Agent-based modeling of vehicle choice in California

September 15, 2025

Prepared For

EER Communications

Prepared by:

Ari Ball-Burack, EER Summer Research Fellow

EVOLVED ENERGY RESEARCH

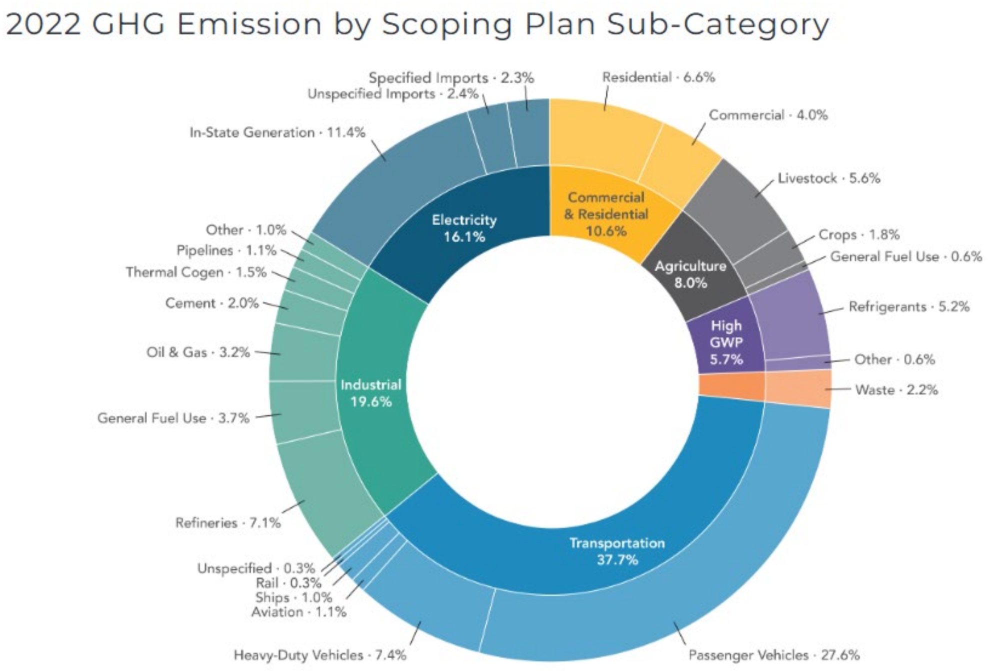
Agenda

1. Why ABM for vehicle choice?
2. Modeling pipeline overview
3. Data description
4. ABM architecture
5. SBI: definition, description, and challenges
6. Usage example: IRA repeal simulations
7. Conclusions

Why ABM for vehicle choice?

Vehicle choice is critical for decarbonization

- Passenger vehicle electrification is critical for decarbonization in California, the US, and worldwide
- For researchers, planners, and policymakers, the question is not only *how many cars sold,* but *who adopts, when, and why*



Source: https://ww2.arb.ca.gov/ghg-inventory-graphs

Today’s forefront of adoption modeling

Approach	Strengths	Drawbacks
Discrete-choice modeling (e.g., multinomial logit)	Link adoption decisions to price, income, and other attributes in a policy-relevant way	Struggle to capture path-dependence and non-linearities (e.g., neighborhood effects)
Diffusion models (e.g., S-curves, Bass diffusion)	Capture system-wide diffusion dynamics incl. path-dependence	Lack granularity: information on *who* adopts *why* is missing
System models (e.g., optimization, equilibrium, integrated assessment, energy-economy)	Capture relationships across the energy system; can express notions of optimality and efficiency	Tend to simplify and homogenize household-level behavior
Expert elicitation	Fills gaps where data are scarce, especially for nascent technologies	Subjective

ABM as a complement to existing efforts

EVOLVED ENERGY RESEARCH

- Agent-based modeling (ABM) fills gaps: combines arbitrarily high granularity with representation of feedbacks, interactions, path-dependence, heterogeneity, and causality
- In each timestep, a set of heterogeneous agents makes adoption decisions based on personal attributes, technological attributes, and contextual conditions
- Drawbacks: data- and computationally-intensive, challenging to calibrate

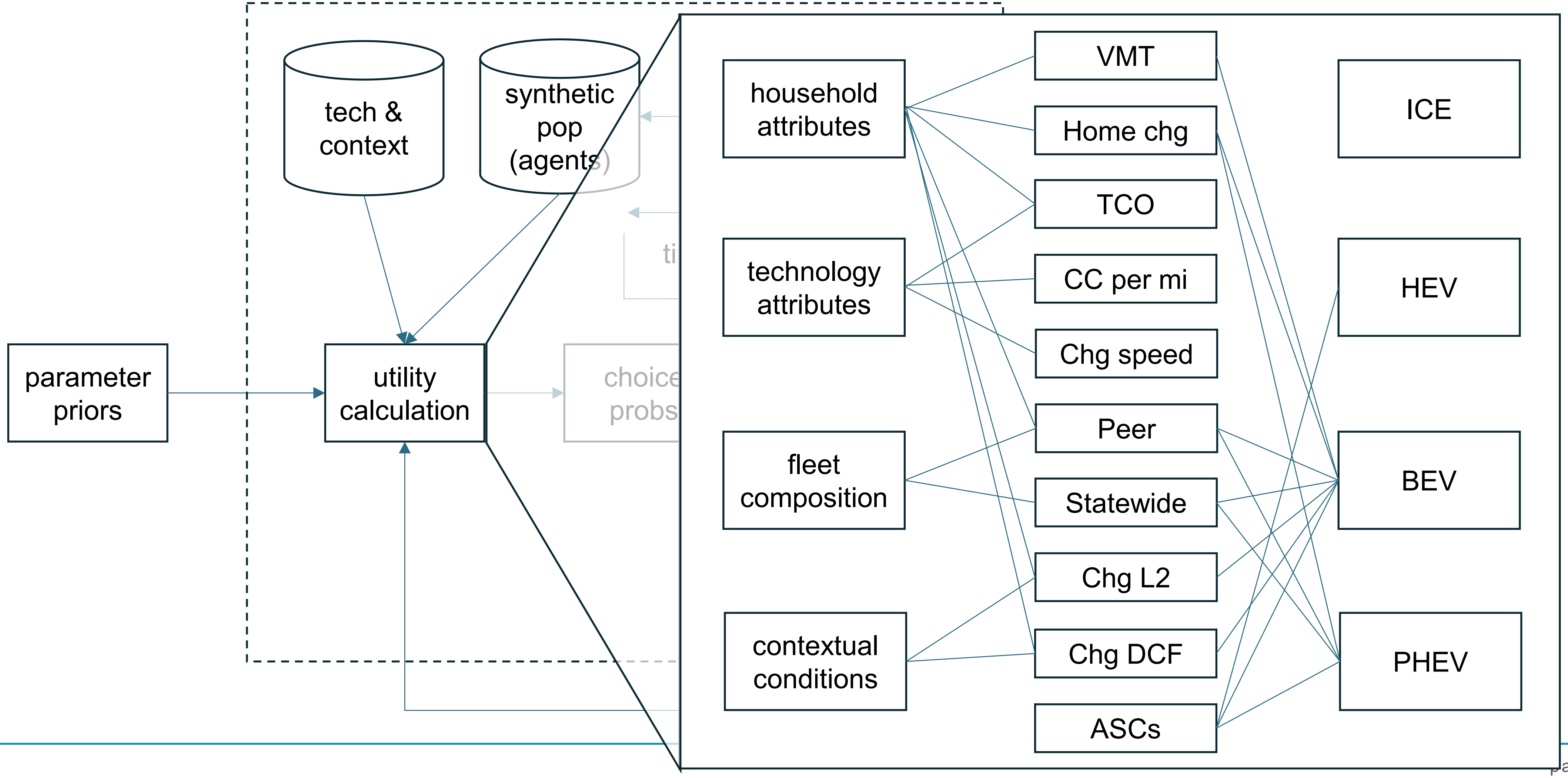
Summer fellowship: project, RQs, deliverables

- Project: empirically calibrated ABM of vehicle choice in California
  1. Data: agents, technologies, ground truth for calibration
  2. ABM: utility construction and choice architecture
  3. Simulation-based inference (SBI): Bayesian empirical calibration
- Research questions
  1. How can publicly-available data be used to build an ABM of vehicle choice in CA?
  2. Can SBI be used to empirically calibrate the ABM, and if so, what does it tell us about vehicle choice decision-making?
  3. How does IRA repeal impact EV adoption in CA through 2050?
- Deliverables: data, code base (ABM and SBI pipeline), blog post, slide deck

EVOLVED
ENERGY
RESEARCH

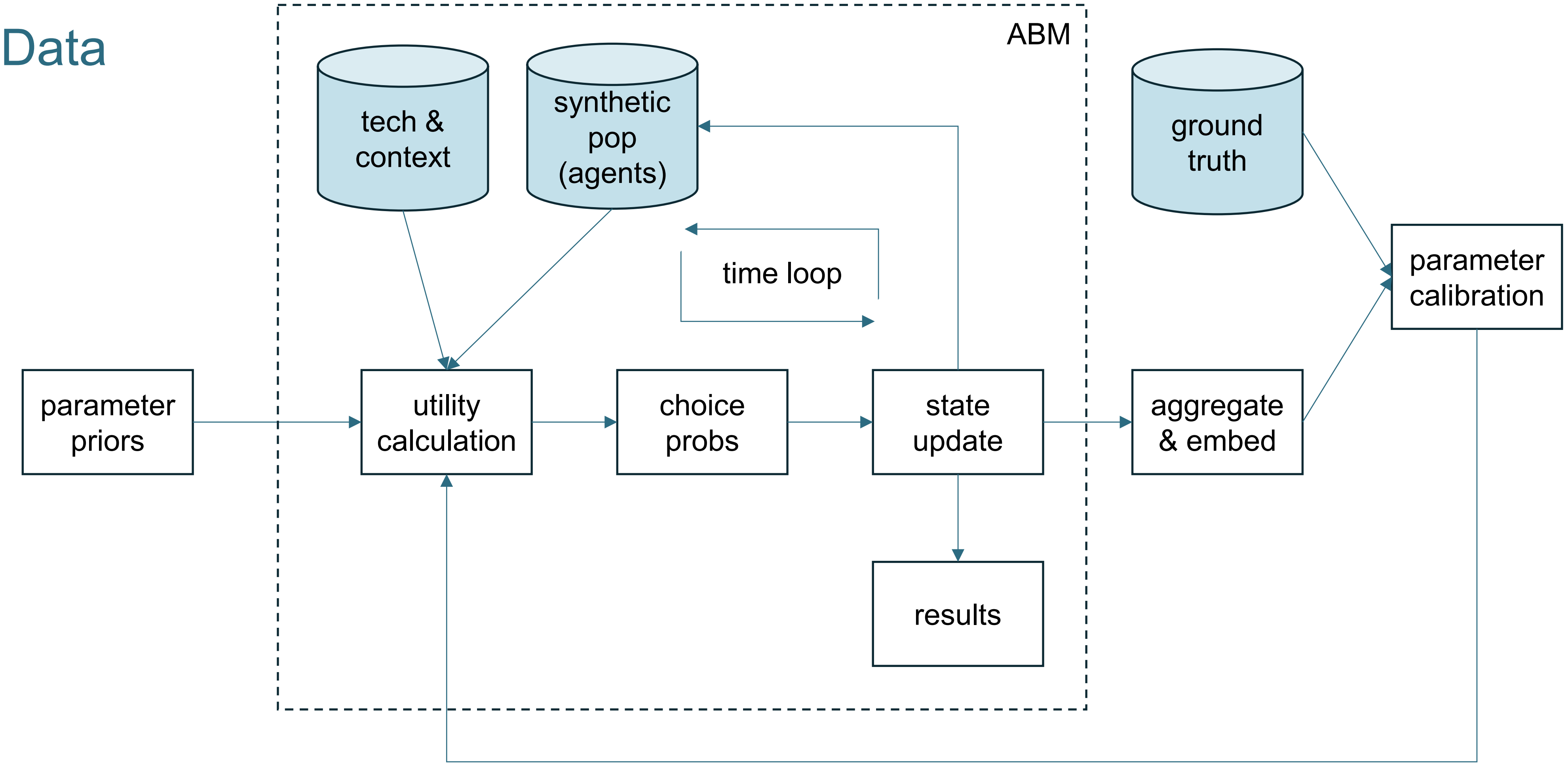
Modeling pipeline overview

Modeling pipeline overview



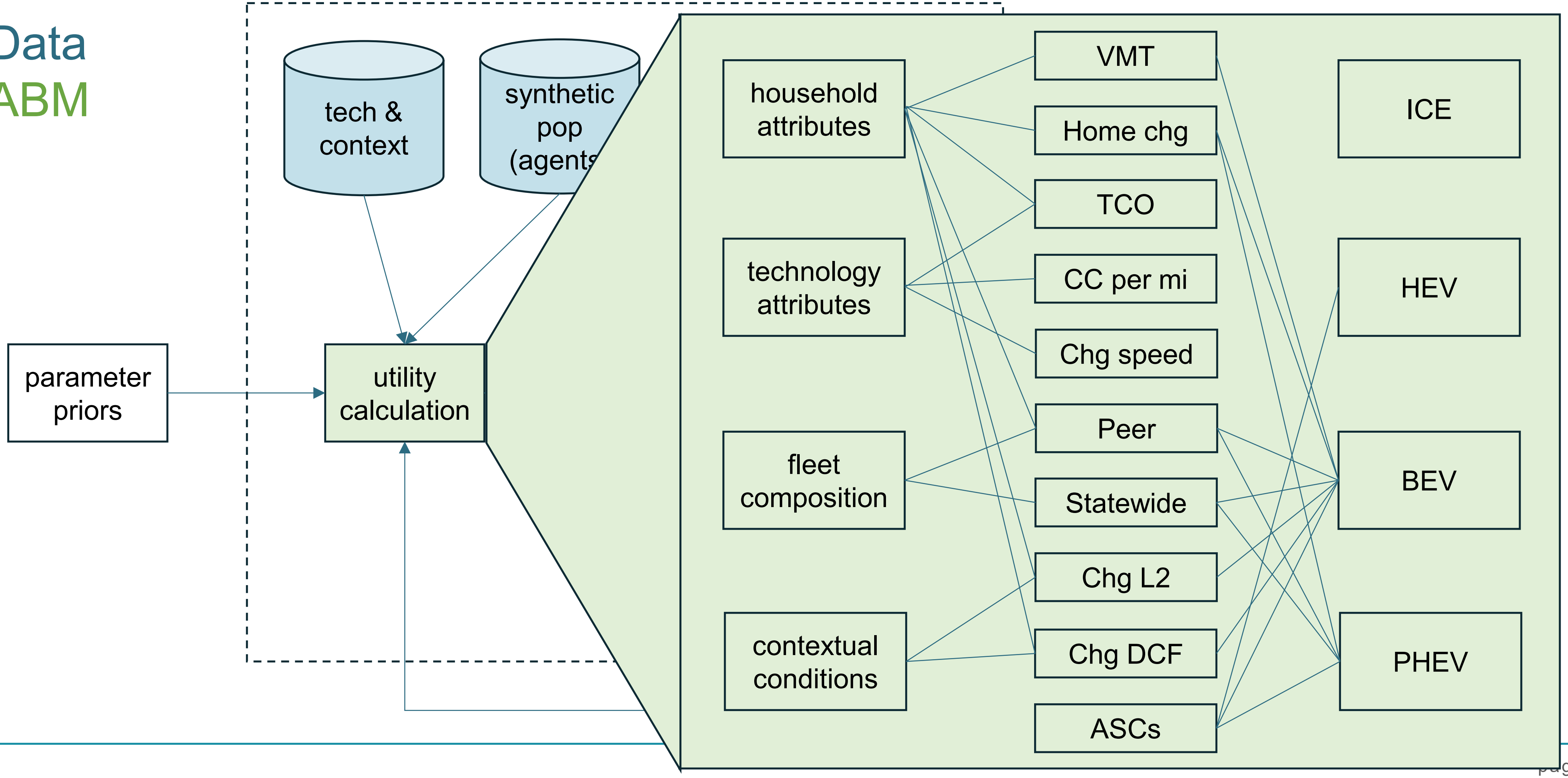
Modeling pipeline overview

1. Data



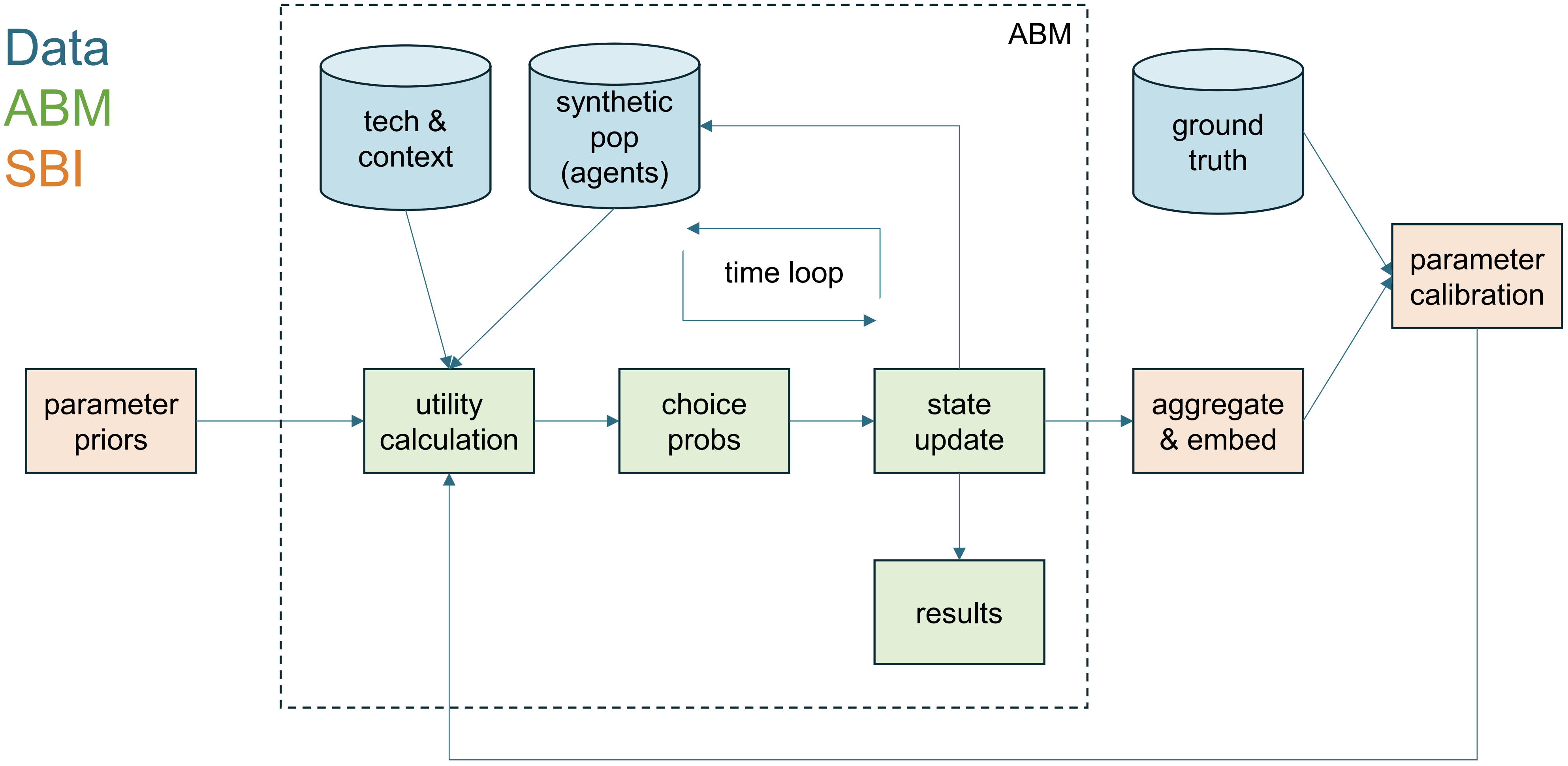
Modeling pipeline overview

1. Data
2. ABM



Modeling pipeline overview

1. Data
2. ABM
3. SBI

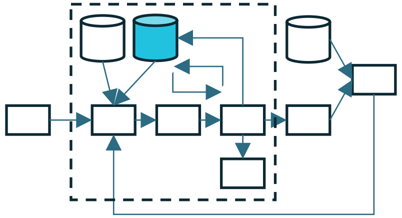


EVOLVED ENERGY RESEARCH

Data description

Synthetic population

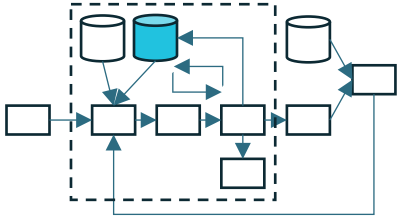
N = 100,000 households (288,065 people, 208,906 vehicles)

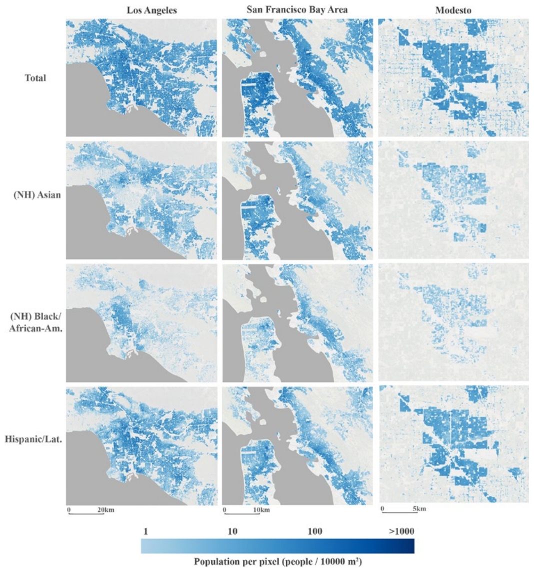


Synthetic population

N = 100,000 households (288,065 people, 208,906 vehicles)

1. Race, housing tenure (rent vs. own), and housing type (single-family vs. multifamily): random choice from population "pixels" and census data (CBG-level)





Depsky, N. J., Cushing, L., & Morello-Frosch, R. (2022). High-resolution gridded estimates of population sociodemographics from the 2020 census in California. *PLoS One*, *17*(7), e0270746.

B25032 | Tenure by Units in Structure

American Community Survey | Universe: Occupied housing units | 2023: ACS 5-Year Estimates Detailed Ta...

Notes Geos Topics

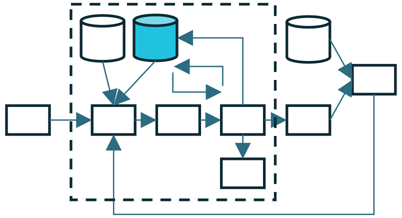
Label	Block Group 2; Census Tract 4237; Alameda County; California	
	Estimate	Margin of Error
Total:	395	±141
Owner-occupied housing units:	308	±136
1, detached	267	±132
1, attached	14	±18
2	10	±14
3 or 4	0	±14
5 to 9	10	±17
10 to 19	7	±12
20 to 49	0	±14
50 or more	0	±14
Mobile home	0	±14
Boat, RV, van, etc.	0	±14
Renter-occupied housing units:	87	±46
1, detached	18	±19
1, attached	0	±14
2	17	±18
3 or 4	0	±14
5 to 9	23	±28
10 to 19	29	±26
20 to 49	0	±14
50 or more	0	±14
Mobile home	0	±14
Boat, RV, van, etc.	0	±14

https://data.census.gov/table?q=b25032

Synthetic population

EVOLVED ENERGY RESEARCH

N = 100,000 households (288,065 people, 208,906 vehicles)



1. Race, housing tenure (rent vs. own), and housing type (single-family vs. multifamily): random choice from population "pixels" and census data (CBG-level)

2. Conditional random choice of household size based on tenure (CBG-level granularity)

B25009 | Tenure by Household Size

American Community Survey | Universe: Occupied housing units | 2023: ACS 5-Year Estimates Detailed Ta...

Notes | Geos | Topics

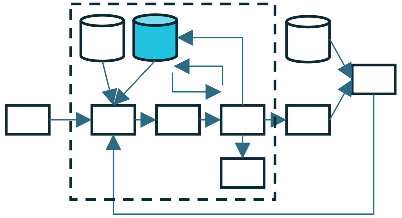
Label	Block Group 2; Census Tract 4237; Alameda County; California	
	Estimate	Margin of Error
Total:	395	±141
Owner occupied:	308	±136
1-person household	39	±46
2-person household	76	±42
3-person household	46	±31
4-person household	66	±38
5-person household	81	±105
6-person household	0	±14
7-or-more person household	0	±14
Renter occupied:	87	±46
1-person household	24	±24
2-person household	48	±30
3-person household	8	±14
4-person household	7	±12
5-person household	0	±14
6-person household	0	±14
7-or-more person household	0	±14

https://data.census.gov/table?q=b25009

Synthetic population

N = 100,000 households (288,065 people, 208,906 vehicles)

1. Race, housing tenure (rent vs. own), and housing type (single-family vs. multifamily): random choice from population “pixels” and census data (CBG-level)
2. Conditional random choice of household size based on tenure (CBG-level granularity)
3. Conditional random choice of vehicle count based on household size (CT-level granularity)



B08201 | Household Size by Vehicles Available

American Community Survey | Universe: Households | 2023: ACS 5-Year Estimates Detailed Tables

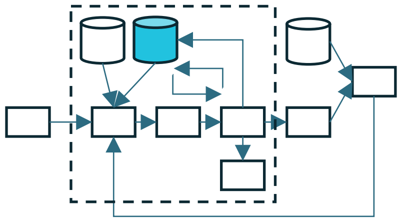
Notes Geos Topics

Label	Census Tract 4237; Alameda County; California	
	Estimate	Margin of Error
Total:	1,143	±178
No vehicle available	261	±85
1 vehicle available	447	±118
2 vehicles available	212	±73
3 vehicles available	213	±121
4 or more vehicles available	10	±15
1-person household:	349	±111
No vehicle available	93	±50
1 vehicle available	216	±89
2 vehicles available	32	±24
3 vehicles available	8	±13
4 or more vehicles available	0	±14
2-person household:	393	±100
No vehicle available	108	±67
1 vehicle available	184	±82
2 vehicles available	75	±42
3 vehicles available	26	±30
4 or more vehicles available	0	±14

https://data.census.gov/table?q=b08201

Synthetic population

N = 100,000 households (288,065 people, 208,906 vehicles)

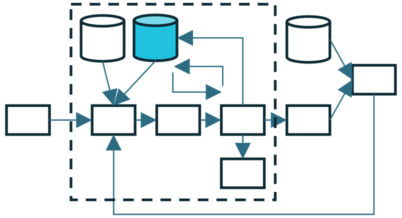


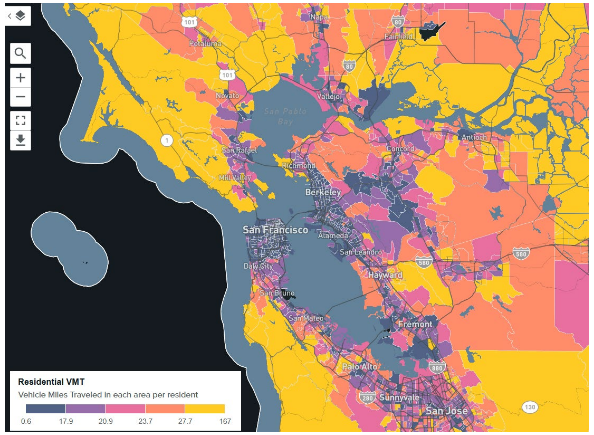
1. Race, housing tenure (rent vs. own), and housing type (single-family vs. multifamily): random choice from population “pixels” and census data (CBG-level)
2. Conditional random choice of household size based on tenure (CBG-level granularity)
3. Conditional random choice of vehicle count based on household size (CT-level granularity)
4. Income: lognormally distributed based on census data (CT-level granularity), with CT-wide exponentially fitted growth rate
   - Limitation: could/should be conditioned on the above. Oh well.

Synthetic population

N = 100,000 households (288,065 people, 208,906 vehicles)

1. Race, housing tenure (rent vs. own), and housing type (single-family vs. multifamily): random choice from population "pixels" and census data (CBG-level)
2. Conditional random choice of household size based on tenure (CBG-level granularity)
3. Conditional random choice of vehicle count based on household size (CT-level granularity)
4. Income: lognormally distributed based on census data (CT-level granularity), with CT-wide exponentially fitted growth rate
5. VMT (CBG-level granularity) from Replica (Thursday and Saturday)

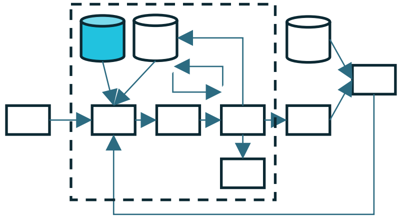




Licensed for UC Berkeley by Replica; do not reproduce

Tech & context: vehicle capital costs

EVOLVED ENERGY RESEARCH

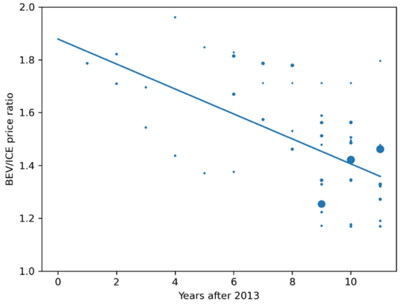


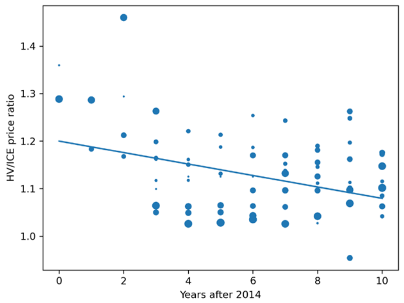
For ICE:

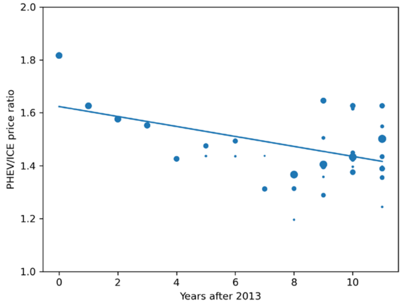
- Sales data 2013-2024 → Scrape KBB starting MSRP → Sales-weighted average

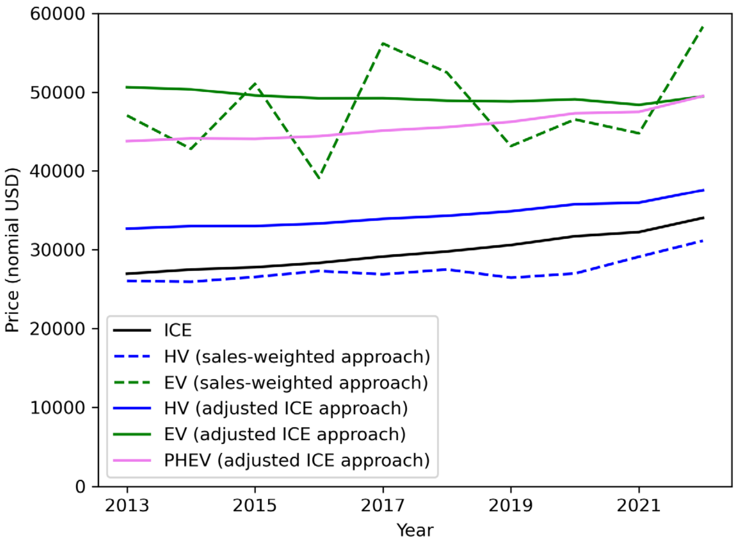
For BEV, PHEV, HEV:

- Scrape KBB for “equivalent” model MSRPs → Sales-weighted regression for X/ICE cost ratio

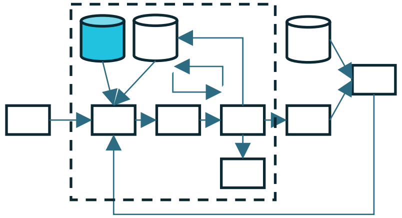






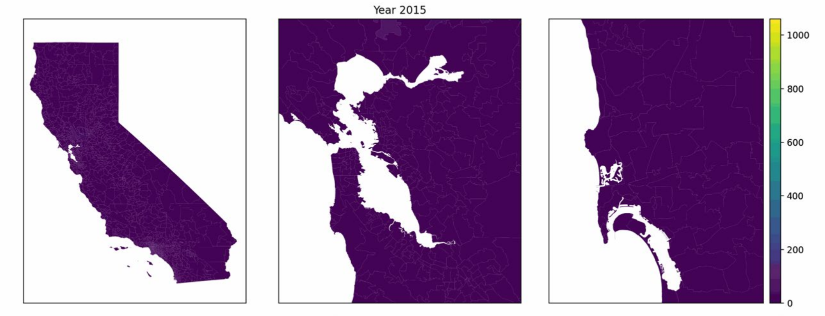


Tech & context: EV charging

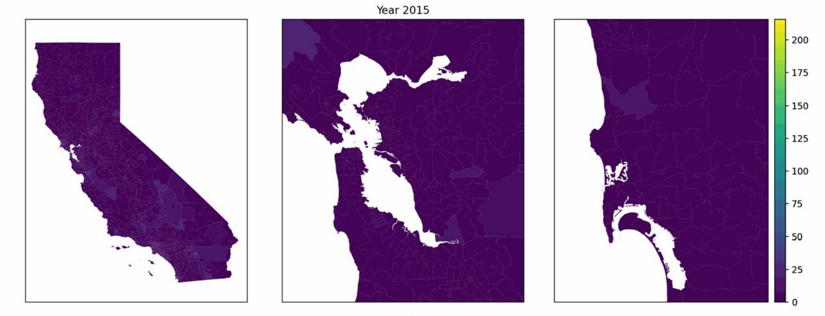


- From Alternative Fuels Data Center: https://afdc.energy.gov/
- Metric: all chargers in ZIP, neighboring zip codes (dist=1), and their neighbors (dist=2)
  - Weighted sum Σ using $w = \frac{1}{(1+d)^2}$ I.e., $w = 1$ for own ZIP, $w = \frac{1}{4}$ for neighbors, $w = \frac{1}{9}$ for their neighbors
  - Model uses $\ln(\Sigma + 1)$

L2:

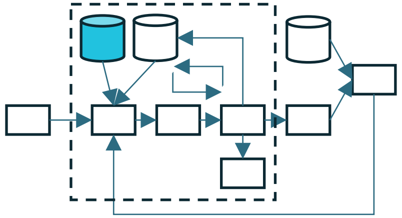


DCF:



Tech & context: capital cost incentives

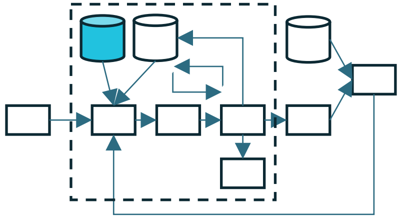
- Federal: pre-Inflation Reduction Act (IRA) federal EV tax credit and IRA EV tax credits
- State: Clean Vehicle Rebate Program (CVRP) and California Clean Fuel Reward (CCFR)
- For simplicity, assumed a single incentive below household income cap of $200,000 (from 2016 for CA, 2023 for USA)



Jurisd	Pwrtrn	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
USA	BEV	7500	7500	7500	7500	7500	7500	7500	7500	7500	7500
USA	PHEV	4000	4000	4000	4000	4000	4000	4000	4000	4000	4000
CA	BEV	2500	2500	2500	2500	2500	2000	3500	2750	2000	0
CA	PHEV	1500	1500	1500	1500	1500	1000	2500	1750	1000	0

(nominal US$)

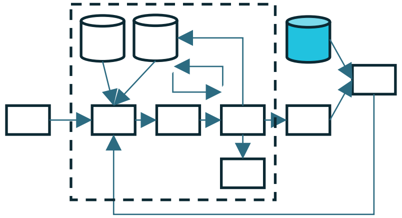
Tech & context: other vehicle attributes

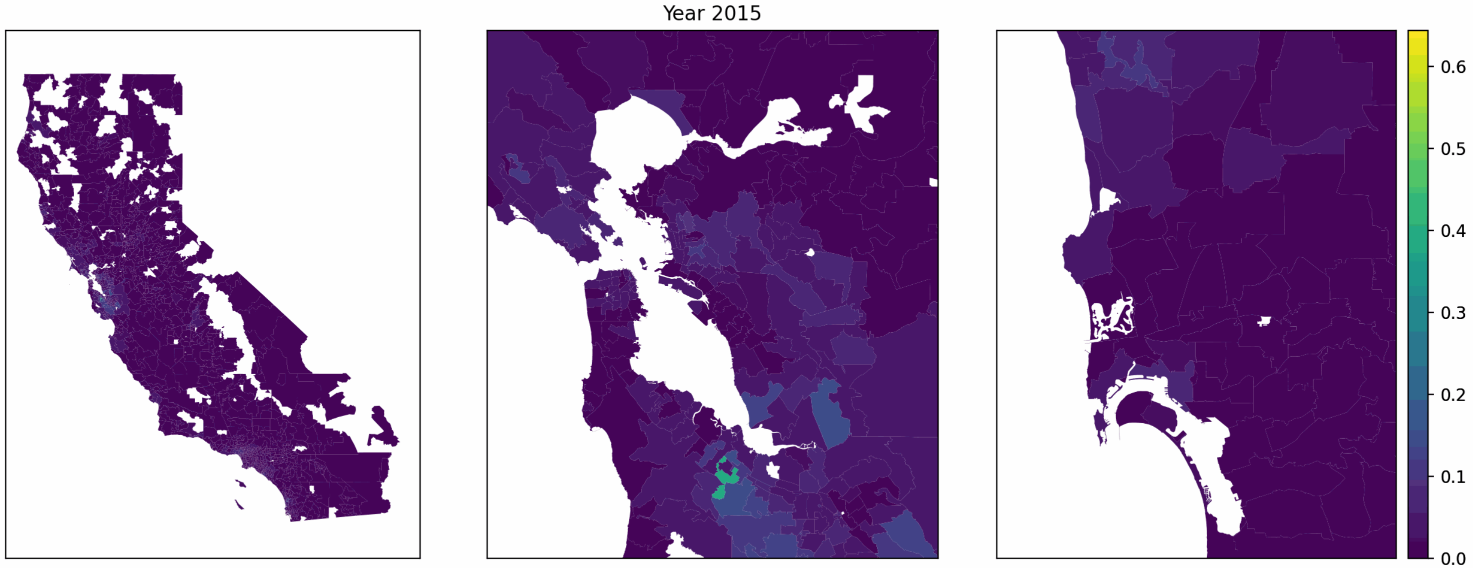


- Electricity and gasoline prices from US EIA (annual CA averages)
- Commodity fuel times: gasoline uses 10 gal/min, electricity uses exponential fit from ICCT data
  - Bauer, G., Hsu, C. W., Nicholas, M., & Lutsey, N. (2021). Charging up america: Assessing the growing need for us charging infrastructure through 2030. *White Paper ICCT*.
- Vehicle efficiency: same methodology as capital cost
- EV range: sales-weighted average range (scraped from KBB). Range for non-BEV pinned at 350 mi

Ground truth: ZIP-level vehicle sales

- Downscaled from combination of:
  - ZIP-level registrations by powertrain
  - County-level sales by ZEV/non-ZEV

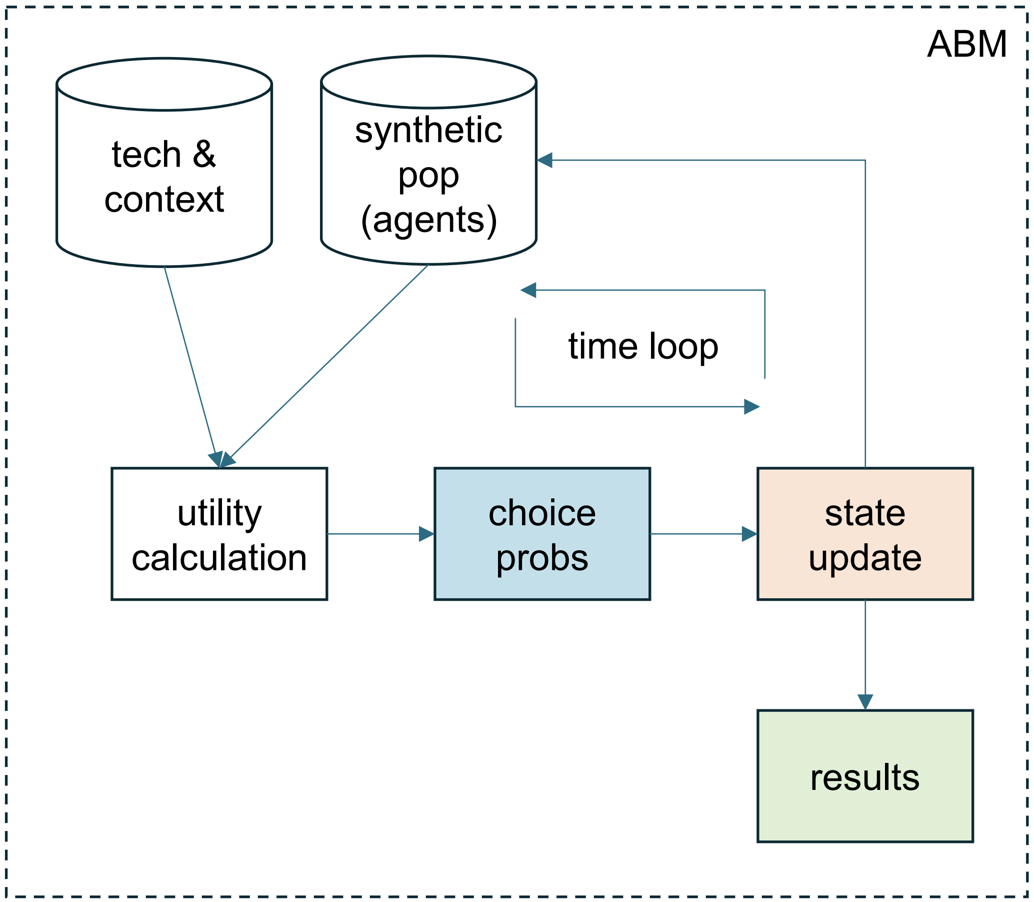


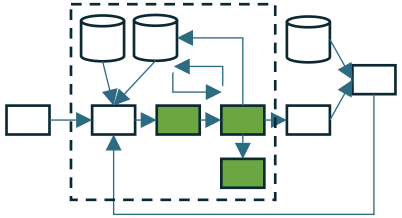


BEV sales share by zip code, 2015-2024

ABM architecture

ABM: the “simple” pieces

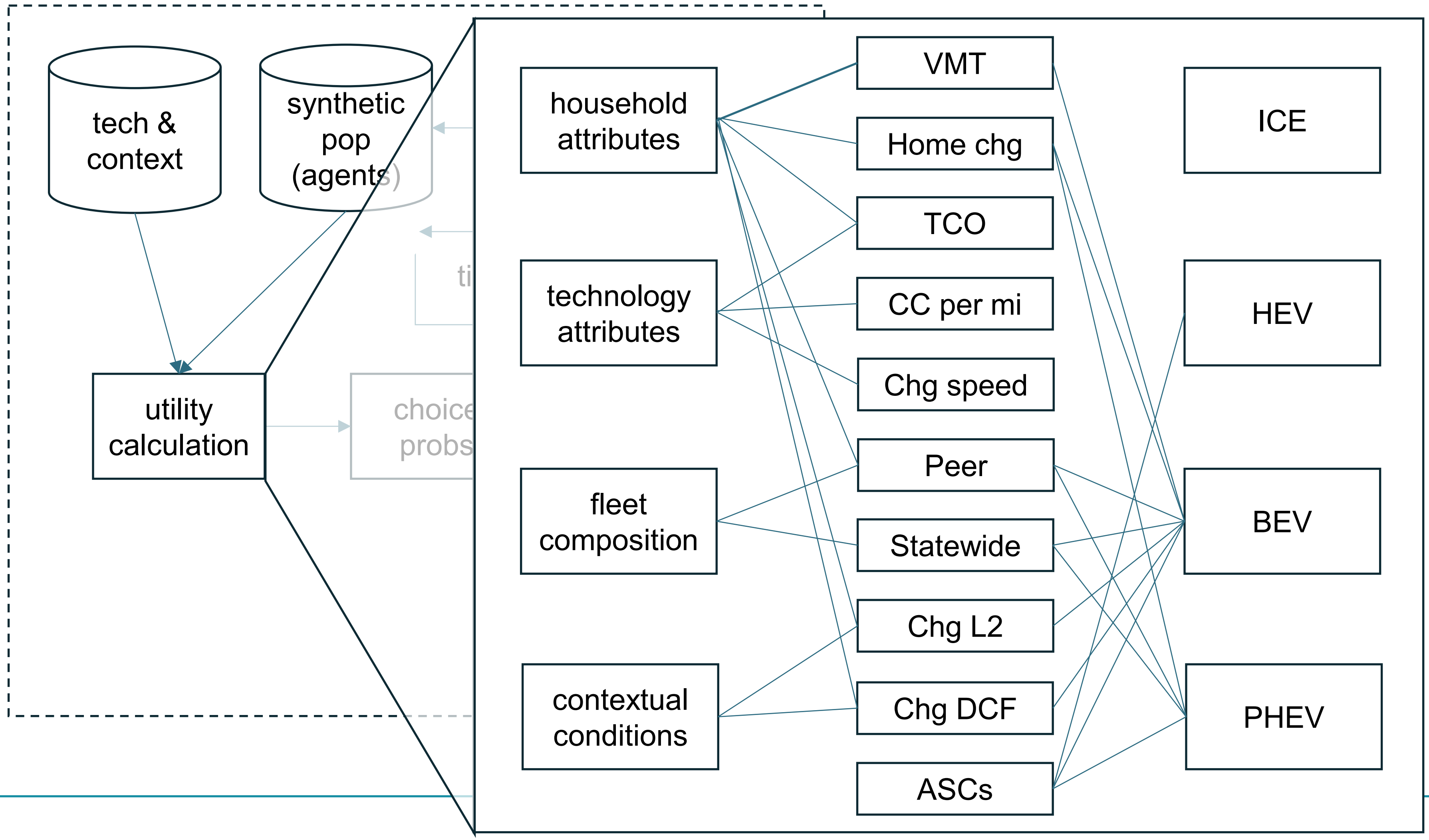


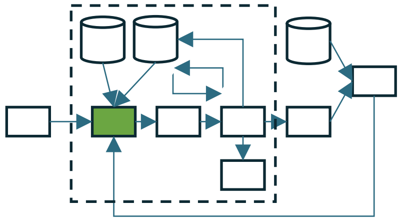


1. Per-household, per-powertrain utilities → softmax and random choice
2. Update fleet (leads to new neighborhood and statewide plug-in prevalence)
3. Summary results (e.g., annual sales or stock by powertrain)

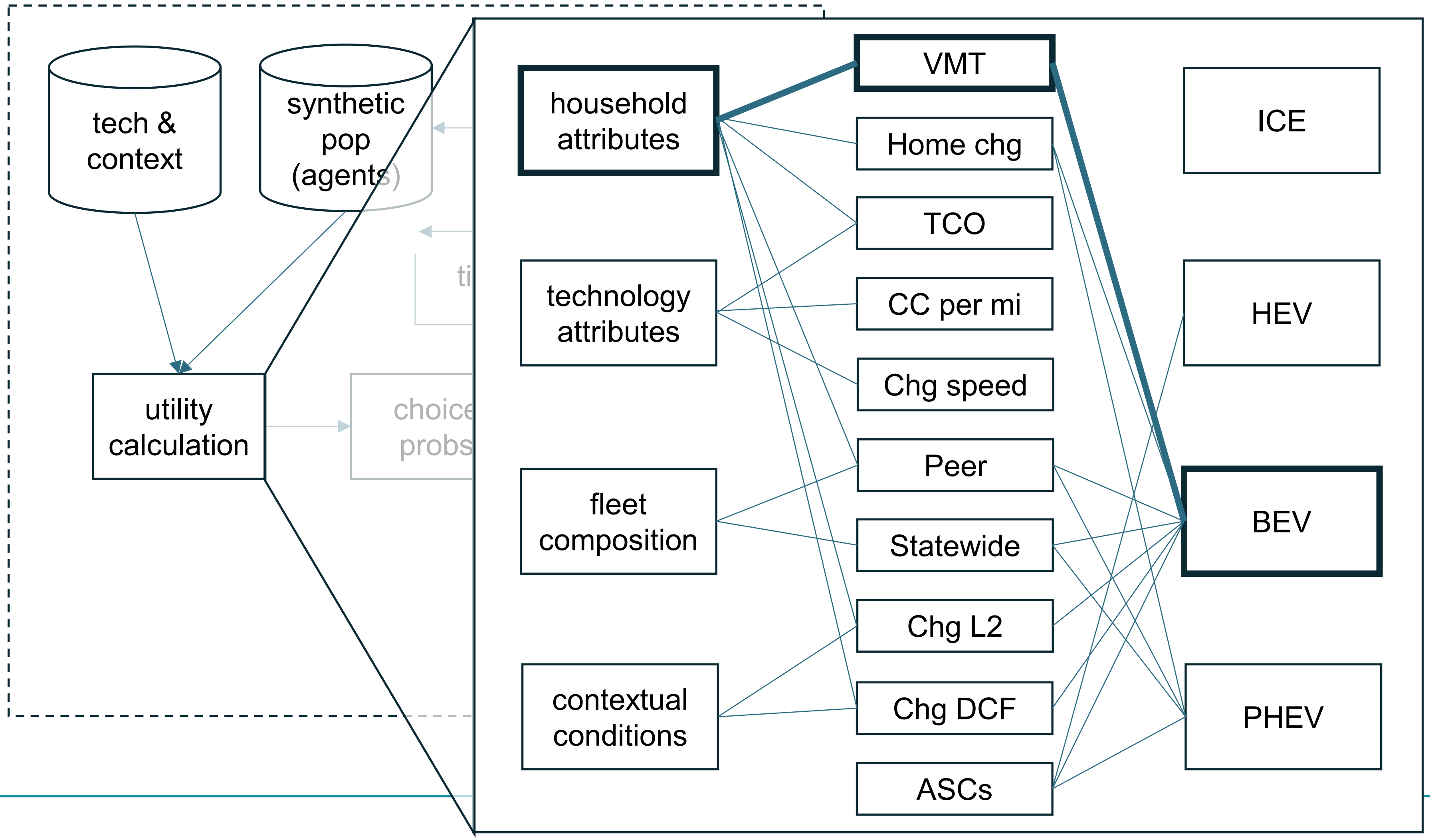
ABM: the utility function

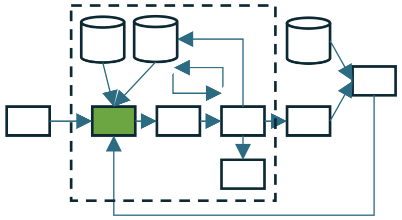
EVOLVED ENERGY RESEARCH





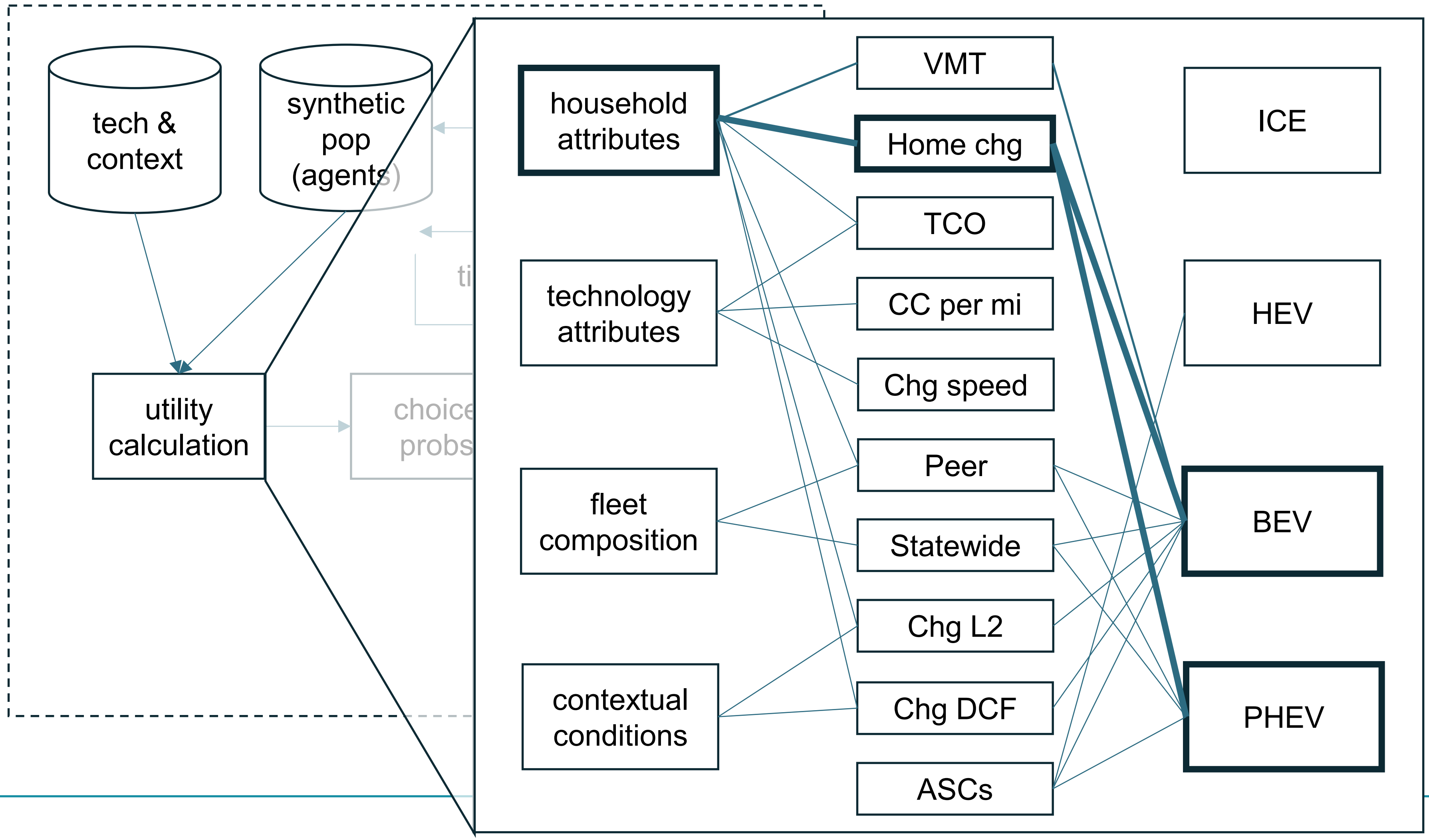
ABM: the utility function

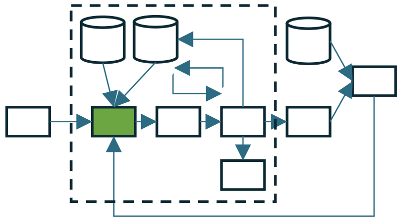




- Per CBG
- Thursday average plus Saturday average
- [H]
- BEV only

ABM: the utility function

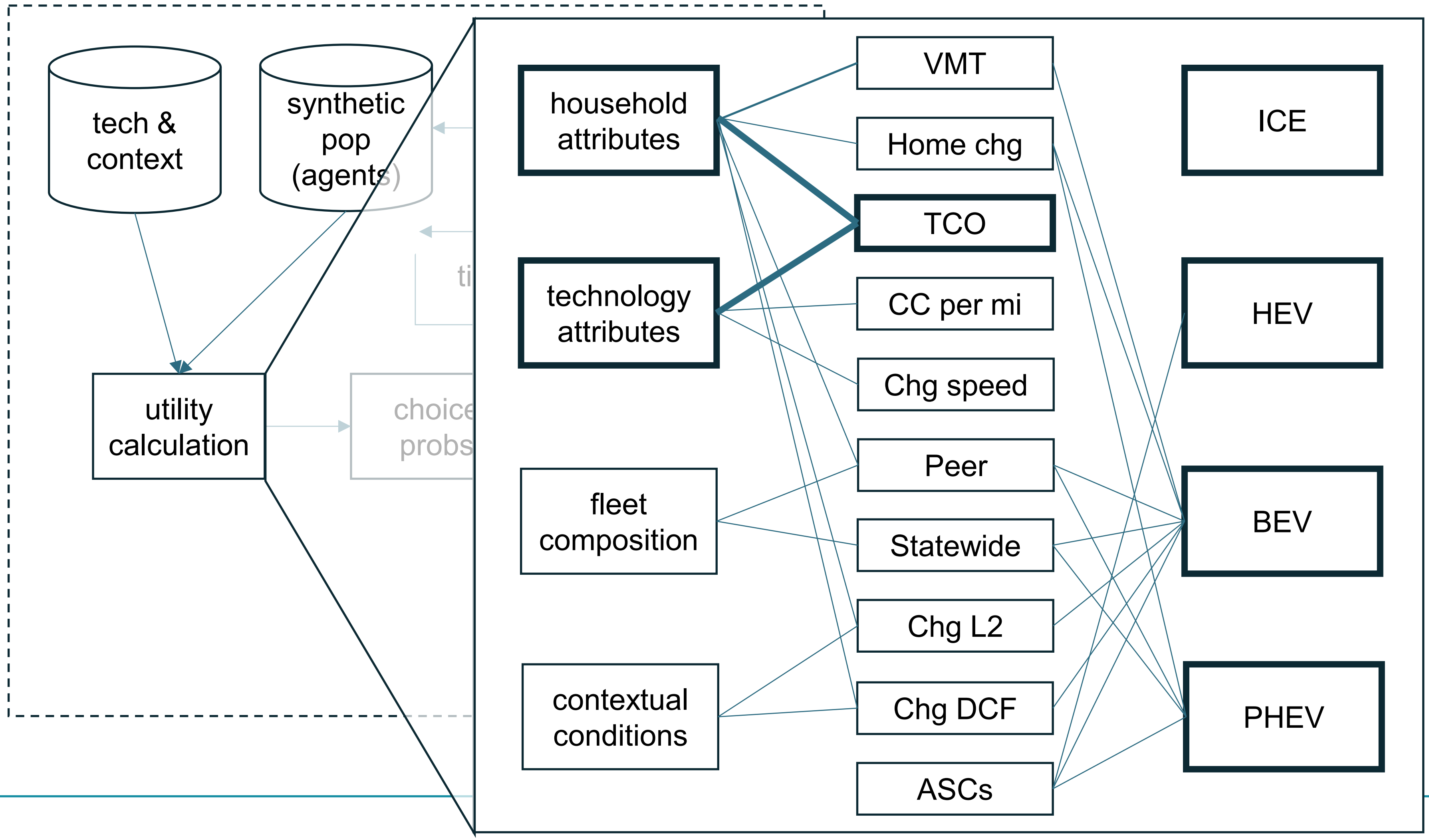


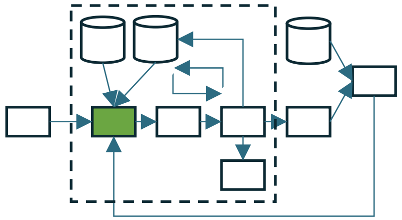


- -1 for MF and renter, 0 for MF or renter, 1 for SF and owner
- [H]
- BEV and PHEV only

ABM: the utility function

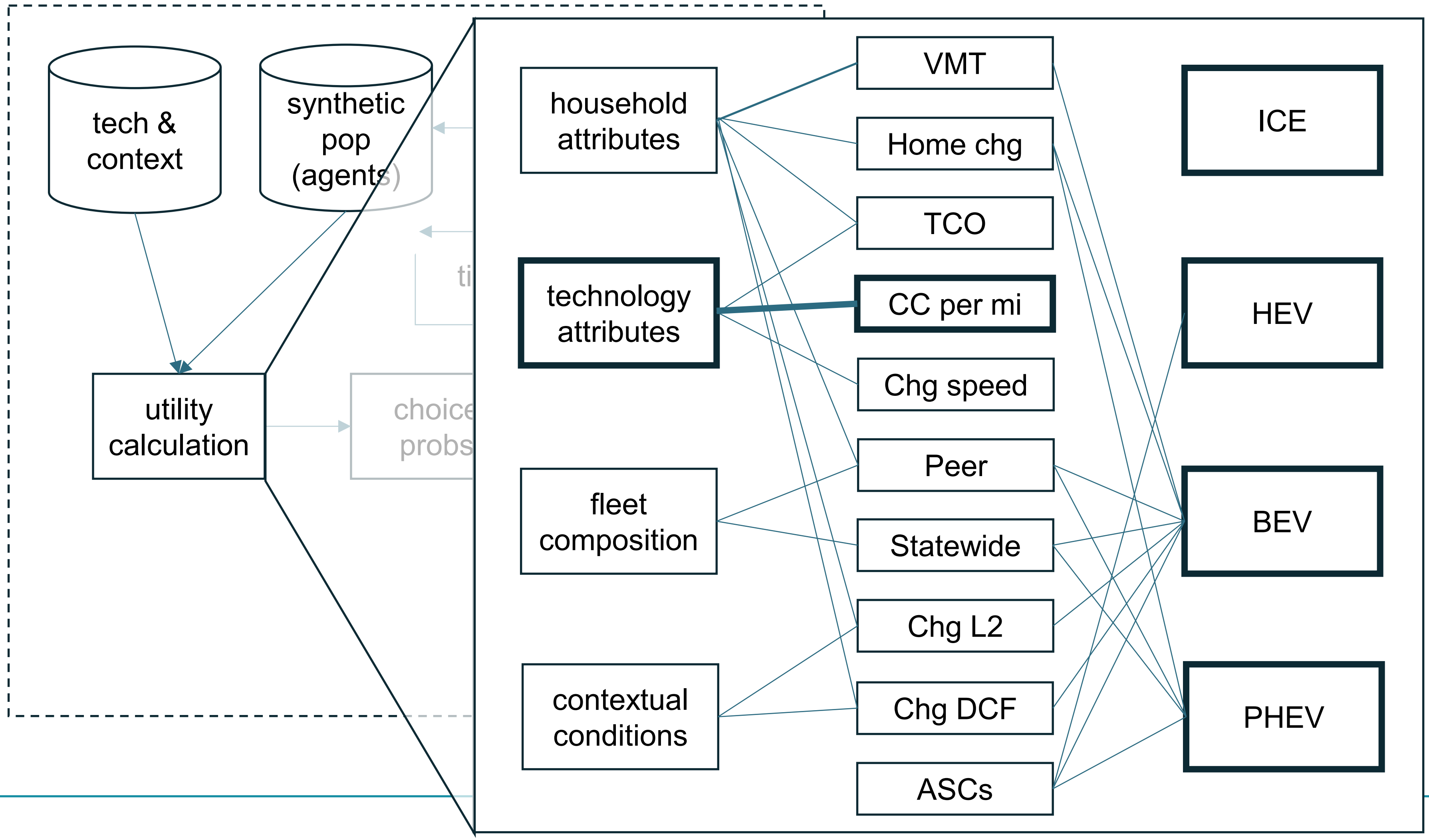
EVOLVED ENERGY RESEARCH

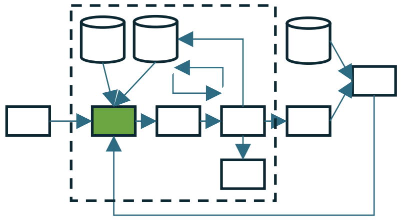




- Based on income-dependent discounting, income-dependent incentives, and HH-specific VMT
- [H x P x T]

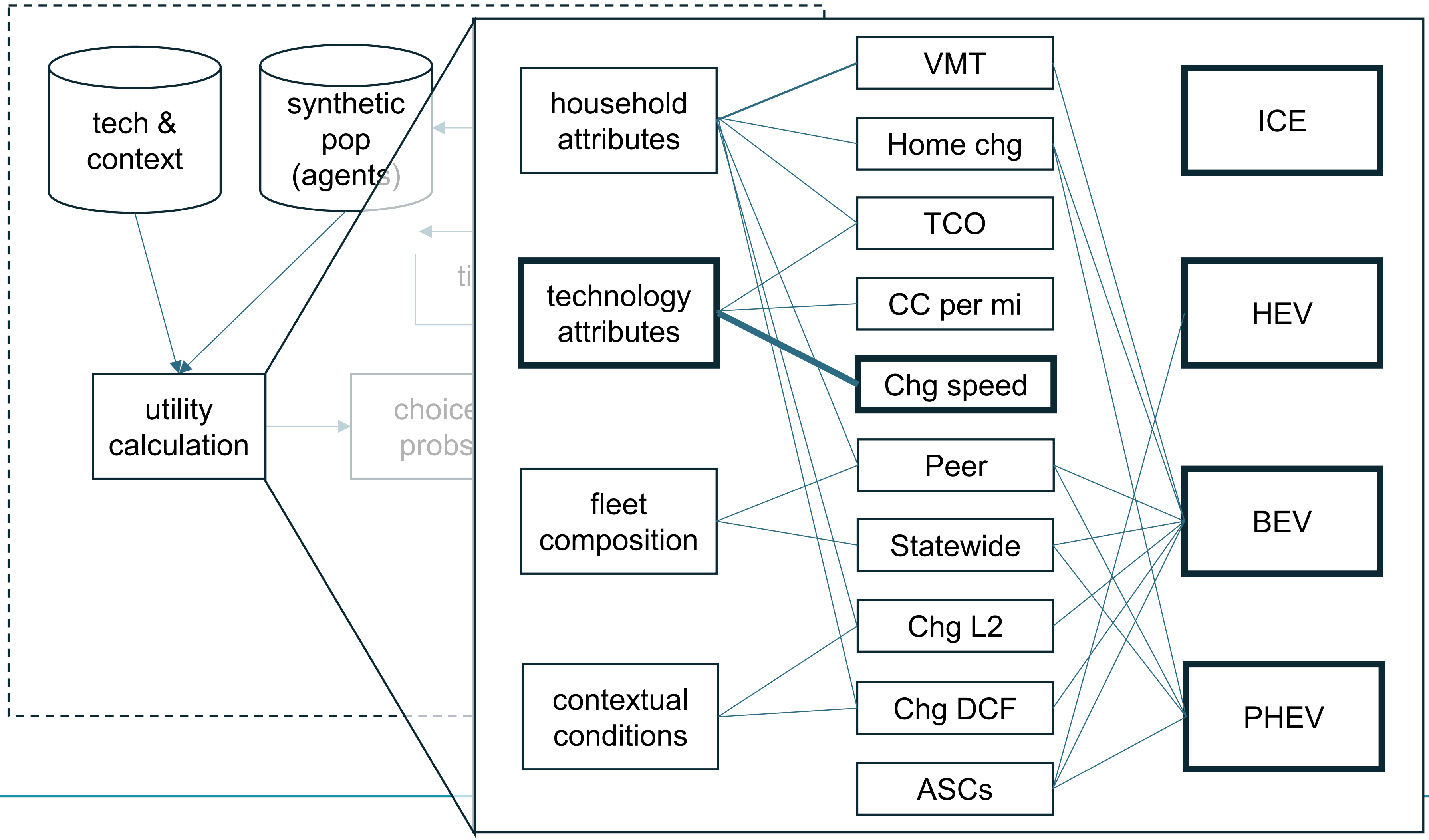
ABM: the utility function

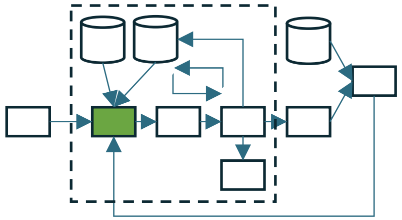




- “Value” of the vehicle
- Capital cost (without incentives) per mi of driving range
- [P x T]

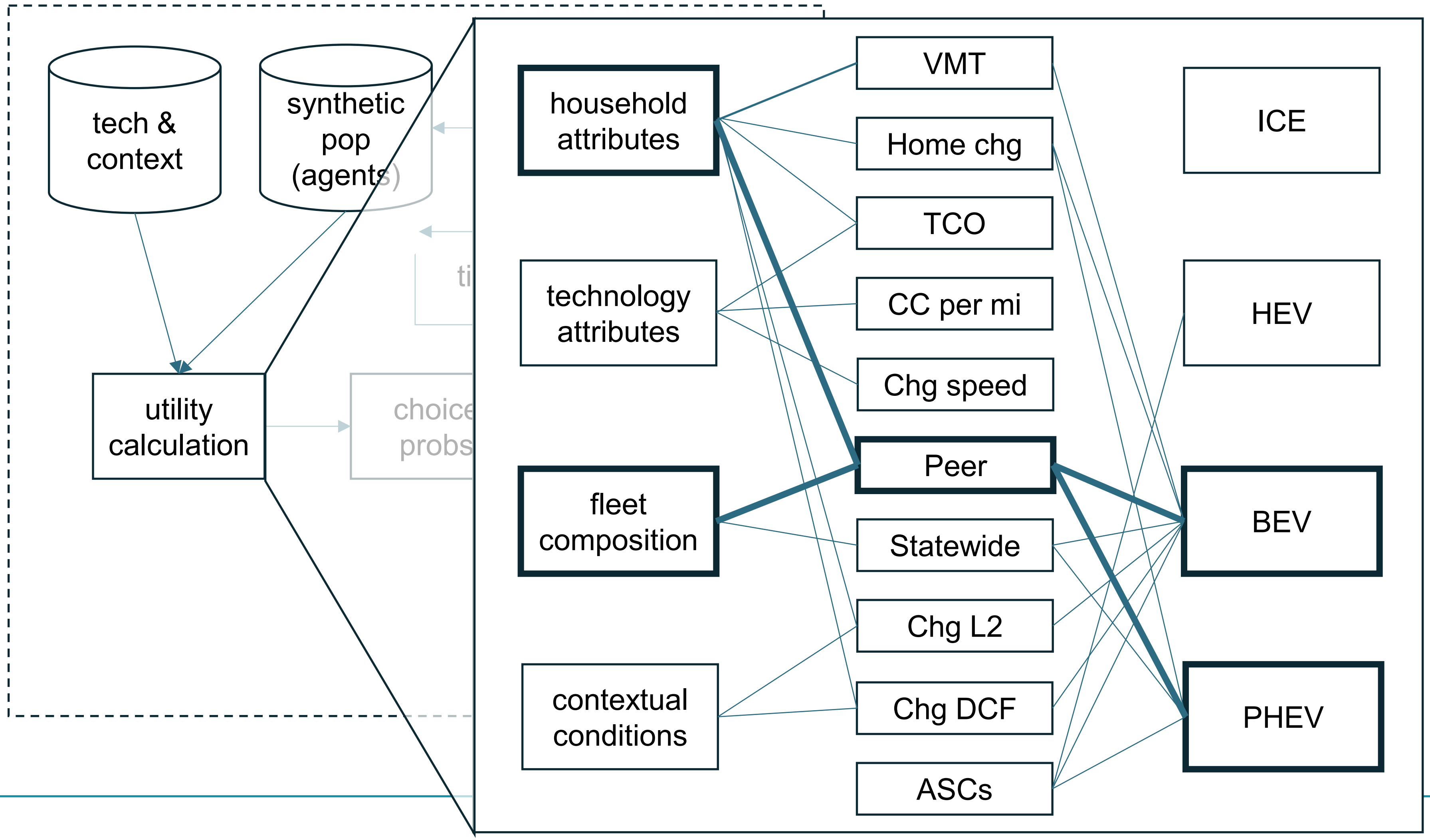
ABM: the utility function

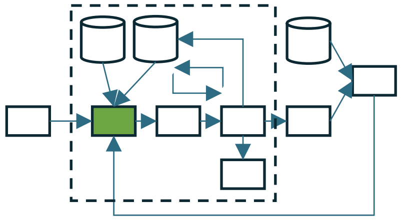




- Miles of range delivered per minute charging
- [P x T]

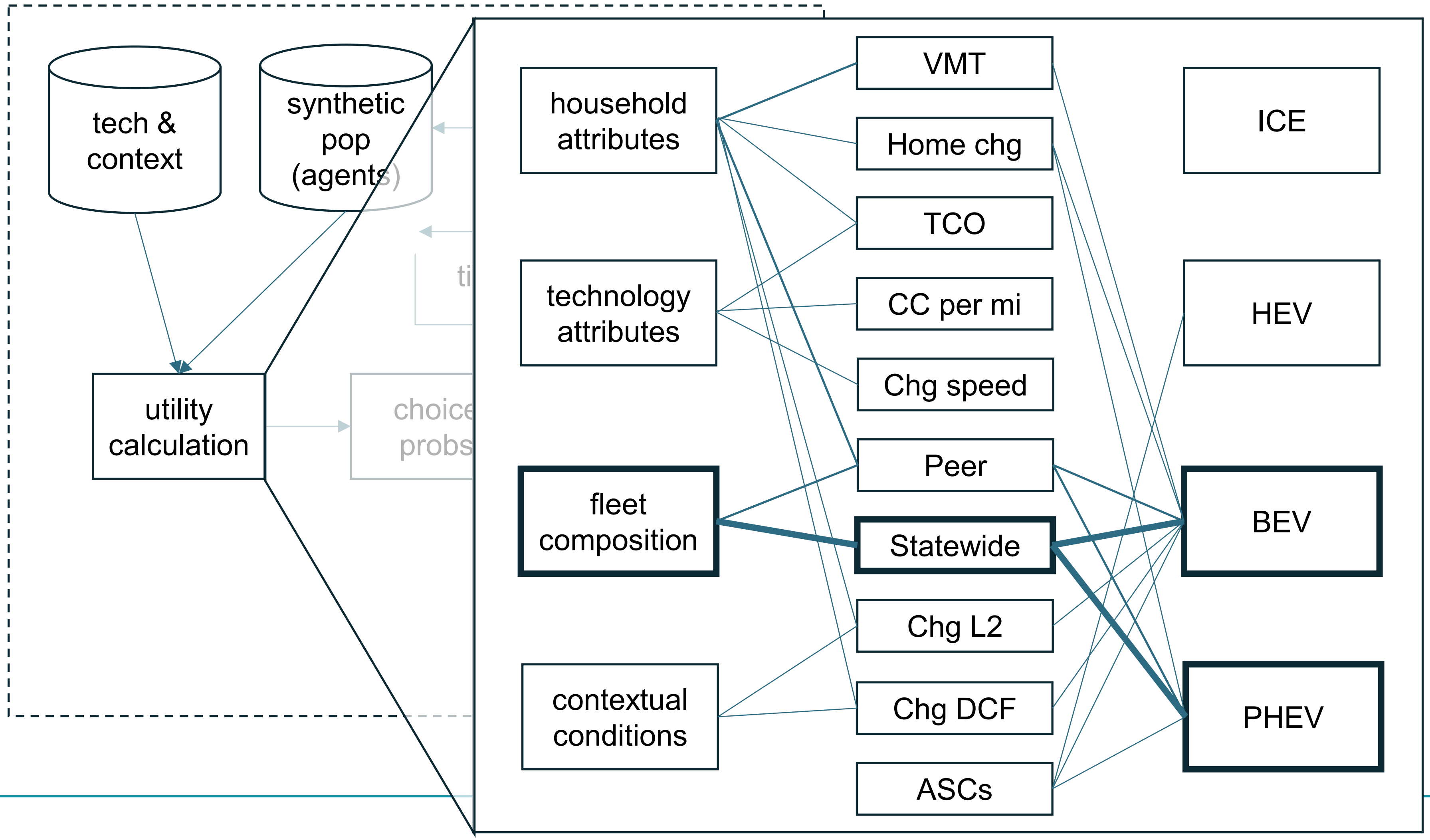
ABM: the utility function





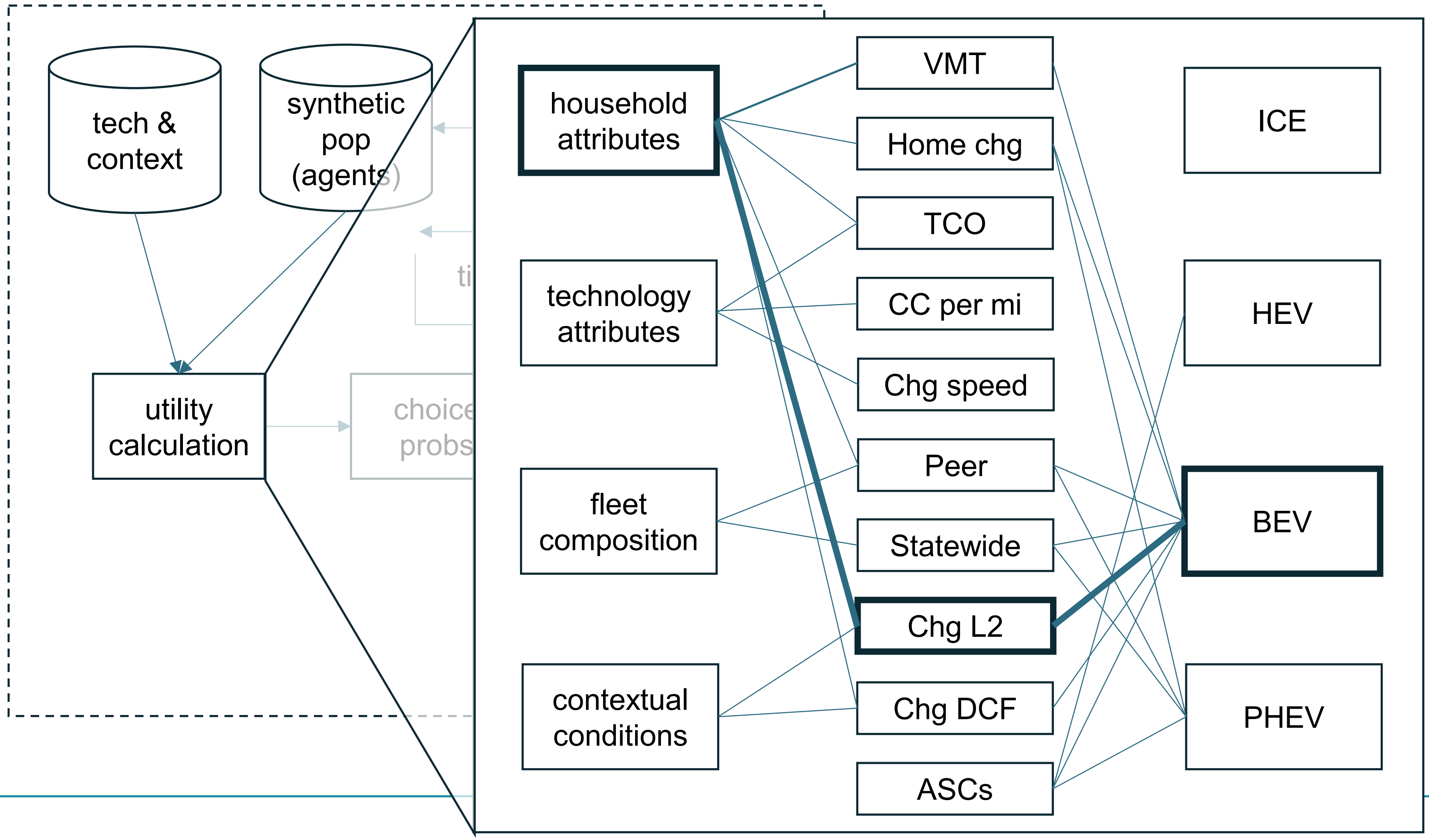
- ZIP code prevalence of plug-in vehicles (%)
- [H x T]
- BEV and PHEV only

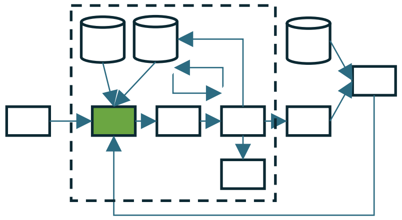
ABM: the utility function



- Statewide prevalence of plug-in vehicles (%)
- [T]
- BEV and PHEV only

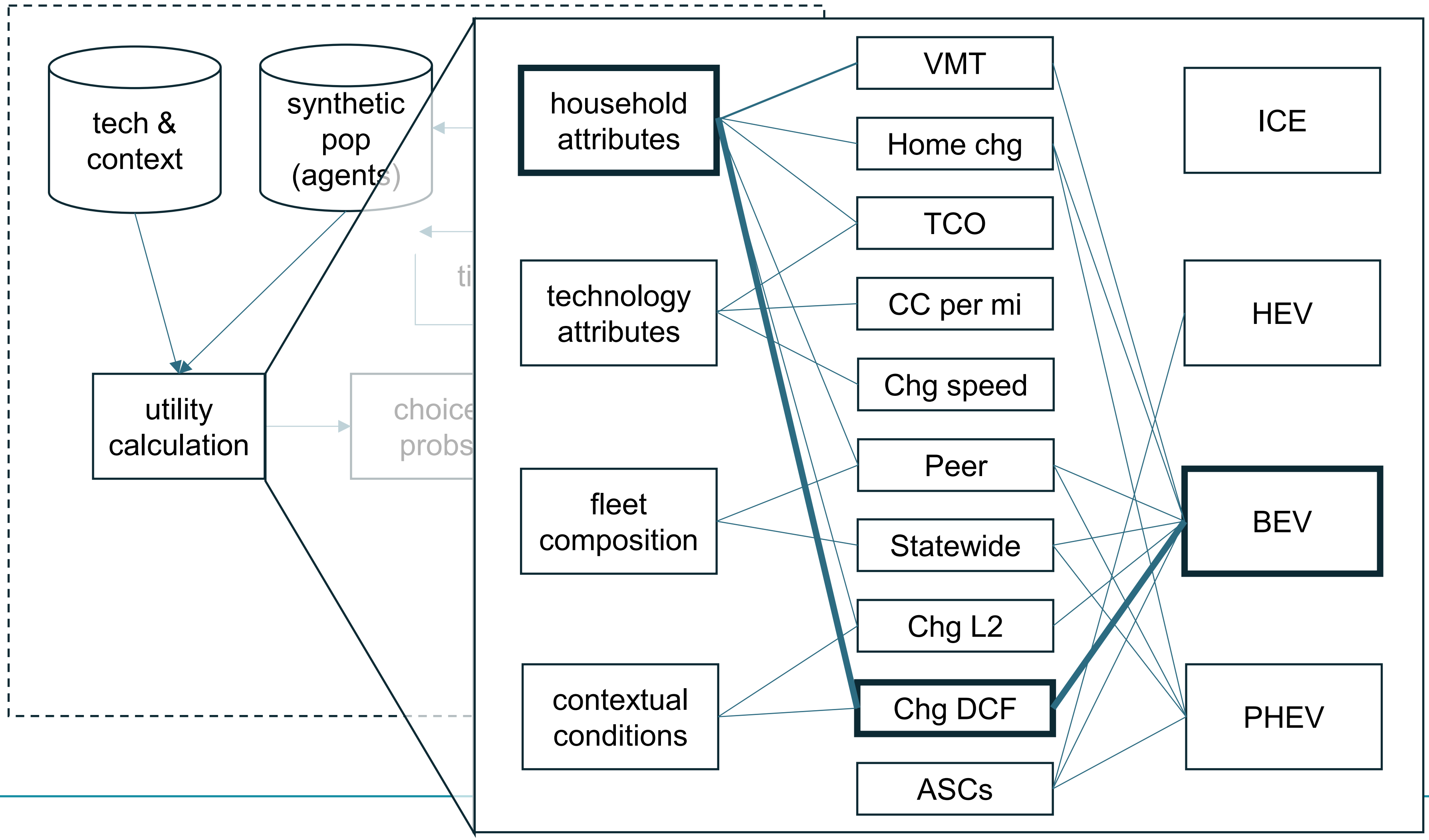
ABM: the utility function

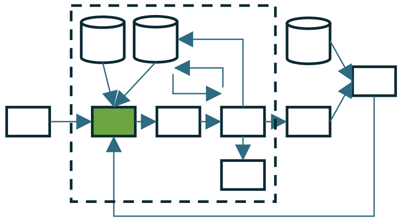




- L2 charge score (as described)
- [H x T]
- BEV only

ABM: the utility function

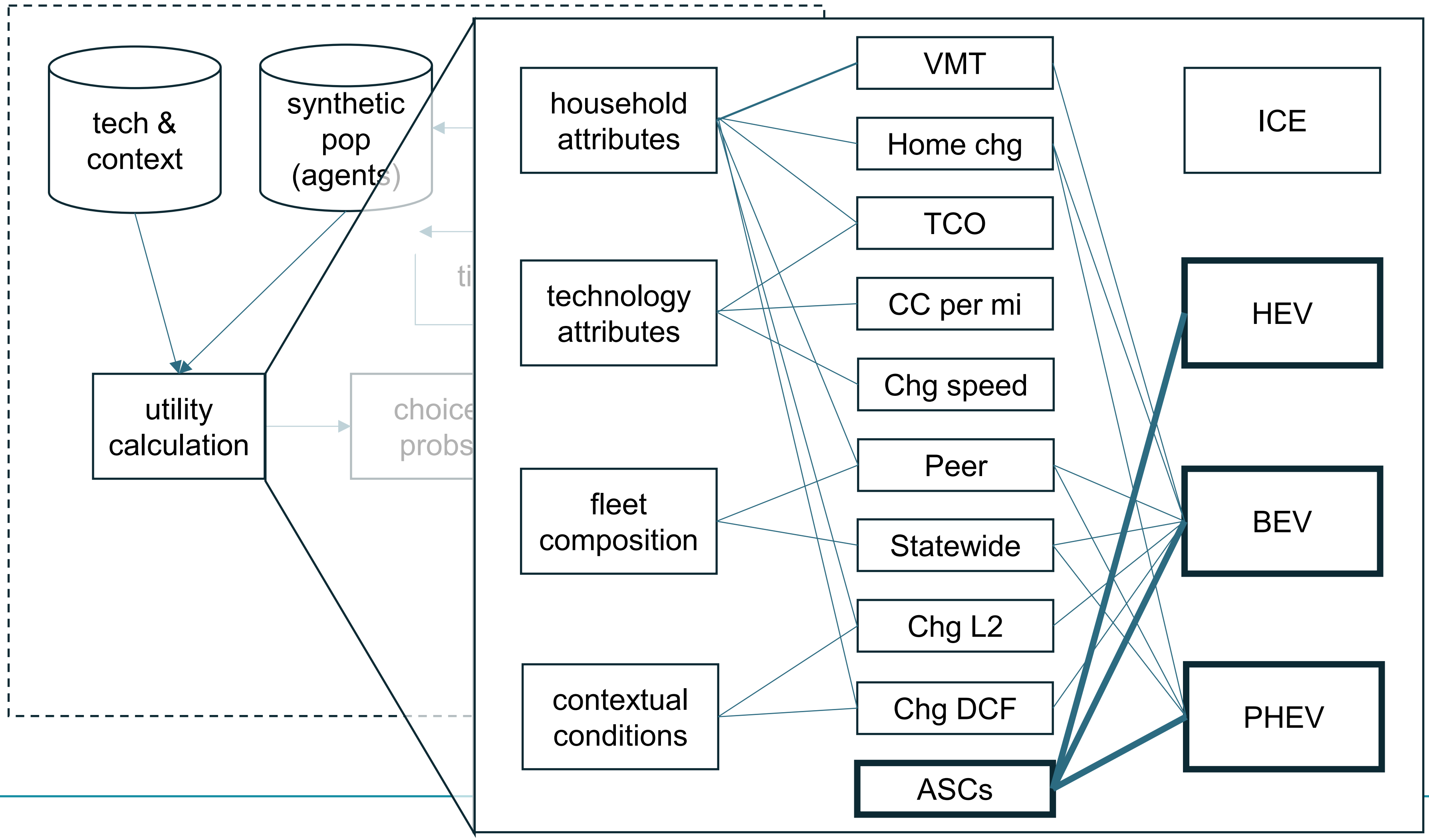


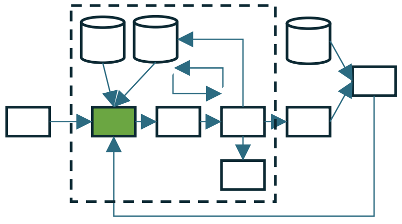


- DCF charge score (as described)
- [H x T]
- BEV only

ABM: the utility function

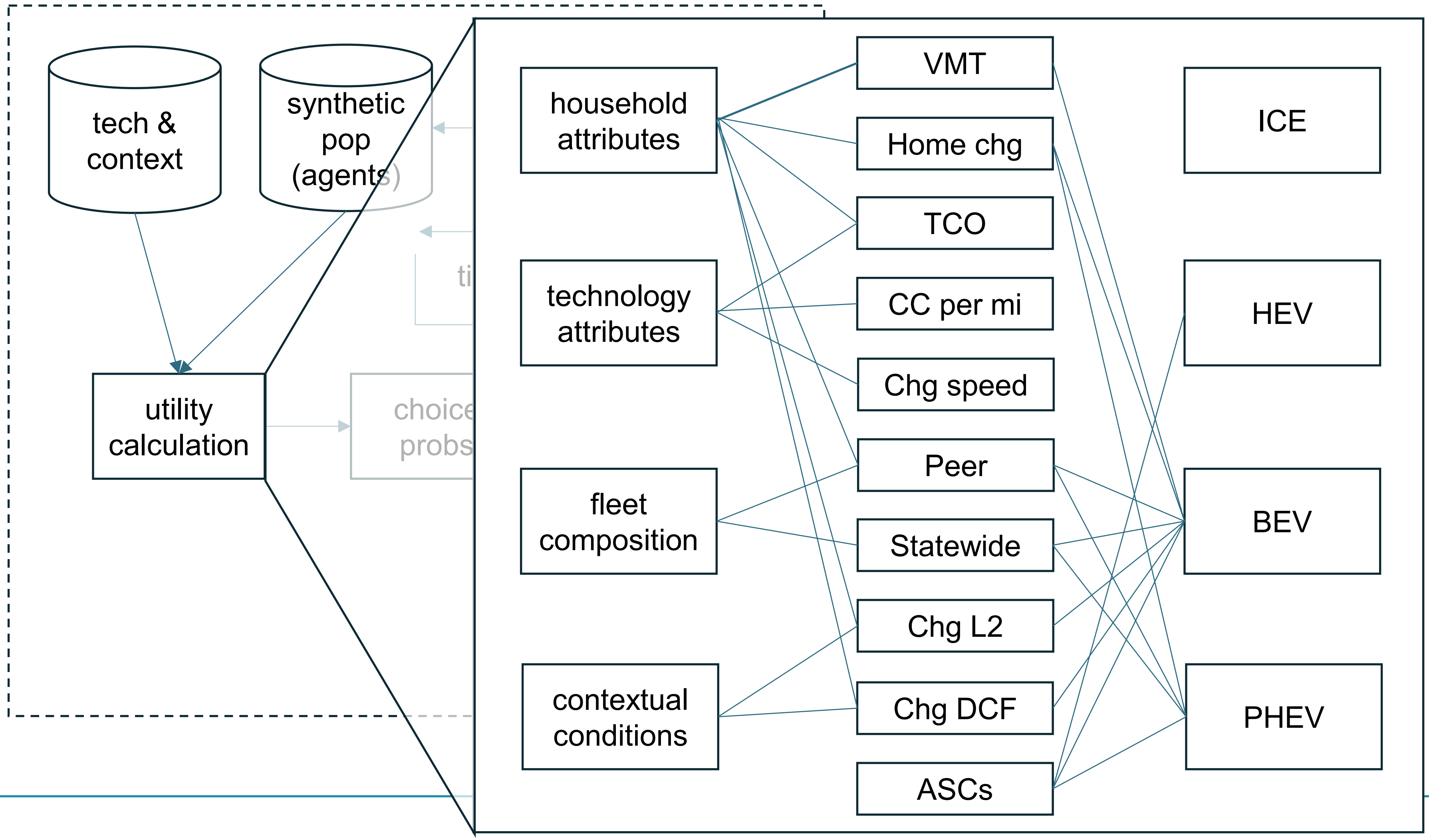
EVOLVED
ENERGY
RESEARCH

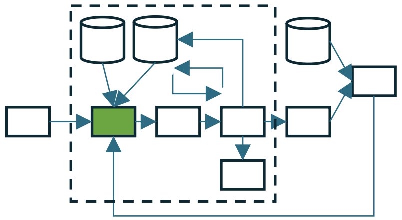




- Alternative-specific constants (think intercepts)
- [1]
- All but ICE

ABM: the utility function





- For a given year, utility = dot product of [H x P x T] tensor (where T is length of parameters θ) and θ
- I.e., θ as *importance* weights

EVOLVED
ENERGY
RESEARCH

SBI: definition, description, and challenges

SBI: What and why

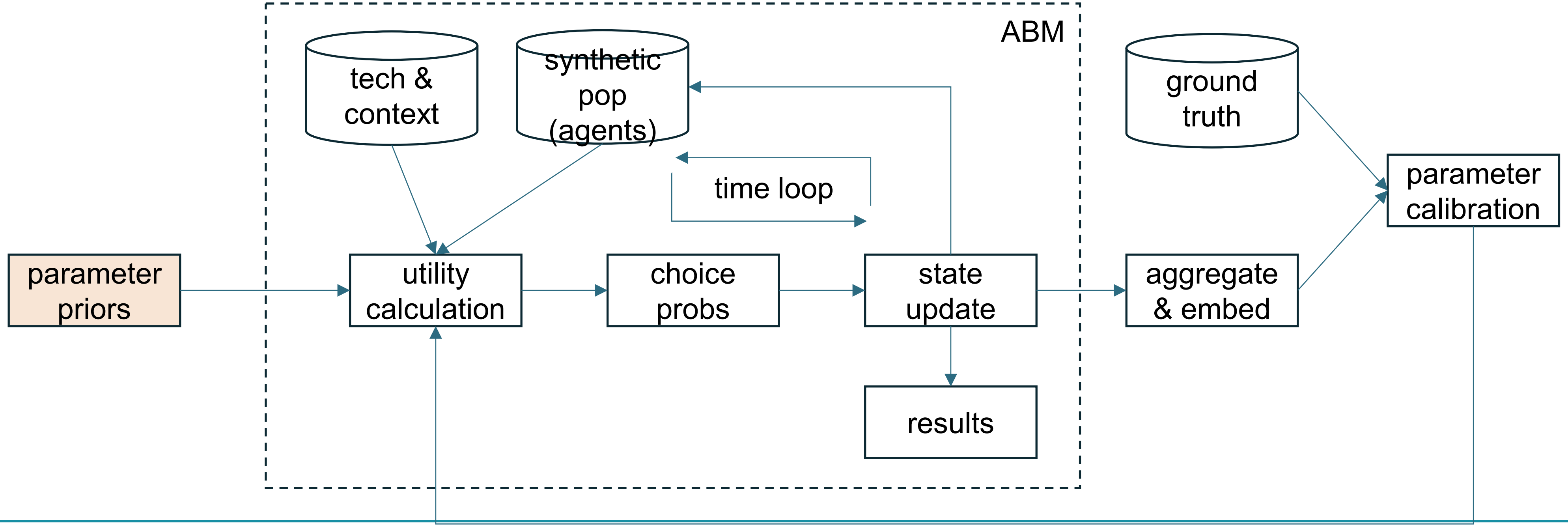
- Bayesian inference where the likelihood *p(x|θ)* is intractable but we can **simulate** data *x* given parameters *θ* – gives not only “best” parameters but entire posterior distribution
  - Recall *θ* is (at least) 12D, model is highly nonlinear and stochastic… no closed-form likelihood
  - But! We can simulate in <2sec (maybe 20sec with larger synthetic population) – “reasonable”
- The idea: “learn” good parameters *θ*, plus uncertainty quantification, by inferring from the results of a (relatively) small set of carefully-chosen simulations
- Key paper and github repo, Dyer et al. 2024:
  - Dyer, J., Cannon, P., Farmer, J. D., & Schmon, S. M. (2024). Black-box Bayesian inference for agent-based models. Journal of Economic Dynamics and Control, 161, 104827.
  - https://github.com/joelnmdyer/sbi4abm

SBI: How

1. Specify a (wide) prior $p(\theta)$
2. Simulate: draw $\theta_i \sim p(\theta)$, run ABM with θ_i to produce x_i, extract summary statistics S_i
3. Embed S_i to lower-dimensional representation s_i
4. Train a NN $q_\phi(\theta|s)$ to approximate posterior $p(\theta|s)$
5. Sequential refinement: draw new θs closer to high-posterior regions, retrain, repeat
6. Evaluate q_ϕ on (embedded) ground-truth s^* to get $q_\phi(\theta|s^*)$, sample to get parameter distribution

SBI: the simple piece

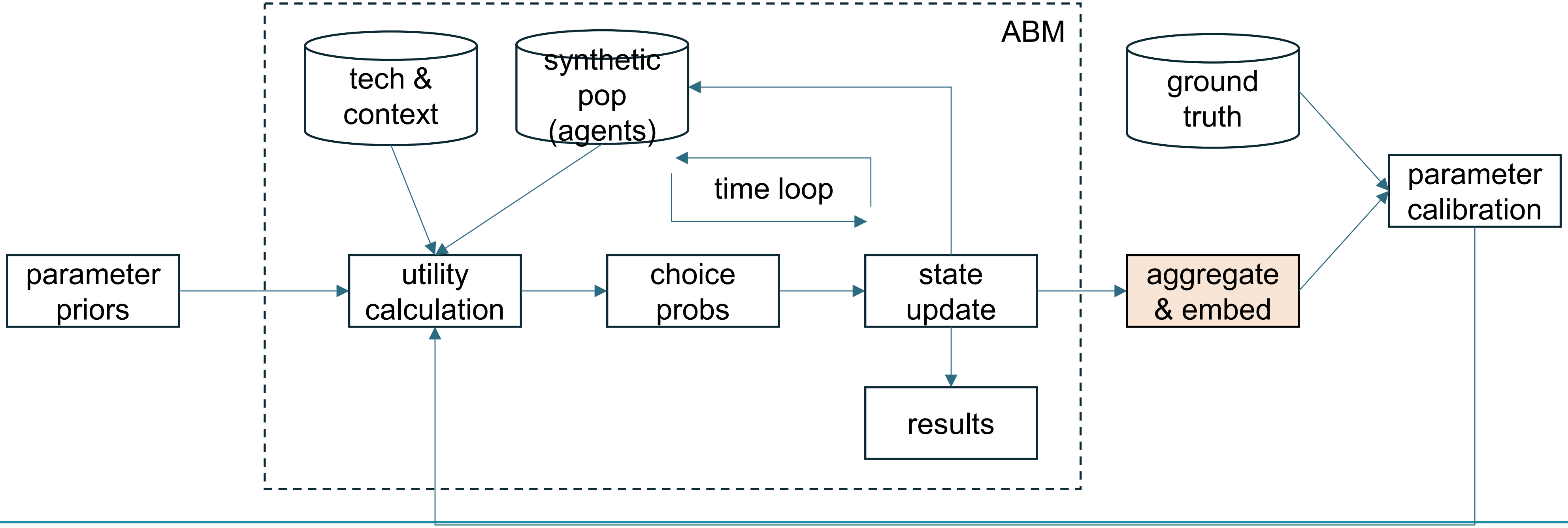
1. Specify a (wide) prior $p(\theta)$
   - BoxUniform distributions for each parameter, with reasonable (approx. [-4,4]) ranges
2. Simulate: draw $\theta_i \sim p(\theta)$, run ABM with θ_i to produce x_i, extract summary statistics S_i
   - ~1000+ simulations per parameter, multiple rounds



SBI: embedding network

Currently using a joint ZIP-state embedding network: 80,000D down to 80D

- 4 powertrain proportions x 1,600 ZIPs x 10 years + "magnitude" channel for weighting
- Joint embedding (hopefully) captures sales-weighted ZIP-level, and statewide, powertrain proportions

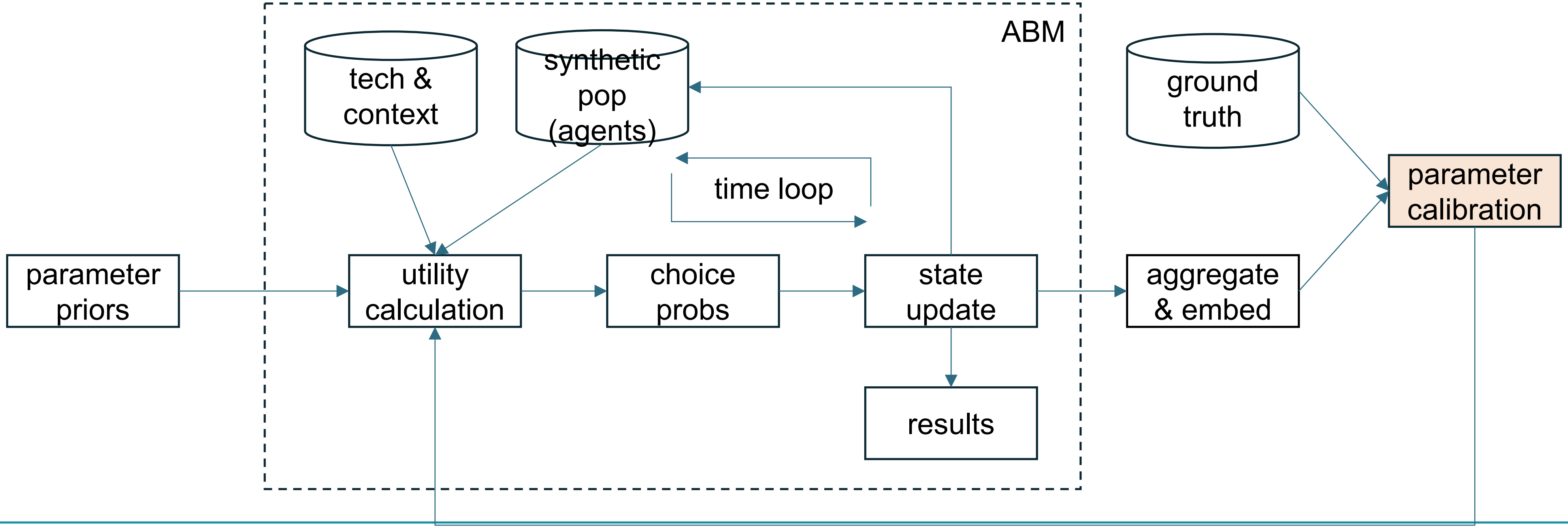


SBI: neural net for inference

EVOLVED ENERGY RESEARCH

Currently using a Normalizing Flow density estimator

- Substantial patching and safeguards for faster sampling, prior bounds enforcement, numerical stability, etc.
- Challenging to make “play well” (leakage issue: predicted posterior mass outside prior range)



SBI difficulties

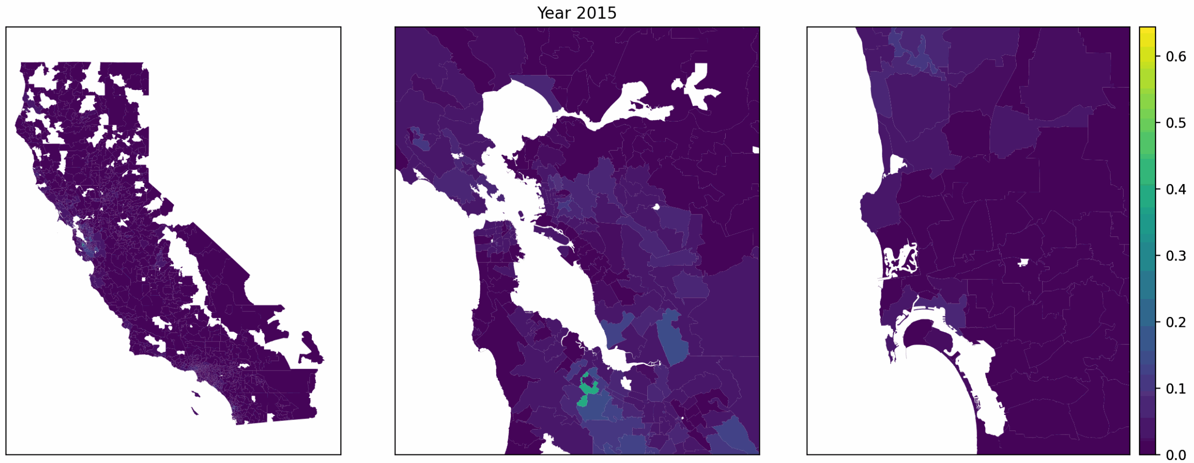
EVOLVED ENERGY RESEARCH

- SBI to empirically calibrate ABM is active research frontier
  - Methodological papers; some simpler, more expensive forms of SBI (e.g., approximate Bayesian computation) on adoption ABM; neural SBI on much simpler ABM
- Substantial progress on SBI pipeline, but still no parameter estimates I'd feel comfortable relying on for policy analysis
- "Manual" SBI: feasible θ based on (eyeball) matching statewide simulated powertrain sales shares with historical trend

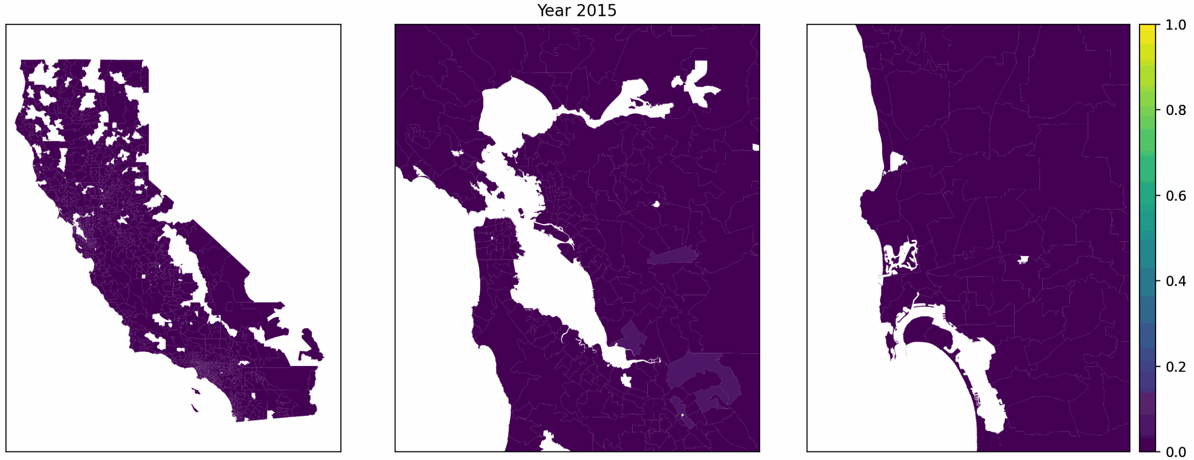
Var.	hchg	vmt	tco	cc_pr	mpm	neigh	state	l2	dcf	a_bev	a_hev	a_ph
Val.	0	-1	-1.5	-0.3	0.5	1.0	0.5	1	1	-3.5	-1.8	-2.2

Candidate θ performace

Ground truth:



ABM:

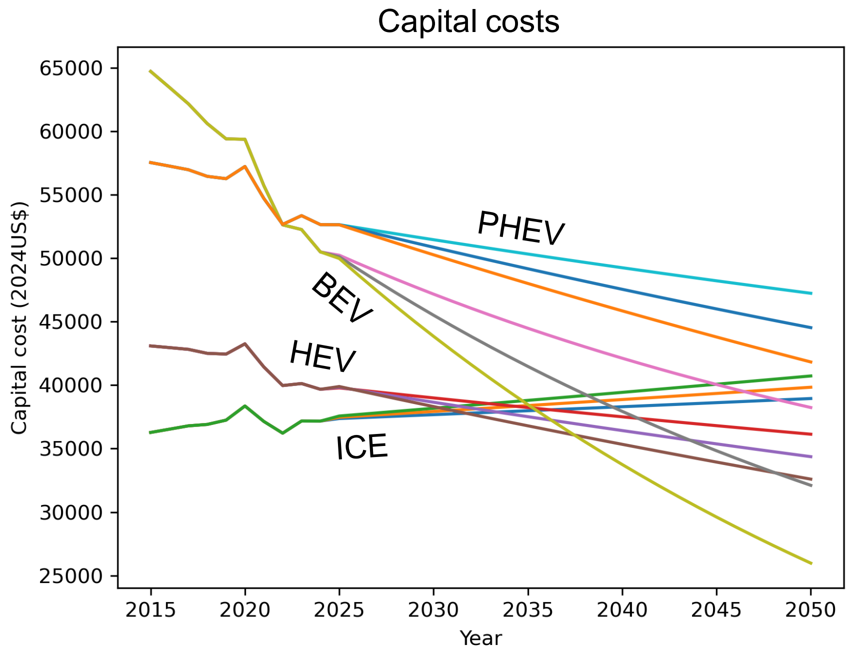


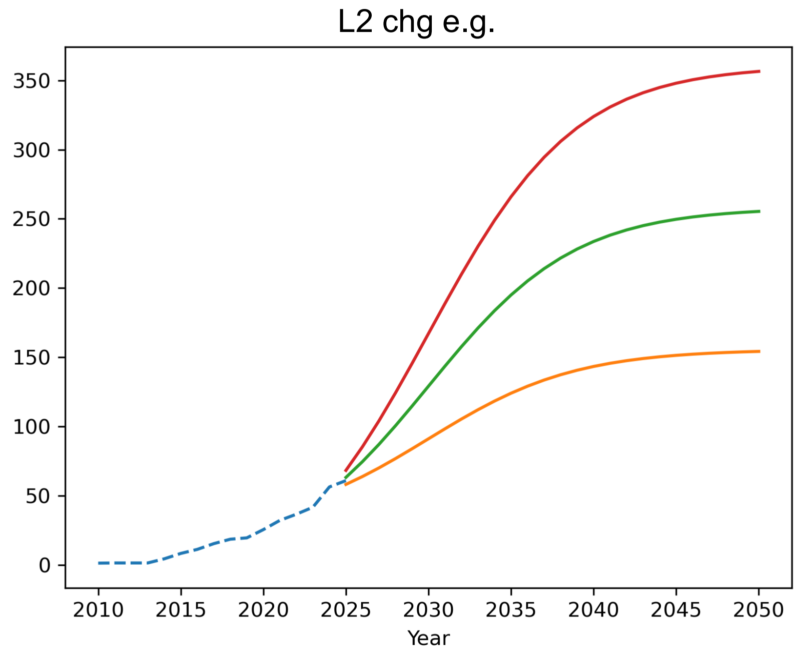
EVOLVED
ENERGY
RESEARCH

Usage example: IRA repeal simulations

Scenario analysis setup

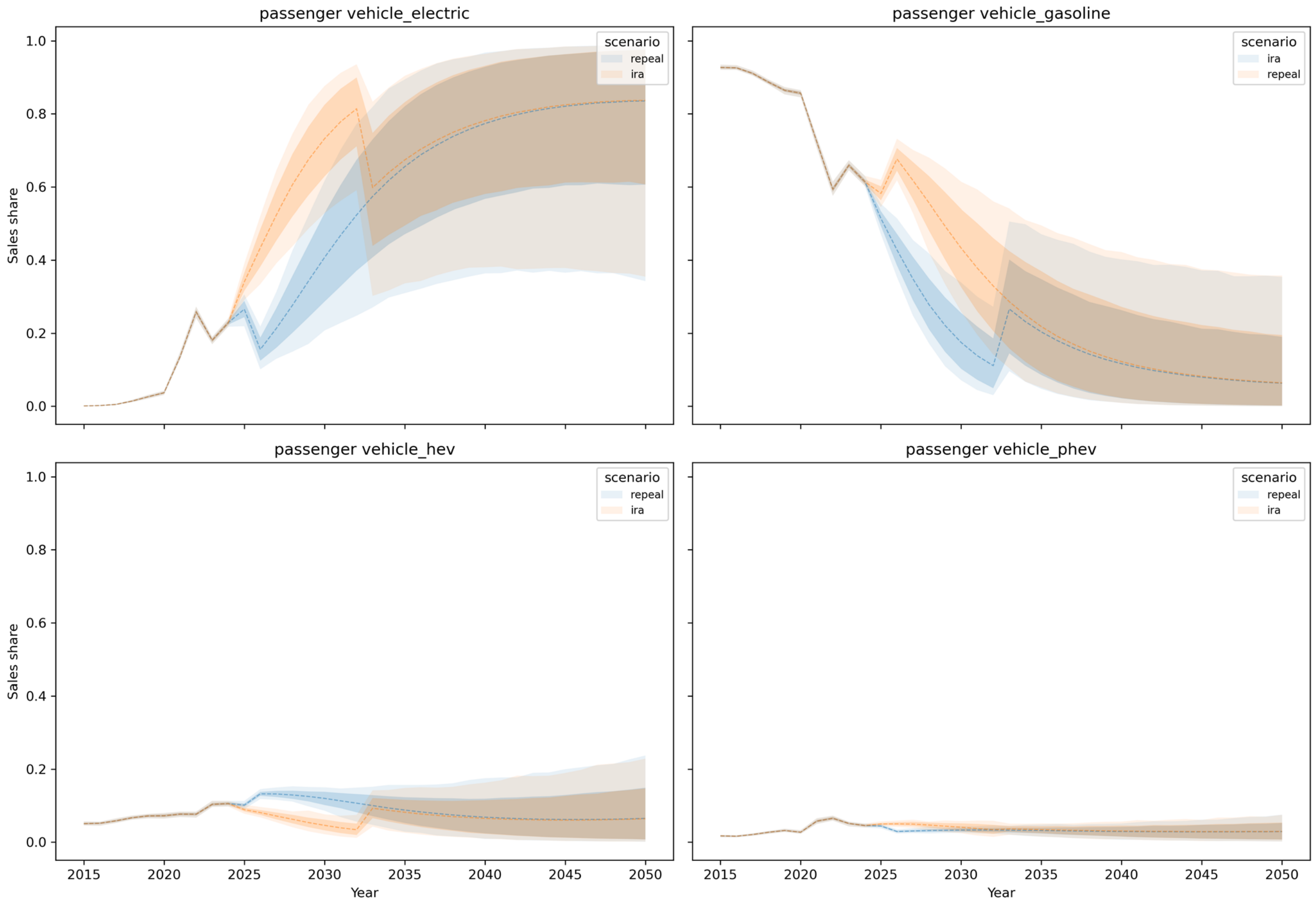
- Test impact of IRA repeal (no federal incentives beginning 2026) vs. counterfactual (federal incentives continue through 2032)
- Sensitivities: commodity cost changes, capital cost changes, EV range improvements, ICE efficiency improvements, charging network build-out



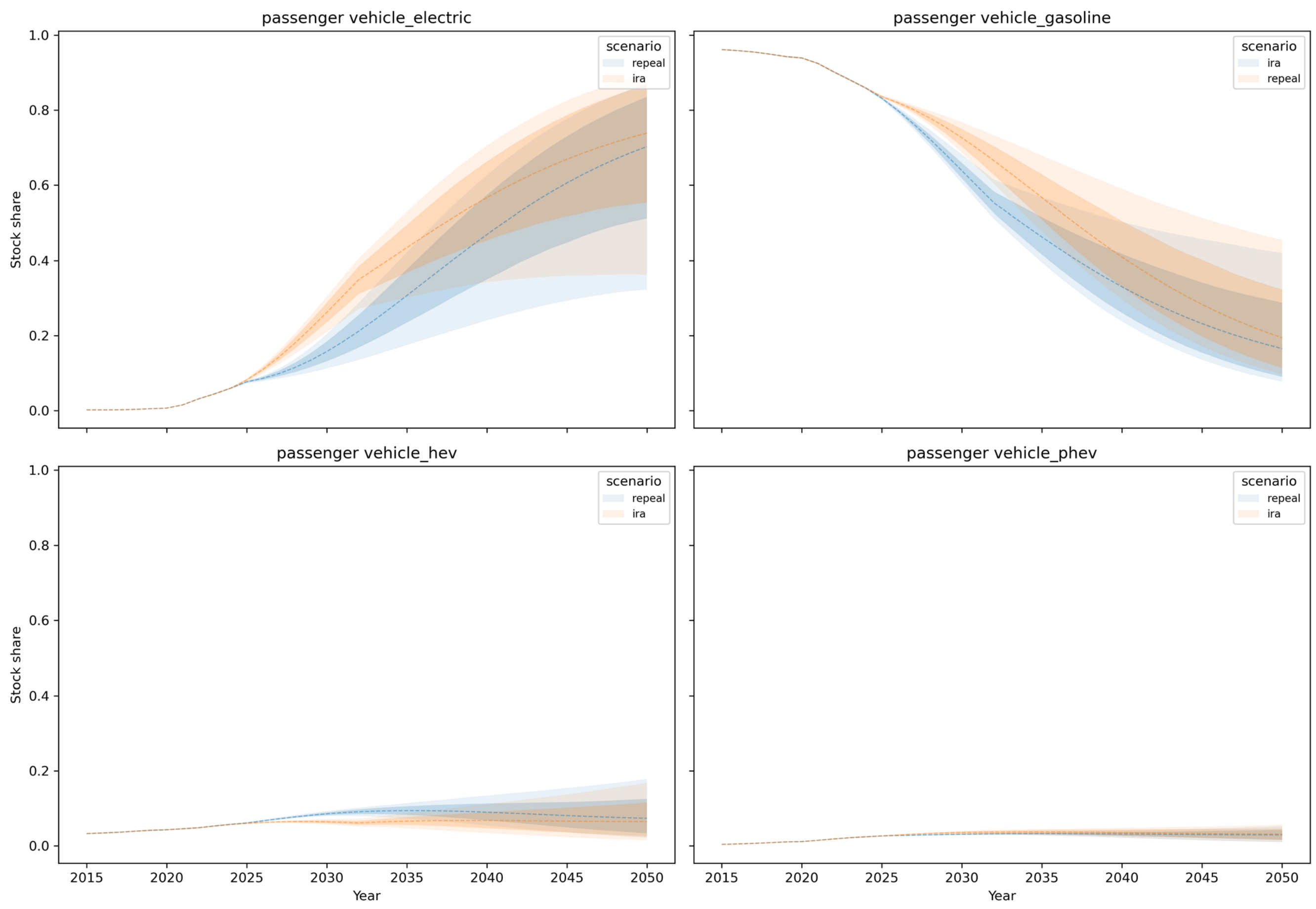


Results

- IRA repeal substantially reduces EV sales shares from 2026-2032
- (sensitivity mean, 10%-90% CI, and full min-max range are shown)



Results

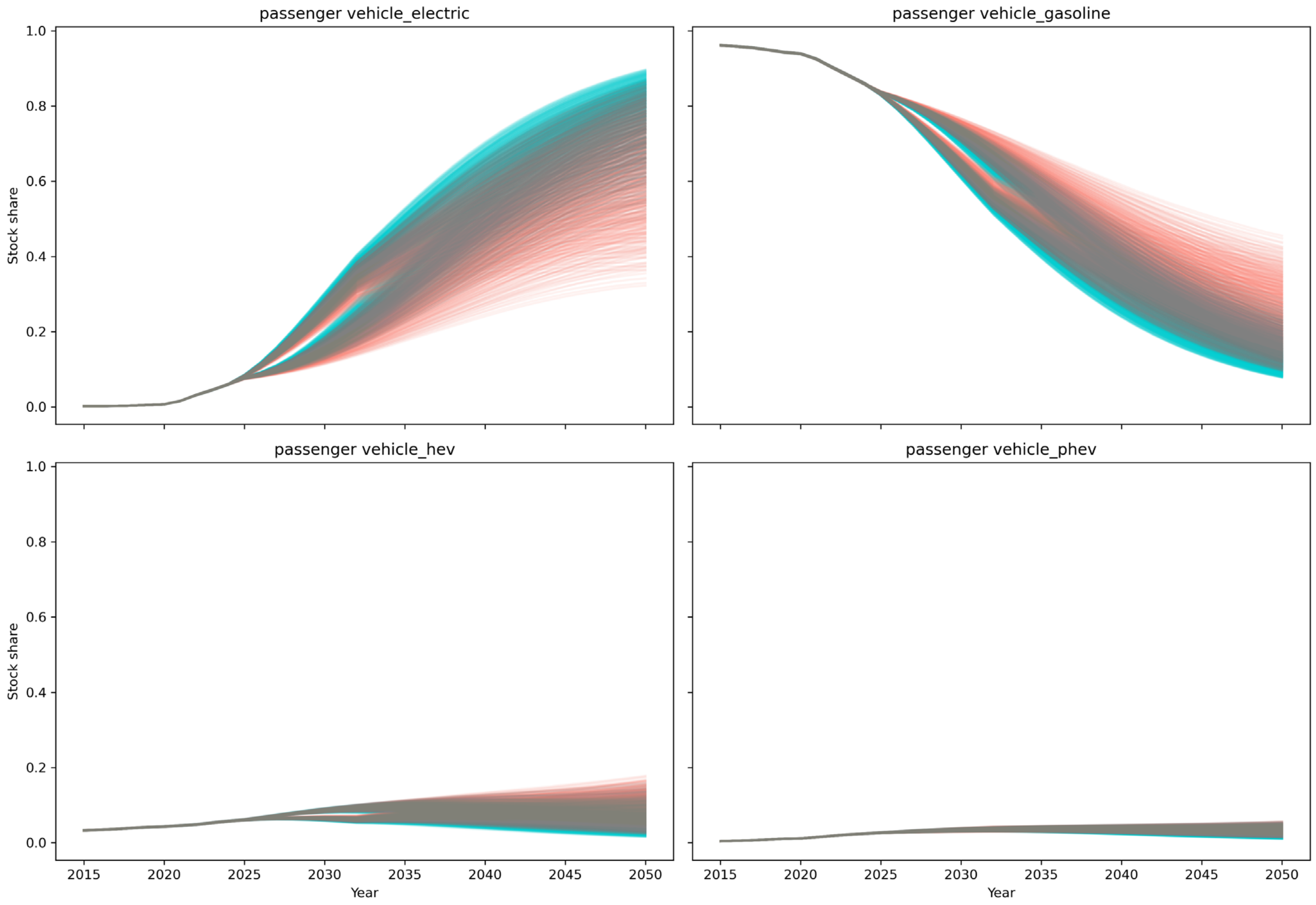


- This has a lasting impact on EV stock shares across sensitivities

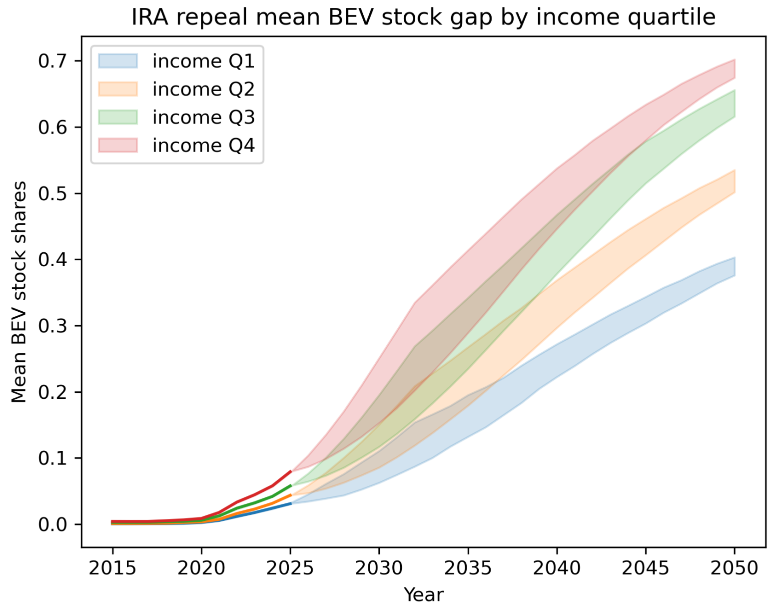
Results

EVOLVED
ENERGY
RESEARCH

- Capital cost changes (particularly EV cost declines) are the most impactful sensitivity
- Opportunity for interactive data visualizations and highly granular analysis
  - (ideally post-SBI, with uncertainty represented by multiple runs for each sensitivity, both due to stochasticity and uncertainty quantification of SBI)

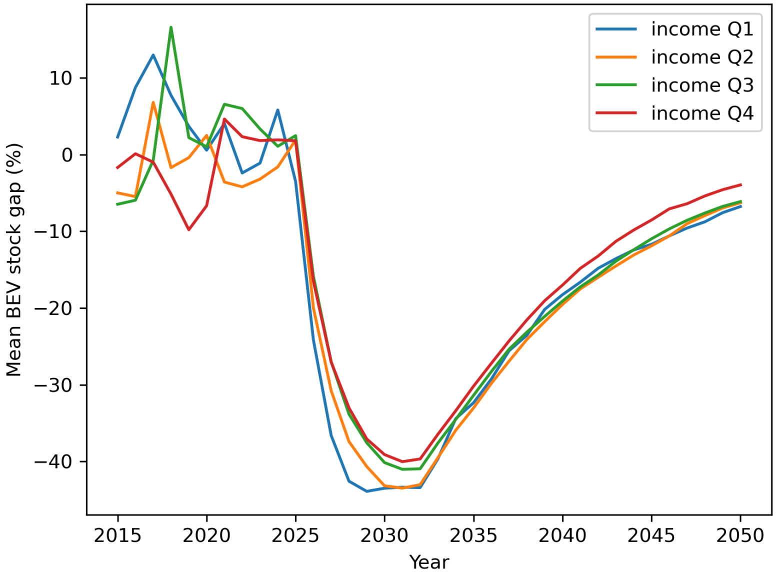


Distributional impacts



*** all sensitivities set to "med"

- All income quartiles impacted by IRA repeal
- Wealthier ZIPs seem to "rebound" to a lower BEV stock gap by 2050
    - Neighborhood effects? Can't currently rule out charging infrastructure, etc., but could look into that



Conclusions

Key takeaways

EVOLVED ENERGY RESEARCH

- ABM presents a promising, if data- and computationally-intense, avenue for modeling vehicle choice in CA
  - Strong representation of feedbacks, interactions, path-dependence, heterogeneity, and causality
  - By and large, data are available for model construction and calibration
- SBI could be used to empirically calibrate such a complex and high-dimensional model, but it is very challenging
  - Existing framework shows promise, but is not yet fully functional/reliable
  - Further work is necessary to extend to such a large-scale model context
- IRA repeal substantially reduces BEV sales shares 2026-2032, which has a lasting impact on BEV stock shares
  - Capital cost changes are a key uncertainty moving forward

THANK YOU

EVOLVED
ENERGY
RESEARCH