choice estimation variables
- Price and location choice model estimations
- REMM sub-model parameter calibration

Panel Discussion: Price and Location Choice Estimation Variables

The agencies noted that choice models boil down to variables and parameters. The agencies had to do a lot of analysis because many of the variables are correlated. Correlated variables are treated as clusters. The agencies reported that they chose variables that could be used to test policy but still provide good estimation results.

The agencies shared the following information:

> **Key Considerations:** There is not a well-defined or standardized set of independent variables for price and location choice models.
>
> **Our Current Solution:** The REMM team developed 175 independent variables for consideration in the price and location choice models. (A handout was shared with the group which can be found in the Appendix under 'Handout 1 – Variables'.)
>
> **Strengths, Weaknesses, Opportunities, Threats Overview:** There is little documentation for the variables other agencies have considered and developed in their UrbanSim implementations.
>
> **Discussion Question:** How does our variable list compare with yours? Are there variables that we are missing? What variables has your agency developed that worked well in your models?

The agencies asked if the panelists would be willing to share their list of variables. MAG-Phoenix sent their variables to the agencies. SEMCOG's variables are available online.

SEMCOG used to go through a very manual process to find the best variables and combination. More recently, SEMCOG and their UrbanSim consultant used Autospec, a cloud-based location choice estimator, to estimate the models. Because the estimation process sometimes gives implausible variables, SEMCOG and their consultant applied logic to limit variables and constrain their coefficient directions/signs.

MRCOG stated their variables came largely from a literature search, but some from historical context and some from knowledge from previous work. Mapping helped to look for variables of interest.

MRCOG and MAG-Phoenix noted the lack of interaction variables in the variables in the handout, particularly in the way of household demographics. For example, the agencies' variables would not allow for younger households to be attracted to other younger households or low car ownership households.

MAG-Phoenix and SEMCOG noted the lack of local accessibility variables and commented that Pandana could be used to provide data for these variables but also acknowledged the difficulty in forecasting some of these variables.

SEMCOG, MAG-Phoenix, and MRCOG commented that they use the same variables for transit accessibility as the local agencies.

SEMCOG commented that looking at the variables others are using is a good start but cautioned that variable transferability from one region to another can be a challenge. SEMCOG's experience showed that counties don't grow in the same manner. It requires quite a lot of work and is a challenge to get the right set of variables to represent each situation. They also noted that often variables are a proxy for history.

Panel Discussion: Price and Location Choice Model Estimation

The agencies shared the following information:

> **Key Considerations:** Little documentation exists for UrbanSim sub-model estimations. We are curious how our estimation compares with others in this field.
>
> **Our Current Solution:** Our current model estimation results are presented in the handout. (A handout was shared with the group which can be found in the Appendix under 'Handout 2 – Modeling Tables'.)
>
> **Strengths, Weaknesses, Opportunities, Threats Overview:** Our current sub-models seem to be performing reasonably but could benefit from peer evaluation.
>
> **Discussion Question:** How do our estimations (variables/specifications, parameters/coefficients, and goodness of fit measures) compare with yours? Do ours seem reasonable? Are there methods you have used to get satisfactory results?

SEMCOG asked who the choosers are in the agencies' location choice model. Do the agencies sample from the whole universe? The agencies reported that the estimation dataset largely comes from the household travel survey. REMM has four household income categories across four counties.

SEMCOG has four household classifications as well in their current model but will be moving to 18 in the new model: three income, two household size, and three age of household head. Control totals for the 18 categories come from REMI. Their UrbanSim model does not have aging cohorts. SEMCOG uses REMI population forecasts and a base year synthesized population to get household control details.

MAG-Phoenix uses three household categories in their UrbanSim model: high, medium and low income. MAG-Phoenix plans to duplicate SEMCOG in creating more categories in the next round.

MRCOG uses four household categories in their UrbanSim model based on income quartiles.

MAG-Phoenix thought it was curious that the volume variable coefficient was negative for some counties and positive for others. They suggested looking into a geographic weighting to deal with this issue. The agencies commented they use to have a geographic weighting variable in their models, but have since removed this.

Panel Discussion: REMM Sub-Model Parameter Calibration

The agencies shared the following information:

> **Key Considerations:** Is there wisdom in adjusting model parameters/coefficients in a calibration exercise?
>
> **Our Current Solution:** To date, no sub-model parameter/coefficient calibration adjustments have been performed.
>
> **Strengths, Weaknesses, Opportunities, Threats Overview:** The strength of this approach is that model results are based entirely on statistically derived model parameters/coefficients. The weakness of this approach is that deficiencies in the estimation dataset can be problematic. Transportation accessibility, for example, doesn't estimate well in location choice or price models yet the link between land use and transportation is a major purpose for the model.
>
> **Discussion Question:** What (if any) sub-model parameter-based calibration adjustments have you found useful? What recommendations do you have for calibrating UrbanSim? Do you advocate calibrating coefficients based on judgment or purely statistical?

SEMCOG does use judgment when calibrating coefficients. MAG-Phoenix bases their calibration more on statistics, but has talked about making more adjustments based on expert judgment as well. MRCOG keeps primarily with statistical estimation but adds extra coefficients at the sub-county level. They then calibrate for specific geographies based on statistics. They have been happy with their estimation which didn't create any variables or results they didn't expect.

The agencies explained that they have run into limitations with their datasets. MAG-Phoenix suggested that it is totally reasonable to change variables or parameters to get reasonable results if there are limitations in the estimation dataset.

SEMCOG uses synthesized data from the PUMS database to estimate their model. They did not use the household survey because the most recent household survey did not include a recent mover question and it also missed some variables from PUMS. MAG-Phoenix uses synthesized data in their model estimation, partly due to the frequency of their household survey. MRCOG also uses synthesized data in their model estimation.

The agencies asked the panelists at what point do they find holding strictly to statistical estimation impacts model sensitivity. SEMCOG's priority of the base line estimation is to get a base year forecast to work. They recognize they have some artifacts that could affect scenario testing. SEMCOG's goal is to have a better model that can produce reasonable forecasts first, then to carry out sensitivity tests. There is some concern that this approach may seem a little like "making the model work", but they rely heavily on input/output processing to validate the results.

Demographic Transition

Panel Discussion: Household Synthesis for Every Year

The agencies shared the following information:

> **Key Considerations:** POPGEN was used to generate base year household list of the region. But every simulation year has many new households.
>
> **Our Current Solution:** Randomly select from the previous year households as the new households.
>
> **Strengths, Weaknesses, Opportunities, Threats Overview:** In the county, the different attributes of households have the same distribution as before. But as GPI forecasted, household size may become smaller.
>
> **Discussion Question:** How do you generate new households in your model? How does it perform?

SEMCOG uses REMI population forecasts to generate control totals (e.g. household size, age, race) for every year. REMI handles the demographic transitions at the control total level. SEMCOG uses UrbanSim's population synthesis (UrbanSim's population synthesis is based on POPGEN). The synthetic population must meet the control total constraints.

MAG-Phoenix receives their control total data from the state demographer. They use UrbanSim's population synthesis. MAG-Phoenix samples from the existing population and compares this to control total. MAG-Phoenix runs the population synthesis each time the base year is updated. They are contemplating synthesizing the population for every year, but this was determined to be too big of a lift to take on at the moment. MAG-Phoenix also has an idea where their domestic migration comes from and have considered loading up the national PUMS data and sampling by immigration proportion. However, MAG-Phoenix is concerned that incomes would vary significantly from different locations across the nation which could impact the dataset if they were adjusted. Also, there is concern if they were to do this that one could say they are duplicating efforts being done by other state agencies.

The agencies also asked the panelists if they use a demographic transition model to age the population by cohorts in the sub-regional demographic allocation. None of the panelist agencies are performing this type of aging of the population.

Outputs and Visualization

The agencies sought specific feedback from the panelists on the following topics:

- Summary statistics, and visuals used for model output
- External review

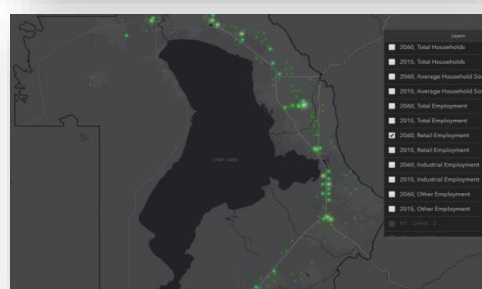
Panel Discussion: Summary Statistics and Visuals Used for Model Output

The agencies reviewed the different levels of REMM output:

- Building Table
- Summary Geography
  - Parcels
  - TAZ
  - Small, Medium, Large Districts
  - ‘City Area’ (TAZ-based aggregations to approximate city boundaries)
- Temporal
  - Phasing Thresholds (10 year)
  - All years

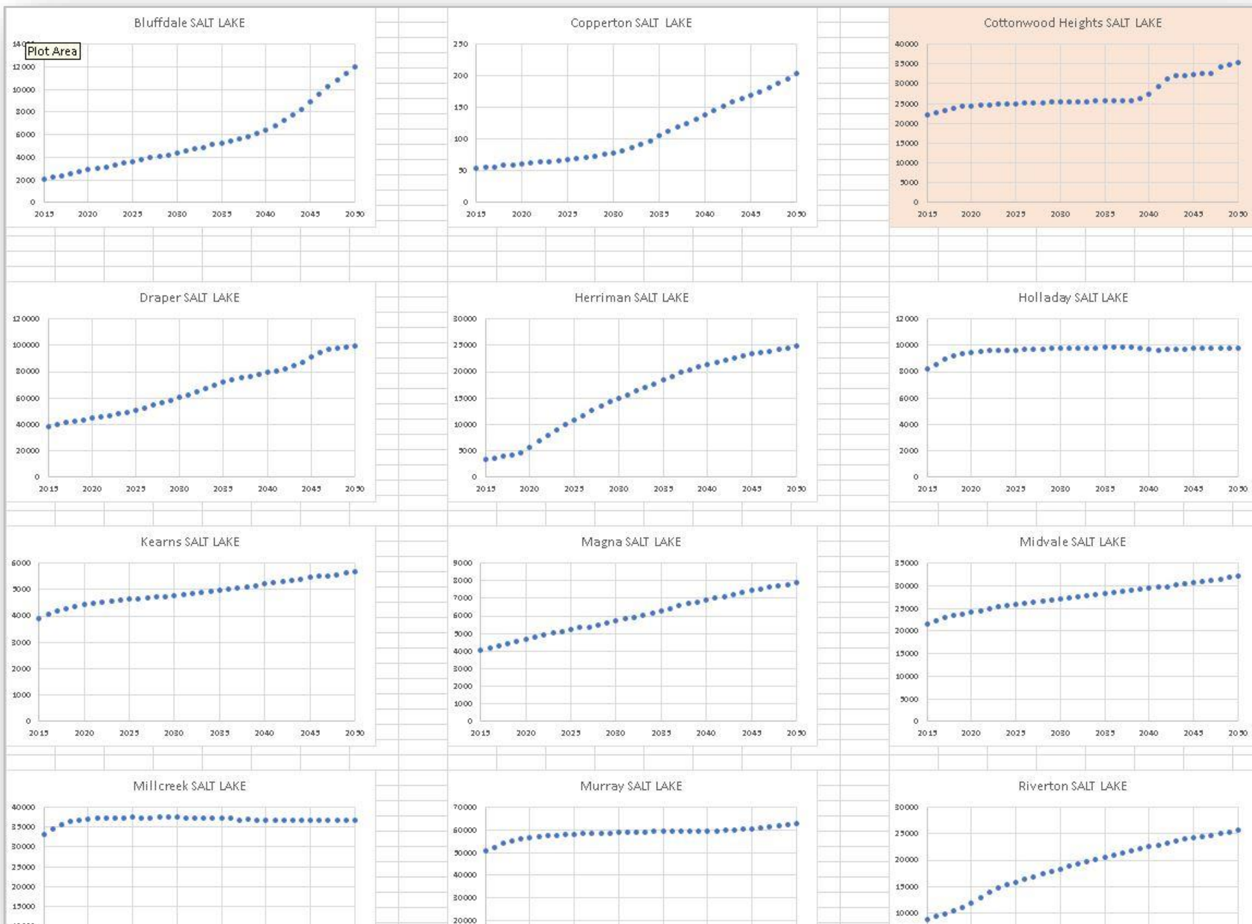
The agencies also reviewed basic validation tests performed on the REMM outputs, many of which are still “one-off”, manual efforts and not yet automated:

- Counts & Growth Charts
  - Population, Household & Job Counts
- Maps
  - Heat Maps of Building Types
  - 3D Distributions
  - Density Comparison
  - Utility Constraint Boundaries
  - Redevelopment Activity
- Ratios
  - Jobs to Household Ratio
  - Retail Jobs to All Jobs Ratio
  - Activity per Developable Acre Ratio

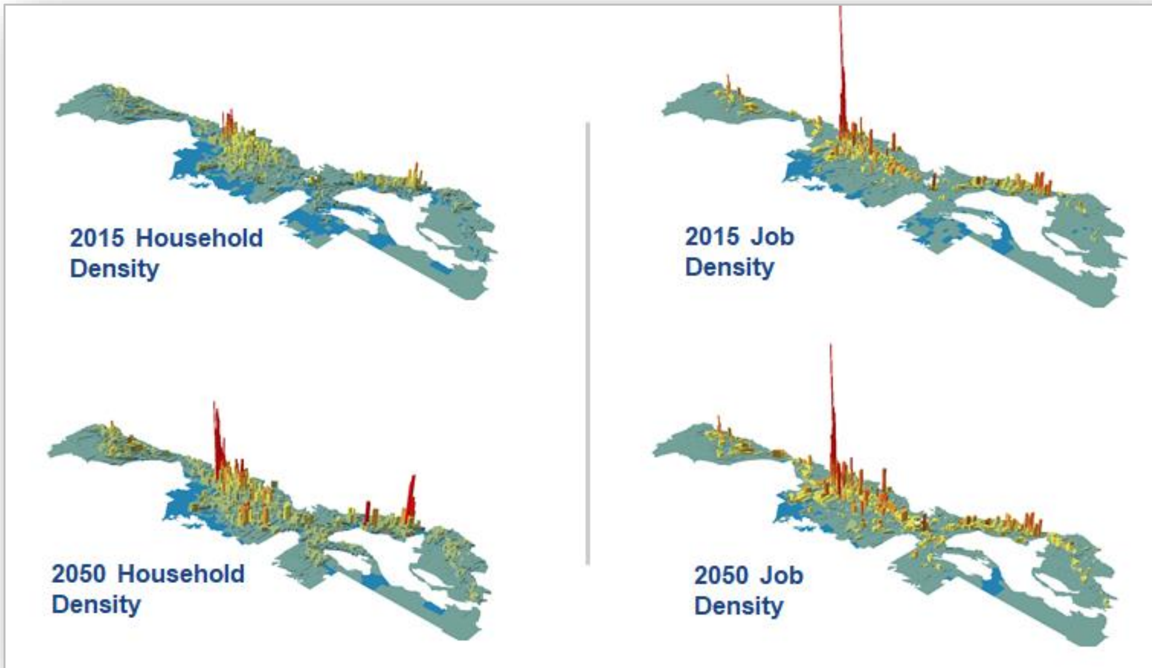


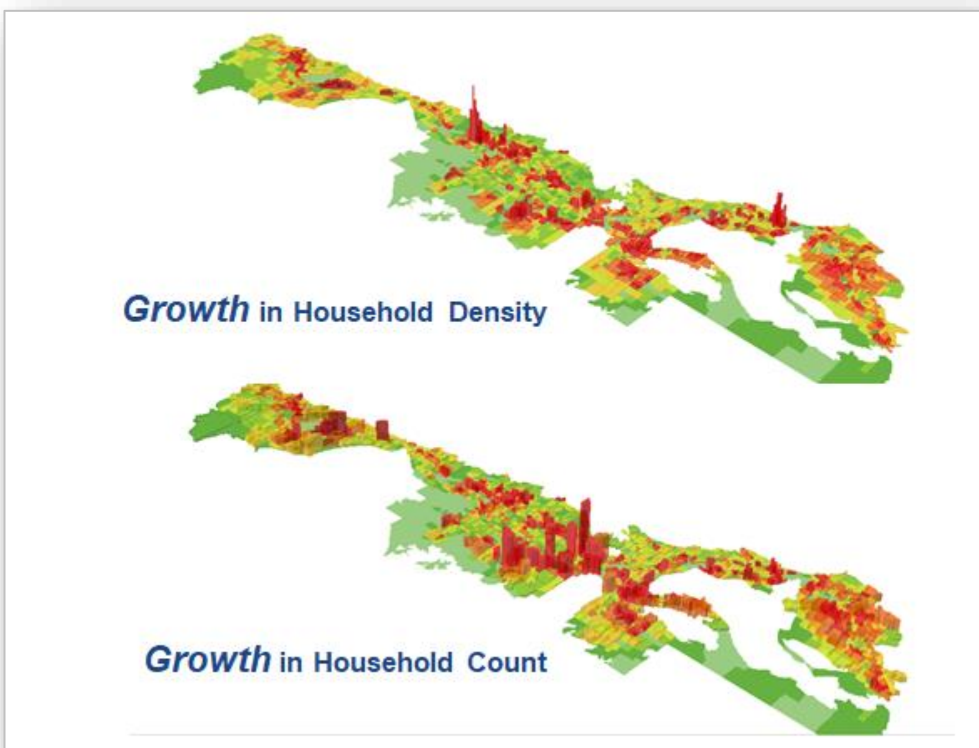
The agencies then showed examples of different validation tests performed on the REMM output:

Example: Charts of Projected Total Employment by 'City Area'

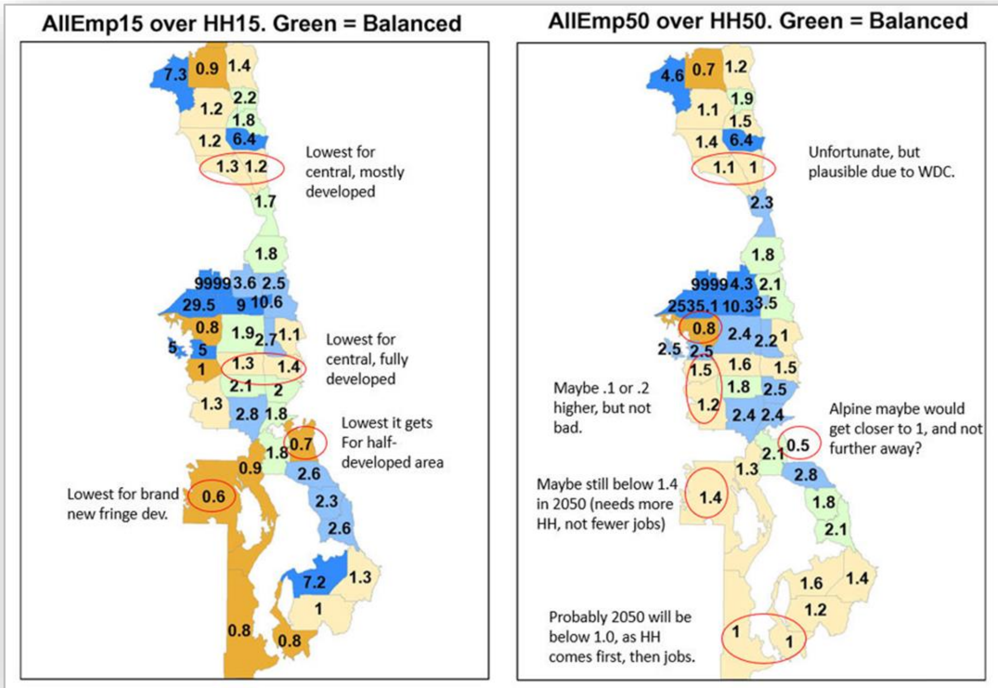


Example: 3D Visuals

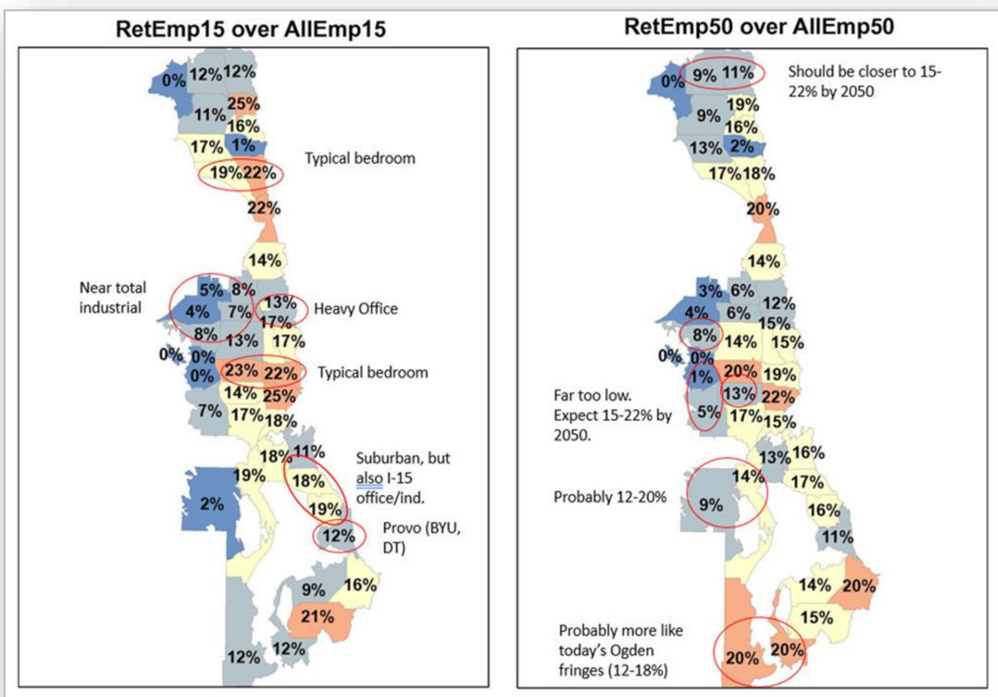




Example: Jobs to Household Ratio



Example: Retail Jobs to All Jobs Ratio

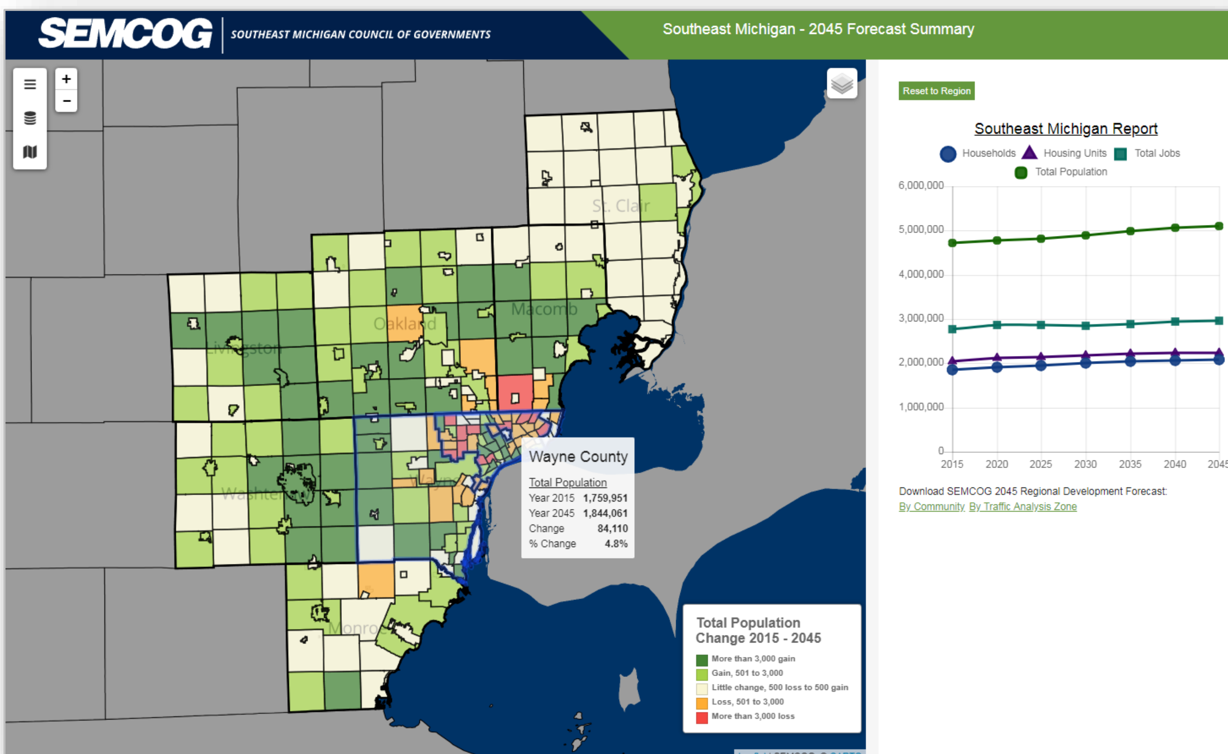


The agencies shared the following information:

> **Key Considerations:** After making adjustments to address validation issues, how should the model results improve or decrease in plausibility? How to assess results more quickly or to compare to previous runs?
>
> **Our Current Solution:** A combination of manual spreadsheet and GIS exploration.
>
> **Strengths, Weaknesses, Opportunities, Threats Overview:** We have good basic assessment tools. May not be comprehensive. Process is manual and time consuming
>
> **Discussion Question:** What do you feel are the 'must haves' to check your output against plausibility? Do we have most things you look at? Are there others?

MAG-Phoenix has used longitudinal, grid cell maps with a manual time slider to display their data. They have found this to be very helpful for their own review as well as for stakeholders. However, MAG-Phoenix has found using both progressive maps and static change maps to be useful. MAG-Phoenix asked Paul Waddell to add better mapping into UrbanSim as part of their contract with the primary purpose to visualize UrbanSim results. MAG-Phoenix puts TAZ-based results and above (e.g. those reviewed by locals) on the web. MAG-Phoenix has not done a lot with 3D visualization, but would like to do more to look at various attributes.

SEMCOG put their final results online in an interactive map (maps.semcog.org/forecast), which they demonstrated for the group.

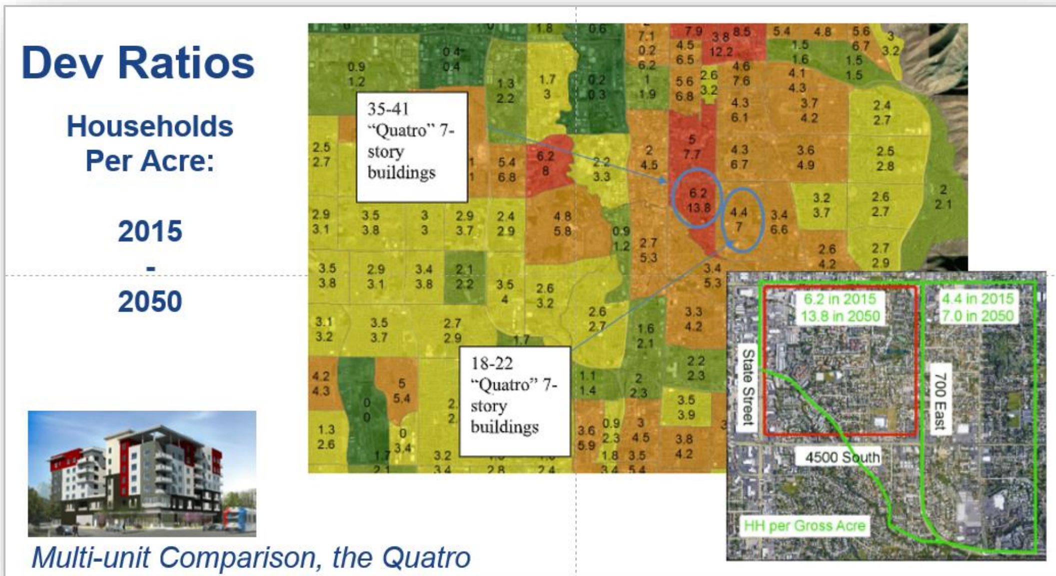


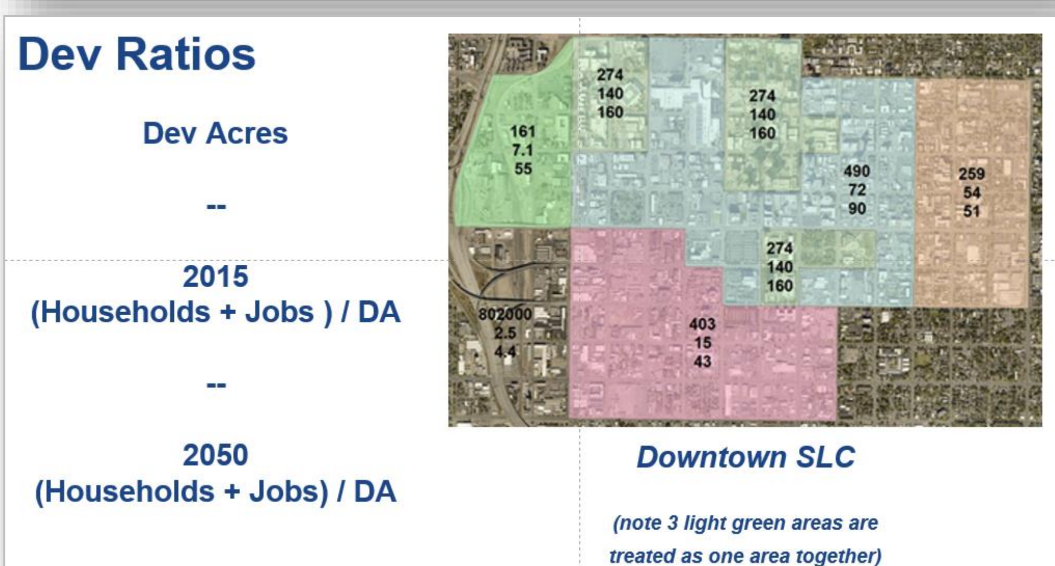
The interactive map includes forecast results with charts. This information is used for public review. SEMCOG has another interactive map with more indicators (around 30-40 indicators) for internal review.

The agencies commented that they have talked about making the policy layer available online. MAG-Phoenix thought this was a good idea.

Panel Discussion: External Review

The agencies hired a consultant to perform an external, third party review of REMM results to help identify data that seemed implausible. The third-party review shared some ratios that helped reveal areas where the socioeconomic forecast deviated from expectations.





The agencies shared the following information:

> **Key Considerations:** Residential and commercial property types redevelop differently and on a different schedule. Are we getting sector-to-sector and job-to-household ratios right?
>
> **Our Current Solution:** Hired external, local consultant to do a plausibility analysis of REMM output. As a result, we are considering institutionalizing development ratio comparisons AND redevelopment constraints.
>
> **Strengths, Weaknesses, Opportunities, Threats Overview:** Can repeat 'ratio analysis' but more concerned with model development to further improve plausibility in this area. One suggestion was to code redevelopment likelihood into a simple likelihood variable.
>
> **Discussion Question:** What types of external review of outputs do you recommend?

SEMCOG has a technical advisory committee, who has some expertise to review data and modeling structure, to provide feedback. The committee's review is more general and is not done at a detailed level. SEMCOG has a pretty intensive internal review process which includes staff from other departments to get input on the area they are most familiar with. A draft of the forecast is always reviewed by local planners. After technical staff approval, SEMCOG has a general assembly meeting to approve the forecasts. SEMCOG emphasized the need to work with and educate partnering agencies and local communities to establish trust in the forecast data and process.

MAG-Phoenix begins their external review process (termed draft zero) by getting a review of the base data and policy layers with the local governments. They have found this to be the most valuable review. MAG-Phoenix commented that so much of the forecast is dependent on the base data layers and if this isn't right the forecast probably isn't going to work out. MAG-Phoenix produces at least two drafts of the forecasts. For the first draft they bring together stakeholders from cities to review the output. Traditionally, they had planners participate in these stakeholder meetings. More and more they are getting traffic engineers as well as professionals from the economic development and water resource areas. The multidiscipline makeup of those reviewing the data has been very helpful because each has a different perspective about what they think is going to happen and is plausible. MAG-Phoenix noted that their stakeholder meetings have changed over time because they have gained a level of trust with the stakeholders, which has made things easier. MAG-Phoenix has also learned to ask the right questions (or to avoid asking the wrong questions) and to do more educating about uncertainty in the stakeholder meetings. After incorporating feedback, MAG-Phoenix prepares a second draft of the forecasts and has another round of meetings, usually online.

MAG-Phoenix brings the control totals, assumptions, and final forecasts to a technical committee. The technical committee approval becomes the official buy-in to forecasts. The regional council eventually approves the forecast totals, but these are done at the Regional Analysis Zone (RAZ) level.

MRCOG follows a similar process to MAG-Phoenix. Draft forecasts are taken to planners around the region. MRCOG and the local planners go back and forth until they get to a point where a broader group of people is involved. MRCOG has a technical board, made up mostly of local planners, that approves the forecasts and passes their recommendation to MRCOG's board. MRCOG commented that scenario planning helped prepare the mindset of the group so that when they got to the final forecasts there is an understanding that assumptions can produce different outcomes.

The agencies asked the panelists how comfortable they were with making land use model data available for other purposes. MAG-Phoenix reported that they were very comfortable and tend toward releasing any data to anyone if they have a good reason and are generally fine with releasing data publicly. SEMCOG reported that they also release their data (e.g. their online map) based on community needs. SEMCOG commented that they make sure the data review is done properly before publishing anything. MRCOG reported they are pretty open with sharing their results. They typically share TAZ-level data plus PDF maps but would like to go to more interactive maps.

Panel Recommendations

After the presentations and challenge discussions, the peer review panelists met separately to discuss and formulate recommendations for the host agencies. The panelists presented their observations and recommendations to agency staff and other attendees at the conclusion of the peer review agenda. A summary of these recommendations is presented below.

Observations

The panelists were very complimentary of the work the local agencies are doing. The panelists commented that the work being done by the agencies is consistent with best practice. The panelists saw no critical flaws in their approach to land use modeling or the process they are following. In fact, much of what the agencies are doing closely matches what the peer panel agencies are doing.

The agencies garnered overall praise from the panelists for making their land use model work and for navigating their way through the arduous base data creation process.

The panelists concurred that the edits made to improve the developer model (e.g. providing utility constraints preventing leap frog development and using zonal level vacancy rates to restrict the developer model) fit within standard practice.

The panelists praised the agencies for their foray into addressing uncertainty resulting from UrbanSim's random seeding (Monte Carlo process). The panelists appreciated seeing how the agencies solved this issue by running the model multiple times and averaging the results.

Recommendations

Recommendation 1: Increase the stratification of household location model

The panelists felt that increasing the number of segments in the household location model would help improve the accuracy of the land use model. This would necessitate creating control totals for each new segment in order for the model to work properly. Since UrbanSim uses a random sampling method, additional controls beyond just household count would be beneficial to help build confidence in the model results.

Recommendation 2: Formalize a quality control and review process of the input data

The panelists emphasized the point that data quality is very important. Establishing a formal review process to ensure the reasonableness and accuracy of up-stream data inputs can save a lot of time and prevent costly errors (such as in additional model re-estimation). The panelists acknowledged that one does not necessarily need to get every parcel exactly right, but one should try to ensure that the major patterns and overall quality are reviewed for plausibility. The panelists also cautioned that if the agencies are developing data in-house, they may need to have data reviewed by different individuals to minimize bias from individual interpretation of the data.

Recommendation 3: Consider review of computed variables instead of only primary variables and using the "all loaded" table

The panelists commented, based on their own experience, that reviewing data together can help detect implausible combinations of attributes. A single variable may seem plausible when viewed in isolation but implausible when viewed in connection with other variables. For instance, a middle-income, one-person household may be plausible for a given land use context. It may also be plausible for a housing unit to have 10,000 square feet. However, it may not make sense for these two to coincide on the same record.

Recommendation 4: Recommend using transactional data instead of assessed data for determining land value

The panelists commented that county assessors are a great source of valuable information for land use models but that the data can have issues. For instance, while the assessor data typically has both land value and improved value, land value data could be problematic. As well, assessor data is often generalized or inconsistent between geographies. Local knowledge of what makes up the assessor data and its level of quality or accuracy is very important. Transactional data, however, is more accurate and reliable than assessed data, and the panelists recommended using transactional data if available. Also, the panelists recommended that the agencies consider adjusting single family, multi-family and commercial land values from what is in assessor data if they find that the assessor data does not make sense.

Recommendation 5: Consider using Pandana to makes estimation easier

The panelists recommended the agencies consider using the capabilities of the Pandana open source network analysis library to measure local accessibility as part of the model's estimation.

Recommendation 6: Make base year data update more efficient and more productive

The panelists commented that taking four to six years to update the base year data is too long. Usually when there is a long data development process, it is challenging to see all parts of the process together. The panelists suggested that the agencies look for bottlenecks and ways to improve efficiencies in their process to make it more productive. The panelists also suggested that the agencies consider adding resources and/or purchasing some of their datasets.

Recommendation 7: Consider making codebase online where others may see/review it

The panelists recommended the agencies consider publishing their UrbanSim codebase online where others may see and review it. This helps make the "black box" portion of the modeling process more transparent. The panelists recommended the agencies publish their model using the GitHub platform and make the model repository public.

Recommendation 8: In general, MPOs working on the UrbanSim platform should have more, regular coordination

The panelists acknowledged the value achieved from having practitioners working on the UrbanSim platform meet together, as demonstrated by this peer review. Meeting together regularly helps increase transparency and collaboration. Also, together MPO's can provide a push to software developers to work on common issues that may need to be fixed, add attributes, or make other improvements in the software.

Recommendation 9: Recommend evaluating risks of census data products changes (especially availability of data content) moving forward

The panelists expressed concern about the future availability of census data. The panelists were uncertain of what data the U.S. Census Bureau will release in the future, what attributes will be included, and if what is released would be undesirably skewed because of privacy protections. The panelists recommended evaluating how the agencies' land use modeling practice might be affected if census data reporting changes.

Appendix

Attendees

Names of the peer review panelists and those attending from the local agencies are presented below in alphabetical order.

Panelists

Jesse Ayers, Maricopa Association of Governments (MAG-Phoenix)

James Kolberg (by phone), Mid-Region Council of Governments (MRCOG)

Guangyu Li, Southeast Michigan Council of Governments (SEMCOG)

Attendees

Wasatch Front Regional Council (WFRC)

- Scott Festin
- Bert Granberg
- Andrew Gruber
- Andy Li
- Nicole Proulx
- Suzie Swim
- Chad Worthen

Mountainland Association of Governments (MAG-Provo)

- Shawn Eliot
- Tim Hereth
- Sarah Lawless

Utah Department of Transportation (UDOT)

- Richard Brockmyer
- Walt Steinvorth
- Khaisy Vonarath

Utah Transportation Authority (UTA)

- Ethan Ray
- Jim Wadley

Brigham Young University (BYU)

- Michael Clay
- Alicia Shum
- Kendall Willardson

University of Utah Gardner Policy Institute (GPI)

- Mallory Bateman
- Mike Hollingshaus

Panelist Biographies

Jesse Ayers, Socioeconomic Modeling Program Manager

Jesse Ayers was hired as Modeling Analyst II in MAG's Information Services Socioeconomic Modeling Program in December 2008. In January of 2011, he was promoted to Modeling Analyst III, and in 2015 was promoted to his current position. His primary focus at MAG is in the development of Arizona's Socioeconomic Modeling, Analysis, and Reporting Toolbox (AZ-SMART). His specialties include urban microsimulation modeling and analysis, Python programming, GIS, and SQL based database work.

Prior to arriving at MAG, Jesse worked for 2 years as a Research Consultant for the Center for Urban Simulation and Policy Analysis (CUSPA) at the University of Washington. There his duties included urban microsimulation modeling and analysis, GIS analysis, database design and administration, and Python programming. While at CUSPA, Jesse earned a MA in Geography from the University of Washington. In addition, Jesse's undergraduate work included a BA in Economics and a BA in Geography, also at the University of Washington. He earned a certificate in Environmental and Natural Resource Economics from the Economics department and focused on Geographic Information Systems and Economic Geography in the Geography department.

James Kolberg, Socioeconomic Analyst

James Kolberg has been working as a Socioeconomic Analyst for the Mid-Region Metropolitan Planning Area (MRMPO) in Albuquerque, New Mexico since early 2017. He has a BA in Economics from Lewis and Clark College and came to MRMPO after a few years working in IT and database administration. He is currently involved in all aspects of land use model development using the UrbanSim parcel model. He has spent the last two years compiling the base year dataset, estimating the models and developing a 2040 forecast for the upcoming Metropolitan Transportation Plan in early 2020. Other than land use model development, James works with local municipalities to model economic development projects using REMI as well as develops population and employment projections for local planning purposes.

Guangyu Li, Chief Land Use Forecast Modeler

Guangyu Li is currently the chief land use forecast modeler at Southeast Michigan Council of Governments (SEMCOG). He specializes in modeling complex demographic, employment and land use changes in modern urban areas and has over 12 years of forecast experience for Southeast Michigan region. He is also responsible for various planning projects including regional transportation accessibility studies, demographic and economic statistical analyses. Guangyu Li holds a Ph.D. degree in land use and transportation from the University of California at Berkeley.

Agenda

Thursday, August 29, 2019		
Continental Breakfast		8:00 am - 8:30 am (30 minutes)
Welcome and Introductions *Andrew Gruber, Executive Director, WFRC*		8:30 am - 8:45 am (15 minutes)
Model Review - Goals and Process *Bert Granberg, WFRC*		8:45 am - 9:00 am (15 minutes)
Model Overview and History *Tim Hereth, MAG and Scott Festin, WFRC*		9:00 am - 10:30 am (90 minutes)
Break		10:30 am - 10:45 am (15 minutes)
Discussion - Model Inputs *Scott Festin and Tim Hereth* • Base year data development • Policy layer inputs		10:45 am - 11:45 am (60 minutes)
Working Lunch - Provided		11:45 am - 1:00 pm (75 minutes)
Discussion - Model Inputs, cont. *Scott Festin and Tim Hereth* • Data quality/consistency issues		1:00 pm - 1:30 pm (30 minutes)
Discussion - Accessibility Sensitivity *Andy Li, WFRC and Tim Hereth* • TDM sensitivity testing • Accessibility factors • Additional variables		1:30 pm - 2:45 pm (75 minutes)
Break		2:45 pm - 3:00 pm (15 minutes)
Discussion - Developer Model *Andy Li and Michael Clay, BYU* • Redevelopment and CBD development • Monte Carlo simulation • Vacancy rate maximums		3:00 pm - 4:30 pm (90 minutes)

Day 1 Wrap Up		4:30 pm - 5:00 pm (30 minutes)
Friday, August 30, 2019		
Continental Breakfast		8:00 am - 8:30 am (30 minutes)
Discussion Catch Up from Day 1		8:30 am - 9:00 am (30 minutes)
Discussion – Developer Model, cont. *Andy Li and Michael Clay* • Model structure • Spatial units		9:00 am - 9:30 am (30 minutes)
Discussion - Model Calibration and Validation *Michael Clay*		9:30 am - 10:30 am (60 minutes)
Break		10:30 am - 10:45 am (15 minutes)
Discussion - Demographic Transition *Scott Festin and Tim Hereth*		10:45 am - 11:15 am (30 minutes)
Discussion - Outputs and Visualization *Tim Hereth and Bert Granberg*		11:15 am - 12:00 pm (45 minutes)
Working Lunch *Chad Worthen, WFRC/MAG/UDOT* • Panel convene to formulate recommendations		12:00 pm - 1:30 pm (90 minutes)
Panel Present Recommendations		1:30 pm - 2:00 pm (30 minutes)

Variables and Specifications Handouts

Handout 1 – Variables

Node Variables *distances are in meters	**TAZ Variables**	**Building Variables**
Train Station (varying distances)	Parcel Count	Building Type Number
Jobs, Industry (varying distances)	Jobs (different categories) within 10 mins	Non-Residential Square Footage
Jobs, Office (varying distances)	Population Density	Notes
Bus Stop (varying distances)	Population within varying distances (miles)	Parcel Id Number
Average Square Feet per Unit (varying distances)	Tdm County Id	Number of Residential Units
Average Lot Size per Unit (varying distances)	Average age of head of household	Stories tall
Average Income (varying distances)	Total Jobs within 30 mins of Transit	Price of non-residential units
Single Family Dwelling Units (varying distances)	Average of household children per zone	Year Built
Multi Family Dwelling Units (varying distances)	Logsum of Jobs (Different Categories)	Residential price per square feet
High Income (varying distances)	Non-residential median price	Residential Price
Low Income (varying distances)	Population within 40 mins of Transit services	Non-residential price
Total Jobs (varying distances)	Residential units within 30 mins	Number of Job Spaces
Jobs Category 1 (varying distances)	Residential Median Price	Is office space
Jobs Category 2 (varying distances)	Industrial Median Price	Distance to Bus Stop
Jobs Category 3 (varying distances)	Population within 20 mins	Elevation
Jobs Category 4 (varying distances)	Total Jobs	Medium District Median Income
Jobs Category 5 (varying distances)	Average Household size per zone	Distance from a trail
Jobs Category 6 (varying distances)	Jobs Category 1	Distance from a rail depot
Jobs Category 7 (varying distances)	Jobs Category 2	Lot Size per unit
Jobs Category 8 (varying distances)	Jobs Category 3	Distance from an airport
Jobs Category 9 (varying distances)	Jobs Category 4	Is Single Family

Jobs Category 10 (varying distances)	Jobs Category 5	Improvement value
Industrial Square Feet, 1000 M	Jobs Category 6	Distance from rail station
Workers, 1500 M	Jobs Category 7	Acres of Parcel
Freeway Exit (varying distances)	Jobs Category 8	Distance from a stream
Industrial Rail Depot	Jobs Category 9	Is government building
Average Children 500 M	Jobs Category 10	Zone Id Number
Sum of Non-Residential Units	Total Jobs within 15 mins	Two-way road volume, non-freeway
Streams 200 M	Sum of Land value	Airport
Trails (varying distances)	Rent Median Price	General Type
	Population within 30 mins	Distance from a freeway exit
	Total Households	Total Job spaces
	Average nonresidential price per zone	Is Multi-Family
	Land Value per Acre	Real FAR
	Household Choice Control	Residential square footage
	Logsum of Household income	Maximum FAR
	Logsum of Total Jobs	Is mixed use
	Logsum of Total Population	Average Building age
	Total Population	Medium District] ID
	Office Median Price	Residential Sales price per square foot
	Commute time (various times)	Square feet per job
	Average household cars per zone	Volume of two-way roads
		Square feet per unit
		Distance from University
		Parcel Volume
		Number of Vacant Residential Units
		County ID number
		Is an industrial type
		Number of Vacant Job Spaces
		Distance from bus route
		Building Age
		Large district median income
		Land Value
		Node ID Number
		Unit Price of Residential
		Distance from a Freeway Exit Tdm output
		Rail Depot
		Is other type of building

Handout 2 – Modeling Tables

Table 11.1: Estimated Single Family Price Model-Salt Lake County

Variable	Salt Lake Coefficient	Salt Lake T-Score
If(avg_building_age < 21)	0.2764	213.9549
If(year_built < 1945)	0.054	22.6937
Intercept	6.2302	224.1397
Ave_income_500	0.1507	62.9217
Commute_time	-0.005	Asserted
In(fwy_exit_dist_tdm_output)	-0.0153	-7.4735
In(rail_stn_dist)	-0.0401	-34.0184
In(sqft_per_unit)	-0.3874	-214.228
In(university_dist)	-0.2568	-62.2276
Parcel_volume	-0.0001	-4.4373

Table 11.2: Estimated Single Family Price Model-Weber County

Variable	Weber Coefficient	Weber T-Score
If(avg_building_age < 21)	0.1153	48.2064
If(year_built < 1945)	-0.0441	-6.4489
Intercept	3.9861	67.0258
Ave_income_500	0.3122	58.049
Commute_time	-0.005	Asserted
In(fwy_exit_dist_tdm_output)	-0.077	-21.4797
In(rail_stn_dist)	0.0133	5.5632
In(sqft_per_unit)	-0.4398	-120.3754
In(university_dist)	0.0659	22.4005

Table 11.3: Estimated Single Family Price Model-Davis County

Variable	Coefficient	T-Score
If(avg_building_age < 21)	0.1356	38.1408
If(year_built < 1945)	-0.0571	-2.615
Intercept	4.0978	49.2753
Ave_income_500	0.382	52.1332
Commute_time	-0.005	Asserted
In(fwy_exit_dist_tdm_output)	-0.0497	-8.8143
In(rail_stn_dist)	0.0519	13.6531
In(sqft_per_unit)	-0.5783	-125.3037
In(university_dist)	0.0217	2.5459
Parcel_volume	0.0004	2.4791

Table 11.4: Estimated Single Family Price Model-Utah

Variable	Coefficient	T-Score
If(avg_building_age < 21)	0.1764	120.8757
If(year_built < 1945)	-0.035	-13.4059

Intercept	7.2749	98.2778
Ave_income_500	0.1026	36.7079
Commute_time	-0.005	Asserted
In(fwy_exit_dist_tdm_output)	-0.0093	-4.6289
In(rail_stn_dist)	-0.1538	-29.3826
In(sqft_per_unit)	-0.3243	-162.0401
In(university_dist)	-0.0894	-42.5518
Parcel_volume	0.0005	7.9508

Table 12: Estimated Multi-Family Price Model

Variable	**Coefficient**	**T-Score**
Intercept	4.5167	52.1371
Ave_income_500	0.1584	20.8317
Avg_building_age	-0.0053	-53.6003
Commute_time	-0.0050	Asserted
Fwy_exit_dist_tdm_output	0.0405	14.1723
Mfdu_500	0.0298	12.5428
Np.log1p(real_far)	0.1126	33.2104
Np.log1p(sqft_per_unit)	-0.2739	-48.8376
Parcel_volume	-0.0003	-2.8235

Table 13: Estimated Industrial Price Model

Variable	**Coefficient**	**T-Score**
Intercept	2.0524	64.9089
Industrial_sqft_1000	0.0000	-7.2685
Jobs_6_1000	0.0247	5.1113
Jobs_8_1000	0.0000	1.8649
Mfdu_1000	-0.0164	-3.7833
Np.log1p(building_age)	-0.0750	-8.9379
Np.log1p(real_far)	-0.0220	-2.0539
Parcel_volume	0.0008	2.5965
Rail_depot	-0.0889	-4.5922

Table 14: Estimated Retail Price Model

Variable	**Coefficient**	**T-Score**
Intercept	5.4746	88.1291
Jobs_1_500	0.0669	6.5721
Jobs_7_500	-0.0550	-4.6322
Np.log1(building_age)	-0.1557	-12.4477
Np.log1p(real_far)	-0.0511	-3.2028
Parcel_volume	0.0019	2.6282
Train_stn_500	0.2760	1.8373

Table 15: Estimated Office Price Model

Variable	**Coefficient**	**T-Score**
Intercept	5.2293	78.0376
Jobs_6_500	0.0308	6.8527
Land_value_per_acre	0.0000	3.7887

Logsum_hhinc4	0.0000	3.4620
Median_income	0.0000	4.6073
Np.log1p(building_age)	-0.1258	-12.8219
Np.log1p(fwy_exit_dist_tdm_output)	-0.0784	-2.6627
Np.log1p(real_far)	0.0403	4.1546
Parcel_volume	0.0008	3.1028

Table 16.1: Estimated Household Location Choice Model-Salt Lake County Income Quartile 1

Variable	Coefficient	T-Score
commute_time	0.046	23.611
np.log1p(is_mf)	1.987	55.887
np.log1p(logsum_jobs9)	0.156	13.377
np.log1p(real_far)	-0.12	-7.024
np.log1p(sqft_per_unit)	-0.131	-8.524
res_price_per_sqft	0.016	127.96

Table 16.2: Salt Lake Income Quartile 2

Variable	Coefficient	T-Score
ave_income_1000	-1.608	-40.962
building_age	-0.214	-71.683
commute_time	0.017	3.316
is_mf	0.936	37.51
np.log1p(logsum_jobs1)	0.302	7.484
np.log1p(sqft_per_unit)	-0.537	-27.456
np.log1p(university_dist)	0.316	7.684

Table 16.3: Salt Lake Income Quartile 3

Variable	Coefficient	T-Score
commute_time	-0.004	-0.786
np.log1p(airport_distance)	0.004	0.104
np.log1p(building_age)	-1.091	-115.534
np.log1p(elevation)	-0.242	-3.441
np.log1p(logsum_jobs6)	-0.586	-10.421
np.log1p(population_within_20_min)	0.199	5.98

Table 16.4: Salt Lake Income Quartile 4

Variable	Coefficient	T-Score
ave_sqft_per_unit_500	0.999	31.279
building_age	-0.185	-66.681
commute_time	0.049	15.823
is_sf	-0.067	-1.759
np.log1p(airport_distance)	0.232	5.097
np.log1p(logsum_hhinc4)	0.447	11.662
np.log1p(res_price_per_sqft)	1.249	27.263

Table 16.5: Weber/Davis County Income Quartile 1

Variable	Coefficient	T-Score
commute_time	-0.0527	-18.1056
mfdu_500	0.168	14.1522
np.log1p(is_mf)	2.3863	34.8139
np.log1p(logsum_hhinc1)	-1.3628	-78.6702
np.log1p(sqft_per_unit)	-0.1176	-6.7304
parcel_volume	-0.013	-11.1352

Table 16.6: Weber/Davis County Income Quartile 2

Variable	Coefficient	T-Score
building_age	-0.1995	-60.3406
elevation	-0.0012	-9.3805
np.log1p(logsumjobs)	0.7091	20.3998
np.log1p(res_price_per_sqft)	-1.4252	-27.4537
np.log1p(sqft_per_unit)	-1.242	-48.6118
np.log1p(university_dist)	-0.0158	-0.448
parcel_volume	0.0013	1.2807

Table 16.7: Weber/Davis County Income Quartile 3

Variable	Coefficient	T-Score
commute_time	-0.0107	-4.1595
np.log1p(building_age)	-0.9243	-96.4301
np.log1p(elevation)	-0.9299	-12.0958
np.log1p(logsum_hhinc1)	-2.2704	-13.7431
np.log1p(logsumjobs)	1.4892	13.7235
np.log1p(university_dist)	0.0434	1.31

Table 16.8: Weber/Davis County Income Quartile 4

Variable	Coefficient	T-Score

np.log1p(logsum_hhinc4)	-0.3384	-9.1397
np.log1p(res_price_per_sqft)	2.4091	48.8049
np.log1p(sqft_per_unit)	1.6577	94.1111
np.log1p(university_dist)	0.6856	24.3586
parcel_volume	-0.0136	-15.8109

Table 16.9: Utah County Income Quartile 1

Variable	**Coefficient**	**T-Score**
commute_time_20	0.0507	16.3174
mfdu_500	0.2924	40.0341
np.log1p(elevation)	-1.4607	-76.506
np.log1p(sqft_per_unit)	-0.5006	-31.2764
parcel_volume	-0.0117	-16.9346

Table 16.10: Utah County Income Quartile 2

Variable	**Coefficient**	**T-Score**
building_age	-0.2204	-46.7065
commute_time_20	0.0101	4.817
np.log1p(logsumjobs)	-0.3417	-24.8259
np.log1p(res_price_per_sqft)	0.9083	25.4808
parcel_volume	-0.0031	-3.9249

Table 16.11: Utah County Income Quartile 3

Variable	**Coefficient**	**T-Score**
commute_time_20	0.0253	14.1759
high_income_1000	0.0009	0.0861
np.log1p(building_age)	-1.4033	-61.9311
np.log1p(res_price_per_sqft)	0.6348	40.5054
parcel_volume	-0.0254	-26.4633

Table 16.12: Utah County Income Quantile 4

Variable	**Coefficient**	**T Score**

Commute_time	0.0856	46.4902
elevation	0.001	13.949
high_income_1000	0.6883	45.6545
np.log1p(logsumjobs)	-0.7146	-8.0164
np.log1p(logsumpopulation)	1.3534	15.915

Table 17.1: Estimated Employment Location Choice Model-Accommodation and Food Jobs

Variable	**Coefficient**	**T-Score**
Is_other	1.0422	7.2003
Is_retail	1.4004	14.0940
Jobs_1_500	1.0106	26.9610
Jobs_9_500	-0.1382	-5.7252
Np.log1p(population_within_20_min)	-0.5519	-32.1869
Np.log1p(unit_price_non_residential)	0.2847	9.4583

Table 17.2: Estimated Employment Location Choice Model-Government and Education Jobs

Variable	**Coefficient**	**T-Score**
Jobs_200	0.9393	33.1218
Np.log1p(jobs_within_20_min_transit)	0.0097	1.0126
Np.log1p(logsumpopulation)	-0.5397	-43.2052
Np.log1(real_far)	-0.7400	-12.8062
Parcel_volume	-0.0291	-17.2404

Table 17.3: Estimated Employment Location Choice Model-Health Care Jobs

Variable	**Coefficient**	**T-Score**
Is_office	1.3036	12.1339
Is_other	0.2112	1.2676
Jobs_4_500	1.0252	28.3133
Np.log1p(ave_nonres_price_zn)	0.7832	9.6160
Np.log1p(jobs_5_within_10_min)	-0.2329	-9.5508
Np.log1p(logsumpopulation)	0.1515	3.9987
Np.log1p(real_far)	-0.6735	-9.0087

Table 17.4: Estimated Employment Location Choice Model-Manufacturing Jobs

Variable	**Coefficient**	**T-Score**
Is_other	-0.3801	-2.6131
Jobs_5_500	1.1803	40.1274
Np.log1(ave_nonres_price_zn)	0.4750	12.3120
Np.log1(real_far)	-0.3394	-6.0237
Parcel_volume	-0.0196	-10.1857

Table 17.5: Estimated Employment Location Choice Model-Office Jobs

Variable	**Coefficient**	**T-Score**

Jobs_6_500	1.0733	41.3855
Np.log1p(is_office)	1.5624	14.4332
Np.log1p(logsumpopulation)	0.2433	21.3569
Np.log1(real_far)	-0.4774	-9.3636

Table 17.6: Estimated Employment Location Choice Model-Other Jobs

Variable	**Coefficient**	**T-Score**
Jobs_7_500	0.9638	37.0170
Np.log1p(is_office)	0.1560	1.4987
Np.log1p(logsumpopulation)	0.1139	10.8138
Np.log1p(real_far)	-0.0989	-2.0469

Table 17.7: Estimated Employment Location Choice Model-Retail Trade Jobs

Variable	**Coefficient**	**T-Score**
Jobs_9_500	1.4782	40.2219
Np.log1p(is_retail)	1.8073	15.3768
Np.log1p(logsumpopulation)	-0.0656	-4.2691
Np.log1p(real_far)	-0.0923	-1.1768

Table 17.8: Estimated Employment Location Choice Model-Wholesale and Transport Jobs

Variable	**Coefficient**	**T-Score**
Jobs_10_500	1.3780	61.1791
Np.log1p(population_within_10_min)	-0.0729	-5.3459
Np.log1p(real_far)	-0.3377	-5.6144
Sfdu_1000	-0.0661	-2.6899

Table 18: Building Costs by Building Height and Type

	Residential	**Industrial**	**Retail**	**Office**
< 15 feet	90	140	160	160
15 – 55 feet	110	175	175	175
55 – 120 feet	120	200	200	200
> 120 feet	140	230	230	230

Table 19: Price to Rent Conversion Factors

Type	**Conversion Factor**
Residential	0.1646
Industrial	0.6488
Retail	0.1111
Office	0.1138

REMM Documentation

An electronic copy of the REMM model documentation was sent to the panelists prior to the peer review. A copy of the model documentation can be found at the following link:

https://drive.google.com/open?id=18yJvzjgN-06PTLS_CxSCu2MzcuY5dQqek0CPLCuaIao