Predictive Model Selection for Forecasting Product Returns

Jungmok Ma
Department of National Defense Science,
Korea National Defense University,
Seoul 10544, South Korea
e-mail: jungmokma@kndu.ac.kr

Harrison M. Kim[1]
Associate Professor
Industrial and Enterprise Systems Engineering,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801
e-mail: hmkim@illinois.edu

As awareness of environmental issues increases, the pressures from the public and policy makers have forced original equipment manufacturers (OEMs) to consider remanufacturing as the key product design option. In order to make the remanufacturing operations more profitable, forecasting product returns is critical due to the uncertainty in quantity and timing. This paper proposes a predictive model selection algorithm to deal with the uncertainty by identifying a better predictive model. Unlike other major approaches in literature such as distributed lag models or DLMs, the predictive model selection algorithm focuses on the predictive power over new or future returns and extends the set of candidate models. The case study of reusable bottles shows that the proposed algorithm can find a better predictive model than the DLM.

[DOI: 10.1115/1.4033086]

Keywords: product return forecasting, predictive modeling, time series analysis, remanufacturing

1 Introduction and Background

Wealth of data is being generated by the users, OEMs, and markets throughout the life cycle of product systems: prelife (product planning and manufacturing), usage, and end-of-life (remanufacturing and disposal) stages. For a successful remanufacturing at the end-of-life stage, various sources of uncertainty (product returns, life cycle length, condition, etc.) should be properly addressed. Among them, the quantity and timing of product returns are perceived as key information in remanufacturing operations [1–3]. The goal of this paper is to present a method to improve the predictive power of existing return forecasting systems.

1.1 Remanufacturing. Remanufacturing is the process that restores used products to a like-new condition and returns them to customers. As remanufacturing reutilizes the materials, parts, or components with the value in original products, it can be a profitable and environmentally friendly option.

From the perspective of supply chain management, remanufacturing is also associated with reverse logistics or closed-loop supply chains. Generally, closed-loop supply chains are uncontrollable due to the uncertainties in quantity, timing, and condition of returned products [3]. Clottey et al. [2] introduced an example of closed-loop supply chains when an OEM is a contract remanufacturer as shown in Fig. 1. Based on the contractual agreement, the OEM should satisfy the monthly orders (D_t) of the contractors. The primary source of remanufacturing is *cores* ($Y_t = \{y_1, \ldots, y_t\}$), which is remanufacturable returned products. Other sources include the volume of cores from inventory (I_{t-1}), core brokers (A_t), and other manufacturers (M_t). Depending on the accuracy of the predicted volume of cores ($\hat{Y}_t$), either more cores are needed (additional acquisition cost) or there are redundant cores (additional inventory cost).

1.2 Forecasting Product Returns. Forecasting product returns is defined as the prediction of the quantity and timing of product returns or cores, which minimizes the additional costs in remanufacturing. Simple and naive approaches include either using the proportion of returns to sales with a known life cycle length [1] or limiting to the special market environments, such as take-back programs [4] or lease contracts [5]. Various advanced approaches were proposed such as causal analysis, simulation/soft-computing, and statistical methods [3]. Among them, the statistical methods are the most popular approach and the main focus of this paper.

Table 1 summarizes the important studies in the statistical methods, especially based on a DLM. All of the studies showed that the key to forecasting returns was the relationship between returns and sales. Therefore, returns could be explained with distributed impacts of previous sales (delay function). However, Goh and Varaprasad [6] and Toktay et al. [1,7] did not conduct prediction tests after building DLMs. Clottey et al. [2] and Clottey and Benton [8] compared different DLMs using simulated data but did not consider other models discussed in this paper.

1.3 Predictive Modeling of Product Returns. When the historical data of returns and sales are available, the DLMs in Table 1 are useful to explain the causal relationships between returns and sales. However, the predictive modeling perspective of forecasting product returns remains widely unexplored.

Shmueli [9] provided an interesting discussion on the differences between explanatory modeling and predictive modeling. Explanatory modeling is "the use of statistical models for testing causal explanations," while predictive modeling is "the process of applying a statistical model or data mining algorithm to data for the purpose of predicting new or future observations" [9]. The key features are summarized as follows: First, R^2 values indicate the explanatory power (explanatory), while the cross-validation process or holdout method is for the predictive power (predictive). Second, for model selection Bayesian information criterion is to measure the goodness-of-fit (explanatory), while Akaike information criterion (AIC) is suited for predictive accuracy (predictive). Third, for predictive purposes multicollinearity will not affect the performance of predictive models (predictive). Fourth, autoregressive integrated moving average (ARIMA)-type models are not suitable for causal explanations, but for prediction (predictive). The proposed algorithm in this paper reflects these points to build a predictive model.

Table 2 shows the position of this paper with regards to the literature. The majority of studies were based on the DLM (bivariate model), which related returns to previous sales. Traditional univariate models (ARIMA and exponential smoothing) were criticized and not used by researchers since the models could not utilize the information of past sales [2,3,10]. In the proposed algorithm, both univariate and bivariate models will be explored along with their mixed components to improve the prediction accuracy.

2 Methodology

The additional cost of the closed-loop supply chain in Fig. 1 can be either $c_o(\hat{Y}_t - Y_t)$ or $c_u(Y_t - \hat{Y}_t)$, where c_o is the cost coefficient of overestimation, and c_u is the cost coefficient of underestimation. In order to minimize the cost (when the cost coefficients are not considered), predictive models for $\hat{Y}_t$ should minimize the absolute deviation between $\hat{Y}_t$ and Y_t. For the validation of

[1]Corresponding author.

Contributed by the Design for Manufacturing Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received September 22, 2015; final manuscript received February 4, 2016; published online April 6, 2016. Assoc. Editor: Carolyn Seepersad.

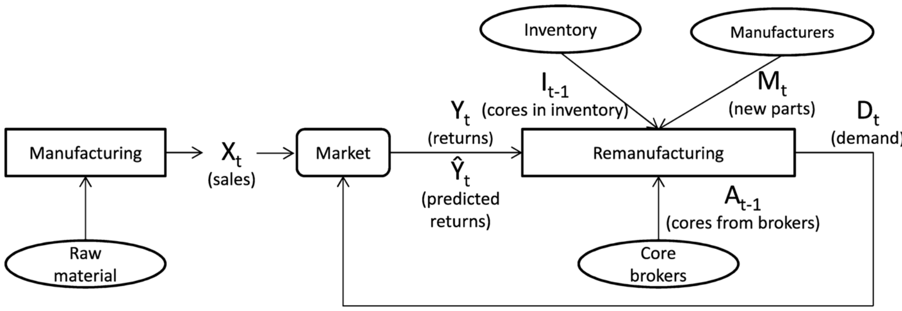

Fig. 1 A closed-loop supply chain

Table 1 DLM-based approaches

Literature	Delay function modeling	Product	Data	Note
Goh and Varaprasad [6]	Transfer function	Reusable bottles	Real (open)	No prediction test
Toktay et al. [1,7]	Bayesian method	Single-use cameras	Real (hidden)	No prediction test
Clottey et al. [2]	Bayesian method	Electronic parts	Simulation (hidden)	Limited prediction tests
Clottey and Benton [8]	Bayesian method	Electronic parts	Simulation (hidden)	Limited prediction tests

predictive models, if multiple one-step-ahead holdout samples can be tested, the prediction performance measure can be expressed as the mean absolute error with a h time horizon

$$\sum_{t=n+1}^{n+h} \frac{|Y_t - \hat{Y}_{t|t-1}|}{h} \tag{1}$$

where n is the index of current time, h is the size of the holdout samples, Y_t is the observations of product returns, and $\hat{Y}_{t|t-1}$ represents a forecast of one-step ahead time based on discrete time series Y_{t-1} and X_{t-1}. The goal is to find a good statistical model such that the prediction performance measure in Eq. (1) is minimized.

2.1 Predictive Models in the Proposed Algorithm

2.1.1 ARIMA. The ARIMA model is a univariate time series model, which relates returns to previous returns. The ARIMA model (ARIMA(p, d, q)) is a combination of three models given as [11]

$$\begin{aligned}(1 - \phi_1 B - \cdots - \phi_p B^p)(1-B)^d Y_t \\ = c + (1 + \theta B + \cdots + \theta_q B^q)e_t\end{aligned} \tag{2}$$

where B represents a backward shift operator, e.g., $BY_t = Y_{t-1}$; the first parenthesis is an autoregressive (AR) model of order p with coefficients ϕ, which is a linear combination of past observations; the second parenthesis is an integration (or differencing operation); and the third parenthesis on the right-hand side is a moving average (MA) model of order q with coefficients θ, which is a linear combination of past forecast errors. The ARIMA model with seasonal terms can be also found in Hyndman and Athanasopoulos [11]. Hyndman and Khandakar [12] provided an automatic forecasting algorithm to handle a large number of univariate time series data.

2.1.2 DLM. DLM relates returns to previous sales with a belief that the effects of previous sales are distributed over future time periods. The distributed lag represents the current and lagged values of the explanatory variable (sales) to predict the current values of the response (returns). DLM is a dynamic model since the effects of the explanatory variables are captured over time. The finite DLM is given as

$$Y_t = \alpha + \sum_{s=0}^{t-1} \beta_s X_{t-s} + e_t \tag{3}$$

where α is the intercept, β_s is the distributed lag weight, and e_t is the white noise. Note that generally x_{n+1} (n is the index of current time) is not available to forecast y_{n+1} [1,2] (i.e., s should start from 1 in Eq. (3) in this case).

In order to estimate the model in Eq. (3), either unrestricted or restricted distributed lag weights can be used. The unrestricted distributed lags can be directly estimated by least squares without any restrictions on the lags. The possible problem of this approach is multicollinearity, which can cause high variances of the estimates. The restricted distributed lags were used by researchers to reduce the effects of multicollinearity: the polynomial distributed lag [13] and the infinite geometric lag [14]. These restricted distributed lags require an assumption of a certain form of lag weights, and this information may be not available to design engineers.

In the predictive model selection algorithm, the unrestricted DLM is used to estimate the lag weights with maximum likelihood estimation. If the algorithm finds that this candidate model is the best, the restricted DLMs can be applied to improve the predictive performance.

2.1.3 Mixed Model. The two candidate models were discussed in Secs. 2.1.1 and 2.1.2. The last model is a mixed model

Table 2 Position of this paper

Model	Data usage	Note
ARIMA (univariate model)	$Y_t = F(Y_{t-1}$ & $e_t)$	Criticized and not used in the literature
DLM (bivariate model)	$Y_t = F(X_{t-1})$	Mainly utilized in the literature
Proposed model	$Y_t = F(X_{t-1}$ & Y_{t-1} & $e_t)$	Search for a better predictive model

Note: Y_t is the returns, X_t is the sales, and e_t is the forecast errors.

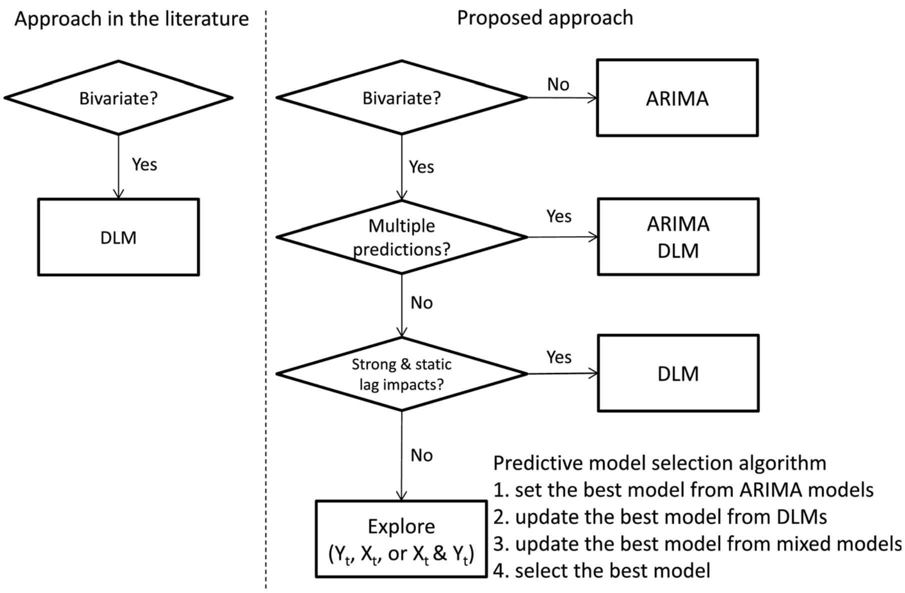

Fig. 2 Difference between DLM and proposed approach

of the two candidate models. The regression model with ARIMA errors is

$$Y_t = \eta(B)X_t + u_t \tag{4}$$

$$\gamma(B)u_t = \delta(B)e_t \tag{5}$$

where $\eta(B) = 1 + \eta_1 B + \cdots + \eta_k B^k$ with k order of regression coefficient η; u_t is an error term in the regression model; $\gamma(B) = 1 - \gamma_1 B - \cdots - \gamma_p B^p$ with coefficient γ, which represents an AR model of order p for u_t; and $\delta(B) = 1 + \delta_1 B + \cdots + \delta_q B^q$ with coefficient δ, which represents an MA model of order q for u_t. u_t can follow an autoregressive moving average process [11], which may require a differencing operation, and e_t is the white noise.

2.2 Predictive Model Selection Algorithm. Figure 2 shows the difference between the approach with the DLM in the literature (left) and the new predictive model selection algorithm (right). While the DLM-based approach requires the strong and static relationships between returns and sales, the predictive model selection algorithm extends this to deal with the case that the lag impact of sales is not strong enough or varying over time. If sales data are not available, the univariate model (ARIMA) can be the only option. If a multistep-ahead prediction is required, the DLM-based approach needs predicted values of sales from the ARIMA model. Finally, if design engineers do not know the existence of strong lag impacts, the following steps should be assumed:

Step 1: Set a candidate model for forecasting future returns: (1) previous Y_t (ARIMA), (2) previous X_t (DLM), or (3) both (mixed model).

Step 2: Fit the model along the candidate model based on AIC $= -2\ln(L) + 2K$, where L is the maximized likelihood value, and K is the number of parameters in the model.

Step 3: Compare the three fitted models based on the prediction performance measure in Eq. (1) with a reasonable h time horizon, and select the best model.

Note that AIC can be used to compare different models but the same data should be used to compute the likelihoods. For example, AIC values cannot be compared with differencing operations and regressors. Step 2 is the model selection, and step 3 is the model validation. The algorithm reflects the important points for predictive modeling as discussed in Sec. 1.3.

3 Case Study

The real data of reusable bottles studied by Goh and Varaprasad [6] were used to test whether the predictive model selection algorithm could find a better predictive model than the DLM. The hypothesis is that a different predictive model can be the best selection depending on data though the DLM was proposed as the single best model in the literature.

Figure 3 shows the sales and returns of reusable bottles in a 60-month period. The original study [6] used all the data to build a model and did not conduct the performance test of the model. This study used the first 50 data points for modeling based on the recommendation of the original study (at least 50 data points for time series models) and the last ten data points (i.e., holdout samples) for the validation of a predictive performance. It is also the conventional validation method which partitions the data into two sets (about 70% for modeling and 30% for validation).

The case study was conducted under three different scenarios. The first scenario (Sec. 3.1) was when sales data at $t = n$ were available for returns at $t = n$ as Goh and Varaprasad [6] assumed. The second scenario (Sec. 3.2) was when sales data at $t = n$ were not available for returns at $t = n$ as other researchers assumed [1,2]. The third scenario (Sec. 3.3) was when there were strong and static relationships between returns and sales, which was the most favorite scenario for the DLM.

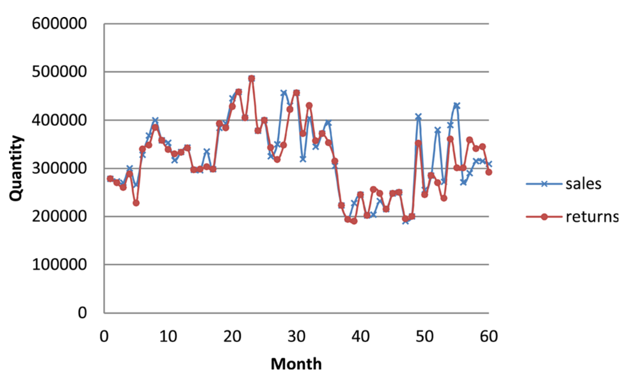

Fig. 3 Sales and returns of reusable bottles (redrawn from Ref. [6])

Table 3 Results of prediction performance measure (best result with underline for each scenario)

	ARIMA	DLM	Mixed model
Sales data from 1 to t are available for return quantity at t (Sec. 3.1)	<u>31,000</u>	38,500	36,000
Sales data from 1 to $t-1$ are available for return quantity at t (Sec. 3.2)	31,000	43,200	<u>27,500</u>
Strong lag impacts exist between sales and returns (Sec. 3.3)	44,400	<u>350</u>	<u>350</u>

Note: Lower numbers are more desirable (A lower limit is zero).

3.1 When Sales Data From 1 to *t* Are Available for Return Quantity at *t*. ARIMA models for time $t = 51$–60 were estimated using the autocorrelation function (ACF) plot, the augmented Dickey–Fuller test, and the extended ACF table [15]. For example, the selected model for $t = 51$ was an ARIMA (1,1,0), which was also the result of the automatic forecasting algorithm [12]. An ACF and a unit root test of the residuals of the fitted ARIMA model showed that the residuals were independent so that the fitted model provided an adequate fit for the data. An ARIMA (1,1,0) was selected for $t = 51$–57, and an ARIMA (2,1,1) was selected for $t = 58$–60. The prediction performance measure in Eq. (1) was calculated in Table 3.

DLMs were also estimated according to the predictive model selection algorithm. Based on the DLM of Goh and Varaprasad [6], DLMs with lag 0, 1, and 2 of sales (i.e., X_t, X_{t-1}, X_{t-2}) were fitted for time $t = 51$–60, and the prediction performance measure was calculated in Table 3. The ARIMA model provided a better prediction performance than the DLM. One possible explanation is that unlike the data at $t = 1$–50 in Fig. 3, the data from $t = 51$–60 show more gaps between Y_t and X_t, which can be viewed as weak or time-varying relations.

The last candidate model is the mixed model. ARIMA models with lag 0, 1, and 2 of sales were estimated. An ARIMA (1,0,0) was selected for $t = 51$ and 52; an ARIMA (0,1,1) for $t = 53$, 59, and 60; an ARIMA (0,1,2) for $t = 54$, 55, and 58; and an ARIMA (1,1,1) for $t = 56$ and 57. The mixed model could not improve the prediction performance measure.

In summary, when sales data at $t = n$ were available for return quantity at $t = n$ with the real reusable bottles data, the proposed algorithm selected the ARIMA model as the best predictive model since the ARIMA model generated the lowest errors.

3.2 When Sales Data From 1 to $t-1$ Are Available for Return Quantity at *t*. The proposed algorithm was applied to the second scenario similar to the first scenario. Since the ARIMA model did not use sales data, it remained the same. After fitting DLMs, the prediction performance measure was calculated as shown in Table 3. The result was worse than the DLM in Sec. 3.1 as the important information (x_{n+1}) was lost.

The interesting result was the mixed model. Table 3 shows that the performance result of the mixed model outperformed that of the ARIMA model. Unlike the mixed models in Sec. 3.1, an ARIMA (0,1,0) with lag 0, 1, and 2 of sales was used mainly (except for an ARIMA (0,1,1) for $t = 53$).

In summary, when sales data at $t = n$ were not available for return quantity at $t = n$ with the real reusable bottles data, the proposed algorithm selected the mixed model as the best predictive model.

3.3 When Strong Lag Impacts Exist. For this scenario, return series were generated based on the real sales data of reusable bottles and the function $0.6X_{t-1} + 0.4X_{t-2} +$ Random error $[-100, 100]$. The ARIMA, DLM, and mixed model were applied to the new data, and the results are shown in Table 3. The DLM provided the better prediction performance than the ARIMA model. Based on the AIC, the mixed model chose the same DLM (i.e., no ARIMA error). When the simulated data had strong and static relationships between returns and sales, the proposed algorithm selected the DLM as the best predictive model. As it is expected, if design engineers have the lag pattern information of the target system from subject-matter experts or economic theories, the DLM can work well for the forecast of product returns.

4 Discussion

In Sec. 3, the predictive model selection algorithm could suggest a better predictive model than the DLM. In the first and second scenarios with the real data, the ARIMA model and the mixed model were selected as the best models. This section discusses the application of the algorithm in remanufacturing and the three issues (required sample size, performance of mixed models, and different time series models) in the predictive models.

4.1 Application in Remanufacturing. The prediction performance measure in Eq. (1) does not include the cost coefficients of overestimation and underestimation since the cost coefficients do not affect the model fitting procedure. If the cost coefficients are used in the prediction performance measure, the measure can provide misleading information of performance (e.g., higher prediction errors might lead to lower additional costs). Once the final model is available, the real cost with the cost coefficients can be calculated. When the cost coefficients are known in advance, the follow-up question is whether the final model can be modified to reduce the real cost. This is one of future works.

In remanufacturing operations, design problems can be formulated [4,5]. For example, different specifications of computers can be remanufactured with a combination of different cores, which will provide different market shares. Design problems can be solved with the sequential or simultaneous consideration of supply chains and products.

4.2 Some Issues in Predictive Models. The first issue is the required sample size for predictive models. It definitely depends on data but generally the univariate time series model (ARIMA) requires more data than the DLM because the DLM can utilize other available data (i.e., sales series). Goh and Varaprasad [6] recommended the data length of 50 for the reusable bottles data in Sec. 3 when time series models were built.

The second issue is whether the mixed model is the best model in the predictive model selection algorithm since it combines the other two models. Table 3 shows that the results from the mixed model can be worse than the other models. This indicates that when the mixed model is fitted, it does not necessarily find the optimal mixture of the other models.

The third issue is whether there are other time series models other than the ARIMA model to improve the result. Depending on the data, other variants of the ARIMA model can be used [16]. For example, the autoregressive conditional heteroscedastic/generalized autoregressive conditional heteroscedastic models can be used for heteroscedastic (nonconstant variance) errors as the name indicates, which is popular in finance. For a long memory (long-range dependence) model, the autoregressive fractionally integrated moving average model can be used. If there exist complex seasonal patterns such as high-frequency seasonality, multiple seasonal periods, noninteger seasonality, and dual-calendar effects, a modeling framework by De Livera et al. [17] can be used.

5 Conclusion

This paper proposed the predictive model selection algorithm to deal with the uncertainty in closed-loop supply chains by identifying a better predictive model. The predictive model selection algorithm focuses on the predictive power over new or future returns and extends the set of candidate models that should be considered. The proposed algorithm can provide a new insight for forecasting returns since many researchers in the literature assumed that there were always relatively strong and static relationships between returns and sales. The case study of reusable bottles also showed that the predictive model selection algorithm could find a better predictive model than the DLM.

In the future, more data can be tested with the predictive model selection algorithm. Note that return series is usually considered as count data so that the Gaussian ARIMA model may be not suitable for very low counts. However, when counts are large (e.g., the reusable bottles data in this paper), a Gaussian distribution generally provides a good fit to the data.

Nomenclature

A_t = volume of cores from core brokers
ARIMA = autoregressive (AR) integrated (I) moving average (MA)
ARIMA(p, d, q) = p order of AR part, d degree of differencing, q order of MA part
B = backward shift operator
c_o = cost coefficient of overestimation
c_u = cost coefficient of underestimation
D_t = monthly orders or demands
DLM = distributed lag model
e_t = forecast errors (white noise)
I_t = volume of cores from inventory
M_t = volume of cores from other manufacturers
n = index of current time
u_t = error term in a mixed model
X_t = observations of product sales
Y_t = observations of remanufacturable returned products or cores
$\hat{Y}_t$ = predicted volume of cores
$\hat{Y}_{t|t-1}$ = one-step ahead time forecast based on Y_{t-1} and X_{t-1}
α = intercept in DLM
β = distributed lag weight in DLM
γ = coefficient of AR model for u_t
δ = coefficient of MA model for u_t
η = regression coefficient in a mixed model
θ = coefficient of MA model
ϕ = coefficient of AR model

References

[1] Toktay, B., 2004, "Forecasting Product Returns," *Business Aspects of Closed-Loop Supply Chains*, V. Guide, and L. Wassenhove, eds., Carnegie Mellon University Press, Pittsburgh, PA, pp. 203–209.
[2] Clottey, T., Benton, W. C., and Srivastava, R., 2012, "Forecasting Product Returns for Remanufacturing Operations," Decis. Sci., **43**(4), pp. 589–614.
[3] Krapp, M., Nebel, J., and Sahamie, R., 2013, "Forecasting Product Returns in Closed-Loop Supply Chains," Int. J. Phys. Distrib. Logist. Manage., **43**(8), pp. 614–637.
[4] Ma, J., Kwak, M., and Kim, H. M., 2014, "Demand Trend Mining for Predictive Life Cycle Design," J. Cleaner Prod., **68**, pp. 189–199.
[5] Ma, J., and Kim, H. M., 2014, "Continuous Preference Trend Mining for Optimal Product Design With Multiple Profit Cycles," ASME J. Mech. Des., **136**(6), p. 061002.
[6] Goh, T. N., and Varaprasad, N., 1986, "A Statistical Methodology for the Analysis of the Life-Cycle of Reusable Containers," IIE Trans., **18**(1), pp. 42–47.
[7] Toktay, L. B., Wein, L. M., and Zenios, S. A., 2000, "Inventory Management of Remanufacturable Products," Manage. Sci., **46**(11), pp. 1412–1426.
[8] Clottey, T., and Benton, W., 2014, "Determining Core Acquisition Quantities When Products Have Long Return Lags," IIE Trans., **46**(9), pp. 880–893.
[9] Shmueli, G., 2011, "To Explain or to Predict?," Stat. Sci., **25**(3), pp. 289–310.
[10] Carrasco-Gallego, R., and Ponce-Cueto, E., 2009, "Forecasting the Returns in Reusable Containers' Closed-Loop Supply Chains. A Case in the LPG Industry," 3rd International Conference on Industrial Engineering and Industrial Management, Barcelona-Terrassa, Spain, Sept. 2–4, pp. 311–320.
[11] Hyndman, R., and Athanasopoulos, G., 2013, "Forecasting: Principles and Practice," Last accessed Jan. 20, 2014, https://www.otexts.org/fpp
[12] Hyndman, R., and Khandakar, Y., 2008, "Automatic Time Series Forecasting: The Forecast Package for R," J. Stat. Software, **27**(3), pp. 1–22.
[13] Almon, S., 1965, "The Distributed Lag Between Capital Appropriations and Expenditures," Econometrica, **33**(1), pp. 178–196.
[14] Koyck, L., 1954, Distributed Lags and Investment Analysis (Contributions to Economic Analysis), 4th ed., North-Holland Publishing Company, Amsterdam, The Netherlands.
[15] Cryer, J., and Chan, K., 2008, Time Series Analysis: With Applications in R, (Springer Texts in Statistics), Springer, Berlin.
[16] Holan, S. H., Lund, R., and Davis, G., 2010, "The ARMA Alphabet Soup: A Tour of ARMA Model Variants," Stat. Surv., **4**(0), pp. 232–274.
[17] De Livera, A., Hyndman, R., and Snyder, R., 2011, "Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing," J. Am. Stat. Assoc., **106**(469), pp. 1513–1527.