# Hierarchical Semi-Markov Models with Duration-Aware Dynamics for Activity Sequences

Rohit Dube<sup>1</sup>, Natarajan Gautam<sup>2</sup>, Amarnath Banerjee<sup>1</sup>, and Harsha Nagarajan<sup>3</sup>

<sup>1</sup>Industrial and Systems Engineering, Texas A&M University, College Station, USA <sup>2</sup>Electrical Engineering and Computer Science, Syracuse University, NY, USA <sup>3</sup>Applied Mathematics and Plasma Physics, Los Alamos National Laboratory, Los Alamos, USA

#### Abstract

Residential electricity demand at granular scales is driven by what people do and for how long. Accurately forecasting this demand for applications like microgrid management and demand response therefore requires generative models that can produce realistic daily activity sequences, capturing both the timing and duration of human behavior. This paper develops a generative model of human activity sequences using nationally representative time-use diaries at a 10-minute resolution. We use this model to quantify which demographic factors are most critical for improving predictive performance.

We propose a hierarchical semi-Markov framework that addresses two key modeling challenges. First, a time-inhomogeneous Markov router learns the patterns of "which activity comes next." Second, a semi-Markov hazard component explicitly models activity durations, capturing "how long" activities realistically last. To ensure statistical stability when data are sparse, the model pools information across related demographic groups and time blocks. The entire framework is trained and evaluated using survey design weights to ensure our findings are representative of the U.S. population.

On a held-out test set, we demonstrate that explicitly modeling durations with the hazard component provides a substantial and statistically significant improvement over purely Markovian models. Furthermore, our analysis reveals a clear hierarchy of demographic factors: Sex, Day-Type, and Household Size provide the largest predictive gains, while Region and Season, though important for energy calculations, contribute little to predicting the activity sequence itself. The result is an interpretable and robust generator of synthetic activity traces, providing a high-fidelity foundation for downstream energy systems modeling.

Keywords: Energy Modeling; Occupant Behavior; Stochastic Load Profile Generation; American Time Use Survey; Duration Modeling; Smart Grid Applications.

### 1 Introduction

The use of distributed energy systems, electrification of transportation with the increasing reliance on renewables are transforming the electric power grid. At the granular scales of buildings, neighborhoods, and microgrids, traditional top-down load forecasting models are becoming increasingly inadequate. Demand at the end-user level is not a smooth, aggregate signal but a highly volatile one, driven directly by the stochastic nature of human behavior; when people wake up, work, cook, or charge their electric vehicles [1, 2, 3, 4, 5]. Accurate planning and control of these edge systems, particularly for applications like demand response and local energy market design, therefore necessitate a deeper, more fundamental understanding of their primary driver: human activity.

To address this, a significant body of research has focused on "bottom-up" activity-based models. These works simulate residential electricity demand by first simulating what people are doing, by using Markov modelling in [6, 7] and bootstrap sampling in [8]. A successful model in this domain must overcome three central challenges: 1) realistically predicting the sequence of activities (what comes next?); 2) accurately capturing their duration (for how long?); 3) maintaining statistical stability even when segmenting the population by demographic factors to build Markov matrices, which often leads to severe data sparsity.

This paper develops a unified framework that simultaneously addresses all three challenges. We propose a hierarchical semi-Markov hazard model that synthesizes a time-inhomogeneous Markov router for sequencing, a hazard-based component for durations, and a hierarchical structure for managing data sparsity while building Markov matrices. Using this model, we conduct a systematic, data-driven evaluation to quantify the effects of hazard and which demographic factors are most critical for predicting activity sequences.

#### 1.1 Related Work

This review covers three strands that frame the study: activity-based residential load modeling (from occupancy-only and Markov simulators to duration-aware methods), time-use diary datasets with emphasis on American Time Use Survey (ATUS) and survey weighting at 10-minute resolution, and hierarchical modeling to stabilize estimates under sparsity. Subsections 1.1.1 to 1.1.3 follow this sequence.

Against this background, the review highlights gaps the present work addresses: understated dwell-time structure in Markov-only models, loss from coarse temporal aggregation, and instability without pooling. These themes connect directly to the research questions on explicit duration modeling and the marginal value of demographic covariates, setting up the modeling Section 3 and results Section 4.

### 1.1.1 Activity-Based Residential Load Modeling

Early work in residential load modeling developed bottom-up frameworks [9, 10] linking occupant activities with appliance demand. A seminal study in [6] proposed one of the first stochastic household load models, simulating appliance usage through Monte Carlo methods informed by occupant behavior. This established the foundation for treating human activity as the driver of residential demand. Later, [11] advanced this line by introducing the models that applied Markov chains to time-use statistics to generate occupancy and activity sequences, then mapped these sequences to appliance use. Their approach demonstrated that even simple Markov structures could capture the diurnal patterns of household demand.

Parallel work by [12, 13] integrated time-use diaries with bottom-up appliance models, particularly for domestic lighting. These studies showed that demand models grounded in activity sequences not only reproduced realistic load curves but also aligned well with measured data. However, as several authors noted, purely Markovian simulators and smaller datasets tend to underestimate the persistence of long events, leading to unrealistic distributions of very short or very long activities.

To address this shortcoming, hybrid models introduced explicit duration components. [14] developed an occupant presence model combining Markov chains with survival analysis to capture how long occupants are present or absent. This duration-aware approach provided more realistic occupancy sequences than memoryless models. Similarly, hazard-based sub-models by [15, 16, 17] have been incorporated in later occupancy research to determine dwell times for room presence or appliance use.

While such hybrid presence models are semi-Markov in spirit—using survival analysis to relax geometric dwell times—they focus on a *binary* presence/absence process and typically assume time-homogeneous or coarsely piecewise-homogeneous dynamics. They do not learn

a time-slot—specific destination distribution over multiple activities. Our work directly builds on these insights: we adopt a semi-Markov formulation with an explicit hazard component, ensuring that both which activity occurs next and how long it persists are modeled not just for occupancy but for various other activities.

Another limitation was often a direct consequence of the smaller datasets available at the time, which constrained the ability to define more detailed activity categories or capture accurate estimates of transition probabilities. In fact, [13] explicitly noted that a "larger and more detailed data set" covering a longer time period would be a key refinement to improve model fidelity. Our work directly addresses this call by leveraging over two decades of the ATUS data (2003–2024), providing the statistical power needed not only to model durations more accurately but also to robustly estimate parameters across numerous demographic subgroups. The full set of ATUS data variables used in our analysis is listed in Appendix C.

### 1.1.2 Time-Use Data and Stochastic Activity Models

Time-use diary surveys (TUS) [18, 19, 20] and electrical end-use datasets [21, 22] have become a central resource for energy-oriented activity modeling. The ability of TUS to represent population-level variability in demand timing has made them especially valuable for demand-side management research.

The ATUS [18] has been particularly influential. The work done by [7] demonstrated how ATUS data can calibrate Markov-chain activity simulators to produce detailed household demand estimates. Building on this, [23] derived typical U.S. occupancy schedules from ATUS records. In related work, [24] used clustering to identify dominant archetypes of daily occupancy. Likewise [25] applied clustering methods to ATUS to derive occupancy profiles for residential building simulations.

These studies reinforce two key premises of our approach: (i) large-scale diary data yield better representative activity sequences, and (ii) fine-grained temporal resolution is critical. Unlike prior work that often reduces diaries to larger intervals and binary occupancy schedules, our framework preserves the full 10-minute resolution of multiple ATUS activities and not just presence and absence of the respondents. This enables mapping of specific activities (e.g., cooking, laundry, leisure) to end-use loads in downstream applications.

A further methodological contribution of our work lies in the rigorous use of survey design weights. While being a best practice in social science, weighting has often been neglected in energy studies, where diaries are treated as simple random samples. We incorporate ATUS weights throughout model training and evaluation, ensuring that estimated transition probabilities and likelihood comparisons reflect the U.S. population. To our knowledge, this population-weighted stochastic modeling is novel in the context of residential load forecasting.

### 1.1.3 Hierarchical Modeling and Demographic Factors

Modeling activity sequences with multiple covariates such as Day-Type, employment, or region poses challenges of sparsity. Stratifying by many attributes quickly fragments the data, reducing the reliability of estimates. In [26] it was emphasized that hierarchical pooling can mitigate this by borrowing strength across groups while retaining group-specific deviations.

In the energy domain, clustering has served as an implicit form of pooling. For instance, [27] applied hierarchical clustering to Belgian TUS data [20], identifying a small set of representative occupancy sequences to stabilize building simulations. Our approach differs in that we retain the full fine-grained resolution of activities and instead shrink slot-level transition probabilities (10-minute resolution) toward block-level priors (6-hour resolution), producing stable yet interpretable daily trajectories even in sparse subgroups.

Demographic and contextual drivers of residential demand have also been studied extensively. The contrast between weekdays and weekends is well documented [23, 28], and employment status has been shown to shift energy use toward evenings [29]. Gender and number of occupants have produced mixed findings: systematic differences were found in [30] for time at home between men and women but with high within-group variability. Seasonal and regional effects, in contrast, appear to affect the intensity of activities (e.g., heating/cooling loads) rather than the sequence of behaviors themselves [13, 31].

#### 1.2 Contributions

The literature on activity-based load modeling spans bottom-up stochastic models, durationaware hazard methods, and the integration of large-scale time-use diaries. Our work synthesizes these strands by combining (i) a hierarchical semi-Markov framework, (ii) explicit hazard modeling of durations, and (iii) survey-weighted estimation on nationally representative ATUS data. This integration addresses gaps between small-scale, high-fidelity behavioral studies and the coarse scheduling assumptions often used in engineering practice. By situating our contributions within established research while extending them with methodological innovations, we demonstrate how realistic, population-representative activity sequences can form the foundation for more accurate downstream energy system applications.

This paper develops an interpretable modeling framework that directly addresses these gaps. Our primary contributions are:

- A Unified Hierarchical Framework: We develop an integrated semi-Markov model that simultaneously captures activity sequencing and durations while ensuring statistical stability through hierarchical shrinkage, even in data-sparse subgroups.
- Quantifying the Value of Durations: We rigorously demonstrate, using a bootstrap analysis, that explicitly modeling durations provides a substantial and statistically significant improvement in predictive accuracy over purely Markovian approaches.
- Identifying Key Demographic Drivers: We systematically measure the predictive value of individual demographic covariates, revealing a clear hierarchy in which some factors are critical for predicting daily activity patterns.
- A Robust Generative Tool: The resulting model is an interpretable and computationally efficient generator of realistic, demographically-conditioned activity sequences, providing a high-fidelity foundation for downstream energy systems modeling.

The remainder of this paper is organized as follows. Section 2 describes our data source, the ATUS, the preprocessing steps used to construct our modeling dataset, and the experiment design. Section 3 provides a detailed walkthrough of the hierarchical semi-Markov model. Section 4 presents the empirical results of our model comparisons, followed by a discussion of their interpretation and implications in Section 5. Finally, Section 6 concludes the paper and suggests directions for future research.

## 2 Data and Preprocessing

The foundation of our model is the ATUS, a nationally representative diary study conducted by the U.S. Census Bureau. Each diary entry details a 24-hour sequence of activities for a respondent aged 15 or older, beginning at 4 a.m. The raw data provides minute-by-minute episodes with 6-digit activity codes, along with rich demographic information and crucial survey design weights that enable population-level inference.

Our data preprocessing pipeline transforms this raw, event-based data into a structured, fixed-grid format suitable for our models. This involves three key steps: data assembly, temporal discretization, and the creation of a purpose-built activity taxonomy described in the following sections.

### 2.1 Data Assembly and Temporal Discretization

First, we merge several raw ATUS files, including the respondent-level, activity-level, and household roster files ([18]), to create a unified dataset for each diary entry. This links each activity episode with the full set of respondent and household characteristics.

While the raw diaries have a minute-by-minute resolution, we discretize each 24-hour diary into a categorical time series of T = 144 fixed 10-minute slots. For each slot t, we identify the primary activity performed by respondent i. This process results in a sequence where the activity in each slot is still represented by one of several hundred raw ATUS codes. To create a tractable and interpretable model, these fine-grained codes must be mapped to a smaller, coherent set of states.

### 2.2 Activity Taxonomy and Rule-Based Mapping

To make our model interpretable and focused on energy-relevant behaviors, we developed the 14-state taxonomy shown in Table 1. The design is guided by separating activities based on location (at home vs. out of home) and their likely electricity consumption.

Mapping the hundreds of raw ATUS codes to our 14 model states is performed using a hierarchical, rule-based procedure. The 6-digit ATUS codes are structured, where the first two digits represent a major activity category (e.g., '01' for Personal Care), the next two define a sub-category, and the final two provide fine-grained detail. Our mapping logic

Table 1: The 14-State Activity Taxonomy for Modeling.

Major Category	$\mathbf{Model\ State}\ (\mathcal{S})$	Description & Examples	
$\overline{At ext{-}Home},$	Cooking	Meal preparation using stove, oven	
(Electric)		or microwave.	
	DISHWASHING	Manual or machine dishwashing	
		and kitchen cleanup.	
	LAUNDRY/IRONING	Using a washer, dryer, or iron.	
	ELECTRIC CLEANING	Using vacuum cleaners, floor pol-	
		ishers, etc.	
	Screens/Leisure	Watching TV, using computers or	
		game consoles, phone calls.	
	Admin on Devices	Household finances, email, using	
		printers or scanners.	
	ELECTRIC APPLIANCE	Using power tools, electric motors	
		etc.	
At-Home,	SLEEP	Includes sleeping and napping.	
(Non Electric)			
	EATING/DRINKING	Consuming meals without active	
		preparation.	
	Personal Care	Bathing, grooming, dressing.	
	CARE AT HOME	Childcare or providing help to	
		other household adults.	
	QUIET/SOCIAL	Reading, relaxing, talking, or other	
		quiet activities.	
	Exercise (No Machine)	Yoga, calisthenics, stretching, etc.	
Out-of-Home	Out-of-Home	Encompasses work, shopping,	
		travel, and all other non-home	
		activities.	

leverages this structure to ensure classifications are transparent and consistent. The rules are applied with the following precedence:

- 1. Exact 6-Digit Code Overrides: We first manually check for specific codes where the energy implication is unambiguous and might otherwise be misclassified. For instance, '020101' ("Interior cleaning") is explicitly mapped to Electric Cleaning.
- 2. **Prefix Rules:** We then use 2- and 4-digit prefixes to classify entire families of activities. This is our primary method for leveraging the ATUS code structure. For example, all codes beginning with the 2-digit prefix '01' (Personal Care) are mapped to the Personal Care state, while codes starting with '0203' are mapped to Laundry/Ironing.
- 3. **Keyword Matching:** For codes not captured by prefixes, we search the activity's text description for specific keywords (e.g., "cook", "oven", "microwave") to assign it to a state like COOKING.
- 4. **Fallback Assignment:** Finally, any remaining unclassified codes are assigned based on coarse, major-category heuristics or default to the safest assignment, Out-of-Home.

This structured mapping process produces a clean, consistent time series  $(S_{i,1}, \ldots, S_{i,144})$  for each respondent i, which forms the core data for training and evaluating our model.

## 2.3 Empirical Motivation for Model Choices

The selection of a time-inhomogeneous semi-Markov model is not merely a theoretical choice but is directly motivated by strong empirical patterns observed in the ATUS data.

Justifying the Semi-Markov Approach. A purely Markovian model assumes a constant probability of leaving that activity at any time step. Figure 1 shows the empirical dwell-time distributions for several key activities. For example, SLEEP shows a strong peak around 8 hours, and COOKING peaks around 20-30 minutes. The probability of leaving an activity is highly dependent on how long it has already been in progress. This non-constant

hazard rate is a clear violation of the Markov assumption and provides a strong justification for employing a semi-Markov model.

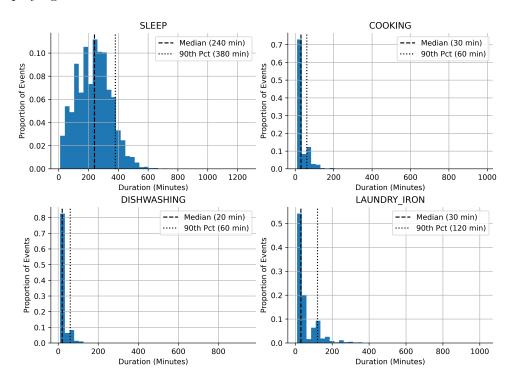


Figure 1: Empirical survival functions for key activities demonstrating that activity durations are not memoryless, motivating the use of a semi-Markov (hazard) model.

Covariate Activity Pattern Differences The importance of the covariates can be understood by visualizing their impact on daily activity patterns. Figure 2 shows the probability of being engaged in selected activities (SLEEP, COOKING, DISHWASHING, LAUNDRY/IRONING) over the course of the day, stratified by three key covariates: Sex, Employment Status, and Day-Type. The separations between the curves highlight why conditioning on these covariates may yield improvements in predictive accuracy.

Justifying the Time-Inhomogeneous Approach. A time-homogeneous model would assume that transition probabilities are constant throughout the day. The data strongly contradicts this assumption. Figure 3 illustrates this by plotting the probability of transitioning from SLEEP to PERSONAL CARE at each 10-minute interval. The probability is near zero for most of the day but exhibits a dramatic spike between 5 a.m. and 10 a.m.

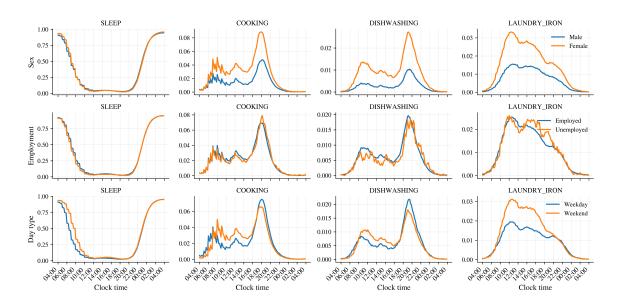


Figure 2: Occupancy/activity probability curves for selected activities (SLEEP, COOKING, DISHWASHING, LAUNDRY/IRONING), stratified by sex, employment status, and Day-Type. The strong separation between groups illustrates why these covariates drive significant predictive improvements.

This strong diurnal pattern, which is present for most activity pairs, necessitates the use of a time-inhomogeneous model with distinct transition matrices for each time slot.

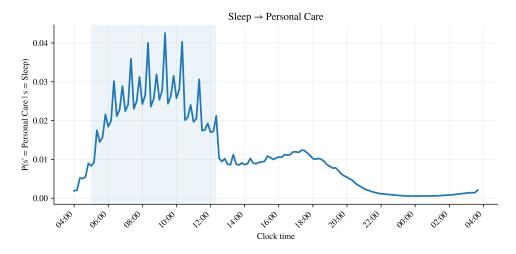


Figure 3: The empirical probability of transitioning from SLEEP to PERSONAL CARE as a function of the time of day. The sharp morning peaks demonstrates the necessity of a time-inhomogeneous model.

### 2.4 Experimental Design

Our analysis uses the full span of publicly available ATUS data from 2003 through 2024 comprising 252, 808 respondents. The complete set of respondent diaries is partitioned into a training set (80%) and a holdout test set (20%). A random split rather than a chronological (last-20% as test) split is used as our goal is to evaluate average-case sequence modeling under the population distribution (by season/quarter, day-type, and demographics), not to forecast across exogenous calendar shocks.

The splits are stratified by the covariates used, ensuring similar composition in train and test for each covariate type. Each respondent contributes exactly one diary, so there is no subject leakage across splits. The model parameters are estimated using only the training data, and all performance metrics are computed on the unseen holdout data to ensure a robust evaluation of predictive performance.

A central goal of this paper is to quantify the predictive value of different demographic factors and the effects of dwell time models. To achieve this, we systematically compare a series of models, each conditioned on different respondent attributes.

The Baseline Model (S0). We first establish a baseline model, denoted S0, which is ungrouped. In this model, all respondents are treated as belonging to a single, homogeneous population. This model captures the average, population-level activity patterns but does not account for any demographic variation.

Single-Covariate Models (S1-S6). We then train a series of models where respondents are partitioned into groups based on a single demographic covariate. These models, labeled S1 through S6, are detailed in Table 2. For each factor, a separate model is trained; for instance, the S2 model partitions the data into two groups  $(g_i \in \{\text{Male}, \text{Female}\})$  and estimates distinct parameters for each.

Model Comparisons. Our analysis focuses on two key comparisons. First, we measure the improvement of each single-covariate model (S1-S6) relative to the ungrouped S0 baseline. This isolates the value of each demographic factor. Second, for every model configuration, we compare the performance of the purely Markovian router model against the full semi-Markov version that includes the hazard component. We denote these enhanced

Table 2: Single-Covariate Models Used for Comparison.

Model ID	Covariate	Groups / Bins
S1	Region	Northeast; Midwest; South; West
S2	Sex	Male; Female
S3	Employment Status	Employed; Not employed; Not in work force
S4	Day-Type	Weekday; Weekend
S5	Household (HH)-Size	1 member; 2 members; 3 members; 4+ members
S6	Quarter/Season	Winter (Q1); Spring (Q2); Summer (Q3); Fall (Q4)

models with an "-H" suffix (e.g., S0-H, S1-H). This allows us to quantify the additional predictive gain from explicitly modeling activity durations.

### 2.5 Survey Weights and Model Inputs

A critical feature of the ATUS dataset is the final person-day weight,  $w_i > 0$ , provided by ATUS for each diary. These weights represent the number of people in the population that each respondent represents. They are useful for correcting biases in the sample that arise from differences in sampling probabilities and response rates across various demographic groups and days of the week. By applying these weights, we can produce estimates that are statistically representative of the entire target population. Thus, we use these weights throughout our analysis to ensure our findings are generalizable and reflect the true composition and behavior of the population.

In training, the statistics for the Markov model and dwell periods are computed as weighted sums. In evaluation, our metrics, such as negative log-likelihood, are also weighted. This ensures that our model's parameters and performance are representative of the target population, not the raw sample composition. The final inputs to train our mathematical model for each respondent i at each time slot t are the activity state  $S_{i,t}$ , their demographic group  $g_i$ , the current run-length of the activity  $\ell_{i,t}$ , and the coarse day-part block b(t).

## 3 Mathematical Model

We model daily activity sequences as a time-inhomogeneous semi-Markov process, discretized into T=144 ten-minute slots. The mathematical notations for this framework are described in the Table 3. This framework is explicitly designed to capture both the fine-grained, time-of-day variations in activity choice and the realistic durations of those activities. Our approach combines a time-inhomogeneous Markov router to model the sequence of activities with a hazard model to control the dwell time in each state.

Table 3: Nomenclature used throughout the model description.

Symbol	Description
K	Number of activity states in the set $S = \{1,, K\}$ .
T	Number of 10-minute slots in a day $(T = 144)$ .
$S_{i,t}$	Activity state of respondent $i$ at slot $t$ .
$g_i \in \mathcal{G}$	Demographic group of respondent $i$ .
b(t)	Day-part block for time slot $t$ .
$w_i$	Survey design weight for respondent $i$ .
$\ell_{i,t}$	Run length (in slots) of the current activity at the start of slot $t$ .
$\mathcal{L} = \{L_1, \dots, L_M\}$	Set of disjoint run-length bins for discretizing duration.
$C_{s,s'}^{(g,t)}$	Weighted count of transitions $s \to s'$ for group $g$ at slot $t$ .
$C_{s,s'}^{(b)}$	Total weighted transition count for $s \to s'$ within block $b$ .
$N_{s,m}^{(g,t)}$	Weighted count of exposures in run-length bin $L_m$ for group $g$ at slot $t$ .
$E_{s,m}^{(g,t)}$	Weighted count of exits from run-length bin $L_m$ for group $g$ at slot $t$ .
$ar{ heta}_{s,s'}^{(b)}$	Block-level prototype probability for the transition $s \to s'$ .
$\widehat{ heta}_{s,s'}^{(g,t)}$	Posterior estimate of the router probability for $s \to s'$ .
$ar{h}_{s,m}^{(b)}$	Block-level prototype hazard probability for run-length bin $L_m$ .
$\widehat{h}_{s,m}^{(g,t)}$	Posterior estimate of the hazard probability.
$arphi_{s,s'}^{(g,t)}$	Conditional destination probability of transitioning to $s'$ .
$ au_b,k$	Hyperparameters for the router model, controlling shrinkage and smoothing.
$\kappa_b$	Hyperparameter for the hazard model, controlling shrinkage.

The primary statistical challenge is data sparsity, magnified by the scale of our time-

inhomogeneous model. For even a single demographic group, estimating the full set of transition probabilities requires fitting 143 distinct  $14 \times 14$  transition matrices that is one for each 10-minute slot of the day which amounts to estimating 28,028 parameters. When the data is further stratified by demographic covariates, the number of observations available for any specific group at a particular time of day becomes extremely thin, making direct estimation from counts highly unstable and prone to overfitting.

To address this, our framework uses a hierarchical approach. We regularize the slotlevel estimates by pooling them toward robust, data-rich estimates computed over coarser day-part blocks. This method allows the model to borrow statistical strength across time, preventing overfitting in sparse cells while preserving the ability to capture specific, timedependent patterns where the data permit.

### 3.1 Day-Part Blocks for Hierarchical Shrinkage

Our approach is the use of four coarse day-part blocks to create stable priors: **Night** (10:00 p.m.–5:50 a.m.), **Morning** (6:00 a.m.–11:50 a.m.), **Afternoon** (12:00 p.m.–5:50 p.m.), and **Evening** (6:00 p.m.–9:50 p.m.).

Within each block b, we compute a single, low-variance prototype transition matrix,  $\bar{\Theta}^{(b)}$ . This is achieved by first aggregating the weighted transition counts,  $C_{s,s'}^{(g,t)}$  (Appendix A), across all demographic groups g and all time slots t that fall within that block:

$$C_{s,s'}^{(b)} = \sum_{g \in \mathcal{G}} \sum_{t \in b} C_{s,s'}^{(g,t)}.$$

The prototype probability,  $\bar{\theta}_{s,s'}^{(b)}$ , is then the simple row-normalization of these aggregated counts:

$$\bar{\theta}_{s,s'}^{(b)} = \frac{C_{s,s'}^{(b)}}{\sum_{j=1}^{K} C_{s,j}^{(b)}}.$$
(1)

This prototype represents the average transition behavior for that part of the day (e.g., a morning matrix) and serves as an informative prior for the models at each specific 10-minute slot. This shrinkage mechanism ensures that when 10-minute slot-level data is sparse, our estimates are pulled toward the reliable block-level average, preventing overfitting and increasing robustness.

### 3.2 The Router Model (S0 – S6)

The router component defines the purely Markovian transition dynamics of our system. It determines the probability,  $\theta_{s,s'}^{(g,t)}$ , of moving from activity s to s'. We model each row of the transition matrix as a draw from a Dirichlet distribution whose parameters are set by the corresponding row of the block prototype,  $\bar{\theta}_{s,\cdot}^{(b(t))}$  from Equation (1):

$$\theta_{s,\cdot}^{(g,t)} \mid \bar{\theta}_{s,\cdot}^{(b(t))} \sim \operatorname{Dir}\left(\tau_b \,\bar{\theta}_{s,\cdot}^{(b(t))} + k \, \frac{1}{K}\right).$$

The design-weighted counts  $C_{s,s'}^{(g,t)}$  calculated by summing the survey weights of all respondents i in group g who transition from s to s' at time t. The posterior mean for the transition probability is then:

$$\widehat{\theta}_{s,s'}^{(g,t)} = \frac{C_{s,s'}^{(g,t)} + \tau_b \, \overline{\theta}_{s,s'}^{(b(t))} + k/K}{\sum_{j=1}^K C_{s,j}^{(g,t)} + \tau_b + k}.$$
(2)

This formulation alone defines the family of purely Markovian models, from the ungrouped baseline (S0) to the single-covariate models (S1-S6). Here,  $\tau_b$  controls the degree of shrinkage towards the block-level prior, with larger values indicating stronger shrinkage. The hyperparameter k provides a small amount of smoothing to avoid zero probabilities for unobserved transitions.

#### 3.3 The Hazard Model for Durations

To achieve realistic durations and define our semi-Markov models (the S-H series), we add a component that explicitly models the probability of an activity ending. For any given respondent, the decision to continue an activity or switch is a binary outcome. We model this using a discrete hazard probability,  $h_{s,m}^{(g,t)}$ , which is the probability of leaving state s during the current time slot, given the activity has lasted for a duration  $\ell_{i,t}$  falling in bin  $L_m$  which is defined as follows:

The run length  $\ell_{i,t}$  is simply how long the current activity has been going on when slot t starts, measured in 10-minute steps (e.g.,  $\ell_{i,t} = 1$  means it just started;  $\ell_{i,t} = 6$  means it has lasted for an hour). Rather than estimate a separate leave-probability for every possible length, we group lengths into a few sensible ranges ("bins") such as short, medium, and long. At each slot, the model looks up which bin  $\ell_{i,t}$  falls into and uses the corresponding parameter for that bin.

We use the Beta-Bernoulli conjugate model. For each cell (s, m, g, t), the sufficient statistics are the weighted count of respondents exposed(N) and the weighted count of those who exit(E) as described in Appendix A. These are calculated by summing the survey weights  $w_i$  of the relevant respondents. We place a Beta prior on the hazard probability, centering it on the robust block-level prototype  $\bar{h}_{s,m}^{(b(t))}$ :

$$h_{s,m}^{(g,t)} \mid \bar{h}_{s,m}^{(b(t))} \sim \text{Beta}\left(\kappa_b \, \bar{h}_{s,m}^{(b(t))}, \, \kappa_b (1 - \bar{h}_{s,m}^{(b(t))})\right).$$

The posterior mean for the hazard is a simple shrinkage estimator combining the local evidence with the block prior:

$$\widehat{h}_{s,m}^{(g,t)} = \frac{E_{s,m}^{(g,t)} + \kappa_b \, \overline{h}_{s,m}^{(b(t))}}{N_{s,m}^{(g,t)} + \kappa_b}.$$
(3)

The hyperparameter  $\kappa_b$  controls the degree of pooling, ensuring stable hazard estimates. The complete derivation of posteriors for the router and the hazard models is showed in the Appendix B.

### 3.4 The Combined Semi-Markov Process (S0-H – S6-H)

The router and hazard components from Equation (2) and 3 are synthesized to form the final transition probabilities for the semi-Markov models (S0-H through S6-H). For a respondent in state s at time t, the model first decides whether to leave with probability  $\hat{h}_{s,m}^{(g_i,t)}$ . If a change occurs, it then decides where to go using the router's conditional destination distribution,  $\varphi_{s,s'}^{(g,t)}$ :

$$\varphi_{s,s'}^{(g,t)} = \frac{\widehat{\theta}_{s,s'}^{(g,t)}}{1 - \widehat{\theta}_{s,s'}^{(g,t)}}, \quad \text{for } s' \neq s.$$
(4)

Using the conditional probability in Equation (4) the complete one-step semi-Markov transition probability is thus:

$$\Pr(S_{i,t+1} = s' \mid S_{i,t} = s, \ \ell_{i,t} \in L_m, \ g_i, \ t) = \begin{cases} 1 - \widehat{h}_{s,m}^{(g_i,t)}, & \text{if } s' = s, \\ \widehat{h}_{s,m}^{(g_i,t)} \varphi_{s,s'}^{(g_i,t)}, & \text{if } s' \neq s. \end{cases}$$
(5)

To begin the generative process for a synthetic diary, the model must first draw an initial state,  $S_1$ , for the first time slot of the day (4:00 a.m.). We estimate a separate initial state

distribution,  $\pi^{(g)}$ , for each demographic group g. This distribution is estimated from the design-weighted frequencies of observed activities at t = 1. The detailed formulation for the estimate of this distribution is provided in Appendix B.3. The generative process for simulating a new 24-hour diary is detailed in Algorithm 1.

```
Algorithm 1 Generative algorithm for simulating a 24-hour activity sequence
 1: Input: Demographic group g, trained parameters \{\widehat{\pi}^{(g)}, \widehat{\Theta}^{(g,t)}, \widehat{h}^{(g,t,m)}\}
 2: Output: Synthetic activity sequence (S_1, \ldots, S_{144})
    Initialization
 3: Draw initial state S_1 \sim \widehat{\pi}^{(g)}
 4: Set run length \ell \leftarrow 1
    Main loop
 5: for t = 1 to T - 1 do
         Set current state s \leftarrow S_t
 6:
         Identify run-length bin m such that \ell \in L_m
 7:
         Retrieve hazard probability \hat{h}_{s,m}^{(g,t)}
 8:
         Draw u \sim U(0,1)
 9:
         if u < \widehat{h}_{s,m}^{(g,t)} then
10:
                                                                                           ▶ Leave current state
              Sample next state S_{t+1} \sim \varphi_{s,\cdot}^{(g,t)}
11:
              Reset run length: \ell \leftarrow 1
12:
         else
                                                                                         > Stay in current state
13:
              Set S_{t+1} \leftarrow s
14:
              Update run length: \ell \leftarrow \ell + 1
15:
         end if
16:
17: end for
18: return (S_1, \ldots, S_{144})
```

## 4 Results

We evaluate our models on a hold-out test set of ATUS diaries to assess their predictive performance. Our analysis is designed to answer two primary questions:

- How much predictive power is gained by explicitly modeling activity durations using the hazard component?
- Which demographic covariates provide the most significant improvements in predictive accuracy in both Markov and semi-Markov models?

#### 4.1 Evaluation Metrics

The performance of each model is quantified on the held-out dataset using two distinct but complementary metrics. Let  $\mathcal{T}_i$  be the set of valid transition times for respondent i, and let  $p(\cdot)$  be the one-step probability from the model.

Negative Log-Likelihood (NLL). The primary evaluation is based on the design-weighted NLL, a proper scoring rule averaged per transition. The NLL provides a rigorous measure of a model's predictive accuracy by calculating the average negative log-probability assigned to the true, observed sequence of activities. A lower NLL score indicates superior performance, as it reflects a higher likelihood assigned to the ground-truth data. As a sensitive metric that evaluates the entire predicted probability distribution, NLL is particularly effective for distinguishing between the subtle performance differences of well-calibrated probabilistic models.

The design-weighted NLL per transition is:

$$NLL = \frac{\sum_{i} w_{i} \sum_{t \in \mathcal{T}_{i}} \left( -\log p(S_{i,t+1} \mid S_{i,t}, \ell_{i,t}, g_{i}, t) \right)}{\sum_{i} w_{i} |\mathcal{T}_{i}|}$$

**Top-1 Next-Activity Accuracy.** To provide a more intuitive measure of performance, we also report the design-weighted Top-1 accuracy. This metric calculates the percentage of 10-minute slots for which the model's single most likely prediction for the next activity was correct. While less sensitive than NLL, Top-1 accuracy offers a straightforward and easily interpretable measure of the model's practical utility in predicting the single most probable outcome at each step.

The design-weighted Top-1 next-activity accuracy is the weighted proportion of time steps where the model's most likely prediction matches the observed activity:

$$Top1 = \frac{\sum_{i} w_{i} \sum_{t \in \mathcal{T}_{i}} \mathbf{1} \left\{ S_{i,t+1} = \arg \max_{s' \in \mathcal{S}} p(S_{i,t+1} = s' \mid S_{i,t}, \ell_{i,t}, g_{i}, t) \right\}}{\sum_{i} w_{i} |\mathcal{T}_{i}|}$$

Using these metrics, the following sections will present a systematic comparison of our model configurations.

### 4.2 Overall Model Performance

The primary findings of our study reveal the following patterns. First, adding the hazard component to explicitly model activity durations provides a substantial and consistent improvement in performance across all model configurations. Second, a hierarchy emerges among the demographic covariates, with 'Sex', 'Day-Type' and 'HH-Size' providing gains in predictive power. Table 3 presents the quantitative results on the hold-out set supporting these conclusions, comparing the NLL and Top-1 next-activity accuracy for the purely Markovian models (S0-S6) and their semi-Markov counterparts (S0-H to S6-H).

### 4.3 Bootstrap Confidence Intervals for Model Comparison

To ensure that the observed differences in performance between models in Section 4.2 are statistically meaningful and not merely due to the specific random sample of diaries in our holdout set, we employ a paired bootstrap procedure. The process is as follows: we generate 2,000 bootstrap replicates of the holdout set with replacement by resampling the respondents (diaries). For each replicate, we compute our metric of interest (e.g., the difference in NLL between model S0 and S0-H). This process yields an empirical distribution of the performance difference. We then calculate the 2.5th and 97.5th percentiles of this distribution to construct a 95% confidence interval (CI). The interpretation of this interval is direct: if the 95% CI for a performance difference does not contain zero, we can conclude with high confidence that the observed improvement is statistically significant.

Bootstrap analysis shows that the most significant performance gain across all models comes from adding the hazard component. Using the bootstrap analysis, we verify that this improvement is statistically robust and not a result of sampling variability. For each model configuration (S0 through S6), we compared the NLL of the purely Markovian model against its semi-Markov counterpart.

Table 4: Predictive Performance of All Models on the holdout set. Average one-step NLL is a measure of predictive error where lower is better. The  $\Delta$ NLL column shows the reduction in error relative to the corresponding baseline (S0 or S0-H). Top-1 Accuracy is the weighted percentage of correctly predicted next activities, where higher is better.

Model	NLL	$\Delta$ NLL (vs.S0 Baseline)	Top-1 Acc.	
Markov Models (Router-Only)				
S0 (Baseline)	0.438585	_	90.2689%	
S1 (Region)	0.438554	0.00003	90.2689%	
S2 (Sex)	0.436974	0.00161	90.2689%	
S3 (Employment)	0.438605	-0.00002	90.2689%	
S4 (Day-Type)	0.437812	0.00077	90.2689%	
S5 (HH-Size)	0.437615	0.00097	90.2689%	
S6 (Quarter/Season)	0.438580	0.00001	90.2689%	
Semi-Markov Models (Router + Hazard)				
S0-H (Baseline)	0.426483	_	90.2733%	
S1-H (Region)	0.426554	-0.00007	90.2730%	
S2-H (Sex)	0.424811	0.00167	90.2727%	
S3-H (Employment)	0.426566	-0.00008	90.2714%	
S4-H (Day-Type)	0.425577	0.00091	90.2733%	
S5-H (HH-Size)	0.425396	0.00109	90.2739%	
S6-H (Quarter/Season)	0.426559	-0.00008	90.2733%	

The results in Figure-4 provides further details on this comparison: explicitly modeling activity durations yields a large and statistically significant improvement in predictive accuracy. For the baseline S0 model, adding the hazard component (S0-H) reduces the NLL by 0.012102 on average, and the 95% confidence interval for this improvement is tightly bound far from zero. This pattern holds for every covariate, confirming that the semi-Markov approach is consistently and significantly superior to the Markovian model for human level activity sequence modeling.

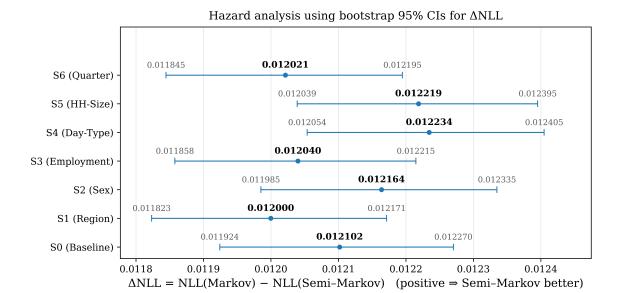
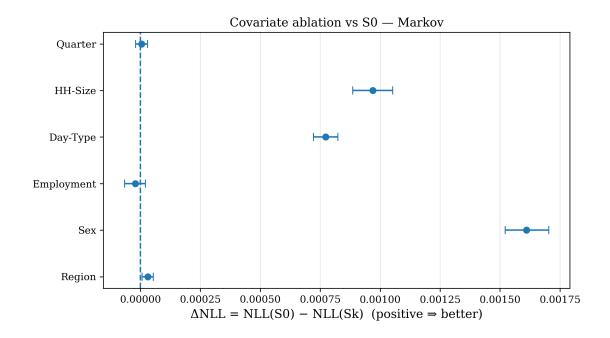


Figure 4: Hazard analysis using bootstrap 95% CIs for  $\Delta$ NLL. We plot  $\Delta$ NLL = NLL(Markov) – NLL(Semi-Markov), so positive values indicate that the Semi-Markov model achieves lower (better) NLL. Points mark the bootstrap mean; horizontal bars show the 95% percentile confidence interval; S0 uses no covariate. All improvements are statistically significant at the 95% confidence level as no intervals contain 0.

### 4.4 Ranking the Predictive Power of Covariates

The reduction in NLL for each single-covariate model relative to the S0/S0-H baseline is visualized in Figure 5. In both the Markov and semi-Markov settings, covariates with positive  $\Delta$ NLL reduce predictive error, while those near or to the left of zero offer little or no benefit:

- Sex (S2 / S2-H) is the most influential covariate. In both Markov and semi-Markov settings it yields the largest reduction in predictive error (ΔNLL ≈ 0.0016), with confidence intervals well above zero, confirming robust predictive benefit.
   while in the semi-Markov model both remain beneficial and statistically significant.
- 2. HH-Size Band (S5 / S5-H) and Day-Type (S4 / S4-H) provide the next largest gains. In the Markov model, HH-Size slightly outperforms Day-Type ( $\Delta$ NLL  $\approx 0.0010$  vs. 0.0008), while in the semi-Markov model both remain beneficial and statistically significant.



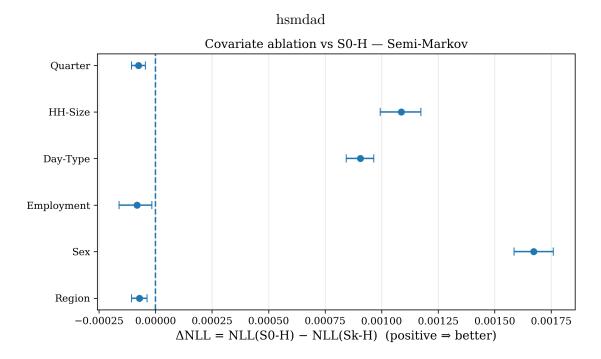


Figure 5: Bootstrap analysis for single-covariate ablations. Top: Markov (S1–S6 vs. S0). Bottom: Semi-Markov (S1-H–S6-H vs. S0-H).  $\Delta$ NLL = NLL(S0-H) – NLL(Sk-H); Error bars: 95% bootstrap CIs.

- 3. Employment Status (S3 / S3-H) does not significantly improve predictions. Its estimates lie close to zero and confidence intervals overlap the baseline, indicating little added value once durations are modelled.
- 4. Region (S1 / S1-H) and Quarter/Season (S6 / S6-H) provide virtually no predictive gain for activity sequencing.

Overall, the results show that Sex is the dominant covariate for next-activity prediction, with HH-Size and Day-Type providing complementary improvements. Other demographic factors such as employment, region, and season add little or no sequencing value once dwell times are incorporated.

### 5 Discussion

Our results demonstrate that a hierarchical semi-Markov models can effectively capture the sequence and duration of human activities, and that a select few demographic factors are disproportionately important for this task. In this section, we interpret these findings, discuss their implications for energy systems modeling, and outline the limitations of our study.

## 5.1 Interpretation of Key Findings

Two central conclusions emerge from our analysis described as follows:

The Dominance of Duration Modeling. Adding dwell time for activities improved the models significantly. This finding underscores a fundamental limitation of memoryless property in Markov models for human behavior. Activities like sleeping or working have characteristic durations that are not geometrically distributed. By incorporating a runlength-dependent hazard function, the semi-Markov model aligns more closely with the temporal logic of human schedules, resulting in a substantial and statistically significant improvement in predictive accuracy.

The Hierarchy of Covariates. The results show that not all demographic factors are equally useful for predicting activity sequences. Sex emerges as the dominant driver of pre-

dictive improvement, reflecting systematic differences in daily routines followed by HH-Size and Day-Type which provided complementary gains, indicating that both family context and whether it is a weekday, weekend, or holiday meaningfully shape activity sequencing. By contrast, Employment Status offers little benefit once durations are accounted for, and Region and Season contribute negligibly—or even slightly degrade performance in the semi-Markov setting. This distinction is important; although Region and Season are critical for downstream energy calculations (e.g., HVAC loads), they do not appear to alter the underlying pattern of human behavior in terms of activity sequencing.

### 5.2 On the Magnitude and Significance of Model Improvements

While the per-transition NLL improvements may appear marginal (e.g., 0.012219 for S1-H vs. SH), it is crucial to recognize that the likelihood of a model generating an entire sequence is exponentially related to the total NLL. A small, consistent improvement over each of the 143 transitions in a day compounds into a substantial increase in the model's overall explanatory power.

An improvement of  $\Delta$ NLL means the better model is  $e^{\Delta$ NLL times more likely to predict the next step correctly. Over a full day, this effect compounds multiplicatively. We can see the significance of this by comparing the two main findings from our results:

First, consider the dominant effect of adding the hazard model. The improvement of the S1-H model over the S1 baseline was approximately 0.2094. For a full day's sequence, the likelihood ratio is:

Likelihood Ratio (Hazard) = 
$$(e^{0.0.012219})^{143} = e^{1.747317} \approx 5.7391$$

This number signifies that the semi-Markov model is more likely to have generated the observed sequences than the purely Markovian model, confirming the critical importance of modeling durations.

Now, consider the moderate improvement from adding the best covariate, Sex (S2-H vs. S0-H), which had a  $\Delta$ NLL of 0.0016. The likelihood ratio for a full day is:

Likelihood Ratio (Covariate) = 
$$(e^{0.0016})^{143} = e^{0.2288} \approx 1.2570$$

This means that for a typical 24-hour diary, the model conditioned on covariate Sex is about 25% more likely to have generated the observed sequence of activities than the ungrouped

semi-Markov baseline. While a smaller figure, it represents a consistent and statistically significant improvement in explanatory power that originates from a seemingly small perstep gain.

### 5.3 Implications for Energy Systems Modeling

While this paper focuses on the prediction of activity sequences, its findings have direct practical implications for the design and operation of modern energy systems. Our model provides the high-fidelity behavioral inputs required for a range of downstream energy analysis tasks.

A Foundation for High-Fidelity Load Profile Generation. The primary down-stream application of our model is to serve as a stochastic activity generator for bottom-up residential load profile simulations. By producing more realistic activity sequences, especially with respect to the timing and duration of high-consumption activities (e.g., Cooking, Screens/Leisure), our framework enables the generation of more accurate load profiles. These high-fidelity profiles are essential for tasks that depend on realistic peak demand characteristics, such as sizing neighborhood-scale battery storage, assessing grid stability, and evaluating the impact of new technologies.

Informing the Electrification Transition. Our model can be used to explore critical scenarios, such as the timing of Electric Vehicle (EV) charging. By identifying realistic and demographically-varying "at-home" windows, the model provides a data-driven basis for simulating charging patterns and assessing their potential grid impact, moving beyond simple assumptions about overnight charging.

#### 5.4 Limitations and Future Research

While our framework demonstrates strong performance, we acknowledge several limitations that point toward avenues for future research. First, our analysis is currently limited to single covariates. Exploring models that capture interaction effects is an important next step. Second, the **14-state activity taxonomy**, while interpretable, aggregates all "Out-of-Home" activities. A more granular taxonomy could provide deeper insights, especially

for modeling EV charging behavior. Finally, our model treats individuals as **independent** agents. Extending the model to a multi-agent framework that captures intra-household dynamics is a challenging but important direction for future work.

### 6 Conclusion

This paper developed and validated a hierarchical semi-Markov model for generating realistic, demographically-conditioned daily activity sequences from nationally representative time-use data. By addressing the key challenges of data sparsity and activity duration modeling, our framework provides a robust foundation for human activity sequencing.

Our empirical evaluation yielded two primary findings. First, explicitly modeling activity durations via the semi-Markov hazard component is not merely an incremental improvement but a critical one, offering a substantial and statistically significant increase in predictive accuracy over purely Markovian approaches. Second, we identified a clear hierarchy in the predictive power of some demographic factors providing meaningful complementary gains while others seem to be negligible.

The resulting model is an interpretable and computationally efficient tool for generating the high-fidelity behavioral inputs needed to understand the impacts of residential electricity consumption and to design more effective energy management strategies. Future work will focus on extending this framework to capture interaction effects between covariates and to model the correlated activities of multiple individuals within a household, further enhancing the realism and policy relevance of the generated sequences.

### Acknowledgements

The authors gratefully acknowledge funding from Triad National Security LLC under the grant from the Department of Energy National Nuclear Security Administration (award no. 89233218CNA000001).

## References

- [1] Matteo Muratori. Impact of uncoordinated plug-in electric vehicle charging on residential power demand. *Nature Energy*, 3(3):193–201, 2018.
- [2] Geoffrey Johnson and Ian Beausoleil-Morrison. Electrical-end-use data from 23 houses

- sampled each minute for simulating micro-generation systems. Applied Thermal Engineering, 114:1449–1456, 2017.
- [3] Weicong Kong, Zhao Yang Dong, David J. Hill, Fengji Luo, and Yan Xu. Short-term residential load forecasting based on resident behaviour learning. *IEEE Transactions on Power Systems*, 33(1):1087–1088, 2018.
- [4] Neil Saldanha and Ian Beausoleil-Morrison. Measured end-use electric load profiles for 12 Canadian houses at high temporal resolution. *Energy and Buildings*, 49:519–530, 2012.
- [5] Xinmei Yuan, Peng Han, Yao Duan, Rosemary E. Alden, Vandana Rallabandi, and Dan M. Ionel. Residential Electrical Load Monitoring and Modeling-State of the Art and Future Trends for Smart Homes and Grids. *Electric Power Components and Systems*, 48(11):1125-1143, 2020.
- [6] Alfonso Capasso, W Grattieri, Regina Lamedica, and A Prudenzi. A bottom-up approach to residential load modeling. *IEEE transactions on power systems*, 9(2):957–964, 2002.
- [7] Matteo Muratori, Matthew C. Roberts, Ramteen Sioshansi, Vincenzo Marano, and Giorgio Rizzoni. A highly resolved modeling technique to simulate residential power demand. *Applied Energy*, 107:465–473, 7 2013.
- [8] Yun Shang Chiou, Kathleen M. Carley, Cliff I. Davidson, and Michael P. Johnson. A high spatial resolution residential energy model based on American Time Use Survey data and the bootstrap sampling method. *Energy and Buildings*, 43(12):3528–3538, 2011.
- [9] Matteo Giacomo Prina, Giampaolo Manzolini, David Moser, Benedetto Nastasi, and Wolfram Sparber. Classification and challenges of bottom-up energy system models-a review. Renewable and Sustainable Energy Reviews, 129:109917, 2020.
- [10] Anjun Zhao, Mengya Chen, Junqi Yu, and Pufang Cui. Simulating appliance-level household electricity data: Accounting for residential behavior and usage patterns in china. *Journal of Building Engineering*, 92:109804, 2024.
- [11] Ian Richardson, Murray Thomson, and David Infield. A high-resolution domestic building occupancy model for energy demand simulations. *Energy and Buildings*, 40(8):1560–1566, 2008.
- [12] Joakim Widén, Magdalena Lundh, Iana Vassileva, Erik Dahlquist, Kajsa Ellegård, and Ewa Wäckelgård. Constructing load profiles for household electricity and hot water from time-use data-Modelling approach and validation. *Energy and Buildings*, 41(7):753–768, 2009.
- [13] Joakim Widén and Ewa Wäckelgård. A high-resolution stochastic model of domestic activity patterns and electricity demand. *Applied Energy*, 87(6):1880–1892, 2010.
- [14] Jessen Page, Darren Robinson, Nicolas Morel, and J-L Scartezzini. A generalised stochastic model for the simulation of occupant presence. *Energy and buildings*, 40(2):83–98, 2008.

- [15] Huiqiao Hou, Jacek Pawlak, Aruna Sivakumar, Bianca Howard, and John Polak. An approach for building occupancy modelling considering the urban context. *Building* and *Environment*, 183:107126, 2020.
- [16] Omar Ahmed, Nurettin Sezer, Mohamed Ouf, Liangzhu (Leon) Wang, and Ibrahim Galal Hassan. State-of-the-art review of occupant behavior modeling and implementation in building performance simulation. Renewable and Sustainable Energy Reviews, 185:113558, 2023.
- [17] Simona D'Oca, H Burak Gunay, Sara Gilani, and William O'Brien. Critical review and illustrative examples of office occupant modelling formalisms. *Building Services Engineering Research and Technology*, 40(6):732-757, 2019.
- [18] U.S. Bureau of Labor Statistics. American time use survey (atus) microdata files, 2003-2024. U.S. Department of Labor.
- [19] Oriel Sullivan and Jonathan Gershuny. United kingdom time use survey, 2014–2015, 2023. [data collection], UK Data Service, SN: 8128.
- [20] Statistics Belgium (Statbel). Belgian time use survey (btus13), 2013–2014, 2015. [data set], Brussels: Statbel. Fieldwork Jan 2013–Feb 2014.
- [21] Electric Vehicle Charging: First of a Kind National Lab Project Will Simulate Fast Charging Station Microgrids INL.
- [22] NEEA. Northwest Energy Efficiency Alliance (NEEA) Home Energy Metering..., 2020.
- [23] Debrudra Mitra, Nicholas Steinmetz, Yiyi Chu, and Kristen S Cetin. Typical occupancy profiles and behaviors in residential buildings in the united states. *Energy and Buildings*, 210:109713, 2020.
- [24] Debrudra Mitra, Yiyi Chu, and Kristen Cetin. Cluster analysis of occupancy schedules in residential buildings in the united states. *Energy and Buildings*, 236:110791, 2021.
- [25] Debrudra Mitra, Yiyi Chu, Kristen Cetin, Yu Wang, and Chien-fei Chen. Variation in residential occupancy profiles in the united states by household income level and characteristics. *Journal of Building Performance Simulation*, 14(6):692–711, 2021.
- [26] Andrew Gelman and Jennifer Hill. Data analysis using regression and multilevel/hierarchical models. Cambridge university press, 2007.
- [27] Dorien Aerts, Joeri Minnen, Ignace Glorieux, Ine Wouters, and Filip Descamps. A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison. *Building and environ*ment, 75:67–78, 2014.
- [28] Debrudra Mitra, Yiyi Chu, Nicholas Steinmetz, Paul Kremer, Jayde Lovejoy, et al. Defining typical occupancy schedules and behaviors in residential buildings using the american time use survey. ASHRAE Transactions, 125:382–390, 2019.

- [29] Máté János Lőrincz, José Luis Ramírez-Mendiola, and Jacopo Torriti. Impact of timeuse behaviour on residential energy consumption in the united kingdom. *Energies*, 14(19):6286, 2021.
- [30] Kiti Suomalainen, David Eyers, Rebecca Ford, Janet Stephenson, Ben Anderson, and Michael Jack. Detailed comparison of energy-related time-use diaries and monitored residential electricity demand. *Energy and Buildings*, 183:418–427, 2019.
- [31] Ben Anderson and Jacopo Torriti. Explaining shifts in uk electricity demand using time use data from 1974 to 2014. *Energy Policy*, 123:544–557, 2018.

## **Appendix**

### A Count Data

Weighted Counts, Exposures, and Exits at the Slot-Level. The weighted count of transitions from state s to s' for group g at time t is:

$$C_{s,s'}^{(g,t)} = \sum_{i:q_i=q} w_i \cdot \mathbf{1} \{ S_{i,t} = s \text{ and } S_{i,t+1} = s' \}.$$

The weighted number of individuals in group g exposed to the risk of leaving state s at time t with a run-length in bin  $L_m$  is:

$$N_{s,m}^{(g,t)} = \sum_{i:q_i=q} w_i \cdot \mathbf{1} \{ S_{i,t} = s \text{ and } \ell_{i,t} \in L_m \}.$$

The weighted number of individuals who exit under these conditions is:

$$E_{s,m}^{(g,t)} = \sum_{i:g_i=g} w_i \cdot \mathbf{1}\{S_{i,t} = s, S_{i,t+1} \neq s, \text{ and } \ell_{i,t} \in L_m\}.$$

**Aggregation to the Block-Level.** The block-level statistics, used to form the priors, are created by summing the slot-level statistics across all time slots t within a block b and across all demographic groups  $\mathcal{G}$ .

The total weighted transition count for  $s \to s'$  within block b is:

$$C_{s,s'}^{(b)} = \sum_{g \in \mathcal{G}} \sum_{t \in b} C_{s,s'}^{(g,t)}.$$

Similarly, the total weighted exposures and exits for the hazard model within a block are:

$$N_{s,m}^{(b)} = \sum_{g \in \mathcal{G}} \sum_{t \in b} N_{s,m}^{(g,t)}$$
 and  $E_{s,m}^{(b)} = \sum_{g \in \mathcal{G}} \sum_{t \in b} E_{s,m}^{(g,t)}$ .

These aggregated counts are then used to calculate the block-level prototype probabilities  $\bar{\theta}^{(b)}$  and  $\bar{h}^{(b)}$ .

## **B** Derivation of Posterior Means

This appendix provides the derivations for the posterior mean estimators used in the main paper for the router, hazard, and initial state models. The conjugate priors are leveraged to obtain simple, closed-form expressions for the posterior distributions.

#### **B.1** Router Model Posterior

The router model estimates the transition probabilities  $\theta_{s,:}^{(g,t)}$  from a state s to all other states s' for a specific group g at a time slot t. This is a multiclass classification problem, for which the Dirichlet-Multinomial conjugacy is well-suited.

• **Likelihood**: The observed data for transitions out of state s are the design-weighted counts  $C_{s,s'}^{(g,t)}$  for each destination state s'. The likelihood of observing these counts, given the transition probabilities  $\theta_{s,\cdot}^{(g,t)}$ , follows a Multinomial distribution:

$$p(C_{s,1}^{(g,t)},\ldots,C_{s,K}^{(g,t)}\mid\theta_{s,\cdot}^{(g,t)})\propto\prod_{s'=1}^{K}\left(\theta_{s,s'}^{(g,t)}\right)^{C_{s,s'}^{(g,t)}}.$$

• **Prior**: As specified in the paper, we place a Dirichlet prior on the vector of transition probabilities  $\theta_{s,\cdot}^{(g,t)}$ . This prior is informed by the block-level prototype  $\bar{\theta}_{s,\cdot}^{(b(t))}$  and smoothed with a small constant k. The prior is:

$$\theta_{s,\cdot}^{(g,t)} \sim \operatorname{Dir}(\alpha_1, \dots, \alpha_K),$$

where the concentration parameters are  $\alpha_{s'} = \tau_b \, \bar{\theta}_{s,s'}^{(b(t))} + k/K$ . The probability density function is:

$$p(\theta_{s,\cdot}^{(g,t)}) \propto \prod_{s'=1}^{K} \left(\theta_{s,s'}^{(g,t)}\right)^{\alpha_{s'}-1}.$$

• **Posterior**: Due to the conjugacy of the Dirichlet prior and Multinomial likelihood, the posterior distribution of  $\theta_{s,\cdot}^{(g,t)}$  is also a Dirichlet distribution. The posterior is found by multiplying the prior and the likelihood:

$$\begin{split} p(\theta_{s,\cdot}^{(g,t)} \mid \text{data}) &\propto p(\text{data} \mid \theta_{s,\cdot}^{(g,t)}) \cdot p(\theta_{s,\cdot}^{(g,t)}), \\ &\propto \left(\prod_{s'=1}^K (\theta_{s,s'}^{(g,t)})^{C_{s,s'}^{(g,t)}}\right) \cdot \left(\prod_{s'=1}^K (\theta_{s,s'}^{(g,t)})^{\alpha_{s'}-1}\right), \\ &\propto \prod_{s'=1}^K (\theta_{s,s'}^{(g,t)})^{C_{s,s'}^{(g,t)}+\alpha_{s'}-1}. \end{split}$$

This is the kernel of a new Dirichlet distribution,  $Dir(\alpha'_1, \ldots, \alpha'_K)$ , with updated parameters:

$$\alpha'_{s'} = C_{s,s'}^{(g,t)} + \alpha_{s'} = C_{s,s'}^{(g,t)} + \tau_b \, \bar{\theta}_{s,s'}^{(b(t))} + k/K$$

• **Posterior Mean**: The mean of a Dirichlet distribution  $Dir(\alpha'_1, \ldots, \alpha'_K)$  is given by  $\mathbb{E}[\theta_{s,s'}] = \frac{\alpha'_{s'}}{\sum_{i=1}^K \alpha'_i}$ . Using this, we get the posterior mean estimate  $\widehat{\theta}_{s,s'}^{(g,t)}$ :

$$\begin{split} \widehat{\theta}_{s,s'}^{(g,t)} &= \frac{C_{s,s'}^{(g,t)} + \tau_b \, \bar{\theta}_{s,s'}^{(b(t))} + k/K}{\sum_{j=1}^{K} (C_{s,j}^{(g,t)} + \tau_b \, \bar{\theta}_{s,j}^{(b(t))} + k/K)}, \\ &= \frac{C_{s,s'}^{(g,t)} + \tau_b \, \bar{\theta}_{s,s'}^{(b(t))} + k/K}{\left(\sum_{j=1}^{K} C_{s,j}^{(g,t)}\right) + \tau_b \left(\sum_{j=1}^{K} \bar{\theta}_{s,j}^{(b(t))}\right) + \left(\sum_{j=1}^{K} k/K\right)}. \end{split}$$

Since  $\sum_{j=1}^K \bar{\theta}_{s,j}^{(b(t))} = 1$ , the denominator simplifies, yielding the expression in the paper:

$$\widehat{\theta}_{s,s'}^{(g,t)} = \frac{C_{s,s'}^{(g,t)} + \tau_b \, \overline{\theta}_{s,s'}^{(b(t))} + k/K}{\sum_{j=1}^K C_{s,j}^{(g,t)} + \tau_b + k}.$$

#### **B.2** Hazard Model Posterior

The hazard model estimates the probability  $h_{s,m}^{(g,t)}$  of leaving a state s, given a run-length in bin  $L_m$ . This is a binary outcome (leave vs. stay), making the Beta-Bernoulli conjugate model the natural choice.

• **Likelihood**: The data consists of  $E_{s,m}^{(g,t)}$  weighted exits and  $N_{s,m}^{(g,t)} - E_{s,m}^{(g,t)}$  weighted stays, out of  $N_{s,m}^{(g,t)}$  total exposures. The likelihood of this outcome follows a Bernoulli process, which for aggregated counts is a Binomial distribution:

$$p(E_{s,m}^{(g,t)} \mid h_{s,m}^{(g,t)}) \propto \left(h_{s,m}^{(g,t)}\right)^{E_{s,m}^{(g,t)}} \left(1 - h_{s,m}^{(g,t)}\right)^{N_{s,m}^{(g,t)} - E_{s,m}^{(g,t)}}.$$

• **Prior**:A Beta prior is placed on the hazard probability  $h_{s,m}^{(g,t)}$ , centered on the block-level prototype  $\bar{h}_{s,m}^{(b(t))}$ . The Beta distribution is defined by two parameters,  $\alpha$  and  $\beta$ :

$$h_{s,m}^{(g,t)} \sim \text{Beta}(\alpha,\beta),$$

where  $\alpha = \kappa_b \, \bar{h}_{s,m}^{(b(t))}$  and  $\beta = \kappa_b (1 - \bar{h}_{s,m}^{(b(t))})$ . The parameter  $\kappa_b$  acts as a pseudo-count, controlling the strength of the prior.

• **Posterior**: The Beta prior is conjugate to the Bernoulli/Binomial likelihood. The posterior distribution is therefore also a Beta distribution with updated parameters  $\alpha'$  and  $\beta'$ :

$$\alpha' = \alpha + (\text{number of successes}) = \kappa_b \, \bar{h}_{s,m}^{(b(t))} + E_{s,m}^{(g,t)},$$
  
 $\beta' = \beta + (\text{number of failures}) = \kappa_b (1 - \bar{h}_{s,m}^{(b(t))}) + (N_{s,m}^{(g,t)} - E_{s,m}^{(g,t)}).$ 

• **Posterior Mean**: The mean of a Beta distribution Beta( $\alpha', \beta'$ ) is  $\frac{\alpha'}{\alpha' + \beta'}$ . Substituting our posterior parameters:

$$\begin{split} \widehat{h}_{s,m}^{(g,t)} &= \frac{\kappa_b \, \bar{h}_{s,m}^{(b(t))} + E_{s,m}^{(g,t)}}{\left(\kappa_b \, \bar{h}_{s,m}^{(b(t))} + E_{s,m}^{(g,t)}\right) + \left(\kappa_b (1 - \bar{h}_{s,m}^{(b(t))}) + N_{s,m}^{(g,t)} - E_{s,m}^{(g,t)}\right)}, \\ &= \frac{E_{s,m}^{(g,t)} + \kappa_b \, \bar{h}_{s,m}^{(b(t))}}{E_{s,m}^{(g,t)} + \kappa_b \, \bar{h}_{s,m}^{(b(t))} + \kappa_b - \kappa_b \, \bar{h}_{s,m}^{(b(t))} + N_{s,m}^{(g,t)} - E_{s,m}^{(g,t)},} \\ &= \frac{E_{s,m}^{(g,t)} + \kappa_b \, \bar{h}_{s,m}^{(b(t))}}{N_{s,m}^{(g,t)} + \kappa_b}. \end{split}$$

This is the shrinkage estimator presented in the main paper.

#### **B.3** Initial State Distribution Posterior

The derivation for the initial state distribution  $\widehat{\pi}^{(g)}$  is nearly identical to the router model, but simpler as it lacks the hierarchical prior from the block-level.

• **Likelihood**: The data are the weighted counts of respondents in group g starting their day in state s, which we denote  $C_s^{(g)} = \sum_{i:g_i=g} w_i \cdot \mathbf{1}\{S_{i,1} = s\}$ . The likelihood follows a Multinomial distribution:

$$p(C_1^{(g)}, \dots, C_K^{(g)} \mid \pi^{(g)}) \propto \prod_{s=1}^K \left(\pi_s^{(g)}\right)^{C_s^{(g)}}.$$

• **Prior**: A symmetric Dirichlet prior is used for smoothing, which corresponds to adding a small pseudo-count to each category.

$$\pi^{(g)} \sim \operatorname{Dir}(k/K, \dots, k/K).$$

• **Posterior**: The posterior is a Dirichlet distribution with parameters updated by the observed counts:

$$\pi^{(g)}\mid \mathrm{data}\sim \mathrm{Dir}(C_1^{(g)}+k/K,\ldots,C_K^{(g)}+k/K).$$

• Posterior Mean: The posterior mean is:

$$\widehat{\pi}_s^{(g)} = \frac{C_s^{(g)} + k/K}{\sum_{j=1}^K (C_j^{(g)} + k/K)} = \frac{\sum_{i:g_i = g} w_i \cdot \mathbf{1} \{ S_{i,1} = s \} + k/K}{\left( \sum_{i:g_i = g} w_i \right) + k}.$$

This matches the expression in Appendix A.

## **C** ATUS Data Columns

Table 5 lists the ATUS microdata variables (columns) that were used in constructing our dataset. These include respondent identifiers, demographic attributes, survey weights, and activity-level details.

## **Code and Data Availability**

The source code for the models, data preprocessing scripts, model weights and matrices, and analysis notebooks presented in this work are publicly available on GitHub at: https://github.com/Rohitd922/atus\_analysis. The ATUS microdata used in this study are available directly from the U.S. Bureau of Labor Statistics.

Table 5: ATUS data columns.

Column Name	Description
TUCASEID	Unique ATUS household identifier for each respondent.
TULINENO	Line number identifying the respondent within the household.
TUDIARYDATE	Date of the diary day (used to derive weekday/weekend and quarter).
TUFINLWGT	Final person-day survey weight for population inference.
TELFS	Labor force status code (employment classification).
TESEX	Respondent's sex (male/female).
TEAGE	Respondent's age in years.
${\tt TUACTIVITY\_N}$	Activity sequence number for each diary episode.
TUACTDUR24	Duration (minutes) of each reported ATUS activity.
TUTIER1CODE	First two digits of the 6-digit activity code (major activity category).
TUTIER2CODE	Middle two digits of the 6-digit activity code (subcategory).
TUTIER3CODE	Last two digits of the 6-digit activity code (fine-grained detail).
HH_SIZE	Household size (constructed from roster file).