Received: 22 August 2023 | Revised: 19 September 2023 | Accepted: 11 October 2023 | Published online: 11 October 2023

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2024, Vol. 2(1) 150–164
DOI: 10.47852/bonviewJDSIS32021569

BON VIEW PUBLISHING

Evaluating Economic Impacts of Automation Using Big Data Approaches

Omid M. Ardakani[1,*] and Mariana Saenz[1]

[1]*Parker College of Business, Georgia Southern University, USA*

Abstract: As automation is increasingly driven by advanced technological integration, quantitatively evaluating its economic impacts becomes crucial. This paper studies the effects of automation on three economic outcomes: transactions, sales, and costs. First, we use big data approaches to distinguish transaction distribution patterns across various temporal segments. These methods employ survival and mean residual functions to cluster transaction distributions and customer traffic data over time. Empirical evidence provides distinct clusters, distinguishing high and low customer traffic. Second, we illustrate how automation can lead to higher forecast accuracy in sales. This approach utilizes stochastic error distance for comparing forecast error distribution functions. Lastly, we study the impact of automation on costs through a probabilistic model. The results suggest that while labor costs increase due to retraining and longer hours, a potential reduction in turnover and waste costs can offset these rises. The impacts of automation and the applicability of methods are demonstrated through Monte Carlo simulations and empirical studies.

Keywords: automation, economic outcomes, forecast accuracy, sampling distribution, stochastic ordering

1. Introduction

Automation has increasingly been recognized as a pivotal driver of economic and industrial transformation. While industries have been integrating automated processes to increase efficiency, reduce costs, and enhance accuracy, the magnitude and scope of its impacts vary (Gunasekaran, 2009; Moore, 2012). This trend is evidenced by the foodservice and food retailing sectors, critical components of the global economy, which have recently seen a surge in automation adoption. According to data from the US Department of Agriculture, these industries supplied approximately $1.69 trillion worth of food in 2020. Driven by a combination of substantial market size, labor force disruptions from unforeseen events such as pandemics, and the promise of technology, the push toward automated food production has become more pronounced. Notably, this automation offers operational efficiency and is regarded as a crucial step toward achieving greater food security (Huang, 2013). Despite their significance, comprehensive studies examining the economic impacts of automation, including sales and costs, are lacking in the literature.

The lack of data with the same frequency and insufficient production data requires employing probabilistic approaches and simulation methods to study the impacts of automation. This paper employs the empirical mean residual function (MRF) introduced by Ardakani et al. (2020) to distinguish customer demand during different times of the day and studies the impacts of automation on customer traffic, sales, and cost. The mean residual (MR) life, defined as the expected additional lifetime given that a component has survived, can be expressed in terms of the failure rate (Gupta & Bradley, 2003) and can be used as a big data tool for distinguishing distributions. The effect on the customer traffic, defined as transactions, is examined by simulating from theoretical distributions best fitted empirical transaction distribution. The simulation exercise consists of estimating the probability distribution of the transaction data and then simulating automation scenarios from known probability densities. The impact of automation on profit is not straightforward since automation affects sales and the total cost differently. An increase in sales and a decline in costs lead to a rise in profit, but if sales and costs both increase, the change in profit depends on the magnitudes of changes in sales and costs. Hence, the two components can be analyzed separately.

This paper identifies the most suitable forecasting model for sales data to assess the impact of automation on sales forecast accuracy. While the autoregressive integrated moving average (ARIMA) is known for its capability to handle non-stationarity in data (Papoulis & Pillai, 2002), various other statistical and machine learning models can also be adopted. The key is ensuring the chosen model provides independent and identically distributed forecast errors. After determining the most suitable model, we extract the forecast error, which is the difference between the actual data and the forecasts. This error provides insight into the model's forecast accuracy. The implications of automation are subsequently examined by comparing the cumulative distribution function (CDF) of the forecast errors from the most suitable model against simulated forecast errors considering automation. The unit step function at zero serves as a reference in this comparison, grounded on the concept of stochastic error distance (SED) (Ardakani et al., 2018; Diebold & Shin, 2017).

This paper also studies the impacts of automation on costs. We consider both labor costs and waste-related expenses. Various automation scenarios are proposed for labor costs based on specified confidence levels. Regarding waste costs, we employ a probabilistic methodology analogous to our approach for customer

*Corresponding author: Omid M. Ardakani, Parker College of Business, Georgia Southern University, USA. Email: oardakani@georgiasouthern.edu

traffic: determining the theoretical distribution from the available data and simulating the potential impacts of automation based on this distribution. The data for the empirical analysis are obtained from a regional foodservice provider. All inside transactions are considered as a proxy for customer traffic. Sales and waste cost data for four items are included in the empirical analysis. The findings suggest that automation increases customer traffic and improves sales forecasting accuracy. Improving forecast accuracy allows practitioners to identify sales factors proactively and highlight potential opportunities to better target demand. The results also indicate that automation may lead to higher labor costs because of retraining costs and longer work hours to meet the higher demand. It is also possible that automation reduces labor costs due to reducing the turnover ratio. Also, automation results in a waste decline for low-waste days and a waste increase for high-waste days. Ideally, we expect higher efficiency and waste reduction to offset higher installation and operating costs.

This paper is organized as follows. Section 2 employs a big data tool to distinguish transaction distribution and proposes a probabilistic method to capture the automation impacts on customer traffic. Section 3 studies the impact of automation on sales by measuring sales forecast accuracy. In this section, we discuss the application of SED for comparing forecast error distribution functions. Section 4 evaluates the impact of automation on costs by considering labor costs and the cost of waste. Section 5 discusses Bayesian methods as a potential alternative and elaborates on their advantages and limitations. Section 6 gives some concluding remarks.

2. Automation and Customer Traffic

Customer traffic, quantified by the number of transactions, directly influences sales and revenues. Understanding this traffic is crucial for automation, as it aids in aligning with customer demands for product freshness and minimizing wait times during peak and off-peak hours. We distinguish these peak and off-peak periods by examining transaction distributions on an hourly basis. The random variable transaction T is defined as the nonnegative, continuous random variable. The distribution of T can be represented by the CDF $F(t) = P(T \le t)$, survival function (SF) $\bar{F}(t) = P(T > t)$, and the probability density function (PDF) $f(t) = dF(t)/dt$. The CDF provides information about the proportion of transactions completed by time t, while the SF offers the likelihood that a transaction time exceeds t.

The estimates for the three representations are as follows. The CDF can be estimated by empirical CDF

$$\hat{F}(T) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(T_i \le t), \tag{1}$$

where $\mathbb{I}(\cdot)$ is the indicator function, and $T_1, \ldots, T_n$ are random samples from the CDF of T. The SF can be estimated by the empirical SF

$$\bar{F}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(T_i > t). \tag{2}$$

A known estimate of the SF is the Kaplan–Meier estimator (Kaplan & Meier, 1958) given by

$$\bar{F}(t) = \prod_{t_i \le t}\left(1 - \frac{d_i}{n_i}\right), \tag{3}$$

where d_i is the number of failed transactions that happened at time t_i, and n_i is the number of successful transactions up to time t_i. Both the empirical CDF and SF are determined by the data. This makes them unbiased in the sense that they are not influenced by any prior assumptions about the distribution of the data. The Kaplan–Meier estimator is a powerful tool in survival analysis. It provides a way to estimate the SF in the presence of censored data (Kleinbaum & Klein, 1996). While the CDF gives the probability that a random variable is less than or equal to a particular value, the SF provides the probability that it exceeds that value. This dual perspective can be essential in many applications. The PDF can be estimated using the nonparametric kernel density estimation. Suppose we observe independent and identically distributed transactions $\{T_j\}_{j=1}^{n}$ with unknown density f. The kernel density estimator can be written as

$$\hat{f}_h(t) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{T_i - t}{h}\right), \tag{4}$$

where h is the bandwidth and controls the degree of smoothing and $K(\cdot)$ is a Gaussian kernel function. The Gaussian kernel function is a nonnegative function that integrates into one. We can calculate the numerical moments of $\hat{f}_h(t)$. The first moment (mean) of the estimated density, $\int_{\infty}^{-\infty} t\hat{f}(t)dt$, is the sample mean, and its second moment, $\int_{\infty}^{-\infty} t^2\hat{f}(t)dt$, is the variance of the estimated density.

Although the CDF, SF, and PDF illustrate the distribution of transactions, these representations have restrictions that limit their ability to distinguish the distribution of T between busy and nonbusy times. An alternative to distinguishing the distribution of T is the MRF (Sun & Zhang, 2009). The MRF is a nonnegative function that provides a more pronounced ordering. The visualization technique for distinguishing distributions following the MRF is called the MR plot introduced by Ardakani et al. (2020) and further discussed in Ardakani et al. (2022). Let the residual of T with distribution function $F(t)$, given that it exceeds a threshold τ, be defined by $T - \tau | T > \tau$. The PDF of the residual of T is given by

$$f(t|\tau) = \frac{f(t)}{\bar{F}(\tau)}, \quad t > \tau, \ \ \bar{F}(\tau) > 0. \tag{5}$$

This equation gives the distribution of transactions that exceed a certain threshold τ. The MRF or mean excess of T with a finite mean is defined as

$$m(\tau) = E_{T>\tau}(T - \tau | T > \tau), \quad \tau \ge 0,$$

where $E_{T>\tau}$ denotes the expectation with respect to the residual PDF introduced in Equation (5). The estimate of the MRF of T defined in Equation (6) is based on the following representation of the MRF:

$$m(\tau) = \frac{1}{\bar{F}(\tau)}\int_{\tau}^{\infty} \bar{F}(t)dt, \tag{6}$$

where $\bar{F}(t)$ can be estimated using the empirical SF defined in Equation 2. The MRF represents the expected amount by which a transaction exceeds a threshold τ. Details are provided in Ardakani et al. (2020).

By plotting the MRF of T, we obtain the MR plots that can help distinguish busy times with respect to the transactions. The transaction data are fundamentally distinct in nature and frequency. We need a method that captures the dynamics of transactions and customer traffic. This involves considering the distribution of transactions over time. Employing MR and SFs allows a nonparametric approach without making unnecessary assumptions. We further illustrate the advantages of MR plots in distinguishing transaction distributions using the following examples. Assume T is normally distributed with PDF

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}, \tag{7}$$

where μ and σ are mean and standard deviation. Figure 1 presents the PDF and MR plots of T for five normal distributions with $\mu = 0$ and $\sigma = 1.4, 1.6, 1.8, 2.0, 2.2$. The values for μ and σ are chosen to illustrate a range of variations around a central mean of zero. Choosing $\mu = 0$ is a conventional decision when presenting the normal distribution (Casella & Berger, 2002). Values of $\sigma = 1.4, 1.6, 1.8, 2.0, 2.2$ are selected to exhibit a systematic spread and its impact on the distribution. This range allows us to observe how the MR plots differentiate distributions even when the PDFs appear very similar. As shown in Figure 1, the PDF plots cannot distinguish or order distributions. The MR plots of T, however, show a stochastic ordering. The MRF is decreasing in τ and increasing in σ.

Similarly, we can assume T follows a Gamma distribution with PDF

$$f(t) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}, \tag{8}$$

where $\Gamma(\cdot)$ is the gamma function, and α and β are the shape and scale parameters. Figure 2 presents the PDF and MR plots for five gamma distributions with $\alpha = 0.5$ and $\beta = 0.8, 0.9, 1.0, 1.1, 1.2$. The shape and scale parameters are selected based on similar reasoning as the normal distribution. Choosing $\alpha = 0.5$ illustrates a scenario where the distribution is right-skewed, common in various applications, such as queueing theory and financial econometrics (Johnson et al., 1995). The values of $\beta = 0.8, 0.9, 1.0, 1.1, 1.2$ display a progression, offering an explicit comparison and interpretation of differences among the MR plots. The MR plots clearly order each gamma distribution.

The data for the empirical analysis are obtained from a regional foodservice provider as a proxy of the foodservice industry. The transaction data span from June 1, 2020, to July 26, 2021, and consist of 20,204 observations. This dataset is sourced from multiple outlets across the US Southeast region, provided by the respective organization. The initial dataset is processed to eliminate incomplete entries. Furthermore, all data have been anonymized and do not contain any personally identifiable information. This dataset offers insights into regional food consumption patterns, assisting the analysis where the number of transactions serves as a proxy for customer traffic. We first confirm that transactions are i.i.d. using the Durbin–Watson test. The data include all inside transactions. Figure 3 plots the transaction distributions per hour using the MR plots. The MR plots reveal three distinct clusters:

- **Morning Cluster (5 am–1 pm):** This cluster illustrates the peak hours for transactions, indicating high customer traffic during the traditional morning to early afternoon hours.
- **Afternoon Cluster (2 pm–5 pm and 4 am):** Interestingly, this cluster captures the 4 am slot, which groups more cohesively with the afternoon transactions than the early morning. One potential explanation for this result is the nature of customers who visit foodservice establishments at 4 am. This timeframe might capture early morning workers, travelers, or night shift employees who have patterns of purchasing that align more with the afternoon crowd than the conventional morning customer traffic.
- **Evening Cluster (6 pm–3 am):** This cluster captures the evening to early morning transactions, generally lower than the morning cluster but still significant.

The MR plot's strength lies in highlighting such non-obvious patterns. While traditional methods may focus on the chronological progression of time, MR plots can reveal customer transaction behaviors that are temporally displaced but statistically similar. The MR plots are linear and suggest that transactions peak during the morning cluster, followed by the afternoon and evening clusters.

Figure 1
PDF and MR plots of normally distributed transactions with $\mu = 0$ and $\sigma = 1.4, 1.6, 1.8, 2.0, 2.2$

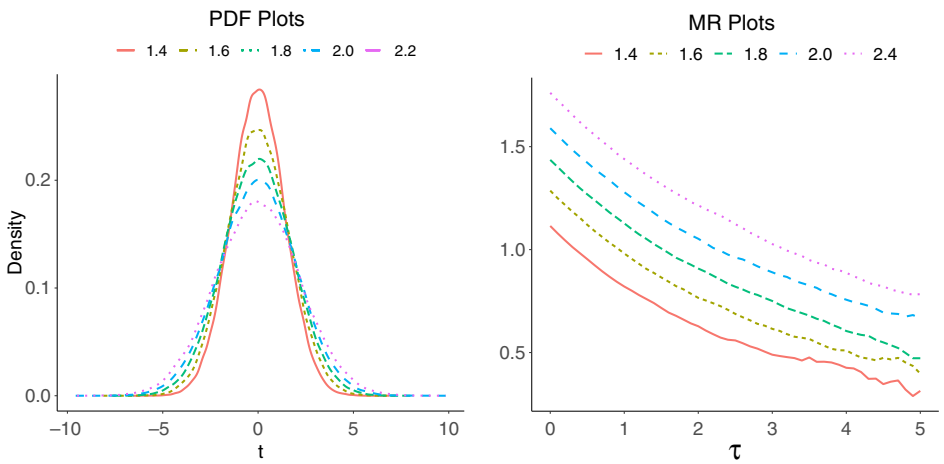


Figure 2
PDF and MR plots of gamma distributions with $\alpha = 0.5$ and $\beta = 0.8,\ 0.9,\ 1,\ 1.1,\ 1.2$

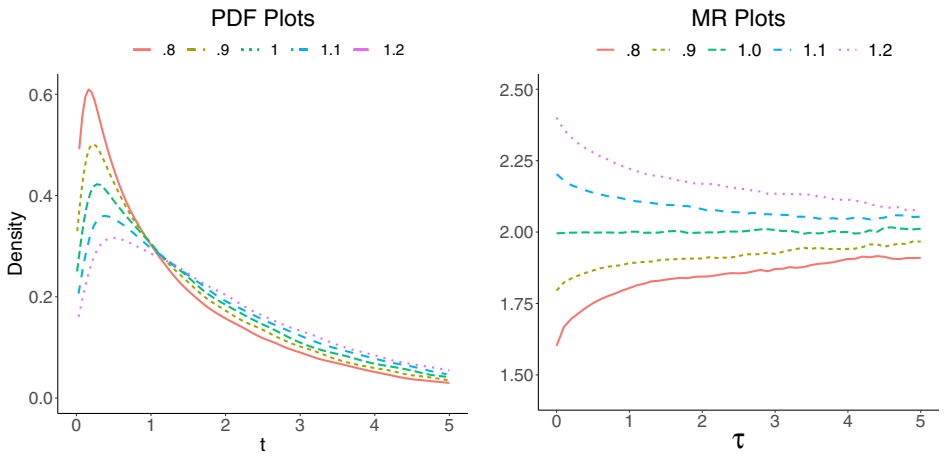
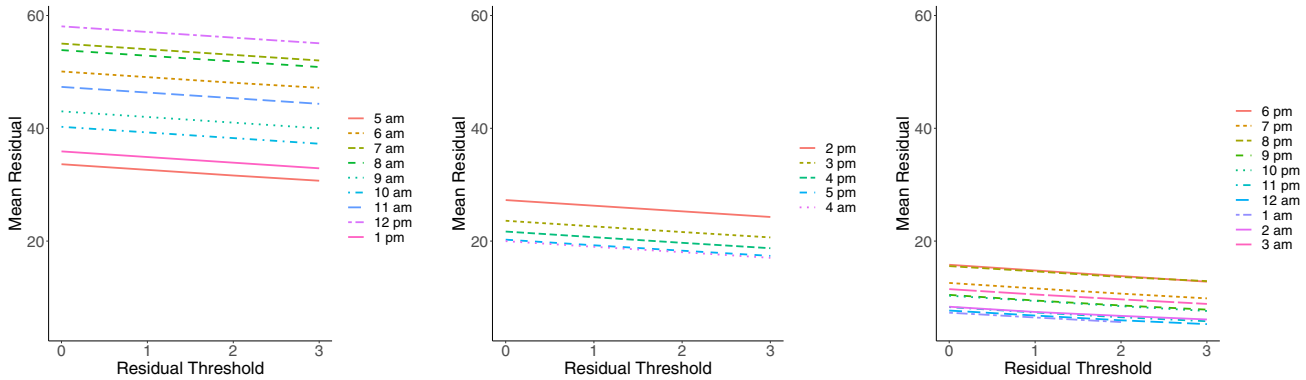


Figure 3
MR plots of transactions distinguished by hour



Automation would affect customer traffic, but measuring its impact before adopting automation strategies can be challenging. We employ a simulation procedure for examining the effects of automation on customer traffic. The automation effect can be captured by estimating the PDF of the continuous random variable transaction T and simulating from the corresponding PDF.

Specifically, the following steps are taken in estimating the PDF. We first estimate the statistical moments to identify potential parametric distributions of the data. This step is followed by examining the PDF and CDF of the transaction data. Next, we use skewness-kurtosis plots to visualize the choice of potential fitted parametric distributions. We then apply the maximum likelihood method to the estimated parametric distribution and assess goodness-of-fit measures for the chosen distribution to select the distribution that best fits the data. We finally use Monte Carlo simulations to analyze the impact of automation.

Although we have primarily utilized frequentist methods, the Bayesian approach offers a compelling alternative that can be particularly beneficial under certain conditions. Unlike the frequentist viewpoint, where parameters are perceived as fixed quantities, the Bayesian paradigm treats these parameters as random variables. In this study, the identified empirical distribution can be equated to a Bayesian prior in some scenarios. This approach offers the advantage of leveraging prior knowledge, which can be valuable when data are limited or prior are particularly insightful. Despite these advantages, Bayesian methods can be computationally demanding. Choosing between the frequentist and Bayesian approaches often depends on objectives, data characteristics, and computational constraints. Refer to Section 5 for a brief discussion on Bayesian methods.

The empirical densities and CDFs of transaction data can be seen in Figure 4. In this example, the transaction density is skewed to the right, and the probability values are concentrated around a low number of transactions. Visualizing PDF and CDF of the transaction data is the first step to identifying the parametric distribution needed in simulations.

To identify a suitable distribution, we compare well-known theoretical distributions to the empirical transaction distribution in the higher-order (skewness-squared, kurtosis) space. Skewness-kurtosis plots proposed by Cullen and Frey (1999) are presented in Figure 5. The theoretical distributions for the normal, logistic, and exponential distributions are represented by a single point in the skewness-kurtosis plot. The theoretical gamma and lognormal distributions are represented by a line, while the theoretical beta distribution is represented by a shared area in the skewness-kurtosis plot. We perform a bootstrap procedure for robustness to

Figure 4
PDF and CDF plots of transactions

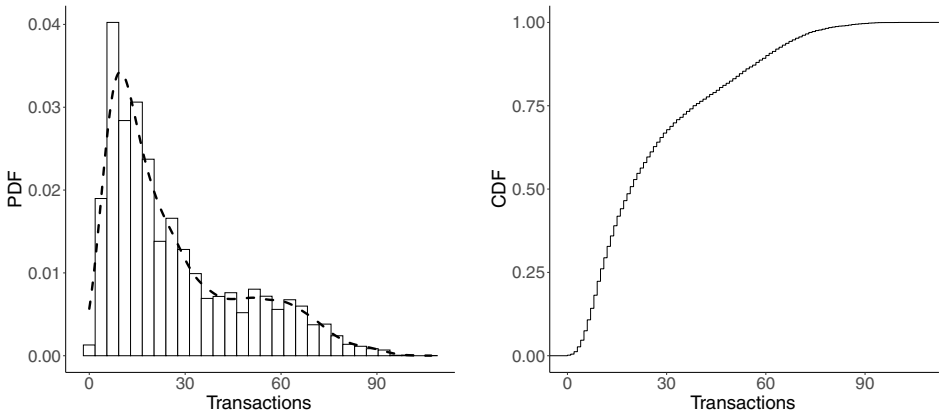
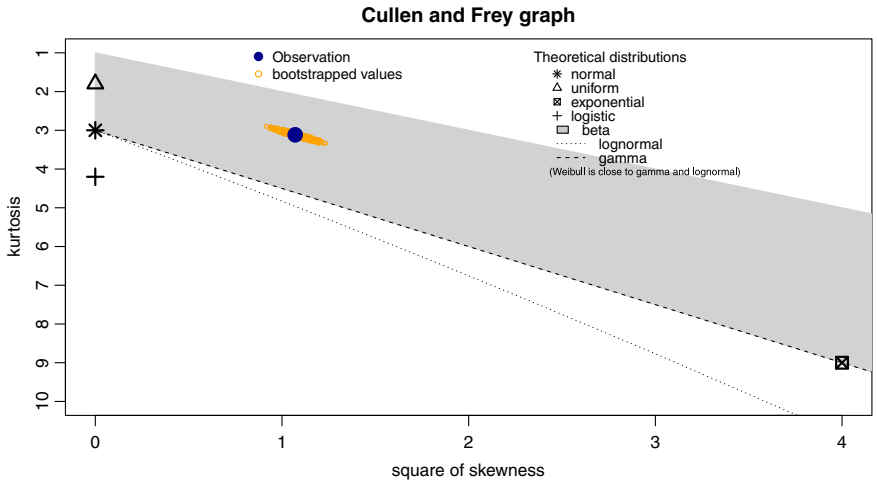


Figure 5
Skewness-kurtosis plot of transactions



compute skewness and kurtosis on samples constructed by 5,000 random samples with replacement from the original dataset, represented by yellow points. While the skewness-kurtosis plot is a preliminary diagnostic tool, the subsequent decisions are driven by a combination of fit, practical significance, and computational feasibility.

We estimate the distribution parameters using the maximum likelihood estimation (MLE) method presented in Delignette-Muller and Dutang (2015). Let $f(t|\theta)$ be the density function of a parametric distribution, where θ are the distribution parameters. Distribution parameters θ can be estimated by maximizing the likelihood function $\mathcal{L}(\theta|t) = \prod_{i=1}^{n} f(t_i|\theta)$ with n observation. The maximum likelihood estimator is then defined as

$$\hat{\theta}(t) = \underset{\theta}{argmax}\ \mathcal{L}(\theta|t). \tag{9}$$

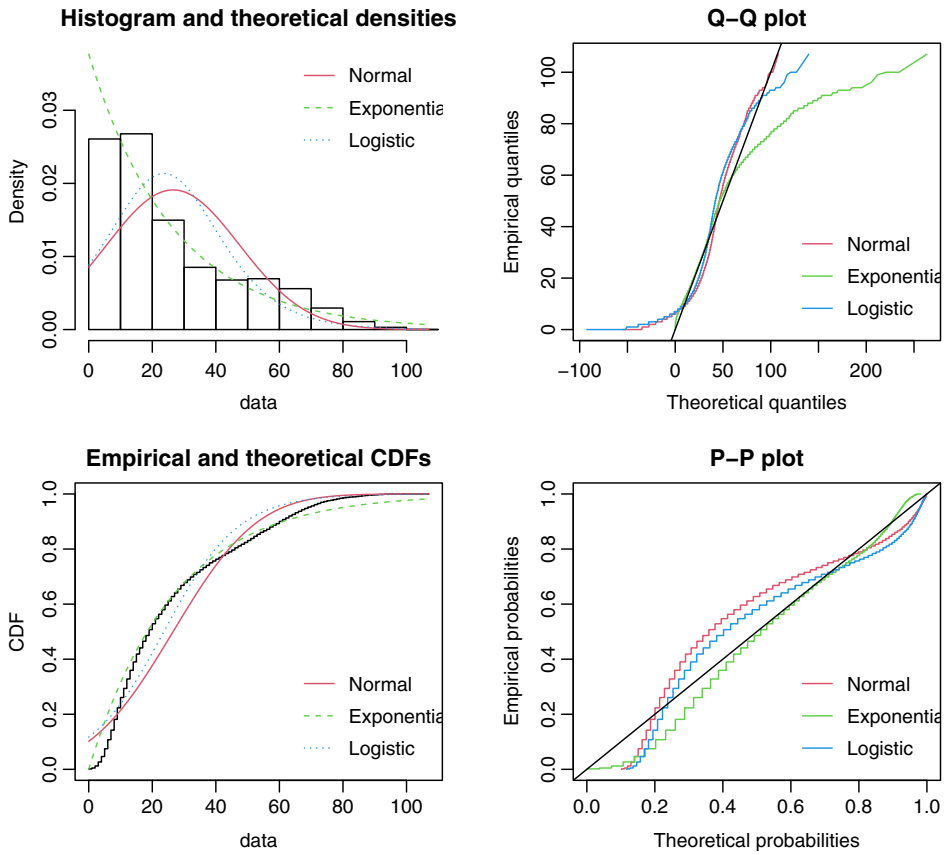
Table 1 gives the parameter estimates and estimated standard errors for the three suitable distributions: normal, exponential, and logistic. The standard errors are reported in parentheses and computed from the Hessian matrix estimate at the maximum likelihood solution. The normal distribution's mean and standard deviation of transactions are estimated to be 26.54 and 20.87. The relatively small standard errors for both parameters, 0.206 and 0.146, indicate these estimates are statistically significant. The exponential distribution is characterized by its rate parameter. The rate parameter captures the speed of occurrence of events. Here,

Table 1
Parameter estimates of the fitted distributions

Normal		Exponential	Logistic	
Mean	SD	Rate	Location	Scale
26.54	20.87	0.037	23.64	11.70
(0.206)	(0.146)	(0.001)	(0.202)	(0.097)

SD denotes the standard deviation. Standard errors are reported in parentheses and are computed from the estimate of the Hessian matrix at the maximum likelihood solution.

Figure 6
Histogram, theoretical densities, empirical CDF, Q-Q, and P-P plots of the fitted distributions



the rate is estimated at 0.037. The small standard error of 0.001 indicates its statistical significance. For the logistic distribution, the location parameter is estimated to be 23.64. It indicates the point where the CDF is 0.5. The scale parameter, measuring the spread of the distribution, is 11.70. Again, the standard errors are relatively small, emphasizing the precision of these estimates.

Once the parameters for the potential empirical distributions are estimated, we decide which fitted distribution is best by considering the histogram and theoretical densities, empirical and theoretical CDFs, and Q-Q and P-P plots. The Q-Q plot represents theoretical quantiles against empirical ones, and the P-P plot represents theoretical probabilities against empirical ones. Figure 6 provides these four plots for the distribution of the transactions fitted to the data. The results suggest that the logistic distribution is preferred for its better description of the empirical distribution. The findings suggest that logistic distribution with location parameter μ and scale parameter s, $\mathbf{L}(\mu, s)$, as the benchmark. $\mathbf{L}(23.64, 11.70)$ represents the transaction density in our example.

Logistic density is a well-known distribution widely used in neural networks and machine learning. Its PDF is symmetric, and its CDF is known as the logistic function. The PDF and CDF of logistic distribution are given by

$$f_T(t) = \frac{e^{-\frac{t-\mu}{s}}}{s(1+e^{-\frac{t-\mu}{s}})^2} \tag{10}$$

and

$$F_T(t) = \frac{1}{1+e^{-\frac{t-\mu}{s}}}. \tag{11}$$

The mean and variance of the logistic density are defined as $\mathbb{E}(T) = \mu$ and $V(T) = s^2\pi^2/3$. For the standard logistic distribution with $\mu = 0$ and $s = 1$, $\mathbb{E}(T) = 0$ and $V(T) = \pi^2/3$. The main property of the logistic distribution is that its PDF can be written in terms of the CDF as $f_T(t) = 1/sF_T(t)(1 - F_T(t))$.

Four scenarios are considered to examine the impact of automation on transactions. Table 2 presents the probability distributions for four different automation scenarios using the logistic distribution. Column one shows the estimated distribution. Columns 2–5 show the different automation scenarios. Scenario 1 represents a lower average transaction and lower transaction variability due to automation. Scenario 2 presents an automation outcome with greater average transaction and lower transaction variability. The reason for lower variability under Scenarios 1 and 2

Table 2
Probability distributions for the simulation study

Estimated	Scenario 1	Scenario 2	Scenario 3	Scenario 4
L(23.64,11.70)	**L**(20,8)	**L**(28,8)	**L**(20,17)	**L**(28,17)

$\mathbf{L}(\alpha, \beta)$ represents logistic distribution with location and scale parameters α and β. For each simulation $n = 10,000$.

is that automation reduces customer wait time, allowing greater customer traffic. Under Scenario 3, automation results in lower average transactions and greater transaction variability. Scenario 4 represents greater average transactions and greater transaction variability due to automation. The reason for higher variability under Scenarios 3 and 4 is that automation allows fresh products during different times of the day. The parameters are estimated using the MLE method. The parameter values for our four different automation scenarios are shown in Table 2.

The automation in Scenario 1 leads to a lower average transaction (reflected in a reduced location parameter of 20) and diminished transaction variability (reflected in a reduced scale parameter of 8). The cause of such changes is automation's ability to decrease customer wait times, leading to more efficient transaction flows. Under Scenario 2, automation increases the average number of transactions, denoted by a location parameter 28. Nevertheless, similar to Scenario 1, transaction variability remains low with a scale parameter of 8. This can be linked to the efficiency gains from automation that reduce waiting times and draw higher customer traffic. For Scenario 3, automation leads to a lower average transaction (location parameter of 20) and higher variability (scale parameter of 17). This increased variability is due to automation's capability to offer fresh products at varied times of the day. In Scenario 4, automation results in a higher average number of transactions (location parameter of 28) and greater transaction variability (scale parameter of 17).

The reason for selecting these specific parameter values across the scenarios is anchored in business-related views regarding how automation might shape transaction averages and variability. The variations across location and scale parameters across these scenarios capture potential real-world consequences of automation, from enhancing operational efficiency (lower wait times) to changing product availability dynamics (freshness throughout the day). To ensure the robustness of our chosen distributions and parameters, we further conducted Monte Carlo simulations. As evidenced in Figure 7, these scenarios provide insights into potential transaction automation outcomes.

Figure 7 presents the Monte Carlo simulations for various automation scenarios. The solid black line represents the estimated empirical distribution, while the four automation scenarios are shown in color. Under Scenarios 1 and 2, automation leads to an increase in the number of transactions when the transaction is lowest. Scenarios 3 and 4 indicate higher probabilities in tails, indicating that automation leads to an increase in the number of transactions when the transaction is highest. While our analysis reveals distinct clusters, our modeling approach takes a holistic view of the entire day, driven by policy relevance. Policymakers and industry leaders often implement strategies encompassing an operational day rather than specific timeframes. We model an entire day's transactions as a unified entity for simulations. This approach offers an overview of transaction patterns.

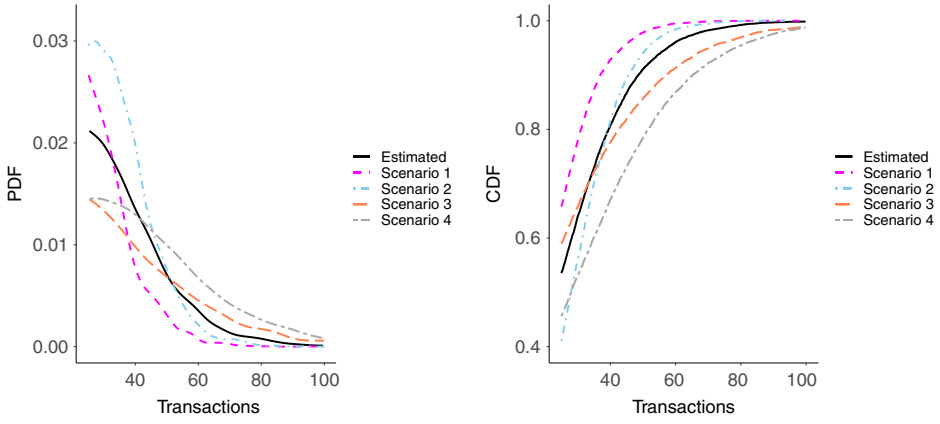
This section examines the importance of automation in minimizing customer waiting time. The distribution of these transactions is divided into peak and off-peak times based on hourly distributions. Transactions are represented using the cumulative distribution, survival, and PDFs. To estimate these functions, the Kaplan–Meier estimator and nonparametric kernel density estimation are employed. While these conventional tools are beneficial, the MRF is introduced as a superior measure for distinguishing between busy and non-busy times. The MR plot serves as a visualization technique for this purpose. The MR plots divide transactions into three clusters, revealing specific patterns in customer behavior. The potential impact of automation on customer traffic is studied using a simulation-based approach. The steps involve moment estimation, diagnostic tools, and MLE.

3. Automation and Sales Forecast Accuracy

Automation can lead to higher sales and profit. This section addresses how automation affects sales by elaborating on forecasting accuracy measures. Forecasting inherently involves projecting future values based on historical data, so a different approach from the transaction analysis is required. For evaluating the impact of automation on sales, SED allows for studying forecast errors directly relevant to sales predictions. An automated tracking system involves more accurate production data collection and can improve forecasting accuracy. Improving sales forecasting accuracy is crucial to being proactive of potential factors affecting sales performances before they happen, and highlight opportunities to better target demand.

We develop a framework to examine the impacts of automation on sales. This framework is aligned with Diebold and Shin (2017), who proposed point forecast accuracy measures based on the distance of the forecast error CDF from the unit step function. Ardakani et al. (2018) further elaborate on this concept by introducing a forecast accuracy measure called mean excess error.

Figure 7
PDF and CDF plots for automation scenarios ($n = 10,000$ for each simulation)

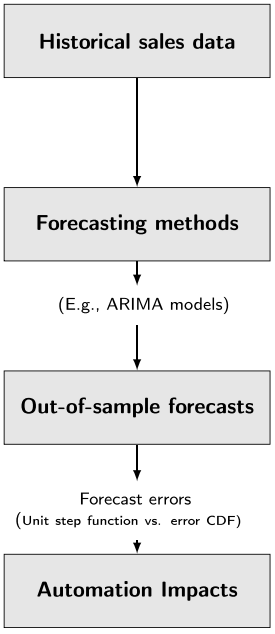


The $SED(F, F^*)$ is given by

$$SED(F, F^*) = \int_{-\infty}^{\infty} |F(e) - F^*(e)| de, \tag{12}$$

where e is the forecast error, $F(e)$ is the CDF of e, and $F^*(e)$ is the unit step function at 0. A smaller SED indicates more accurate forecasts. Forecast models can be ranked using the notion of SED. We use this formulation for finding the forecast accuracy due to automation. The steps are as follows: First, the historical sales data are used to determine the most suitable forecasting model. Then, out-of-sample forecasts are computed on a rolling scheme, and forecast errors are obtained. After estimating the error density, the automation forecast errors are simulated from the distribution. Automation scenarios are defined based on the simulations. Finally, the error CDF from the most suitable model and the error CDF from the simulation are compared to the unit step function. Figure 8 summarizes the steps above.

Figure 8
Sales forecasting and automation impacts



The forecasting procedure involves two phases. First, we evaluate multiple forecasting models using in-sample data (training data). After selecting the initial model, we employ a rolling window forecasting approach for out-of-sample predictions (test data). Specifically, the following steps are taken:

1. **Initial split:** The entire dataset is divided into an in-sample training set and an out-of-sample test set.
2. **Model fit:** The chosen forecasting model is trained using the initial in-sample data.
3. **Forecasting:** Values for the subsequent period are forecasted in the out-of-sample set.
4. **Rolling the window:** After one-step-ahead forecasting, the in-sample data are expanded, and the out-of-sample set is reduced by the same duration.
5. **Iterative update:** Steps 2–4 are repeated till the entire out-of-sample set is forecasted.

The data must be stationary time series to obtain h-step ahead forecasts for sale items. Stationarity indicates that sales distribution does not change over time (i.e., stable mean and covariance structure). The autocorrelation function (ACF) and the partial autocorrelation function (PACF) are used to determine if the data are stationary. The ACF and PACF will also be used to determine potential forecasting models for sales. Let $\{s\}_{t=1}^{T}$ be a sales sequence. We want a stable mean and covariance structure of the series over time to meet the covariance stationarity property. The ACF gives the simple and regular correlation between s_t and $s_{t-\tau}$ and can be defined as

$$\rho(\tau) = \frac{cov(s_t, s_{t-1})}{\sqrt{var(s_t)}\sqrt{var(s_{t-\tau})}}, \tag{13}$$

where *var* and *cov* are the variance and covariance. The PACF provides the partial correlation, which is the association between s_t and $s_{t-\tau}$ after controlling for the effect of $s_{t-1}, \cdots, s_{t-\tau+1}$. The sales data are recorded hourly and consist of 250,250 observations. The ACFs and PACFs for this dataset are shown in Figure 9. Damped wave patterns indicate the highly autocorrelated sales for four sale items. The serially correlated data provide biased and inconsistent forecasts.

One way to overcome the serial correlation problem is to use the first difference series defined as $\Delta s = s_t - s_{t-1}$, which is integrated into ARIMA models. The standard ARIMA(p, d, q) is defined as

$$\left(1 - L^d\right) s_t = c + \phi(L) s_t + \theta(L)\varepsilon_t, \tag{14}$$

where L is the lag operator, $\{\varepsilon\}_{t=1}^{T}$ is a serially uncorrelated white noise sequence, and $\phi(\cdot)$ and $\theta(\cdot)$ are polynomials of order p and q. To select the optimal orders p, d, q, we follow the algorithm provided by Hyndman and Khandakar (2008), which takes into account unit root tests, Akaike information criterion minimization, and MLE. The algorithm involves (1) performing a unit root test based on successive KPSS (Kwiatkowski et al., 1992) for finding the integrated order d and (2) selecting the autoregressive and moving average orders p, q via $AIC = -2\log(\mathcal{L}) + 2(p + q)$, where $\mathcal{L}$ is the maximized likelihood. The most suitable ARIMA model estimates are presented in Table 3.

It is worth noting that the primary goal of this section is not to benchmark the forecasting power of ARIMA against other models for store sales predictions. The primary objective of using ARIMA in this study is not purely for forecasting but to produce independent and identically distributed forecast errors. ARIMA is chosen given its ability to transform non-stationary time series data into stationary series. In the context of our study, achieving independent and identically distributed forecast errors is crucial because it allows for a comparison of error CDF with an ideal forecast counterpart and also aids in generating realistic simulation scenarios.

Based on the forecasting models described above, the one-step-ahead forecasts are computed on a rolling scheme, and the standardized forecast errors are obtained by splitting the entire sample into 80% training and 20% test sets. The standardized forecast error, also known as studentized forecast error, results from dividing forecast error by its standard deviation. This normalization satisfies the equal variance assumption.

We fit univariate distributions to standardized forecast errors by the maximum likelihood of finding automation impacts on sales. Numerical optimization is used to find the best values. Once the distribution parameters are estimated, the standard errors are calculated from the Hessian matrix. Table 4 presents the

Figure 9
Autocorrelation and partial autocorrelation functions of sales for four items

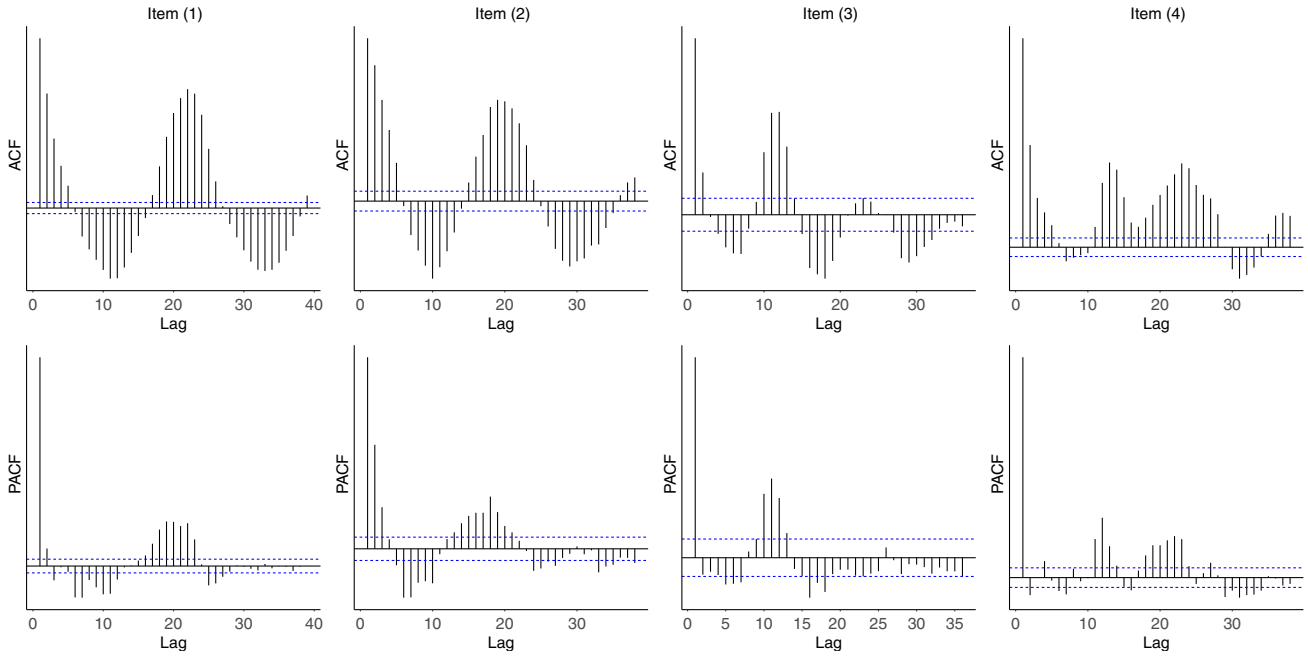


Table 3
ARIMA model estimates for the four sales items

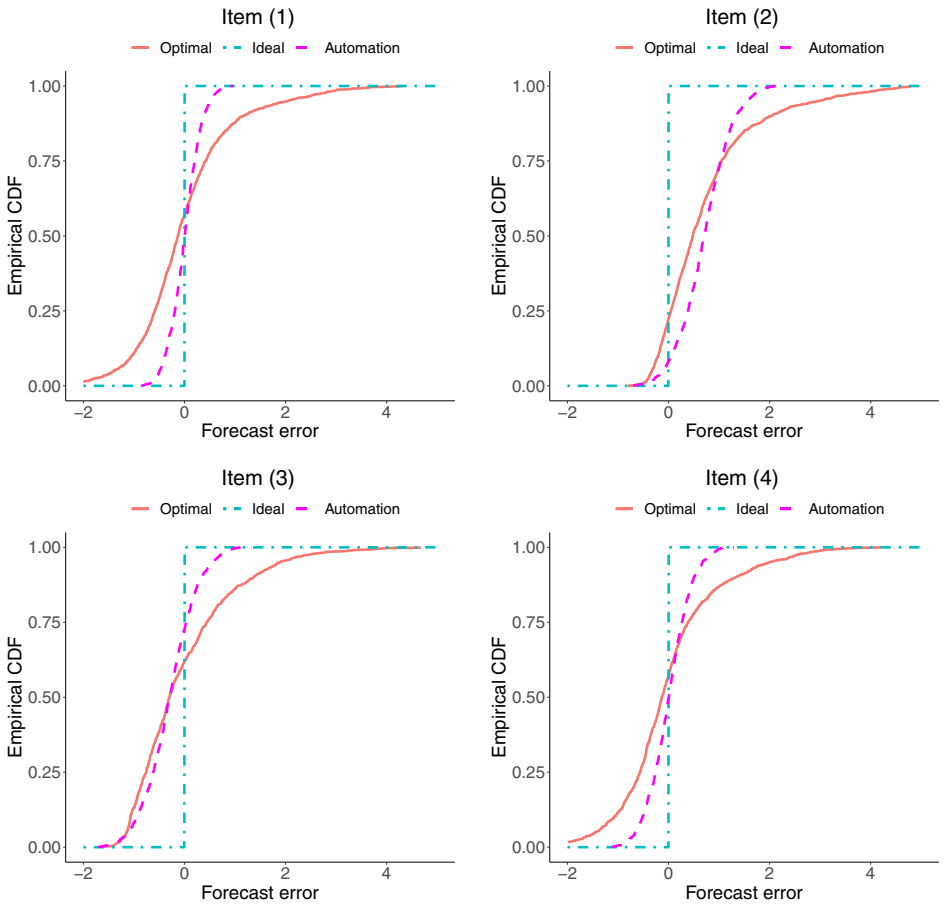
Model	Item (1) ARIMA(0,1,2)	Item (2) ARIMA(0,1,2)	Item (3) ARIMA(4,0,4)	Item (4) ARIMA(0,1,1)
ϕ_1			0.95	
			(0.20)	
ϕ_2			0.33	
			(0.35)	
ϕ_3			−0.79	
			(0.27)	
ϕ_4			0.06	
			(0.11)	
θ_1	−0.31	−0.71	−0.66	−0.39
	(0.01)	(0.01)	(0.20)	(0.01)
θ_2	−0.14	−0.22	−0.56	
	(0.01)	(0.01)	(0.29)	
θ_3			0.62	
			(0.19)	
θ_4			0.12	
			(0.08)	
AIC	86,434	59,307	33,479	46,903
MAPE	87.05	78.56	80.10	∞
RMSE	26.70	10.48	10.52	8.60
$\log\mathcal{L}$	−43,214	−29,650	−16,729	−23,449

Standard errors are reported in the parentheses. ϕ_i and θ_i are the autoregressive and moving average coefficients. The AIC is the Akaike information criterion. The MAPE and RMSE are the mean absolute percentage error and root mean squared error and $\log\mathcal{L}$ is the log-likelihood.

parameter estimates and standard errors. The results show a consistent standard deviation of 1.0 across all items, which suggests uniform variability in forecast errors. This consistency can be attributed to the standardization of forecast errors. The best-fitted distribution for the four items is normal. This result is intuitive because of the central limit theorem for which the forecast error distribution in large samples tends toward a normal distribution.

To examine the impact of automation on sales, we simulate forecast errors from more concentrated normal densities with lower standard deviations. Table 5 provides the estimated and simulated error densities. The impacts of automation are

Figure 10
Empirical CDF plots of most suitable, ideal, and simulated forecast errors



represented by $\mathbf{N}(0.0, 0.3)$, $\mathbf{N}(0.7, 0.5)$, $\mathbf{N}(-0.3, 0.5)$, and $\mathbf{N}(0.0, 0.4)$ for Items (1) to (4), respectively. Figure 10 shows empirical CDF plots of most suitable (red solid), ideal (blue dot-dashed), and simulated (pink dashed) forecast errors. An error CDF closer to the idea indicates a higher forecast accuracy. All the simulated error CDFs are ranked higher, indicating automation would lead to higher forecast accuracy.

This simulation highlights potential scenarios where automation can lead to more accurate forecasting. The parameters for the simulated normal distributions are selected based on a theoretical understanding that automation reduces the variability in forecast errors, making the errors more concentrated. The standard deviations selected for the simulation represent potential improvements due to automation. These simulated standard deviations are smaller than the estimated standard deviations from Table 4. We emphasize the hypothetical nature of these simulations. These are illustrative scenarios rather than empirically derived outcomes. A few points are noteworthy.

1. The simulation parameters are chosen to study a range of scenarios. While these parameters might seem arbitrary, they are selected to represent potential scenarios.
2. The results in Table 5 and Figure 10 are outcomes of the simulations to illustrate scenarios where automation might lead to more accurate forecasting.
3. Alternatively, we can simulate sales data that take into account the impact of automation and then apply the forecasting scheme. Such an approach would also introduce additional assumptions. While our approach models changes in forecast error due to automation, we acknowledge the value of studying alternative methods.

Green and Weaver (2008) study various qualitative methodologies adopted in the foodservice industry for sales forecasting. They evaluate the IT systems in place and their role

Table 4
Parameter estimates of the fitted error densities

Item (1)		Item (2)		Item (3)		Item (4)	
Mean	SD	Mean	SD	Mean	SD	Mean	SD
0.0	1.0	0.7	1.0	−0.3	1.0	0.0	1.0
(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)

The best-fitted distribution for the four items is normal. SD denotes the standard deviation. Standard errors are reported in parentheses and computed from the Hessian matrix estimate.

Table 5
Estimated and simulated forecast error densities

	Item (1)	Item (2)	Item (3)	Item (4)
Estimated	**N**(0.0,1.0)	**N**(0.7,1.0)	**N**(-0.3,1.0)	**N**(0.0,1.0)
Simulated	**N**(0.0,0.3)	**N**(0.7,0.5)	**N**(-0.3,0.5)	**N**(0.0,0.4)

$\mathbf{N}(\mu, \sigma)$ is the normal density with mean μ and standard deviation σ.

in forecasting and show a growing trend toward adopting IT systems. Accurate sales forecasting directly impacted the restaurant's ability to manage inventory efficiently and optimize labor costs. Our study also studies automated tools, complementing by adding quantitative data and performance metrics of various forecasting models. Tuomi and Ascencao (2023) also study automation effects in the hospitality industry through observational studies, expert interviews, and system evaluations. They argue that the introduction of automation can augment the efficiency of some roles, but entirely replacing human staff, especially in luxury settings, could detract from the overall customer experience. This section focuses on forecasting automation, which tackles the backend processes and predictive models. In contrast, Tuomi and Ascencao (2023) focus on the frontend tasks in the hospitality industry, especially those involving direct customer interaction.

4. Cost Analysis of Automation

Profit is computed by subtracting total cost from sales. The impact of automation on profit may be complex since automation affects both sales and cost. Increases in sales and lower costs lead to a rise in profit, but if sales and costs increase, the change in profit depends on the magnitude of changes in each component. The sales component is discussed in the previous section. This section elaborates on the other element of profit. Ideally, examining the marginal impact of automation on cost should be done by including the impact on the cost function. Estimating the cost function involves regressing cost on input prices, the number of items produced, and information on the quantity of fixed inputs used. However, the lack of data on the quantity of all fixed inputs (e.g., materials, technology, utilities) and existing different frequencies of the available data make the regression approach unfeasible. Therefore, the cost analysis involves estimating the impact of automation separately on labor costs and the cost of waste.

The cost data are hourly, with 16,851 observations for labor cost and 18,921 observations for the waste data. Given the importance of costs in determining profit and the variability associated with different cost components and data availability, the methodology for this section is customized. Factors such as wages, benefits, and payroll taxes influence labor costs. The scenarios formulated around standard deviations from the mean are motivated by the literature and the need to capture a broad range of potential outcomes post-automation (Davenport & Ronanki, 2018). The range allows for both short-term and long-term effects.

Four scenarios are considered to assess the impact of automation on labor costs. Let $\mathbf{l}_t$ be the hourly labor cost at time t with mean $\mu_{\mathbf{l}}$ and standard deviation $\sigma_{\mathbf{l}}$. Scenarios are estimated by assuming that automation increases (or decreases) labor costs by one and two standard deviations. The parameters for different scenarios are based on the existing literature. Utilizing one and two standard deviations provides a balanced approach to capture moderate and more extreme scenarios of the effects of automation. Including these specific scenarios helps emphasize the dual nature of automation's impact on labor costs—it can increase and decrease. By considering these standard deviations, we aim to capture this outcome variability.

After automation, labor costs may increase for two reasons: (1) The employee may be retrained for a different position and (2) greater customer traffic because of lower wait time, which may impact the cost at other levels. Automation may also reduce labor costs by reducing the worker turnover ratio. Specifically, Scenarios 1 and 2 are shown by $\mu_{\mathbf{l}} \pm \sigma_{\mathbf{l}}$ and automation Scenarios 3 and 4 are shown by $\mu_{\mathbf{l}} \pm 2\sigma_{\mathbf{l}}$. Figure 11 presents the scatter plots and histogram of the labor cost per hour and automation impacts. The vertical dashed lines represent the mean for each case. Although higher average labor costs (Scenarios 1 and 3) are attainable in the short run, lower average labor costs (Scenarios 2 and 4) are expected in the long run. Automation reduction in labor costs incentivizes taking steps towards an automated system.

Table 6 compares hourly labor costs under different automation scenarios. The actual labor costs per hour range between \$10.24 and \$18.16. Accounting for a one standard deviation increase in labor costs due to automation (Scenario 1), the range lies between \$11.60 and \$19.52. A decrease by one standard deviation (Scenario 2) yields costs ranging from \$8.89 to \$16.81. Considering a more extreme case of a two standard deviation increase (Scenario 3), the labor costs span from \$12.96 to \$20.88. A two standard deviation decrease (Scenario 4) projects the costs between \$7.53 and \$15.45. These findings suggest that the long-term expectation of automation is to reduce labor costs, with Scenario 4 indicating the most substantial potential reduction. This is consistent with the view that automation, in the long run, could lead to more efficient operations, thus reducing labor costs.

The approach of examining the effects of automation on labor costs using standard deviations is in line with the methods proposed by Davenport and Ronanki (2018). They emphasize the need for businesses to understand the impact of artificial intelligence and automation on their operations. However, our study extends their work by using hourly labor costs (big data) to show how automation can influence costs in real-world operational environments. This adds an empirical dimension to the conceptual frameworks presented in existing literature. In addition, by assessing both increases and decreases in labor costs, this study offers a balanced view of potential outcomes, filling a gap in the literature that often tends to focus on either the positive or negative implications of automation.

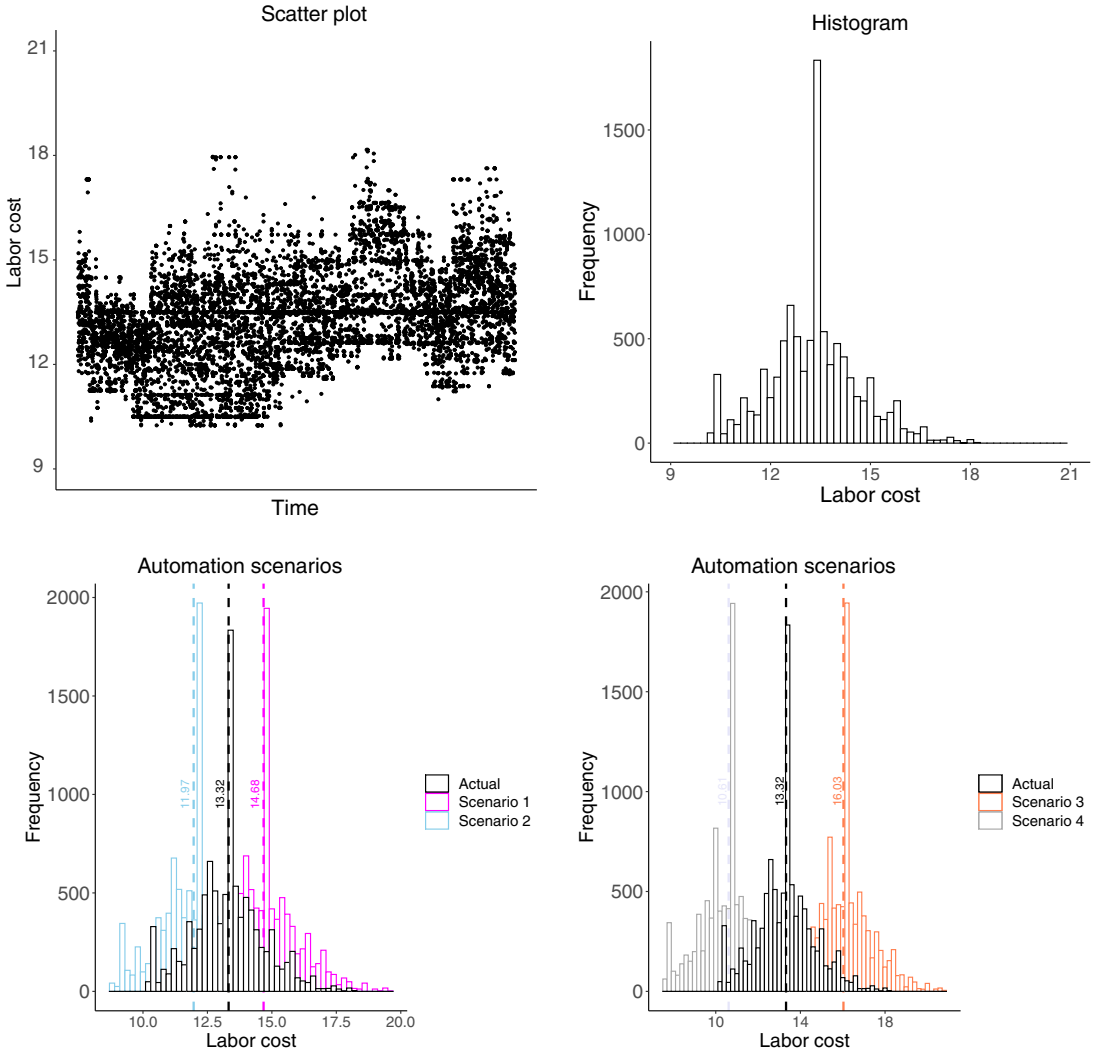
The relationship between automation and labor costs is studied in the economics and business literature. For instance, Acemoglu and Restrepo (2019) theoretically examine how automation technologies might displace labor in certain tasks while potentially creating new roles, leading to mixed effects on labor demand. Our findings also empirically focus on hourly labor costs and consider the real-world challenges businesses face when data on all factors affecting costs are not uniformly available. This approach complements the theoretical frameworks and bridges the gap between theory and practice.

It is equally essential to evaluate how automation influences other operational costs. One such significant cost, particularly in

Table 6
Min and max of hourly labor cost for the automation scenarios

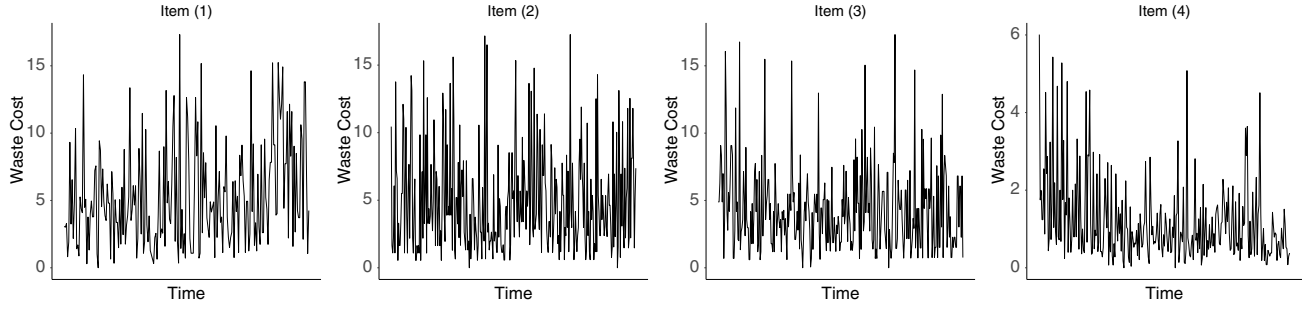
Actual		Scenario 1		Scenario 2		Scenario 3		Scenario 4	
Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
10.24	18.16	11.60	19.52	8.89	16.81	12.96	20.88	7.53	15.45

Figure 11
Scatter plots and histograms of labor cost per hour along with automation impacts



manufacturing and production environments, is the cost associated with waste. The cost of waste, measured as daily waste cost per item, is usually recorded manually. Figure 12 plots the waste costs over time. To understand the potential effects of automation on waste, we identify the theoretical distribution of the data and then simulate the impact of automation. Specifically, we apply moment matching estimation (MME) to discern the theoretical distributions of waste cost. For a detailed explanation of this methodology, see Venables and Ripley (2013) and Vose (1996). Under the MME, the moment parameters θ are determined by setting the theoretical moment equal to the corresponding empirical moment. This matching is followed by the law of large numbers. The theoretical moments

Figure 12
Waste costs for the four items



is given by $m_k(\theta)$ and the empirical (sample) moments are defined as

$$\bar{W}^k = n^{-1} \sum_i w_i^k, \tag{15}$$

where W is the waste cost, and k is the number of parameters to estimate. The MME estimates the parameters θ by minimizing the sum of squared errors

$$\underset{\theta}{\operatorname{argmin}} \sum_k \sum_i (m_k(\theta) - w_i^k)^2. \tag{16}$$

Consistency and asymptotic normality are the large sample properties of estimator $\hat{\theta}$. See Newey and McFadden (1994) for proof of consistency and statistical inference.

The theoretical distribution of the waste cost data is identified by comparing Q-Q and P-P plots for well-known theoretical distributions to the empirical waste cost distribution. The Q-Q plot represents theoretical quantiles against empirical ones, and the P-P plot represents theoretical probabilities against empirical ones. Figure 13 presents the histogram and theoretical densities along with Q-Q and P-P plots for the fitted distributions of waste cost. The findings suggest that the gamma distribution with shape and scale parameters α and β best fit the waste cost. The PDF of the gamma distribution can be written as

$$f(w) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{\alpha-1} e^{-\beta w}, \tag{17}$$

where $\Gamma(\cdot)$ is the gamma function.

Table 7 presents the estimated gamma distribution for the waste cost data and four different automation scenarios estimated through Monte Carlo simulations. We assume automation increases waste cost but with different variability in the data. Automation increases average daily waste but produces higher quality products. Figure 14 plots the estimated distribution for the waste cost data (solid black line) and the automation scenarios (colored dashed lines). The findings suggest that automation decreases waste for low-waste-cost days (i.e., Scenarios 1 and 2) and increases for high-waste-cost days (i.e., Scenarios 3 and 4).

The four automation scenarios presented in Figure 14 capture a range of potential impacts automation could have on waste costs.

Table 7
Automation impacts on simulated gamma densities

Estimated	Scenario 1	Scenario 2	Scenario 3	Scenario 4
$\mathbf{G}(1.2, 0.3)$	$\mathbf{G}((3.0, 0.2)$	$\mathbf{G}(5.0, 0.2)$	$\mathbf{G}(3.0, 0.4)$	$\mathbf{G}(5.0, 0.4)$

$\mathbf{G}(\alpha, \beta)$ is the gamma density with shape and scale parameters α and β.

Figure 13
Histogram, density, Q-Q, and P-P plots for the fitted distributions of waste cost

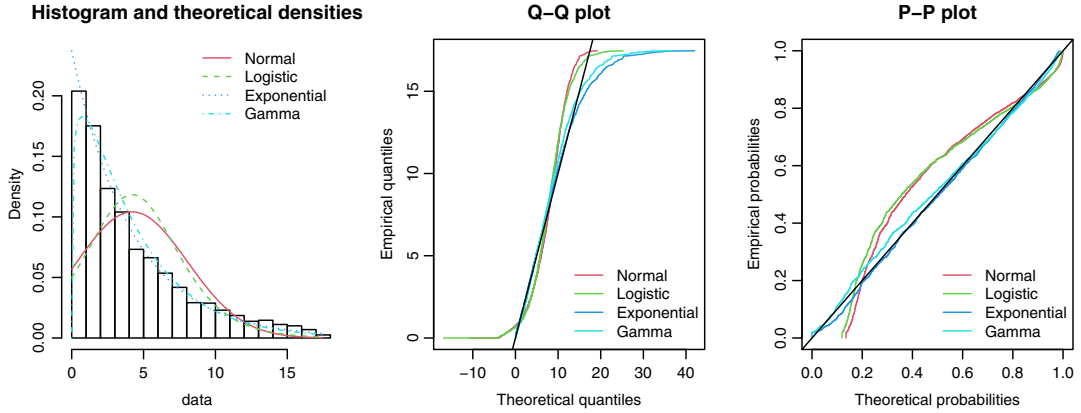
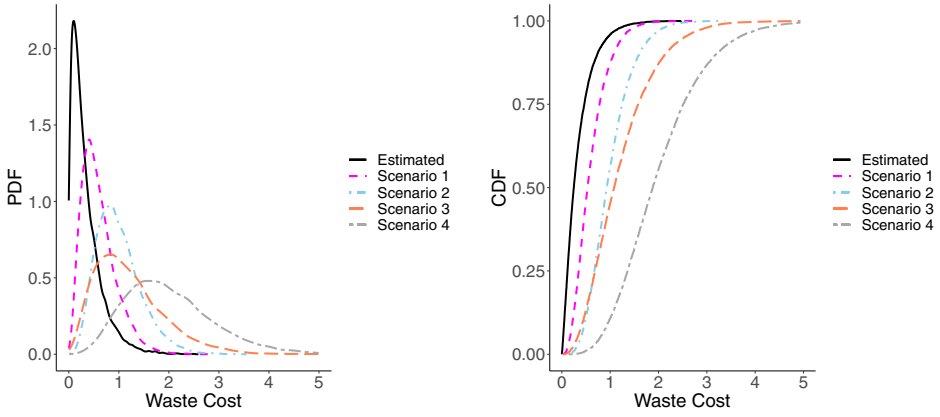


Figure 14
PDF and CDF plots for automation scenarios ($n = 10,000$ for each simulation)



The choices for shape and scale parameters in the gamma distributions for each scenario, as listed in Table 7, are driven by preliminary analyses, empirical observations, and their correspondence to realistic automation outcomes. Specifically, Scenarios 1 and 2 focus on the effects of automation on days with low-waste costs, while Scenarios 3 and 4 emphasize its impact on days with high waste costs.

Labor and waste significantly influence overall costs. Labor costs often represent the most significant recurring expense for foodservice businesses. As the industry faces fluctuating wage rates, employee turnover, and the constant need for training and development, labor costs remain at the forefront of financial considerations. On the other hand, waste costs are critical in determining operational efficiency and sustainability. Potential extensions to this research could encompass a more holistic assessment of automation's impact, including inventory management, energy consumption, and customer experience.

This section focuses on analyzing the cost of automation, mainly labor and waste expenses. Our findings show the potential cost-saving aspects of automation. This result is consistent with Rydzik and Kissoon (2022). They study the implications of automation for tourism workers and emphasize the ethical considerations and potential socio-economic disparities resulting from widespread automation in the tourism sector. The recent literature has employed survey-based approaches to study behavioral patterns regarding automation's impact on production costs. See Mui et al. (2022) for another example in the foodservice industry. Our study uses quantitative methods and simulations to cover a broad range of outcomes, addressing a gap in the literature.

5. Frequentist vs. Bayesian

The frequentist methodology has been the primary choice for analyzing and inferring distributions in this study. However, given the interdisciplinary and evolving landscape of statistical methodologies, it is essential to compare and contrast this with the Bayesian perspective, which offers a different but complementary approach to inference (Gelman, 2006; Gelman et al., 2013). In the frequentist paradigm, parameters are assumed to be fixed quantities. The data are considered a random sample drawn from a particular population, and our goal is to use the sample data to make inferences about these fixed but unknown parameters (Lehmann & Casella, 2006). In a Bayesian approach, parameters are random variables, and prior beliefs about these parameters are expressed through a prior distribution, $p(\theta)$. When new data, x, are observed, the Bayesian updates this prior to forming a posterior distribution, $p(\theta|x)$, using Bayes' theorem (Bayes, 1958)

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}.$$

The empirical distribution identified in our study could be analogous to a Bayesian prior in specific contexts. In our case, this distribution serves as an empirical representation of our sample. However, in the Bayesian paradigm, this would be the prior belief about the transaction distribution (Bernardo & Smith, 2009). The frequentist techniques we used, including fitting distributions to empirical data, directly reflect the observed data patterns and are conventional in many industries, ensuring ease of interpretation for stakeholders (Efron, 2012). Bayesian methods, however, allow for the incorporation of prior knowledge and are particularly powerful when data are sparse or when prior information is significant (Robert, 2007). They offer a full probability distribution over the parameters, which can provide richer insights than point estimates.

The frequentist methods may sometimes not fully leverage prior information, while Bayesian methods can be computationally demanding, especially with non-conjugate priors, and may require advanced sampling methods (Gilks et al., 1995). In industry settings where rapid decisions are essential, computational feasibility becomes necessary. Our frequentist methods are designed for computational efficiency. In contrast, specific Bayesian approaches, especially with complex priors or large datasets, can become computationally intensive and may necessitate specialized algorithms and hardware (Brooks et al., 2011). The choice between frequentist and Bayesian approaches often depends on the research objectives, data nature, computational considerations, and the target audience's preference.

6. Concluding Remarks

The rise of technology urges industries to adopt automated systems to target demand and efficiently utilize resources. One way to increase productivity measured by output per hour is by adopting automated systems. Automated systems help redistribute resources and increase efficiency by lowering total costs. In the foodservice industry, automated systems can use dispensing, breading, and frying machines with a tracking system. Retrospectively, automation can be measured by analyzing the historical data. Historical data, however, are not available before implementing automation. This paper introduces a probabilistic view to capture automation impacts prospectively. This forward-looking view accommodates predictive approaches in achieving the targets.

This paper examines automation impacts on customer traffic, sales, and cost through big data approaches and simulation studies. First, univariate distributions are fitted to the empirical data to determine parameter estimates of the distributions. Distribution parameters are estimated by maximizing likelihood and MME. After selecting the best-fitted distribution, Monte Carlo simulations are constructed to find automation impacts. Automation impacts on sales are captured through ranking forecast models and SED. SED is the distance of the forecast error CDF from the unit step function. Automation is captured by simulated forecast errors, which provide a smaller SED, implying improved forecast accuracy.

Funding Support

This research was funded by the Parker College of Business at Georgia Southern University.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability

The data utilized in this research originate from a consultancy agreement with a private firm. Due to confidentiality commitments,

the data cannot be made publicly available. In the interest of maintaining proprietary and strategic advantages, the firm has opted to keep the data private. However, all relevant methodologies and analyses employed in this study are provided to ensure the replicability of the research with similar datasets.

References

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, *33*(2), 3–30.

Ardakani, O., Asadi, M., Ebrahimi, N., & Soofi, E. (2022). Variants of mixtures: Information properties and applications. *Journal of the Iranian Statistical Society*, *20*(1), 27–59.

Ardakani, O. M., Asadi, M., Ebrahimi, N., & Soofi, E. S. (2020). MR plot: A big data tool for distinguishing distributions. *Statistical Analysis and Data Mining*, *13*(4), 405–418.

Ardakani, O. M., Ebrahimi, N., & Soofi, E. S. (2018). Ranking forecasts by stochastic error distance, information and reliability measures. *International Statistical Review*, *86*(3), 442–468.

Bayes, T. (1958). An essay towards solving a problem in the doctrine of chances. *Biometrika, 45*(3–4), 296–315.

Bernardo, J. M., & Smith, A. F. (2009). *Bayesian theory*. USA: Wiley.

Brooks, S., Gelman, A., Jones, G., & Meng, X. L. (2011). *Handbook of Markov chain Monte Carlo*. USA: CRC Press.

Casella, G., & Berger, R. L. (2002). *Statistical inference*. USA: Duxbury Thomson Learning.

Cullen, A., & Frey, H. (1999). *Probabilistic techniques in exposure assessment*. USA: Springer US.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*. 108–116.

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, *64*(4), 1–34.

Diebold, F. X., & Shin, M. (2017). Assessing point forecast accuracy by stochastic error distance. *Econometric Reviews*, *36*(6-9), 588–598.

Efron, B. (2012). *Large-scale inference: Empirical Bayes methods for estimation, testing, and prediction*. UK: Cambridge University Press.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534.

Gelman, A., Carlin, J. B., Stern, H. S., Duson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. USA: CRC Press.

Gilks, W.R., Richardson, S., & Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. USA: CRC Press.

Green, Y. N., & Weaver, P. A. (2008). Approaches, techniques, and information technology systems in the restaurants and foodservice industry: A qualitative study in sales forecasting. *International Journal of Hospitality & Tourism Administration*, *9*(2), 164–191.

Gunasekaran, S. (2009). Food engineering. In G. V. Barbosa-Canovas (Eds.), *Automation of food processing*. Eolss Publishers.

Gupta, R., & Bradley, D. (2003). Representing the mean residual life in terms of the failure rate. *Mathematical and Computer Modelling*, *37*(12-13), 1271–1280.

Huang, Y. (2013). Automatic process control for the food industry: An introduction. In D. G. Caldwell (Eds.) *Robotics and automation in the food industry: Current and future technologies* (pp. 3–20). Woodhead Publishing.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *27*(3), 1–22.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions*. USA: Wiley.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457–481.

Kleinbaum, D. G., & Klein, M. (1996). *Survival analysis a self-learning text*. Germany: Springer.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, *54*(1-3), 159–178.

Lehmann, E. L., & Casella, G. (2006). *Theory of point estimation*. Germany: Springer.

Moore, C. A. (2012). *Automation in the food industry*. Germany: Springer.

Mui, K. W., Wong, L. T., Tsang, T. W., Chiu, Y. H., & Lai, K. W. (2022). Food waste generation in a university and the handling efficiency of a university catering facility-scale automatic collection system. *Facilities*, *40*(5–6), 297–315.

Newey, W.K., & McFadden, D. (1994). Chapter 36 Large sample estimation and hypothesis testing. *Handbook of Econometrics*, *4*, 2111–2245.

Papoulis, A., & Pillai, S. U. (2002). *Probability, random variables, and stochastic processes*. USA: McGraw-Hill.

Robert, C. P. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*. Germany: Springer.

Rydzik, A., & Kissoon, C. S. (2022). Decent work and tourism workers in the age of intelligent automation and digital surveillance. *Journal of Sustainable Tourism*, *30*(12), 2860–2877.

Sun, L., & Zhang, Z. (2009). A class of transformed mean residual life models with censored survival data. *Journal of the American Statistical Association*, *104*(486), 803–815.

Tuomi, A., & Ascencao, M. P. (2023). Intelligent automation in hospitality: Exploring the relative automatability of frontline food service tasks. *Journal of Hospitality and Tourism Insights*, *6*(1), 151–173.

Venables, W. N., & Ripley, B. D. (2013). *Modern applied statistics with S-PLUS*. Germany: Springer.

Vose, D. (1996). *Quantitative risk analysis: A guide to Monte Carlo simulation modelling*. USA: Wiley.

How to Cite: Ardakani, O. M. & Saenz, M. (2024). Evaluating Economic Impacts of Automation Using Big Data Approaches. *Journal of Data Science and Intelligent Systems* 2(1), 150–164, https://doi.org/10.47852/bonviewJDSIS32021569