ARCA: Forecasting Demand for Device Accessories at Amazon

Kyriaki Dimitriadou
kikid@amazon.com
Amazon
Seattle, Washington, USA

Vicky Yu
vickyyu@amazon.com
Amazon
Seattle, Washington, USA

Michael Behrman
mbehrman@amazon.com
Amazon
Seattle, Washington, USA

Lihong Xie
lihxie@amazon.com
Amazon
Sunnyvale, California, USA

Kunal Bhosale
bhkuna@amazon.com
Amazon
Sunnyvale, California, USA

Lynn Teresi
lteresi@amazon.com
Amazon
Sunnyvale, California, USA

Abstract

Multi-horizon and multi-lead time forecasting is a well-established area in machine learning, particularly in demand forecasting, which plays a critical role in a product's lifecycle. Accurate forecasts support key operational functions, including inventory management, financial planning, promotion planning, and supply chain optimization. Traditional demand forecasting methods typically rely on learning sales patterns directly from historical data of a given product. However, forecasting demand for accessories—products that are purchased in conjunction with main devices (e.g., covers or headphones for tablets)—introduces additional complexities. In this paper, we propose a novel forecasting technique for accessories that leverages their inherent attach rate patterns to main devices. Additionally, we introduce a correction module to mitigate biases in the forecasts of the main products, thereby improving the accuracy of accessory predictions. While our primary focus is forecasting accessories for Amazon devices, the proposed methodology is broadly applicable to any product that exhibits an attach rate dependency. The proposed model was deployed in production within Amazon in July 2024 and has since been generating daily accessory forecasts across 17 countries and two channels: Online (Amazon website) and Offline (third-party retailers).

ACM Reference Format:
Kyriaki Dimitriadou, Vicky Yu, Michael Behrman, Lihong Xie, Kunal Bhosale, and Lynn Teresi. 2025. ARCA: Forecasting Demand for Device Accessories at Amazon. In *Proceedings of the 1st Workshop on "AI for Supply Chain: Today and Future" @ 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25), August 3, 2025, Toronto, ON, Canada.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/XXXXXX.XXXXXX

1 Introduction

At Amazon, accessories—such as cover cases, chargers, keyboards, and headphones—play an important role in enhancing the customer experience by complementing main devices and improving their functionality. Despite their importance, forecasting demand for accessories is often overlooked or treated as a secondary concern in traditional demand models. Accurate forecasting for accessories is valuable for optimizing inventory, planning promotions, and making effective product allocation decisions across both online and offline channels.

Forecasting accessory demand presents unique challenges due to their inherent dependency on the sales and lifecycle of the corresponding main devices. Unlike primary products, which follow independent demand trends driven by market forces, seasonality, and promotions, accessory sales are closely tied to the sales of main devices through attach rates. This creates complexities, such as variable attach rates, purchase timing lags between accessory and main device sales, and dependencies across accessory types (e.g., customers purchasing a tablet cover and a stylus together). Additionally, misaligned product lifecycles, new device launches, and external factors like competition and supply chain disruptions further complicate forecasting efforts. Traditional time-series forecasting models, often used for main devices, cannot adequately capture all of these dynamics for accessories.

To address these challenges, this paper proposes an attach rate-based forecasting model for accessory demand, aligning with how product managers and planners think about and manage accessory sales. By focusing on the relationship between main devices and their accessories, this approach offers greater interpretability and transparency, as planners can directly relate the forecast to historical attach rate patterns and promotional activities. In contrast to traditional forecasting models, which may be perceived as a "black box" to non-technical stakeholders, our approach provides a more intuitive and actionable solution.

We introduce a novel forecasting framework, named **ARCA** (Attach Rate with Correction for Accessories) that combines an attach rate prediction model with a residual correction model. The attach rate model predicts future attach rate of each accessory based on historical sales of the accessory itself, as well as the main device, while the Residual Model corrects biases introduced by inaccuracies in the main device forecasts. Deployed in production at Amazon in July 2024, this model has shown promising results, including high adoption and a significant reduction in forecast errors.

The contributions of this paper include:

(1) We propose a dynamic attach rate-based forecasting model that leverages machine learning to predict attach rates while addressing outliers with a dedicated approach. This model improves forecasting accuracy by capturing demand variations beyond static attach rate assumptions. Additionally, it aligns closely with how product managers and planners

KDD '25, Toronto, ON, Canada.

ACM ISBN 979-8-4007-1454-2/25/08
https://doi.org/10.1145/XXXXXX.XXXXXX

think about accessory demand, making the forecasts more interpretable, actionable, and consistent with real-world decision-making.

(2) Our framework improves forecast precision by incorporating a residual correction model that adjusts for biases in attach rate estimations and main device forecasts, ensuring adaptability to shifting demand patterns.
(3) We introduce a correlation-weighted mechanism that dynamically adjusts the reliance on attach rate predictions based on the strength of accessory-to-main-device relationships. This approach enhances forecasting accuracy for products with varying levels of dependency, offering a flexible and robust solution.

The rest of the paper is organized as follows: in Section 2 we present the related work on accessory forecasting. In Section 3 we describe our framework including details on the attach rate model, as well as the residual correction module. In Section 4 we present our experimental results. Finally, we conclude our work in section 5 and end with the customer problem statement.

2 Related work

Forecasting accessory demand, particularly in relation to their associated main devices, remains an underexplored domain within demand forecasting research. Traditional approaches typically focus on standalone products and rely on time-series models applied to historical sales data to generate demand predictions [6, 7]. More recent advancements have leveraged deep learning techniques for direct product-level forecasting. For instance, Salinas et al. [11] propose an RNN-based method that produces probabilistic forecasts by learning from a large collection of related time series, enabling accurate item-level demand predictions even in cases of sparse data. Rangapuram et al. [10] combine state space models with deep learning to develop a probabilistic forecasting framework that balances interpretability and data efficiency while capturing complex temporal patterns.

In the automotive sector, [4] employ Deep Neural Networks (DNNs) and Recurrent Neural Networks (RNNs) to predict vehicle accessory demand by modeling complex temporal dependencies in sales data, addressing the absence of standardized industry approaches for accessory demand forecasting. However, this study does not incorporate human-in-the-loop elements such as product planners' input, nor does it exploit attach rate relationships between accessories and vehicle sales. Similarly, Ramosaj et al. [9] compare statistical methods and machine learning algorithms for accessory sales forecasting in a medium-sized Swiss enterprise, finding that SARIMAX models [1] enhanced with human expert input outperform purely data-driven approaches. Unlike our work, their study does not utilize or evaluate attach rate–based modeling approaches, which are central to our proposed methodology.

Additionally, Arvan et al. [2] provide a systematic review of judgmental demand forecasting, emphasizing the integration of expert judgment with quantitative models. In contrast, our paper introduces a novel forecasting approach for accessory demand that leverages data-driven, attach rate–based machine learning models combined with a bias correction module, distinguishing it from existing reviews and prior works.

Overall, accessory demand forecasting requires novel approaches that account for product dependencies, lifecycle misalignment, and cross-product interactions. ARCA advances the field by integrating machine learning with correction mechanisms in a structured and interpretable way, improving forecast accuracy while enhancing explainability.

3 Modeling Approach

3.1 Framework Overview

The Accessory Prediction Framework, named ARCA (figure 1) outlines a structured approach to forecasting accessory demand by integrating historical sales data, pricing information, and main device forecasts. The framework consists of two primary components: the **Accessory Attach Rate Model** and the **Residual Model**, each addressing distinct stages of the prediction process.

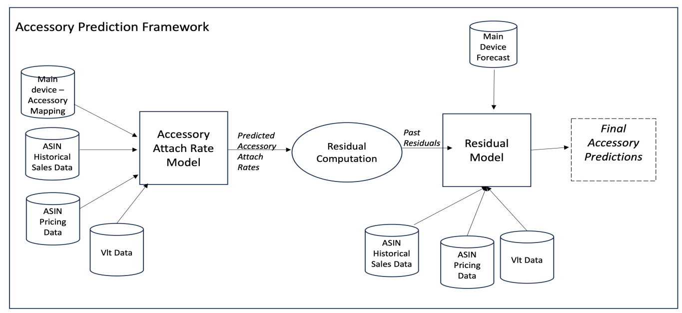

Figure 1: ARCA: forecasting accessory demand using an attach rate-based model with correction for accessories.

The prediction pipeline begins with multiple input data sources, including main device–accessory mapping which provides information about what accessory is compatible with each main device, historical sales data, pricing data and vendor lead time (VLT) data (time from when you place an order to when you receive the product) which serve as foundational inputs to the Accessory Attach Rate model. This model estimates the expected attach rate for each accessory based on historical relationships between accessory and main device sales. The attach rate quantifies the proportional relationship between an accessory's demand and its corresponding main device sales. Specifically, we define the attach rate as the ratio of accessory units sold to the attached main device units sold in a day.

The attach rate-based predictions are then passed to the **Residual Computation** module, which calculates residuals—discrepancies between predicted accessory sales (computed as the predicted attach rate multiplied by forecasted main device sales) and observed sales. These residuals account for variability in accessory demand that is not captured by the attach rate model.

To refine predictions, the computed residuals are incorporated into the **Residual Model**, which integrates additional input features, including historical sales, pricing, and main device forecasts. The final accessory predictions represent the optimized demand forecast for accessories, combining the structured attach rate approach with residual-based adjustments. This framework enhances both the interpretability of the demand model and its adaptability to fluctuations in sales patterns.

Our forecasting methodology follows a factor-based approach, where accessory demand is predicted at the *CCAD level*: Country/Channel/ASIN/Day level, where an ASIN corresponds to a unique product on Amazon, using the following equation:

$$x_i = y_m \times a_{i,m} + e_i \tag{1}$$

where for each day in our forecast horizon x_i represents the forecasted sales for accessory ASIN i, $y_m$is the forecasted sales volume for the main device m attached to ASIN i,$a_{i,m}$is the attach rate for accessory ASIN i relative to main device m, e_i is the error term capturing the deviation between the forecasted and actual sales. Next, we delve into each component of the framework, detailing its functionality and the specific role it plays in the prediction process. In the following sections, we provide a detailed breakdown of each component within the framework, highlighting its functionality and contribution to the overall prediction accuracy.

Accessory Attach Rate Model The Accessory Attach Rate Model (AR [1]) is central to the ARCA framework, designed to provide dynamic and accurate forecasts of accessory demand by predicting the attach rate for each accessory relative to its corresponding main device. The proposed approach uses a forecasting model that leverages a comprehensive set of features and recent data trends to predict attach rates with great precision.

The **target variable** for the AR model is the **daily attach rate** ($a_{i,m,t}$), which is defined as the ratio of accessory sales to attached main device sales on a given day. Specifically, the attach rate for accessory ASIN i, relative to the attached main device m at time t, is given by:

$$a_{i,m,t} = \frac{s_{i,t}}{s_{m,t}} \tag{2}$$

where $a_{i,m,t}$ is the attach rate for accessory ASIN i relative to main device m on day t, $s_{i,t}$ is the accessory sales for ASIN i on day t and $s_{m,t}$ is the sales of the main device m on day t. In cases where an accessory is associated with multiple main devices (e.g., headphones compatible with several tablet generations), the main device sales $s_{m,t}$ are calculated as the sum of sales across all compatible devices on day t.

The model aims to predict this attach rate for each accessory ASIN, using historical sales data, price information, and other relevant features. To account for the dynamic nature of accessory demand, the model utilizes several **predictor variables**, which can be represented as:

$$a_{i,m,t} = f(p_{i,t}, p_{m,t}, v_{i,t}, s_t, h_t, \text{ASIN}_i) \tag{3}$$

Where $p_{i,t}$ and $p_{m,t}$ are the price related features for the accessory and main device at time t, $v_{i,t}$ is the binary VLT feature indicating the availability to ship (value for the feature is set to 0) or not (value is set to 1) at the day of the purchase t, s_t represents seasonal indicators (e.g., month and day of the week features), h_t are holiday-related feature and ASIN_i are the dummy variable for the accessory ASIN.

[1]We use the acronym AR to refer to the attach rate model in our work. This is not to be confused with the acronym for autoregressive models which are common in time series bibliography.

To develop the AR model, we employ a Random Forest model, which is known for its ability to capture complex, nonlinear relationships within the data. The relationship between accessory demand and main device sales is rarely linear, and the attach rate can fluctuate based on various dynamic factors, therefore we selected this model as it is a good fit for our forecasting needs.

The model is trained using historical data from the past 365 days leading up to the forecast period. For each accessory i, the Random Forest model predicts the attach rate for each day t as:

$$\hat{a}_{i,m,t} = \text{RF}(p_{i,t}, p_{m,t}, v_{i,t}, s_t, h_t, \text{ASIN}_i) \tag{4}$$

Where $\hat{a}_{r,i,m,t}$ represents the predicted attach rate for accessory ASIN i relative to main device m on day t, and RF denotes the Random Forest model that takes into account the various input features.

The predicted attach rate is then used in the ARCA framework to estimate accessory sales by multiplying the predicted attach rate by the forecasted sales of the main device:

$$\hat{s}_{i,t} = \hat{a}_{i,m,t} \times \hat{s}_{m,t} \tag{5}$$

Where $\hat{s}_{i,t}$ is the predicted sales for accessory i on day t, $\hat{a}_{i,m,t}$ is the predicted attach rate for accessory i relative to main device m on day t and $\hat{s}_{m,t}$ is the forecasted sales of the main device m on day t.This dynamic approach to forecasting the attach rate ensures that the model can adapt to changing market conditions, including seasonal demand shifts, pricing changes, and promotional activities, improving the accuracy of accessory demand predictions.

3.2 Outlier Correction

To ensure that the model remains stable and accurate in the presence of erratic predictions we employ an Outlier Correction methodology. The purpose of this methodology is to identify and mitigate significant discrepancies between the predicted and historical observed attach rates for accessories. The approach first computes the average actual attach rate ($a_{i,m,t}^{\text{avg}}$) from historical data over a specified period, and then compares it with the predicted attach rate ($\hat{a}_{i,m,t}$) for each accessory. The prediction is flagged as an outlier if the percentage deviation between the predicted and average attach rate exceeds a predefined threshold τ, i.e.,

$$\left| \frac{\hat{a}_{i,m,t} - a_{i,m,t}^{\text{avg}}}{a_{i,m,t}^{\text{avg}}} \right| > \tau \tag{6}$$

where $\hat{a}_{i,m,t}$ is the predicted attach rate for accessory ASIN i relative to main device m at time t, $a_{i,m,t}^{\text{avg}}$ is the average actual attach rate for accessory i over historical data and τ is the threshold that defines the acceptable range of deviation between the predicted and average actual attach rate.

Once an outlier is detected, the flagged prediction is replaced with a fallback estimate, derived from the median actual attach rate ($a_{i,m,t}^{\text{med}}$) observed over the past 120 days, as follows:

$$\hat{a}_{i,m,t}^{\text{corr}} = a_{i,m,t}^{\text{med}} \tag{7}$$

where $a_{i,m,t}^{\text{med}}$ is the median actual attach rate for accessory ASIN i relative to main device m.

While the median offers a more robust estimate in the presence of anomalies, it lacks responsiveness to recent trends and seasonality captured by the model. Therefore, we only use it as a correction mechanism when the model's output deviates substantially from historical norms. This selective correction balances robustness with adaptability, ensuring more stable and reliable demand forecasts.

3.3 Residual Computation

The residuals in the AR model represent the discrepancy between actual accessory sales and the sales predicted using the attach rate model. These discrepancies arise from two primary sources: (a) the reliance on forecasted main device sales, which may contain inherent biases, and (b) the use of predicted attach rates, which, while generally robust, can exhibit minor fluctuations over time. Accurately modeling and adjusting for these residuals is essential to improving the overall forecasting performance of the model. Accurately modeling these residuals is crucial for improving the overall forecast accuracy. To achieve this, we explored multiple modeling approaches, including linear regression and Random Forest, to predict future residuals at various levels of granularity. The best results are obtained using a Random Forest model trained at the CC level.

To train the residual model, we first compute the residuals from historical data. These residuals quantify the discrepancy between observed accessory sales and the estimated sales derived from the predicted attach rate and forecasted main device sales. Specifically, for each accessory ASIN i, the residual e_i at a day level is computed as:

$$e_i = \text{actualSales}_i - \text{forecastedSales}_m \times \hat{a}_{i,m} \tag{8}$$

where e_i represents the residual for accessory ASIN i, actualSales_i denotes the observed sales of accessory i, forecastedSales_m denotes the observed sales of the corresponding main device m, $\hat{a}_{i,m}$ is the predicted attach rate for accessory i relative to main device m.

The residual forecasting model incorporates a comprehensive set of features that capture product attributes, pricing values, seasonal effects, and promotional activities to improve the accuracy of the final predictions. A key predictor is the *forecasted unit sales* for the main device, serving as the basis for estimating accessory demand. Pricing-related features include the *average selling price* (ASP) for both the accessory and main device. The *days on sale* feature tracks the number of days an accessory is available at a discount. The model accounts for major holidays such as Prime Day, Christmas, and Black Friday through binary indicators (*is_major_holiday*) and temporal counters for days before and after each holiday (*days_before_*_holiday, days_after_*_holiday*). Additionally, seasonal effects are captured through binary indicators for the day of the week and month. To refine predictions at the product level, the model includes *ASIN dummy variables* (*asin_X*) and *Program-dummy variables* (*program_Y*) to account for individual product effects. To account for *within-program cannibalization*, the model tracks the *average, minimum, and maximum discounts* for related ASINs within the same program, leveraging historical pricing data. By incorporating these diverse features, the model effectively captures key demand drivers, improving the accuracy and stability of accessory attach rate predictions.

Once the residuals are computed, they are integrated into the forecasting framework to refine sales predictions. The final accessory sales prediction is adjusted by incorporating the predicted residuals into the calculation. Specifically, the corrected forecasted sales for accessory i on day t is given by:

$$\hat{s}_{i,t}^{\text{final}} = (\hat{a}_{i,m,t} \times \hat{s}_{m,t}) + \hat{e}_{i,t} \tag{9}$$

where $\hat{s}_{i,t}^{\text{final}}$ represents the final predicted sales for accessory i on day t, $\hat{a}_{i,m,t}$ is the predicted attach rate for accessory i relative to the main device m, $\hat{s}_{m,t}$ is the forecasted sales of the main device m on day t and $\hat{e}_{i,t}$ is the predicted residual for accessory i, capturing systematic deviations not accounted for by the attach rate model or the systematic biases in the forecast for the main devices.

By incorporating residual forecasts into the demand estimation process, the model improves its robustness and accuracy, mitigating errors introduced by forecasted main device sales and attach rate fluctuations. This correction enhances the reliability and precision of accessory demand predictions.

3.4 Correlation Weighed Forecasting

We have observed that approximately 90% of accessory sales exhibit a strong correlation with main device sales (higher than 0.54 Pearson correlation coefficient [8]). However, around 10% of accessory demand does not follow a typical attach rate pattern, making it challenging to accurately forecast using a standard attach rate methodology. To address this, we developed a dynamic forecasting approach that adjusts predictions based on the correlation between accessory and main device sales. The main idea here is that the forecast for accessory sales (ASIN i) on a given day t is dynamically adjusted based on the correlation between the accessory sales and the main device sales. Specifically, we leverage an **attach rate model** (ARCA's prediction) with a higher weight when the accessory's sales are strongly correlated with the sales of the main device. When the correlation is weak, a forecasting model that directly forecasts for the accessory ASIN sales is used, providing a more accurate prediction. Following this methodology, the forecast for ASIN i on day t is computed as follows:

$$\text{forecast}_i = w_{\text{ARCA}} \times \text{forecast}_{\text{ARCA}} + w_{\text{other-model}} \times \text{forecast}_{\text{other-model}} \tag{10}$$

where the weight for the ARCA model, w_{ARCA}, is determined by the **normalized Pearson correlation** between accessory sales and main device sales, as $w_{\text{ARCA}} = \frac{\text{Pearson Correlation}+1}{2}$ and the weight for the alternative forecasting model, $w_{\text{other-model}}$, is given by $w_{\text{other-model}} = 1 - w_{\text{ARCA}}$.

The Pearson correlation reflects the strength of the relationship between the sales of the accessory and the main device, where a higher correlation results in a greater reliance on the ARCA-based forecast. For accessories with lower correlations, the weight shifts towards the alternative model, which directly forecasts the accessory's sales without relying on the attach rate. This methodology allows for flexible forecasting, adjusting dynamically to the correlation structure between products.

4 Experimental Results

In this section, we present our experimental results. For every experiment we conducted, we take an average of 12 past forecast versions: 12 trained models and their forecasts in the past, in order to draw conclusions. Specifically, we create one forecast version per month since January of 2024 until January 2025 for the following experiments unless otherwise noted.

4.1 Evaluation Metrics Description

To evaluate our results, we used the standard metrics of wMAPE and wBIAS which we have adapted so that we can use for multiple forecast versions and different time horizons (evaluation periods). wMAPE stands for weighted Mean Absolute Percentage Error and is defined as:

$wMAPE_V = \frac{\sum_A A_{A,V,i} * |PE_{A,V}|}{\sum_A A_{A,V}}$ where $PE_{A,V} = \frac{\sum_{i \in h} F_{A,V,i} - \sum_{i \in h} A_{A,V,i}}{\sum_{i \in h} A_{A,V,i}}$.

$F_{A,V,i}$ are the forecasted units for a given ASIN A on a given forecast version V on a given day i, $A_{A,V,i}$ are the actual units sold on that day for that ASIN and h is the evaluation horizon.

Furthermore, we also use the metric of Bias in order to refer to the persistent forecast error which is a component of the total calculated forecast error. We define Bias as: $wBIAS_V = \frac{\sum_A A_{A,V,i} * PE_{A,V}}{\sum_A A_{A,V}}$

The wMAPE metric provides insight into the accuracy of the forecasts, with lower values indicating better model performance. Meanwhile, wBIAS serves to evaluate the directional accuracy of the models, where values close to zero indicate minimal forecast bias, and negative or positive values reflect underestimation or overestimation tendencies, respectively.

4.2 Attach Rate Predictions

In this experiment, we evaluated wMAPE and wBIAS of different forecasting approaches for the attach rate across 12 historical forecast versions. Our analysis focuses on specific horizon and lead time combinations—horizons of 5 and 15 days, and lead times of 0, 5, and 10 days—which are the most relevant for the allocation use case of our model. The reported results, presented in Table 1, focus on forecasts for the US ONLINE market (products sold on amazon.com in the US), which is the country/channel with the highest historical accessory sales. We compare multiple approaches for predicting future attach rates. The Median AR model, used as a baseline, calculates the median attach rate over the past 60 days for each accessory ASIN and applies this value as the forecasted attach rate across the entire horizon. This method results in a high AVG wMAPE of 70.33% and a substantial bias of 53.74%, indicating a consistent tendency to overestimate demand.

Model	AVG wMAPE	AVG wBIAS
1: Median AR	70.33	53.74
2: Model-based AR - CC - Week	69.35	8.94
3: Model-based AR - CC - Day	64.79	0.36
4: Model-based AR - Program - Week	69.13	11.86
5: Model-based AR - Program - Day	**58.94**	**1.74**
6: Model-based AR - Device - Week	69.57	1.09
7: Model-based AR - Device - Day	62.88	-31.26

Table 1: Performance comparison of different AR models.

We also present results from several model-based approaches. For these experiments, we employ a Random Forest model [3] as the ML approach, as it is readily available in our testing platform and effectively captures nonlinear patterns—unlike linear models, which performed worse than the baseline due to their inability to capture attach rate trends. We present results for attach rate models built at different levels: a) the Country/Channel level (CC): a single model trained on all ASINs at the country/channel level, b) the Program level: a model trained separately for each accessory program (e.g., all headphone ASINs connecting to a specific tablet) and c) the Device level: A model trained for each device type (e.g., eReader, SMP, Alexa etc.), encompassing multiple accessory programs. For each modeling approach, we test building the model using daily vs. weekly data aggregation, analyzing whether forecasting at the day level yields better results than week-level aggregation.

Our results indicate that model-based approaches work better than a static attach rate (approach 1). Among the model-based approaches, approach 5: Model-based AR - Program - Day performs best in terms of accuracy, achieving the lowest AVG wMAPE (58.94% wMAPE) and a low bias (1.74%), making it the most balanced model. In contrast, approach 2: Model-based AR - CC - Week performs the worst, with the highest wMAPE (69.35%) and a high bias (8.94%), demonstrating poor forecast accuracy. Notably, one variation of Model-based AR - Program - Day exhibits a negative bias (-31.26%), indicating a tendency to underpredict demand. Overall, daily-level models outperform their weekly counterparts, indicating that using more data for training the model (at the day level) helps learn patterns more accurately. Furthermore, building the model at the program level achieves the best accuracy indicating that accessory programs have unique attach rate behaviors, and a model trained specifically on a program's data can better learn its patterns compared to a model that generalizes across all programs (either on the country/channel level or the device level).

4.3 Residual Model Performance

Next, we compare the performance of the pure attach rate model with that of the attach rate model augmented by the residual model. For this experiment, we utilize the best-performing attach rate model approach from the previous experiment, specifically Approach 5: Model-based AR - Program - Day. We present the results of the wMAPE across 12 forecast versions for the US ONLINE channel.

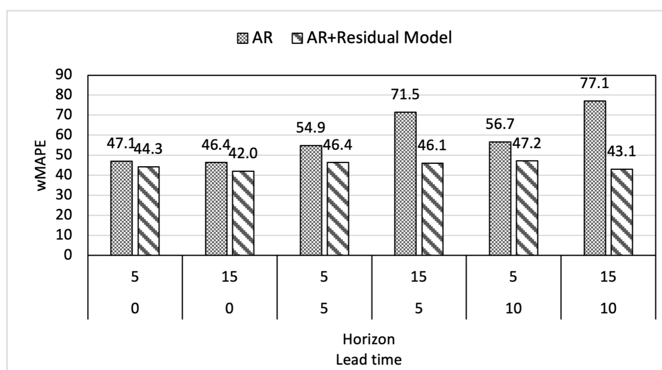

Figure 2: Impact of the Residual Model on wMAPE.

The comparison between the AR model and the AR model with residuals (AR+Residual) in Figure 2, reveals that the latter consistently outperforms the former across different lead times and

Lead Time	Horizon	wMAPE (ARCA)	wBIAS (ARCA)	wMAPE (Correlation Weighted ARCA)	wBIAS (Correlation Weighted ARCA)
0	5	44.3	-23.2	42.3	-18.4
0	15	42.0	-20.3	40.0	-21.3
5	5	46.4	-22.3	42.3	-12.5
5	15	46.1	-19.6	43.0	-21.6
10	5	47.2	-22.4	46.1	-15.6
10	15	43.1	-20.2	42.1	-23.3

Table 2: Impact of the Correlation Weighted Scheme on ARCA's Performance.

forecasting horizons, as measured by the weighted Mean Absolute Percentage Error (wMAPE). The AR+Residual model shows improvements in forecast accuracy, ranging from 2.8% to 34.0%, depending on the lead time and horizon. These results highlight the effectiveness of incorporating residuals in refining predictions, particularly for longer-term forecasts, where the AR model tends to exhibit larger errors. Overall, the AR+Residual model demonstrates superior performance, with an average wMAPE of 44.9% compared to the AR model's 58.9,% underscoring its ability to capture deviations and provide more reliable forecasts.

4.4 Correlation Weighted Forecasting

Next, we present the results of the Correlation Weighted ARCA and compare it with the full ARCA performance (AR + Residual Model) presented above. As our "other-model" when directly forecasting for accessories, we used an XGBoost model [5] utilizing the same feature set as described in section 3.3. Note that we do not use this model directly for forecasting accessories, as it does not align with the attach rate–based paradigm that planners use to evaluate forecasts.

The results, presented in Table 2, reveal that, for a lead time of 0, the Correlation Weighted ARCA model outperforms the standard ARCA model in terms of both wMAPE and wBIAS across all horizon periods. Specifically, at a 5-period horizon, the wMAPE for ARCA is 44.3%, whereas for Correlation Weighted ARCA, it is 42.3%. Similarly, at the 15-period horizon, ARCA's wMAPE is 42.0%, compared to 40.0% for the Correlation Weighted ARCA. This demonstrates the efficacy of incorporating correlation weighting, as it leads to a reduction in forecast error. However, despite the improvements in accuracy, both models exhibit notable bias, with ARCA showing consistently negative wBIAS values, indicating a tendency to underestimate the forecast. The Correlation Weighted ARCA model also exhibits a similar bias but to a lesser extent, suggesting that the correlation weighting improves the calibration of the forecast by reducing the overall underestimation observed in the ARCA model.

5 Conclusion

In this paper, we presented a novel machine learning-based framework for forecasting accessory demand using attach rates, addressing the unique challenges of dependency on main device forecasts. By integrating a dynamic attach rate model with a residual correction mechanism and a correlation-weighted forecasting approach, we improved forecasting accuracy and adoption of the forecasts by product planners. Our approach offers greater interpretability and alignment with how product managers and planners approach accessory demand, making the forecasts more actionable. The model, deployed in production at Amazon in 2024, demonstrated significant improvements in forecast precision, benefiting inventory management, promotion planning, and operational strategies. Our performance evaluations indicate that the model achieves an average wMAPE of 42% in our largest selling country/channel based on historical sales data. This framework represents an effective solution for forecasting accessory demand and can be extended to other product categories with similar attach rate dependencies.

References

[1] N. S. Arunraj, D. Ahrens, and M. Fernandes. Application of sarimax model to forecast daily sales in food retail industry. *International Journal of Operations Research and Information Systems*, pages 1–21, 2016.
[2] M. Arvan, B. Fahimnia, M. Reisi Ardali, and E. Siemsen. Integrating human judgement into quantitative forecasting methods: A review. *Omega*, 86:237–252, 2019.
[3] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
[4] Eduardo Cadena, Kevin Albright, Hui Wang, and Sanjay Ajmera. Forecasting accessory demand in the automotive industry. *SMU Data Science Review*, 7(2), 2023.
[5] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.
[6] Rob J. Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, 2018.
[7] Bryan Lim, Sercan Ömer Arık, Niels Loeff, and Stefan Zohren. Time series forecasting with deep learning: A survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194):20200209, 2021.
[8] Karl Pearson. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58(347-352):240–242, 1895.
[9] Agneta Ramosaj, Nicolas Ramosaj, and Marino Widmer. How to improve accessories sales forecasting of a medium-sized swiss enterprise? a comparison between statistical methods and machine learning algorithms. *Journal of Applied Business and Economics*, 26(4):153–172, 2024.
[10] Sanjay S. Rangapuram, Matthias W. Seeger, Jan Gasthaus, Lukas Stella, Yingzhen Wang, and Tim Januschowski. Deep state space models for time series forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 31, pages 7796–7805, 2018.
[11] David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.