ESTIMATING ON-DEMAND DELIVERY DEMANDS WITH MACHINE LEARNING AND SYNTHETIC POPULATION

Yangyang Wang
University of Michigan, School of Information
wyy@umich.edu

Tayo Fabusuyi
University of Michigan Transportation Research Institute
fabusuyi@umich.edu

Word Count: 4316 words + 2 table(s) × 250 = 4816 words

Submission Date: October 27, 2023

ABSTRACT

The rapid growth of on-demand delivery in the last decade, further accelerated by the Covid-19 pandemic, has presented challenges in predicting online shopping demand for city logistics and urban planning. However, the scarcity of comprehensive delivery data poses obstacles to developing accurate estimation methods. In this study, we propose a novel approach that combines raking-generated synthetic population with machine learning techniques to predict on-demand delivery. Leveraging publicly available datasets, we create a robust and scalable framework for estimating delivery demand at a granular geographical level. The Iterative Proportional Fitting (IPF) algorithm enables us to generate synthetic populations that capture delivery heterogeneity and spatial dynamics. Among various machine learning models, Gradient Boosting Regression emerges as the preferred choice for delivery estimation. Notably, our analysis reveals shifting feature importance over time, reflecting the impact of the pandemic on consumer behavior. Families with children assume greater significance during the pandemic, while the influence of higher income levels diminishes, signaling a democratization of on-demand services. Validation against real-world data further confirms the reliability of our approach. Our research empowers proactive city planning and resilient urban environments by providing valuable insights into on-demand delivery patterns. We acknowledge limitations and envision future work integrating real-time data sources and refining models for even more accurate estimations. The study offers a roadmap for understanding the post-pandemic dynamics of on-demand services and their implications for urban logistics and planning.

Keywords: Synthetic Population, Delivery Estimation, Statistical Methods, Machine Learning

INTRODUCTION AND BACKGROUND CONTEXT

The last decade has witnessed substantial growth in the volume of on-demand delivery. The 2017 National Household Travel Survey (NHTS) estimated that more than half of all US residents made at least one online purchase in the month the survey was carried out in 2017, representing a 12% increase compared to the previous NHTS figure carried out in 2009. The trend has accelerated significantly since the Covid-19 pandemic. Figures for the average household on-demand package delivery during the pandemic (2021) from the Puget Sound Regional Council (PSRC) Household Travel Survey (HTS) show an increase of more than 100% compared to the 2019 average household figure. The increase is even more noticeable for on-demand food and groceries delivery. On-demand delivery for food in 2021 was 625% of that observed in 2019 while demand for groceries in 2021 was 530% of the 2019 figure.

This development has implications for brick-and-mortar transactions and in planning for city logistics infrastructure, including the active management of parking and curb spaces. However, obtaining data that can influence planning activities has often been a major obstacle given that on-demand delivery data are typically owned by private sector entities who are very protective of such data. This constraint and the need to address it informs the present study. Our approach relies exclusively on the use of publicly available data in developing robust and scalable approaches for predicting demand for online shopping activities. The estimates provide information not only on the magnitude of the demand, but information is also provided that captures the heterogeneity of the goods delivered and the spatial resolution of the delivery.

Publicly available datasets utilized during the study include the PRSC HTS and the US Census Public Use Microdata Sample (PUMS) dataset from which the American Community Survey tables were generated. Using these datasets, we integrate both raking and machine learning methods to create a replicable and reproducible approach for estimating the demand for at-home delivery online purchases in granular geographical areas. These estimates, associated with synthetic households obtained through an iterative proportional fitting (IPF) method, will be generated in a manner that explicitly addresses the heterogeneity and the spatial dimension of the purchases. A key innovation is developing a framework that enables the estimation of the volume of online purchases and determining the characteristics of the household and/or person making these purchases and the spatial variation observed for specific locales. This provides potential policy levers, both in terms of the variables of interest and incentives that could be directed to specific segments of the population.

The information provided by the study could be invaluable in assisting municipalities to understand the rapidly evolving patterns and trends in on-demand delivery, and enabling them to be proactive and resilient in managing the upsurge in e-commerce activities. The insight affords a richer assessment of the local e-commerce economy – particularly what the current landscape is; what the primary drivers of change are, in what direction they are headed, and by what magnitude as well as what factors may be amenable to policy modification. From these, policy responses that may ensure sustainable and resilient cities could subsequently be put in place. These policies may address the location of micro-fulfillment facilities to maximize coverage within the municipality or in the real-time management of curb spaces for fulfilling orders.

The balance of the paper is organized as follows: The review of existing work, primarily on the methodology, including the innovative approach used in our study is covered in Section 2. Data issues are discussed in Section 3 while the study's empirical framework is discussed in Section 4. Computational results and the discussion of the findings are presented in Section 5.

Section 6 concludes and provides some direction for future research.

LITERATURE REVIEW

Existing Works

Synthetic populations emerge as guardians of data confidentiality, as demonstrated by Stojanovic and Smith (*1*) in their exploration of the raking method's efficacy in generating realistic populations suitable for analysis. Wilson and Whelan (*2*) further underscore the raking method's practicality in estimating small-area statistics, effectively surmounting data limitations that often hinder such analyses. Meanwhile, Deac et al. (*3*) contribute a valuable comparative analysis of the raking method and iterative proportional fitting, guiding researchers and policymakers toward the most appropriate method for accurately representing original data distributions through synthetic populations. Additionally, Beard et al. (*4*) delve into the raking method's intricacies within the context of small-area estimation, shedding light on its advantages and limitations for researchers seeking detailed data at localized geographical scales. Lastly, the comprehensive review by Ballas and Tatem (*5*) illuminates diverse applications of synthetic populations for agent-based modeling, unveiling the strengths and weaknesses of various methods, including the raking technique, and providing essential guidance for researchers in their explorations across multifaceted domains.

Delivery demand estimation is a crucial area of research that has gained significant attention in recent years due to the rapid growth of e-commerce and the increasing popularity of on-demand delivery services. Several studies have been conducted to estimate and understand the demand for various types of deliveries, ranging from online package deliveries to on-demand food delivery services. Fabusuyi (*6*) addresses the challenge of estimating online delivery purchases for small geographic areas, utilizing publicly available data sets to generate demand estimates. Similarly, Zhu et al. (*7*) focus on predicting the Order Fulfillment Cycle Time (OFCT) for on-demand food delivery platforms, crucial for customer satisfaction and efficient planning decisions. Liu et al. (*8*) explore the impacts of collection-delivery points as a greener last-mile alternative, shedding light on mode choice and trip-chaining decisions in delivery trips. These studies collectively demonstrate the significance of accurate delivery demand estimation in optimizing logistics operations, improving customer experiences, and mitigating environmental impacts. By leveraging advanced methodologies, such as deep neural networks and panel cross-nested logit models, these research efforts provide valuable insights for policymakers and stakeholders seeking to enhance delivery services while promoting sustainable transportation options. However, a need remains for further research in this area, particularly in addressing the challenges posed by the evolving delivery landscape and the integration of emerging technologies.

Identified Gaps

Uncharted territories: Despite the valuable insights provided by existing studies, there remains a crucial need for more comprehensive validation of synthetic population and raking methods. To ensure the accuracy and reliability of these methods across diverse contexts, researchers must rigorously compare the characteristics of synthetic populations with real-world data. Such validation exercises would shed light on the effectiveness and applicability of synthetic population generation techniques, helping to address potential limitations and uncover uncharted territories for further exploration.

Choosing the right tool: The field of synthetic population generation lacks sufficient comparative studies that thoroughly assess the performance of different methods, including the raking technique. Without such comparative analyses, researchers may encounter challenges in selecting the most appropriate method for their specific research or practical applications. To empower decision-makers and researchers, it is imperative to conduct in-depth comparative studies that thoroughly evaluate the strengths and weaknesses of various synthetic population generation approaches. Armed with this knowledge, researchers can confidently choose the right tool for their analytical needs and enhance the validity of their findings.

Embracing uncertainty: Despite the advancements in synthetic population generation, the literature often falls short in adequately addressing the uncertainty associated with these methods, particularly concerning the raking technique. Quantifying and understanding this uncertainty are vital to improve the reliability of synthetic populations and their application in decision-making processes. Embracing uncertainty allows researchers and decision-makers to appreciate the potential limitations and risks associated with synthetic population data, guiding them in making more informed and cautious decisions based on the available information. By recognizing and embracing uncertainty, researchers can work towards refining synthetic population generation methods and enhancing the overall credibility of their results.

DATA

Data Source

The American Community Survey (ACS) Public Use Microdata Sample (PUMS) files contain untabulated records concerning individuals or housing units, allowing access to marginal sums based on various dimensions like sex, age, and housing size, at the granularity of the census block. However, these files do not record detailed information about each individual or household, as they only offer partial samples. Therefore, generating synthetic populations becomes essential, particularly for specific geographical areas.

To estimate deliveries, we utilize the Puget Sound Regional Council Travel Survey 2017 (PSRC), which collected responses from 2665 households in the Seattle-Tacoma-Bellevue metropolitan statistical area. Each household provided information about their annual income, employment status, education level, online orders, and more. Another crucial data source is the National Household Travel Survey 2017 (NHTS). We focus on specific features from both datasets that are associated with a particular region and time-segment, particularly the NHTS samples in the Seattle area with Zip Code "42660" (NHTS-cut), referenced in both PSRC and NHTS datasets. Although the households in these two sets of observations might not be identical, they offer two sample pools from the same population. Figure 1 illustrates the representation of a temporally-restricted portion of PSRC data and a spatially-restricted section of NHTS data.

Explanatory Variables of Interest

6 dummy variables are used in this research:

- HHINC: Whether the annual income of the household is greater than $\$100,000$.
- HHCHILD: whether the household has at least one member less than 18 years old.
- HHVEH: whether the household has at least one vehicle for daily usage.
- FEMALE: whether the individual is a female.
- SENIOR: whether the individual is older than 65.
- COLLEGE: whether the individual has completed a college degree at least.

To address the imputation of household-level variables into individual-level data, six dummy variables are of interest, with three at the household level (HHINC, HHCHILD, HHVEH) and three

Data Source		PSRC 2017	NHTS-cut 2017	PSRC 2021
Income:	> \$100,000:	1171	88	340
	\$75,000 - \$100,000:	369	70	154
	< \$75,000:	941	106	557
	Missing:	184	8	557
	Total households:	2665	272	2020
Age:	< 25:	940	67	72
	25 - 45:	2282	161	414
	45 - 65:	1122	156	347
	> 65:	580	104	284
	Missing:	0	0	2056
	Total individuals:	4924	488	3173
Education:	< high school:	40	17	18
	High school grad.:	203	46	116
	Technical training:	98	0	38
	Associate degree:	644	99	352
	Bachelor degree:	1762	162	346
	Graduate degree:	1472	122	247
	Missing:	705	42	2056
	Total individuals:	4924	488	3173
Gender:	Female:	2440	257	593
	Male:	2396	231	502
	Missing:	88	0	2078
	Total individuals:	4924	488	3173
Life cycle:	2 adults, no children:	1254	124	1
	1 adult, no children:	880	73	1
	1 adult, with children:	531	6	1
	2 adults, with children:	0	69	1113
	Missing:	0	0	904
	Total households:	2665	272	2020
Employment:	Full time:	2600	236	527
	Retired:	538	0	254
	Part time:	344	45	100
	Freelancer:	268	0	59
	Not employed:	239	0	99
	Homemaker:	197	0	47
	Volunteer(intern):	33	0	8
	Missing:	705	207	2079
	Total individuals:	4924	488	3173

TABLE 1 Statistics on the size of our data sources

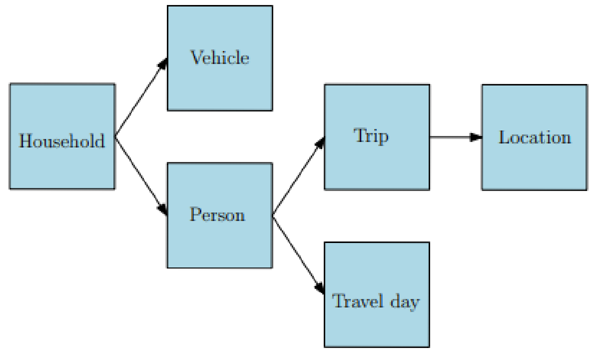

FIGURE 1 Hierarchical Structure of NHTS and PSRC dataset

at the individual level (SENIOR, FEMALE, COLLEGE). Since the delivery data to be predicted pertains to individuals, methods are proposed to estimate the missing household-level variables at the individual level.

Harmonization

For the original data, this task is relatively straightforward because both the household dataset and the individual dataset share a common key that indicates the household number. By merging or joining these two datasets using the key, all the household-level variables can be filled for each individual based on the household they belong to.

However, dealing with marginal sums poses a more complex challenge. In this case, only sums at the household or individual level are available for each variable. To convert these sums to the individual level, a set of unique imputation factors is used. These imputation factors are calculated as the average of individual-level data within each type of household.

By applying these imputation factors to the household-level sums, they can be appropriately converted to individual-level estimates, thus enabling the utilization of the complete dataset for delivery data prediction.

- For example, if 5 households are in **HHCHILD = True** category, with individual number $2,3,4,3,4$, respectively, then the imputation factor for **HHCHILD = True** would be $(2+3+4+3+4)/5 = 3.2$, which means the marginal sum for **HHCHILD = True** would need to be multiplied by 3.2 to be converted to individual level.

EMPIRICAL FRAMEWORK

Raking Method

Iterative Proportional Fitting

With original data matrix M and marginal sum in 6 dimensions $(u_1, u_2, ..., u_6)$, the Iterative Proportional Fitting (IPF) algorithm could be performed to generate synthesized data N.

It is essential to acknowledge that, due to the imputation of 3 out of 6 marginal sums from household level to individual level, the total sum N might not exactly represent the population number of the area. To address this issue, all the values are scaled to ensure that the total sum corresponds precisely to the population number.

After scaling, the final result is a matrix containing decimal numbers, which must be inte-

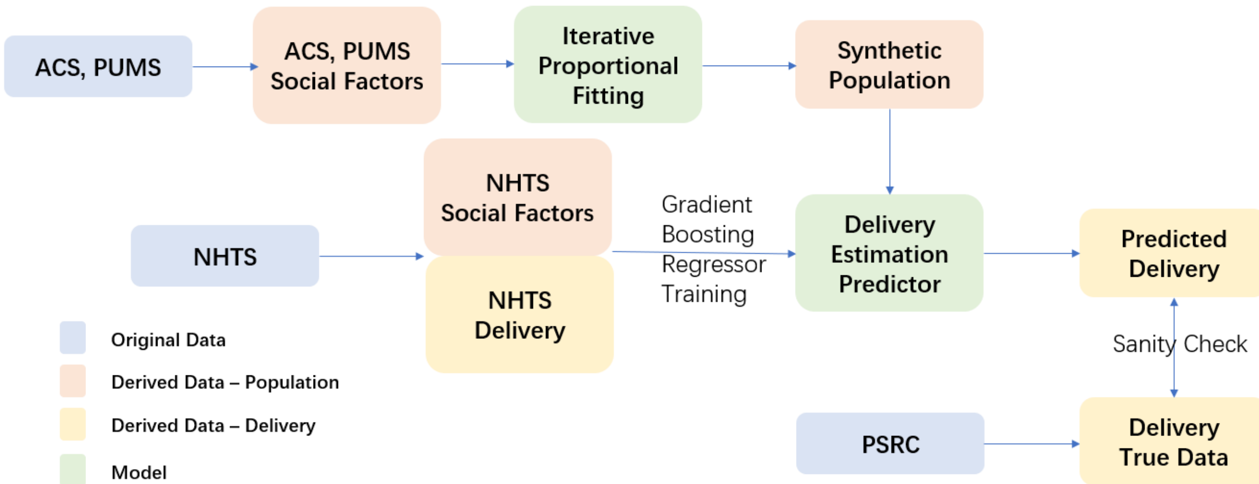

FIGURE 2 Panoramic Chart

Algorithm 1 IPF(M, u, ε)

Input M: original matrix; $(u_1, u_2, \ldots, u_D)$: marginal sum in each dimension
ε: error tolerance
Output N: generated matrix

0: N = empty array with same shape as M
0: δ = Euclidean distance between N and M
0: **while** $\delta > \varepsilon$ **do**
0: **for** d in number of dimension of M **do**
0: d_sums = sum of each unit of M in dimension d
0: **for** each element in N (represented as N[i,j,...,k,...], where k is the ordinal in dimension d) **do**
0: N[i,j,...,k,...] = M[i,j,...,k,...] * u_d[k] / d_sums[k]
0: **end for**
0: **end for**
0: **end while**=0

gerized since the population should be represented as integers. The integerization process is crucial to maintain the discrete nature of the population count, and it is typically achieved using appropriate rounding or truncation methods. By integerizing the values, the results will be consistent with the expected whole-number population representation.

Integerization

The method "Truncate, replicate, sample" (TRS) proposed by Robin Lovelace and Dimitris Ballas is adopted to integerize N. By treating decimal as probability estimation, TSR unfolds into 3 steps.

1. **Truncate**: Take the integer part of N as N'. Leave the residual part $n = N - N'$.
2. **Replicate**: Take N' as the base of the final result because integers are confirmed numbers generated by IPF.
3. **Sample**: For the decimal part n, treat it as a probability distribution matrix and make random sampling. After the first sample, which might fall in any cells in the matrix, is generated, add it to N' in the corresponding cell. Repeat the sampling process until N' gets to the same total as N has.

After this process, N' is now an integer matrix generated using integerized IPF method that represents the synthetic population in this area. It could be used in different application scenes such as estimating deliveries.

Estimating Deliveries

Prediction (Gradient Boosting)

Social factors, a 6-dimensional array with True/False value, is treated as x variable. The delivery data is treated as y variable. A machine learning algorithm could be used to train for predicting y based on x using regression. Traditional regression methods have been conducted and evaluated in terms of accuracy.

Gradient boosting is a powerful ensemble method that combines weak learners to create a strong learner in an iterative manner. Its concept is easiest to understand in the context of least-squares regression, where the objective is to build a model F to predict values $\hat{y} = F(x)$ by minimizing the mean squared error $\frac{1}{n}\sum_i (\hat{y}_i - y_i)^2$, where i indexes over a training set of size n with actual values of the output variable y:

- $\hat{y}_i =$ the predicted value $F(x_i)$
- $y_i =$ the observed value
- $n =$ the number of samples in y

In the gradient boosting algorithm with M stages, at each stage m (where $1 \leq m \leq M$), a model F_m is already in place. To improve F_m, the algorithm introduces a new estimator, $h_m(x)$, and updates F_m as follows:

$$F_{m+1}(x_i) = F_m(x_i) + h_m(x_i) = y_i$$

or, equivalently,

$$h_m(x_i) = y_i - F_m(x_i)$$

In this process, gradient boosting fits h_m to the residual $y_i - F_m(x_i)$. The key idea is that each F_{m+1} aims to correct the errors made by its predecessor F_m. This concept can be extended to other loss functions beyond squared error and applied to classification and ranking problems. The generalization involves noticing that the residuals $h_m(x_i)$ for a given model are proportional to the negative gradients of the mean squared error (MSE) loss function with respect to $F(x_i)$:

$$L_{\text{MSE}} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - F(x_i)\right)^2$$

$$-\frac{\partial L_{\text{MSE}}}{\partial F(x_i)} = \frac{2}{n}\left(y_i - F(x_i)\right) = \frac{2}{n}h_m(x_i).$$

This observation shows that gradient boosting can be specialized to a gradient descent algorithm, and its generalization involves incorporating different loss functions and their gradients.

Integerization (Permutation)

To address the issue of converting the delivery values to integers, we utilize the Permutation method, given that Gradient Boosting Regression yields decimal values.

In consideration of our categorical features, which are associated with bins, we treat each sample as a binary vector in a d-dimensional space, where d represents the number of features. Given two binary vectors $A = \left(0, 1^d\right)$ and $B = \left(0, 1^d\right)$ with the same number of coordinates, we define the Hamming distance, denoted as HD, as the metric to assess their dissimilarity.

$$\text{HD}(A,B) := \frac{\text{number of unequal corresponding coordinates}}{\text{number of coordinates in } A \text{ or } B}.$$

We expand the training dataset by including a constant number, n, of training samples $x^{i_1,i_2,\ldots,i_n}train$ for each testing sample $x^i test$. While the corresponding delivery values $y^{i_1,i_2,\ldots,i_n}train$ may vary, we use the mode value, denoted as *yanchor*, as the reference value for this testing sample.

After obtaining predictions from the Gradient Boosting Regression, if the predicted value $\hat{y}$ is greater than the anchor value, i.e., $\hat{y} > y_{anchor}$, the prediction is rounded down. Conversely, if $\hat{y} < y_{anchor}$, the prediction is rounded up. This rounding process ensures that the predicted delivery values are integers, as required.

COMPUTATIONAL RESULTS AND DISCUSSIONS

Outcomes and Model Comparison

In order to select the most suitable machine learning algorithm for delivery estimation, we conduct an experiment using different types of models. The selected models include:

- Linear model (Linear Regression)
- Tree-based model (Random Forest Regression)
- Neighbor-based model (K Nearest Neighbors Regression)
- Kernel-based model (Support Vector Regression)
- Ensemble model (Gradient Boosting Regression)

Bias is the critical factor in our algorithm's prediction accuracy. To assess the bias component (excluding the variance part) in the prediction errors, we employ 5-fold cross-validation. This method involves dividing the data into five subsets, using four subsets for training, and the remaining subset for testing. This process is repeated five times, each time using a different subset for testing. The average of the five sets of predicted values, denoted as $\bar{\hat{y}}$, is considered the final result.

To evaluate the performance of the models, we use a metric that captures the average squared difference between $\bar{\hat{y}}$ and the true value y. This metric helps us measure the accuracy of the predictions and enables us to compare the performance of different algorithms in terms of bias in the delivery estimation task.

The result from Table 2 indicates that Gradient Boosting Regression is the preferred model.

Model	Bias Square
Linear Regression	18.5706
Random Forest Regression	17.5199
K Nearest Neighbors Regression	19.5086
Support Vector Regression	18.0322
Gradient Boosting Regression	17.5185

TABLE 2 Bias of Different Models

Robustness and Validation

In this study, the prediction results are validated using the PSRC dataset, which collects survey data from the Seattle area in 2017. The data to be validated corresponds to the delivery estimation for the Seattle area during the same year.

From the PSRC dataset, two specific sub-datasets, namely the Person and Days datasets, are selected and processed for this validation task. The Days dataset provides valuable information related to deliveries, which can be summarized for each individual during workday surveys. The delivery data is then aggregated for each individual based on their unique individual ID. During the aggregation process, each delivery data record is weighted according to the workday's weight specified in the Days dataset. Furthermore, each individual is assigned to the corresponding census tract based on their household's location. By performing these steps, delivery data for all individuals in each census tract can be calculated.

It is worth noting that the distribution of the delivery data is not symmetric. Certain census tracts exhibit a more significant proportion of small values or even 0 values, which could result from missing records or insufficient data points. Consequently, the data distribution exhibits characteristics of a Lognormal Distribution. To address this, all 0 values are removed, and a logarithm transformation with a base of 2 is applied to the data. This transformation ensures that the processed data now follows a normal distribution, which facilitates the validation and comparison of prediction results.

Using the properties in Normal Distribution, we could easily calculate the 99% confidence interval for each value in the census tract using sample variance s=2.2 and sample size n according to the record number. Suppose random variable $M = \log_2 X$, where X denotes the delivery data. The confidence interval could be expressed as

$$\hat{\mu}_L = \bar{m} - z_{\alpha/2} \cdot \sigma/\sqrt{n},\ \hat{\mu}_U = \bar{m} + z_{\alpha/2} \cdot \sigma/\sqrt{n}$$

After that, the confidence interval for delivery data X would be $(2^{\hat{\mu}_L}, 2^{\hat{\mu}_U})$.

The proportion of census tract where the estimated value lies in the 99% confidence interval is 63.6%.

Reflections

Feature Importance over Time

The employed machine learning technique in section 5.2 offers the opportunity to investigate the significance of features. Given the structure of the NHTS dataset comprising three distinct temporal sections, namely 2017, 2019, and 2021, each segment can be modeled independently. To evaluate the feature importance in each year's model, a tree-based Gini Impurity method is utilized.

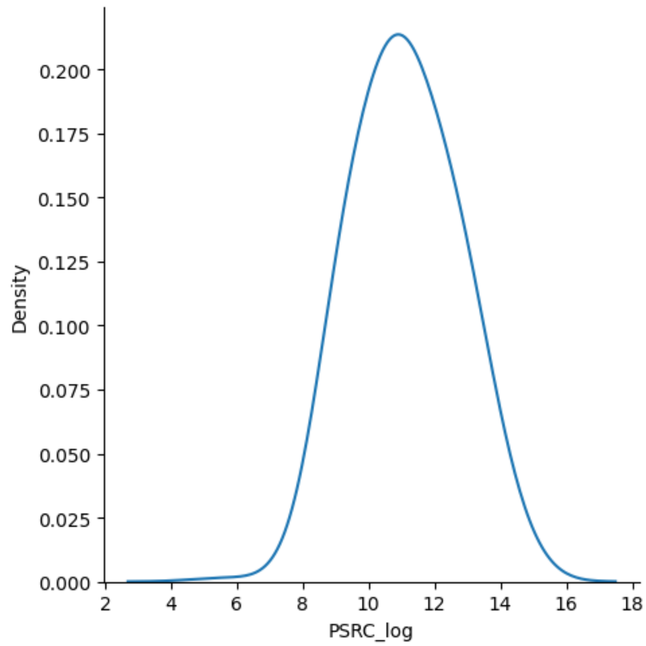

FIGURE 3 Distribution of Delivery Values after Log Transformation

Importantly, it is essential to acknowledge that discernible disparities exist between the pre-pandemic period (2018-2019) and the post-pandemic era, which could impact the results and interpretations of the models.

In the post-pandemic context, the factor of "Individuals whose household has child/children" has assumed heightened significance in the estimation of deliveries. This observation is likely attributed to the fact that during this period, certain special needs of children can only be fulfilled through deliveries, considering the limitations on visiting on-site places. Conversely, in the pre-pandemic era, there were more opportunities for families to access on-site establishments, reducing the reliance on deliveries.

In contrast, the factor of "Individuals whose household has monthly salary larger than $\$100,000$" has exhibited reduced importance in the estimation of deliveries. This trend is possibly linked to the widespread closure of on-site establishments during the pandemic. Consequently, families, regardless of their income level, increasingly turned to deliveries as a preferred means of obtaining goods and services, thereby diminishing the impact of higher income on delivery choices.

Equivalency of NHTS and PSRC Dataset

To verify the cross-validation and common attributes between the PSRC2017 and NHTS2017 datasets, the mean delivery values for different categories of individuals were computed in both datasets. Out of the 64 different categories, 34 categories showed matching mean delivery values in both datasets.

To further assess their similarity, the delivery values of these 34 common categories in each dataset were ranked from highest to lowest, resulting in rankings from 1 to 34. The Mean Absolute Error (MAE) was then calculated by measuring the discrepancy between the two sets of rankings,

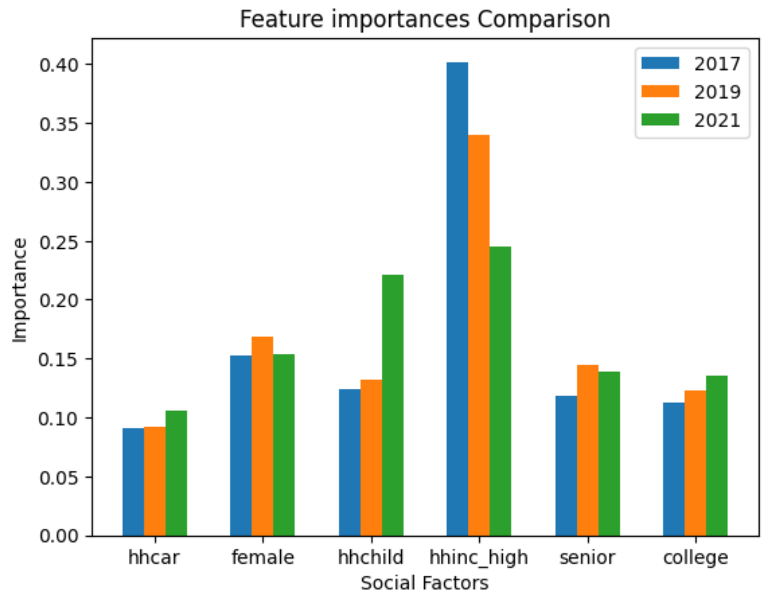

FIGURE 4 Feature Importance over the Years

yielding a value of 9.243.

To demonstrate the significance of this result, 100,000 bootstrap experiments were conducted. In each experiment, two random sequences of integers from 1 to 34 were generated, and their respective MAE values were calculated. The distribution of 10,000 MAE values from these experiments displayed an average of 11.321.

Remarkably, the computed MAE value of 9.243 was found to lie approximately within the lowest 5% quantile of the bootstrap-generated MAE distribution. This compelling evidence confirms the statistical significance of the result, providing strong support for the similarity between the two datasets in terms of the relationship between social factors and delivery values.

Future Research Directions

The "Feature Importance" section in our study provides insights into the importance trends of specific features over time. On the other hand, the "Equivalency of NHTS and PSRC Dataset" section justifies the possibility of interpolating data between the PSRC and NHTS datasets. In the current situation, our PSRC dataset has data only in the Seattle area, covering the period from 2017 to 2021. Conversely, the NHTS dataset includes data from a nationwide Metropolitan Area but is limited to the year 2017.

To overcome the data limitations and deduce delivery values nationwide for the year 2017, we utilize the available NHTS dataset. We assume that the discovered trend of feature importance in the "Feature Importance" section is applicable to different geographic areas, a justification supported by the findings in the "Equivalency of NHTS and PSRC Dataset" section.

Based on this assumption, we extend the application of the identified feature importance trend to other Metropolitan Areas. Consequently, we can interpolate the nationwide delivery data for the years 2019 and 2021. This process allows for the chronological observation of the evolution of delivery data in various geographical areas. By adopting this methodology, we effectively fill the data gaps and gain a comprehensive understanding of the temporal trends in delivery data across different regions, despite the limitations posed by the individual datasets.

CONCLUSIONS

Our study introduces a novel approach for on-demand delivery estimation, merging raking-generated synthetic population with powerful machine learning techniques. The rapid growth of on-demand delivery, intensified by the Covid-19 pandemic, demands scalable and robust prediction methods. Leveraging publicly available datasets, we create a reliable framework for estimating online shopping demand at a granular geographical level.

The Iterative Proportional Fitting (IPF) algorithm enables us to generate synthetic populations that capture delivery heterogeneity and spatial dynamics. Our analysis of various machine learning models identifies Gradient Boosting Regression as the superior choice for delivery estimation.

Remarkably, we uncover shifting feature importance over time, reflecting the profound impact of the pandemic on consumer behavior. Families with children show heightened significance during the pandemic, while the influence of higher income levels diminishes, indicating a democratization of on-demand services. Validation against the PSRC dataset further confirms the robustness of our approach.

In conclusion, our research pioneers on-demand delivery estimation, enabling proactive city planning and resilient urban environments. We acknowledge limitations and envision integrating real-time data sources and refining our models for even more accurate estimations. Post-pandemic behavioral dynamics warrant further exploration, shaping the future of urban logistics and planning.

REFERENCES

1. Stojanovic, J. and M. Smith, Exploration of the raking method's efficacy in generating realistic populations suitable for analysis. *Journal of Data Analysis*, Year.
2. Wilson, J. and S. Whelan, Underscore the raking method's practicality in estimating small-area statistics. *Statistical Methods*, Year.
3. Deac, R., J. Author2, and M. Author3, Comparative analysis of the raking method and iterative proportional fitting for accurately representing original data distributions through synthetic populations. *Population Studies*, Year.
4. Beard, D., S. Author2, and T. Author3, In-depth exploration of the raking method's intricacies within the context of small-area estimation. *Journal of Spatial Analysis*, Year.
5. Ballas, M. and J. Tatem, Comprehensive review of diverse applications of synthetic populations for agent-based modeling, including the strengths and weaknesses of various methods, such as the raking technique. *Agent-Based Modeling Journal*, Year.
6. Fabusuyi, T., R. Twumasi-Boakye, A. Broaddus, J. Fishelson, and R. C. Hampshire, Estimating small area demand for online package delivery. *Journal of Transport Geography*, Vol. 88, 2020, p. 102864.
7. Zhu, L., W. Yu, K. Zhou, X. Wang, W. Feng, P. Wang, N. Chen, and P. Lee, Order Fulfillment Cycle Time Estimation for On-Demand Food Delivery. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Association for Computing Machinery, New York, NY, USA, 2020, KDD '20, p. 2571–2580.
8. Liu, C., Q. Wang, and Y. O. Susilo, Assessing the impacts of collection-delivery points to individual's activity-travel patterns: A greener last mile alternative? *Transportation Research Part E: Logistics and Transportation Review*, Vol. 121, 2019, pp. 84–99, special Issue on Supply chain elements in freight transport modelling; Guest Edited by Lóránt Tavasszy and Hanno Friedrich.
9. Bierlaire, M., BIOGEME: A free package for the estimation of discrete choice models. In *3rd Swiss Transportation Research Conference*, Ascona, Switzerland, 2003.
10. Bierlaire, M., *An Introduction to BIOGEME Version 1.6*, 2008.
11. Garrow, L. A., T. D. Bodea, and M. Lee, Generation of synthetic datasets for discrete choice analysis. *Transportation*, Vol. 37, No. 2, 2009, pp. 183–202.
12. Koppelman, F. S. and L. A. Garrow, Efficiently Estimating Nested Logit Models with Choice-Based Samples: Example Applications. *Transportation Research Record, Journal of the Transportation Research Board*, Vol. 1921, No. 1, 2005, pp. 63–69.
13. Transportation Research Board, *Information for Authors: a guide for preparing and submitting manuscripts for presentation at the TRB Annual Meeting and for Publication in TRB's Journal*. Washington, D.C., 2012.
14. Donders, A. R. T., G. J. Van Der Heijden, T. Stijnen, and K. G. Moons, A gentle introduction to imputation of missing values. *Journal of clinical epidemiology*, Vol. 59, No. 10, 2006, pp. 1087–1091.
15. Smith, L., R. Beckman, and K. Baggerly, *TRANSIMS: Transportation analysis and simulation system*. Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 1995.
16. Williamson, P., M. Birkin, and P. H. Rees, The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environment and Planning A*, Vol. 30, No. 5, 1998, pp. 785–816.

17. Lovelace, R. and D. Ballas, 'Truncate, replicate, sample': A method for creating integer weights for spatial microsimulation. *Computers, Environment and Urban Systems*, Vol. 41, 2013, pp. 1–11.

18. Natekin, A. and A. Knoll, Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, Vol. 7, 2013, p. 21.

19. Bentéjac, C., A. Csörgő, and G. Martínez-Muñoz, A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, Vol. 54, 2021, pp. 1937–1967.

20. Fabusuyi, T., R. Twumasi-Boakye, A. Broaddus, J. Fishelson, and R. Hampshire, Estimating small area demand for online package delivery. *Journal of transport geography*, Vol. 88, 2020, p. 102864.

21. Disha, R. A. and S. Waheed, Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. *Cybersecurity*, Vol. 5, No. 1, 2022, p. 1.