Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

ESTIMATION AND INFERENCE FOR NON-STATIONARY ARRIVAL MODELS WITH A LINEAR TREND

Peter W. Glynn

Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, USA

Zeyu Zheng

Department of Industrial Engineering
and Operations Research
University of California, Berkeley
Berkeley, CA 94720, USA

ABSTRACT

This paper is concerned with building statistical models for non-stationary input processes with a linear trend. Under a Poisson assumption, we investigate the use of the maximum likelihood (ML) method to estimate the model and establish limiting behavior for the ML estimator in an asymptotic regime that naturally arises in applications with high-volume inputs. We also develop likelihood ratio tests for the presence of a linear trend and discuss the asymptotic efficiency. Change-point detection procedures are discussed to identify an unknown point when the model switches from a stationary mode to non-stationarity with a linear trend. Numerical experiments on an e-commerce data set are included. Incorporating a linear trend into an input model can improve prediction accuracy and potentially enhance associated performance evaluations and decision making.

1 INTRODUCTION

Simulation is often used as a planning tool to help determine optimal production capacities or to help assess the number of staffing representatives needed at a service operations facility. In such settings, there often is an observed time series available to help statistically model the exogenous inputs to the simulation model. For example, the time series may describe historical demand for a product or historical arrivals at the service facility. When the time series is of long duration, the time series may exhibit some non-stationarity. This paper is concerned with fitting the simplest such models, namely in describing time series in which the non-stationarity is governed by a linear trend model.

In particular, many discrete-event simulation models are fed by point processes describing customer arrivals, customer demands, etc. In such settings, a natural starting point is to assume that the underlying point process is a non-stationary Poisson process with an affine intensity function. In this paper, we discuss some of the key statistical issues that arise in building such statistical models. First, we use the maximum likelihood (ML) method to estimate parameters for the linear trend model, which naturally handles the heteroskedasticity due to the non-stationarity. We study the arising optimization problem when the ML method is applied. One challenge is that there is generally no closed form optimizer of the likelihood function. We show that the optimization problem is convex and identify mild conditions for the uniqueness of the ML optimizer.

The study of the limiting behavior for the ML estimator and its associated tests has been focused on the asymptotic regime in which the sample size goes to infinity, a regime that is particularly popular in the contexts where the observed samples are independent and identically distributed (iid); see Lehmann and Casella (2006). In the non-stationary modeling context with a linear trend in presence, we consider an alternative new asymptotic regime in which the magnitude of the baseline intensity goes to infinity, while

978-1-7281-3283-9/19/$31.00 ©2019 IEEE

the relative change in intensity due to the trend is fixed. This regime is especially suitable in scenarios when the observed input is high-volume.

The ML formulation permits the construction of confidence intervals, goodness-of-fit tests, and hypothesis tests. In particular, we establish the limiting behavior (consistency and central limit theorem) for the ML estimator in the new asymptotic regime. Asymptotically valid confidence intervals can therefore be constructed for the ML estimators. Section 3 discusses a likelihood ratio (LR) test on whether the observed data supports the hypothesis of a statistically significant linear trend. We argue that the proposed LR test has optimal efficiency in the new asymptotic regime. Massey et al. (1996) also consider the ML estimation problem for a Poisson model with linear rate, with a focus on comparing the ML method to least square approaches in terms of estimation efficiency. Zheng and Glynn (2017) discuss the ML estimation problem for a nonhomogeneous Poisson process with a continuous piecewise linear intensity and solve the estimation problem via convex optimization. See also Chapter 6 of Nelder and Wedderburn (1972) for a comprehensive study of estimation problems for Poisson models. Kuhl et al. (1997) is the first paper to address the trend estimation issue in Poisson modeling and simulation. They discuss estimation and simulation methods for Poisson processes with trend and cyclic features; their definition of trend is based on the cumulative intensity function, which is different from ours and therefore the two methods fit different sets of applications.

For scenarios in which the input time series switches from a stationary model to a non-stationary linear trend model at an unknown change point, we provide in Section 4 an estimation and inference procedure to identify and validate such a change point. The limiting distribution of the test statistic is provided. Akman and Raftery (1986) and subsequent papers discuss inference procedures for a change point in Poisson processes that involves a single-shot change to the intensity model (e.g., a switch between two constant rates). Perry et al. (2006) discuss the estimation of the change point in the same setting as ours, but no inference or limiting behavior were covered.

Our main contributions in this paper are:

- Development of a convex formulation of the maximum likelihood (ML) estimator for a Poisson model with linear trend, as well as analysis of the uniqueness of the ML solution; see Section 2.
- Analysis of the limiting behavior (consistency and central limit theorem) of the ML estimator when the baseline intensity tends to infinity; see Section 2.
- Construction of a likelihood ratio test for the significance of a linear trend, as well as discussion on its asymptotic efficiency; see Section 3.
- Development of change-point estimation and inference when the intensity process switches from stationary to non-stationary with a linear trend; see Section 4.
- A numerical study of the ML estimator and the test for trend implemented on a real-world data set; see Section 5.

2 A POISSON MODEL WITH LINEAR TREND

Suppose that we observe a sequence of arrival counts in m consecutive time intervals (not necessarily equally-spaced), namely $N_1, N_2, \dots, N_m$. With a linear trend in presence, assume that N_i's are independent Poisson random variables (rv's), with

$$\mathbb{E}N_i = \lambda(1+b^* t_i),$$

in which $\lambda > 0$ is the base rate parameter and b^* is the slope parameter. Note that the t_i's are fixed instruments and without loss of generality we assume that

$$0 \le t_1 < t_2 < \cdots < t_m.$$

For example, if $t_i = i-1$, the model can be used on arrival count data that are observed over consecutive operational periods (e.g., days). Therefore N_i can model the arrival count on the i-th day. If $t_1 = 0$, the relative trend over the full time horizon is given by

$$\frac{\mathbb{E}N_m - \mathbb{E}N_1}{\mathbb{E}N_1} = b^* t_m.$$

2.1 Maximum Likelihood Estimation

We estimate λ and b^* using the maximum likelihood (ML) method. The likelihood function is given by

$$L(\gamma, b) = \prod_{i=1}^{m} e^{-\gamma(1+bt_i)} \frac{(\gamma(1+bt_i))^{N_i}}{N_i!}$$

and the log-likelihood function is

$$\mathscr{L}(\gamma, b) = \log L(\gamma, b) = -\sum_{i=1}^{m} (\gamma(1+bt_i)) + \sum_{i=1}^{m} N_i \log(\gamma(1+bt_i)).$$

The maximum likelihood estimator $\hat{\lambda}$, $\hat{b}$ is therefore given by the solution to the optimization problem

$$\begin{aligned} & \max_{\gamma,b} \mathscr{L}(\gamma, b) \\ & \text{s.t. } \gamma \geq 0 \\ & \quad 1 + bt_i \geq 0, \text{ for } 1 \leq i \leq m. \end{aligned} \tag{1}$$

Note that the gradient of $\mathscr{L}$ is given by

$$\begin{pmatrix} \frac{\partial \mathscr{L}}{\partial \gamma} \\ \frac{\partial \mathscr{L}}{\partial b} \end{pmatrix} = \begin{pmatrix} -\sum_{i=1}^{m}(1+bt_i) + \sum_{i=1}^{m} N_i \frac{1}{\gamma} \\ -\sum_{i=1}^{m} \gamma t_i + \sum_{i=1}^{m} \frac{N_i t_i}{1+bt_i} \end{pmatrix}.$$

So, if the following equation has a unique solution that satisfies the non-negativity constraints as given in (1), that solution will be the ML estimator.

$$\begin{aligned} \hat{\lambda} &= \frac{\sum_{i=1}^{m} N_i}{\sum_{i=1}^{m}(1+\hat{b}t_i)} \\ \frac{\sum_{i=1}^{m} t_i \sum_{i=1}^{m} N_i}{\sum_{i=1}^{m}(1+\hat{b}t_i)} &= \sum_{i=1}^{m} \frac{N_i t_i}{1+\hat{b}t_i}. \end{aligned} \tag{2}$$

In cases where the solution to (2) does not satisfy the non-negativity constraints, we provide a uniqueness result for the general ML optimization problem (1).

Theorem 1 If $\sum_{i=1}^{m} 1_{\{N_i>0\}} \geq 2$, the ML optimization problem (1) is equivalent to a convex optimization problem and has a unique solution.

2.2 Limiting Behavior of the Maximum Likelihood Estimator

We next derive the limiting behavior of the ML estimators when λ^* tends to infinity. This asymptotic regime naturally arises in the contexts where the volume of input processes is high.

Theorem 2 Suppose that $b > 0$. Denote $Z_1, Z_2, \cdots, Z_m$ as a sequence of iid standard normal random variables. When $\lambda \to \infty$,

$$\begin{aligned} \frac{\hat{\lambda}}{\lambda} &\to 1 \\ \hat{b} - b^* &\to 0 \end{aligned}$$

with probability one, and

$$\begin{pmatrix} 1 & \frac{\sum_{i=1}^m t_i}{\sum_{i=1}^m (1+b^* t_i)} \\ 1 & \frac{1}{\sum_{i=1}^m t_i}\sum_{i=1}^m \frac{t_i^2}{1+b^* t_i} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda}(\frac{\hat{\lambda}}{\lambda}-1) \\ \sqrt{\lambda}(\hat{b}-b^*) \end{pmatrix} \Rightarrow \begin{pmatrix} \frac{\sum_{i=1}^m \sqrt{1+b^* t_i}\cdot Z_i}{\sum_{i=1}^m (1+b^* t_i)} \\ \frac{1}{\sum_{i=1}^m t_i}\sum_{i=1}^m \frac{t_i \cdot Z_i}{\sqrt{1+b^* t_i}} \end{pmatrix}.$$

We discuss briefly the outline of the proof and defer the full proof to an extended version of this paper. The consistency of the ML estimator is mainly due to the uniqueness of ML optimizer (so that the local optimizer is also global), that the distribution of $(N_1, N_2, \cdots, N_m)$ is distinct for different parameters, and that the true value of parameter b lies in the interior of the parameter space. The central limit theorem is built on the consistency of the ML estimator and on the smoothness conditions of the log-likelihood function. Suppose that the ML estimator satisfies the gradient condition (2), we have

$$\hat{\lambda} - \lambda - \Big(\frac{\sum_{i=1}^m N_i}{\sum_{i=1}^m (1+\hat{b}t_i)} - \frac{\sum_{i=1}^m N_i}{\sum_{i=1}^m (1+b^* t_i)}\Big) - \Big(\frac{\sum_{i=1}^m N_i}{\sum_{i=1}^m (1+b^* t_i)} - \lambda\Big) = 0$$

$$\hat{\lambda} - \lambda - \Big(\frac{1}{\sum_{i=1}^m t_i}\sum_{i=1}^m \frac{N_i t_i}{1+\hat{b}t_i} - \frac{1}{\sum_{i=1}^m t_i}\sum_{i=1}^m \frac{N_i t_i}{1+b^* t_i}\Big) - \Big(\frac{1}{\sum_{i=1}^m t_i}\sum_{i=1}^m \frac{N_i t_i}{1+b^* t_i} - \lambda\Big) = 0.$$

This set of equations, is approximately (intended to carry no rigorous meaning in this proof outline) equivalent to

$$\frac{1}{\sqrt{\lambda}}(\hat{\lambda}-\lambda) + \Big(\frac{1}{\lambda}\frac{\sum_{i=1}^m N_i \sum_{i=1}^m t_i}{(\sum_{i=1}^m (1+b^* t_i))^2}\Big)\sqrt{\lambda}(\hat{b}-b^*) = \frac{1}{\sqrt{\lambda}}\Big(\frac{\sum_{i=1}^m N_i}{\sum_{i=1}^m (1+b^* t_i)} - \lambda\Big)$$

$$\frac{1}{\sqrt{\lambda}}(\hat{\lambda}-\lambda) + \Big(\frac{1}{\lambda}\frac{1}{\sum_{i=1}^m t_i^2}\sum_{i=1}^m \frac{N_i t_i}{(1+b^* t_i)^2}\Big)\sqrt{\lambda}(\hat{b}-b^*) = \frac{1}{\sqrt{\lambda}}\Big(\frac{1}{\sum_{i=1}^m t_i}\sum_{i=1}^m \frac{N_i t_i}{1+b^* t_i} - \lambda\Big).$$

Note that the coefficients of $\sqrt{\lambda}(\hat{b}-b^*)$ in both equations converge almost surely respectively to $\sum_{i=1}^m t_i$ and $\frac{1}{\sum_{i=1}^m t_i}\sum_{i=1}^m \frac{t_i^2}{1+b^* t_i}$ as $\lambda \to \infty$. The right-hand-side of the two equations is approximately normal. For instance,

$$\frac{1}{\sqrt{\lambda}}\Big(\frac{\sum_{i=1}^m N_i}{\sum_{i=1}^m (1+b^* t_i)} - \lambda\Big) = \frac{1}{\sum_{i=1}^m (1+b^* t_i)}\Big(\sum_{i=1}^m \frac{(N_i - \mathbb{E}N_i)}{\sqrt{\mathrm{Var}(N_i)}}\sqrt{1+b^* y_i}\Big),$$

which is asymptotically normal at a proper scale as $\lambda \to \infty$.

We next discuss the construction of confidence intervals using the central limit theorem in Theorem 2. Let

$$\Sigma_1 = \begin{pmatrix} 1 & \frac{\sum_{i=1}^m t_i}{\sum_{i=1}^m (1+b^* t_i)} \\ 1 & \frac{1}{\sum_{i=1}^m t_i}\sum_{i=1}^m \frac{t_i^2}{1+b^* t_i} \end{pmatrix}, \Sigma_2 = \begin{pmatrix} \frac{1}{\sum_{i=1}^m (1+b^* t_i)} & \frac{1}{\sum_{i=1}^m (1+b^* t_i)} \\ \frac{1}{\sum_{i=1}^m (1+b^* t_i)} & \Big(\frac{1}{\sum_{i=1}^m t_i}\Big)^2 \sum_{i=1}^m \frac{t_i^2}{1+b^* t_i} \end{pmatrix}.$$

The limiting joint distribution of

$$\begin{pmatrix} \sqrt{\lambda}(\frac{\hat{\lambda}}{\lambda}-1) \\ \sqrt{\lambda}(\hat{b}-b^*) \end{pmatrix}$$

is

$$\mathcal{N}\big((0,0)^\top, \Sigma_1^{-1}\Sigma_2(\Sigma_1^{-1})^\top\big).$$

It is easy to notice that Σ_1 is invertible when $b^* > 0$. With the central limit theorem in hand, we can conveniently construct asymptotically valid confidence intervals for λ, b^* or confidence regions for (λ, b^*). For example, denoting $\tilde{\Sigma} = \Sigma_1^{-1}\Sigma_2(\Sigma_1^{-1})^\top$, we can construct an asymptotic $100(1-\alpha)\%$ confidence interval

$$\left[\hat{b} - z\sqrt{\frac{\tilde{\Sigma}_{11}}{\hat{\lambda}}}, \hat{b} + z\sqrt{\frac{\tilde{\Sigma}_{11}}{\hat{\lambda}}}\right]$$

in which z is selected as $P(-z \le N(0,1) \le z) = 1-\alpha$.

3 LIKELIHOOD RATIO TEST FOR THE SIGNIFICANCE OF TREND

In this section, we consider the construction of a likelihood ratio statistic for the significance of a linear trend, namely a test of $b^* = 0$ versus $b^* \neq 0$. When $b^* \neq 0$, the underlying model indeed contains a linear trend, while when $b^* = 0$, the underlying model is stationary. Fitting a Poisson model with linear trend to an observed time series of data often gives us a non-zero estimator $\hat{b}$. The question is whether or not the estimated non-zero linear trend is statistically meaningful. If the test however shows that the linear trend is not significant, then a stationary model could be sufficient and may also reduce a possible over-fitting.

This hypothesis test aims at inferring from the data whether there is a significant trend. The test statistic T is constructed by

$$T = 2\log\Big(\frac{\max_{\gamma,b} L(\gamma,b)}{\max_{\gamma} L(\gamma,0)}\Big).$$

The likelihood ratio test rejects the null hypothesis for a large value of T. We next show the limiting distribution for the test statistic under the null hypothesis.

Theorem 3 Under the null hypothesis $b^* = 0$, when $\lambda \to \infty$,

$$T \Rightarrow \chi_1^2$$

in which χ_1^2 denotes the chi-squared distribution with one degree of freedom.

The proofs of Theorem 3 borrows ideas from Chernoff (1954) and Lehmann and Romano (2006). With Theorem 3 in hand, we can choose the critical value z_α for the likelihood ratio test at confidence level α such that $P(\chi_1^2 > z_\alpha) = \alpha$.

We next discuss the optimality of the likelihood ratio test as constructed above. Uniformly most powerful tests (UMP) and its variants optimality principles UMP unbiased and UMP invariant are almost restricted to one-parameter families, which do not fit our case. We consider another optimality principle, the *scale asymptotically uniformly most powerful* (SAUMP).

Definition 1 For testing $b^* = 0$ versus $b^* > 0$, a sequence of tests $\{\phi_{\lambda^*}\}$ indexed by $\{\lambda^* \in \mathbb{Z}^+\}$ is called *scale asymptotically uniformly most powerful* (SAUMP) at asymptotic level α if $\limsup_{\lambda^* \to \infty} \mathbb{E}_{b^*=0}(\phi_{\lambda^*}) \leq \alpha$ under the null hypothesis and if for other sequence of test functions $\{\psi_{\lambda^*}\}$ satisfying $\limsup_{\lambda^* \to \infty} \mathbb{E}_{b^*=0}(\psi_{\lambda^*}) \leq \alpha$,

$$\limsup_{\lambda^* \to \infty} \sup\{\mathbb{E}_{b^*}(\psi_{\lambda^*}) - \mathbb{E}_{b^*}(\phi_{\lambda^*}) \,:\, b^* > 0\} \leq 0.$$

This definition of asymptotic optimality is analogous to that of *asymptotically uniformly most powerful* (AUMP). See section 13.3 of Lehmann and Romano (2006) for reference. In AUMP, the asymptotic limit is considered when the size m of iid samples goes to infinity, while in SAUMP, the asymptotic limit is considered when the scale λ^* is sent to infinity.

Theorem 4 The likelihood ratio test based on the test statistic T is SAUMP.

There are scenarios in which the one-sided test

$$H_0 \,:\, b^* = 0, \quad \text{vs} \quad H_1 \,:\, b^* > 0$$

is preferred over the two-sided test. In these scenarios, it is of interest to test whether there is a statistically significant positive linear trend. For this one-sided test, we construct the associated likelihood ratio statistic

$$T^o = 2\log\Big(\frac{\max_{\gamma,b\geq 0} L(\gamma,b)}{\max_{\gamma} L(\gamma,0)}\Big).$$

Theorem 5 Under the null hypothesis $b^* = 0$, when $\lambda^* \to \infty$,

$$T^o \Rightarrow W$$

in which W is a rv that takes value 0 with probability $1/2$ and follows the χ_1^2 distribution with probability $1/2$.

4 CHANGE POINT DETECTION FOR A LINEAR TREND

Let $N_1, N_2, \ldots, N_m$ be a sequence of independent Poisson random variables with rate parameters $\mu_1, \mu_2, \ldots, \mu_m$. This section derives estimation and test for a change point, at which the Poisson rates change from a constant sequence to a sequence with linear trend. Specifically, the model is specified by

$$\mu_1 = \lambda^*,\ \frac{\mu_i}{\mu_1} = \begin{cases} 1 & i \le \tau^* \\ 1 + b^*(i - \tau^*) & \tau^* < i \le m \end{cases},$$

in which $1 \le \tau^* \le m$, λ^*, b^* are unknown positive parameters. The likelihood function is given by

$$L^C(\tau, \gamma, b) = \prod_{i=1}^{\tau} e^{-\gamma} \frac{\gamma^{N_i}}{N_i!} \prod_{i=\tau+1}^{m} e^{-\gamma(1+b(i-\tau))} \frac{(\gamma(1 + b(i-\tau)))^{N_i}}{N_i!}.$$

The maximum likelihood estimators $\hat{\tau}$, $\hat{\lambda}$, $\hat{b}$ are given by

$$\max_{\tau,\gamma,b} L^C(\tau, \gamma, b).$$

We construct a likelihood ratio statistic to test the hypothesis

$$H_0 : \frac{\mu_i}{\mu_1} = 1,\ 1 \le i \le m \quad \text{vs} \quad H_1 : \frac{\mu_i}{\mu_1} = \begin{cases} 1 & i \le \tau^* \\ 1 + b^*(i - \tau^*) & \tau^* < i \le m \end{cases}, \text{ for some } \tau^* < m \text{ and } b^* > 0.$$

Denote the likelihood ratio R as

$$R = \frac{L^C(\hat{\tau}, \hat{\gamma}, \hat{b})}{\max_\gamma L^C(m, \gamma, 0)}.$$

The tests rejects the null hypothesis for large values of R. Given the data, for small intensities, a bootstrapping procedure can be employed to generate the confidence region of the test. For large intensities, we show the following result.

Theorem 6 Under the null hypothesis,

$$2\log(R) \Rightarrow \max_{1 \le j \le m-1} \frac{(\sum_{i=j+1}^{m} (i-j) Z_i)^2}{\sum_{i=j+1}^{m} (i-j)^2},$$

as $\lambda \to \infty$, where $Z_1, Z_2, \cdots, Z_m$ are iid $\mathcal{N}(0,1)$ random variables.

5 EXPERIMENTS

We show implementations of the methods developed in this paper on a real data set.

5.1 Estimation and Tests

To illustrate the estimation and test procedures, we implement the proposed procedures on a data set from a large online e-commerce company. Thousands of products are sold on the online platform. We focus on the estimation and test for the trend in the daily demand time series for some of the best-selling products. The daily demand for these products varies from the level of 50 to $1,000$. The time range of the data is a half year, and in the data pre-processing holidays are viewed as outliers and removed simultaneously for all products. The length of the demand time series is $m = 161$ days after the data pre-processing.

We present results in Table 1 for the demand sequence of six representative products. For each product, the size of demand sequence is $m = 161$. Therefore, in our model formulation, $N_1, N_2, \cdots, N_m$ are given by the data, and the time stamps $t_1, t_2, \cdots, t_m$ are given by the actual time sequence starting by $t_1 = 1$. We report the estimators $(\hat{\lambda}, \hat{b})$, the estimated cumulative trend across the full time horizon $t_m \cdot \hat{b}$, and the test statistic T for the significance of trend. The 95% quantile of the χ_1^2 distribution 3.8415 is provided as the critical value. Specifically, $P(\chi_1^2 > 3.8415) \approx 0.05$.

Figure 1 to Figure 6 visualizes the daily demand sequence for Product 1 to Product 6. The tests suggest that the demand time series for Product 1 and 2 is stationary, in the sense that a linear trend is not significant in the data. The tests also suggest that a linear trend is a significant modeling element of the demand time series for Product 3 to Product 6.

Table 1: Results for parameter estimation and significance tests.

	Base-Rate $\hat{\lambda}$	Slope $\hat{b}$	Cumulative Trend $\hat{b} \cdot t_m$	LR Test Statistic T	Trend Significance $T > 3.84$?
Product 1	74.85	1.88×10^{-5}	0.31%	9.10×10^{-6}	Not Significant
Product 2	66.06	-2.6×10^{-4}	−4.2%	0.0033	Not Significant
Product 3	78.40	0.028	466%	22.04	Significant
Product 4	149.12	0.013	209%	46.37	Significant
Product 5	625.51	−0.0046	−75%	28.02	Significant
Product 6	466.20	−0.0051	−85%	136.70	Significant

5.2 Out-of-sample Prediction Performances

In this subsection, we compare the out-of-sample prediction performance using our non-stationary Poisson model with a linear trend to that using a standard Poisson model without trend. Note that the full demand time series is $N_1, N_2, \cdots, N_m$ with $m = 161$. We presume that one has observed the demand for the first k days and predicts the demand for the next l days, i.e., the $k+1, k+2, \ldots, k+l$-th day. For a given $k \in [90, m-l]$, we choose a look-back window of w days to fit our model with a linear trend. Specifically, the demand data in days $k-w+1, k-w+2, \ldots, k$ are used to fit the model and infer the parameters. The look-back window size is fixed such that for the prediction task for each $k \in [w, m-l]$, the size of data used is the same. Denote the fitted parameters as $\hat{\lambda}_k$ and $\hat{b}_k$ for a given $k \in [w, m-l]$. Then we use the fitted model to predict the demand for the next l days, that is, $\hat{N}_{k+1}, \hat{N}_{k+2}, \cdots, \hat{N}_{k+l}$. Specifically,

$$\hat{N}_{k+j} = \hat{\lambda}_k (1 + \hat{b}_k (k+j))$$

for $j = 1, 2, \ldots, l$. The relative prediction error e_{kl} is defined as

$$e_{kl} = \frac{1}{l} \sum_{i=1}^{l} \frac{|\hat{N}_{k+i} - N_{k+i}|}{N_{k+i}}. \tag{3}$$

We then compute the averaged relative prediction error $\hat{e}_l$ by rolling over k from w to $m-l$. That is

$$\hat{e}_l = \frac{1}{m-l-w+1} \sum_{k=1}^{m-l} e_{kl}. \tag{4}$$

In the experiments using our trend model, we fix $w = 90$. Therefore, $\hat{e}_l$ represents the error level for predicting the future l days of demand using our model and input data of last 90 days. If otherwise the trend is ignored, a stationary Poisson model is used to fit the historical data and predict the demand for the future l days of demand. We denote w as the size of look-back window to fit the stationary model. Specifically, if the historical demand in the first k days are available, the demand data that lies in the look-back window

$(N_{k-w+1}, N_{k-w+2}, \cdots, N_k)$ are used to fit the stationary model and predict the future demand. Specifically, the stationary prediction for each of the daily demand on days $k+1, k+2, \ldots, k+l$ is the same, given by

$$\frac{1}{w}\sum_{j=k-w+1}^{k} N_j.$$

The computation of relative prediction error is the same as in (3) and (4).

Table 2 shows the results by selecting $l = 7$, such that the prediction of demand in the future 7 days is of interest. A fixed size of look-back window 90 is selected for our trend model. Three different sizes of look-back windows ($w = 90, 30, 7$) are considered for the stationary model. When $w = 90$, the stationary model uses exactly the same data as the trend model. The choice of $w = 7$ reflects the idea that one only uses very recent data to predict future demands, in order to mitigate nonstationary effects that might present in a longer history. One possible downside of selecting a short look-back window is that the data size is small and may cause statistical instability. The choice of $w = 30$ balances the other two choices. The numerical results show that our trend model has a stronger out-of-sample performance than the stationary model with any of the three choices of the look-back window size. For Product 1 and Product 2, as shown in the table, the out-of-sample prediction performance is comparable for all models. This is consistent with the hypothesis test result that there is no statistically significant trend in the demand data for Product 1 and 2. Meanwhile, the choice of look-back window size does not matter much for the demand prediction of Product 1 and Product 2. For Product 3 to Product 6, the linear trend model demonstrates significantly better out-of-sample prediction power compared to the stationary model with different look-back window sizes. When there is a significant trend present in the data, a small look-back window mitigates some nonstationarity and improves the prediction performance, but cannot match the performance of the Poisson model with a linear trend.

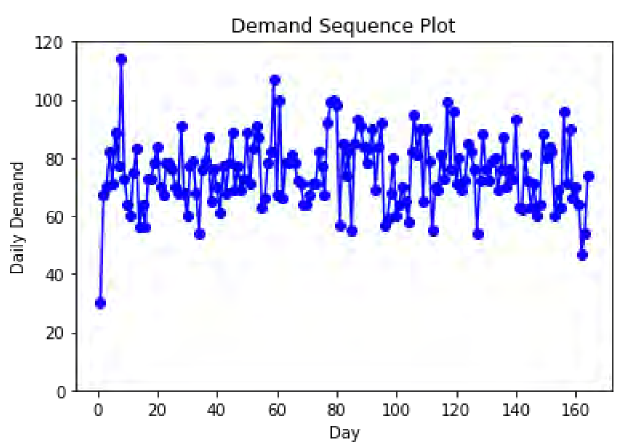

Figure 1: Product 1.

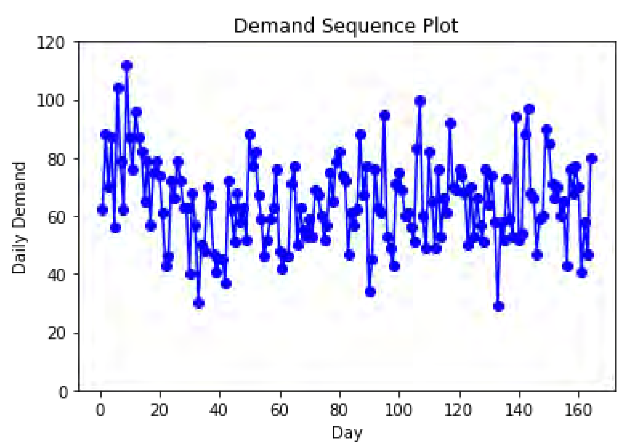

Figure 2: Product 2.

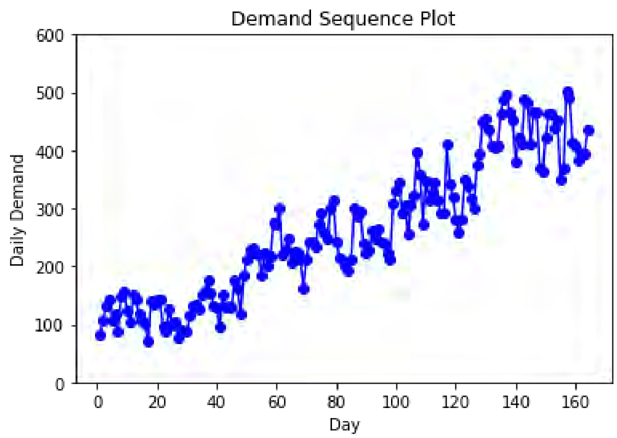

Figure 3: Product 3.

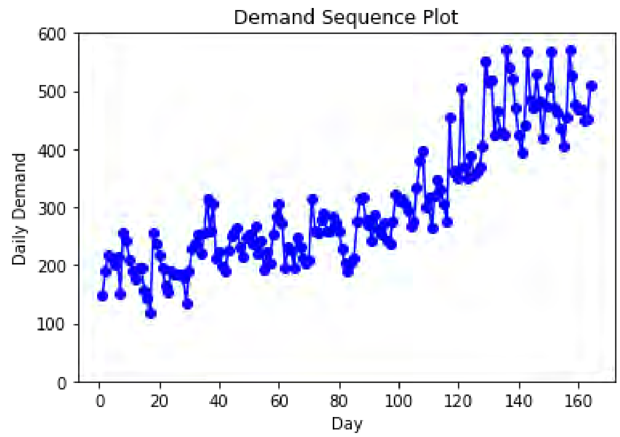

Figure 4: Product 4.

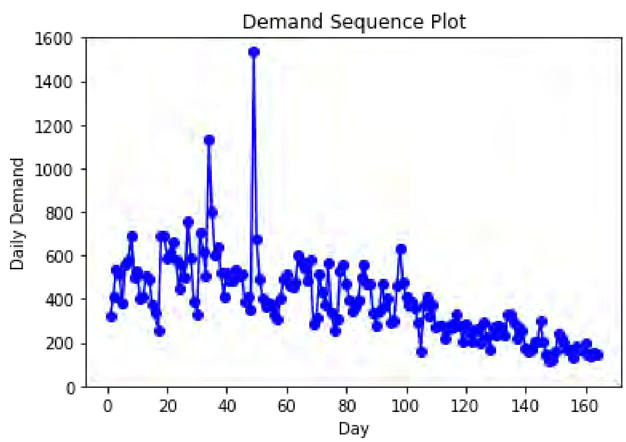

Figure 5: Product 5.

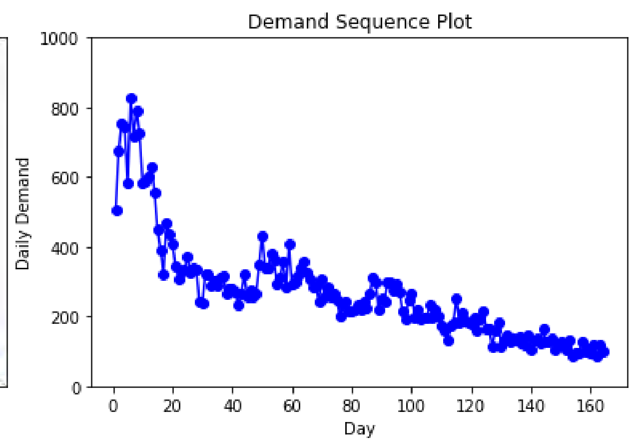

Figure 6: Product 6.

Table 2: Relative errors for out-of-sample prediction with different models.

	Out-of-sample Prediction Relative Error			
	Trend Model	Stationary $w = 7$	Stationary $w = 30$	Stationary $w = 90$
Product 1	0.171	0.144	0.167	0.154
Product 2	0.163	0.158	0.167	0.159
Product 3	0.124	0.203	0.238	0.392
Product 4	0.109	0.127	0.143	0.239
Product 5	0.329	0.386	0.410	0.582
Product 6	0.185	0.217	0.353	0.561

ACKNOWLEDGMENT

We thank the referees for their helpful comments and the track chairs for handling our paper. We also thank Yufeng Zheng for his assistance with data pre-processing.

REFERENCES

Akman, V., and A. Raftery. 1986. "Asymptotic Inference for a Change-point Poisson Process". *The Annals of Statistics* 14(4):1583–1590.

Chernoff, H. 1954. "On the Distribution of the Likelihood Ratio". *The Annals of Mathematical Statistics* 25(3):573–578.

Kuhl, M., H. Damerdji, and J. Wilson. 1997. "Estimating and Simulating Poisson Processes with Trends or Asymmetric Cyclic Effects". In *Proceedings of the 1997 Winter Simulation Conference*, 287–295. IEEE. edited by S. Andradbttir, K. J. Healy, D. H. Withers, and B. I,. Nelson, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lehmann, E. L., and G. Casella. 2006. *Theory of Point Estimation*. New York: Springer-Verlag.

Lehmann, E. L., and J. P. Romano. 2006. *Testing Statistical Hypotheses*. New York: Springer-Verlag.

Massey, W. A., G. A. Parker, and W. Whitt. 1996. "Estimating the Parameters of a Nonhomogeneous Poisson Process with Linear Rate". *Telecommunication Systems* 5(2):361–388.

Nelder, J. A., and R. W. Wedderburn. 1972. "Generalized Linear Models". *Journal of the Royal Statistical Society: Series A (General)* 135(3):370–384.

Perry, M. B., J. J. Pignatiello Jr, and J. R. Simpson. 2006. "Estimating the Change Point of a Poisson Rate Parameter with a Linear Trend Disturbance". *Quality and Reliability Engineering International* 22(4):371–384.

Zheng, Z., and P. W. Glynn. 2017. "Fitting Continuous Piecewise Linear Poisson Intensities via Maximum Likelihood and Least Squares". In *Proceedings of the 2017 Winter Simulation Conference*, 1740–1749. IEEE. edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

PETER W. GLYNN is the Thomas Ford Professor in the Department of Management Science and Engineering (MS&E) at Stanford University. He is a Fellow of INFORMS and of the Institute of Mathematical Statistics, has been co-winner of Best Publication Awards from the INFORMS Simulation Society in 1993, 2008, and 2016, and was the co-winner of the John von Neumann Theory Prize from INFORMS in 2010. In 2012, he was elected to the National Academy of Engineering. His research interests lie in stochastic simulation, queueing theory, and statistical inference for stochastic processes. His email address is glynn@stanford.edu. His website is https://web.stanford.edu/~glynn/.

ZEYU ZHENG is an Assistant Professor in the Department of Industrial Engineering and Operations Research (IEOR) at the University of California, Berkeley. He received his Ph.D. in Management Science and Engineering, Ph.D. minor in Statistics, and M.A. in Economics, from Stanford University in 2018, and B.S. in Mathematics from Peking University in 2012. His has research interests in simulation, stochastic modeling, data-oriented decision making, and financial technologies. His email address is zyzheng@berkeley.edu. His website is https://zhengzeyu.com.