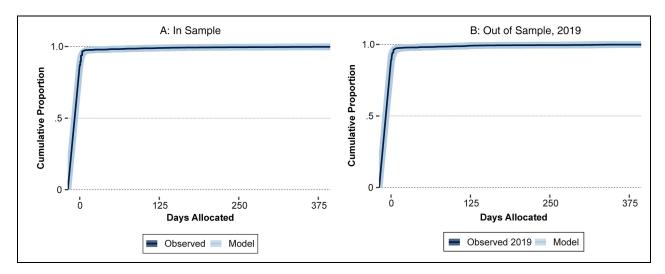**Figure 5.** Structural Model Fits the Data Distribution.
*Notes*: The dark line shows observed data ECDF. The overlapping lighter shaded lines show 20 realizations from the posterior predictive distribution. Plot axes are truncated at 375 to show detail.

Given the characteristics of practices and locations, and conditional on the known values of the error terms ($\epsilon_{mlt}$), the optimality conditions for this constrained maximization problem result in a system of equations that consist of equality conditions for positive quantities and inequality conditions for zeros. This enables us to map an observed allocation decision into its corresponding region in the error space, which forms the basis for deriving the model likelihood. Because we, the researchers, do not observe the error term, we assume it is Gumbel distributed (following Bhat 2005), thus producing a closed form for the likelihood function.

Without loss of generality, suppose the first K locations are allocated positive time, and the rest are not visited. Then the model likelihood for an observed optimal allocation ($a_1^* > 0, \cdots, a_K^* > 0, a_{K+1}^*=0, \cdots, a_L^*=0$) is given by the following:

$$\ell(a_1^*, a_2^*, \cdots, a_K^*, 0, \cdots, 0) = \left[\prod_{l=1}^{K} c_l\right]\left[\sum_{l=1}^{K} \frac{1}{c_l}\right]$$

$$\times \left[\frac{\prod_{l=1}^{K} e^{V(a_l^*)}}{\left(\sum_{l=1}^{L} e^{V(a_l^*)}\right)^K}\right] (K-1)!, \quad (4)$$

where $c_l = \left(\frac{1-\gamma_l}{a_l^*+\tau_l}\right)$ and $V(a_l^*) = \beta'X_l + \log(\gamma_l) + (\gamma_l - 1)\log(a_l^*+\tau_l)$. We provide more details for the likelihood derivation in the Web Appendix.

Practices might place different weights in evaluating the utility of their allocation decisions. To incorporate practice heterogeneity, we allow for the baseline preference and diminishing returns parameters, which determine the shape of utility function, to be practice-specific. We use a random-effects specification for the practice-level parameters as follows:

$$\theta_m \equiv (\beta_{0a}, \gamma_h^*, \gamma_a^*)_m \sim MVN(Dz_m, \Lambda), \quad (5)$$

where $z_m = (1, age_m, percent\ female_m, percent\ MD_m, practice\ size_m)$ is a vector of practice-level characteristics that shift the mean of the random-effect distribution. The matrix D contains regression coefficients, and $\Lambda$ is the covariance matrix of unobserved heterogeneity. Based on previous literature, we chose four characteristics: average physician age ($age_m$), the proportion of female physicians at the practice (percent $female_m$), the proportion of physicians that had MD degrees (vs. Doctor of Osteopathic Medicine [DO] degrees) (percent $MD_m$), and practice size (practice $size_m$). Age is associated with less frequent participation in rural outreach (O'Sullivan, Joyce, and McGrail 2014), as is being a female physician (O'Sullivan, Stoelwinder, and McGrail 2015). In general, DOs, as opposed to MDs, are more likely to practice in a rural location (Ahmed and Carmody 2022). Practice size should have an effect on the diminishing returns of utility of time spent in the home market. Larger practices require a larger patient base. Thus, spending more time at home does not necessarily grow the practice base since the number of patients, even in urban areas, is limited.

Note that $\theta_m$ does not include the baseline parameter for the home location ($\beta_{0h}$). Because one of the baseline parameters is required to be fixed for model identification, we fix the baseline for the home location to be zero (i.e., $\beta_{0h} = 0$). In addition, to ensure the diminishing returns parameters fall between 0 and 1, we reparameterize $\gamma_h$ as $1/(1 + e^{-\gamma_h^*})$ and analogously for $\gamma_a$.

## Estimation Results

Given the hierarchical structure, we estimated the model using Bayesian Markov chain Monte Carlo methods with proper but relatively diffuse priors (details in the Web Appendix). We ran the chain for 500,000 iterations, with the first 100,000 draws discarded as burn-in. We thinned the chain by retaining every tenth iteration, thus leaving 40,000 draws to summarize the posterior distribution.