Bias/Variance Trade-Off in Estimates of a Process Parameter Based on Temporal Data

PATRICIA L. COOPER BARFOOT, STEFAN H. STEINER, and R. JOCK MACKAY

University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

We recommend an approach to estimate a process performance measure (or parameter) at the present time from a stream of data where the performance may drift slowly over time. It is common practice to estimate current process performance using either present-time data only or including all historical data. When sample sizes by time period are small, an estimate based only on present-time data is imprecise. When the performance changes over time, including historical data in estimation trades more bias for less variability. We propose to regulate the bias/variance trade-off using estimating equations that down-weight past data. We derive approximations for the variance of the estimator and the distribution of a test statistic involving the estimator. The work is motivated by estimation of a customer loyalty measure where realistic data demonstrates the proposed approach.

Key Words: Net Promoter Score (NPS); Process Monitoring; Weighted Estimating Equations.

1. Motivation and General Problem

DECISION MAKERS responsible for managing the performance of a process often monitor an estimate of a process parameter such as the mean or a rate over time. For example, one popular business-management philosophy prioritizes actions for driving growth around improving a mean value of a customer loyalty measure (Reichheld and Markey (2011)). Customer loyalty data are observed on subgroups of customers at regular time intervals and managers need an estimate of the measure summarizing the present performance. Commonly there is one or more subject-level covariates that have an effect on the subject-level outcome that is observed. We may want to divide the subjects into multiple subgroups of interest, which we refer to as streams. Based on the data observed over time, a manager may want to

- monitor estimates of the current mean value of performance over time. For example, the manager tracks the trend in the customer loyalty measure for customers across various versions of their product in order to validate the impact of design enhancements or to track the measure relative to competitive benchmarks.
- monitor a test of a statistical hypothesis involving present estimates of performance across multiple streams. For example, the manager assesses differences in the present customer loyalty measure between customers from various geographic regions in order to plan future marketing efforts.

In a further example, an agency responsible for the quality of laboratories that perform medical testing monitors estimates of present performance by lab (which we could refer to as a stream) based on data observed from regular operation of the various labs. The agency assesses the proficiency of a lab by tracking the trend in its mean measure of performance and tests hypotheses to compare its performance to its

Dr. Cooper Barfoot is a new graduate from the Department of Statistics and Actuarial Science. Her email address is plcooper@uwaterloo.ca.

Dr. MacKay is Professor Emeritus in the Department of Statistic and Actuarial Science. His email address is rjmackay @uwaterloo.ca.

Dr. Steiner is Chair and Professor in the Department of Statistics and Actuarial Science. He is a Fellow of ASQ. His email address is shsteine@uwaterloo.ca.

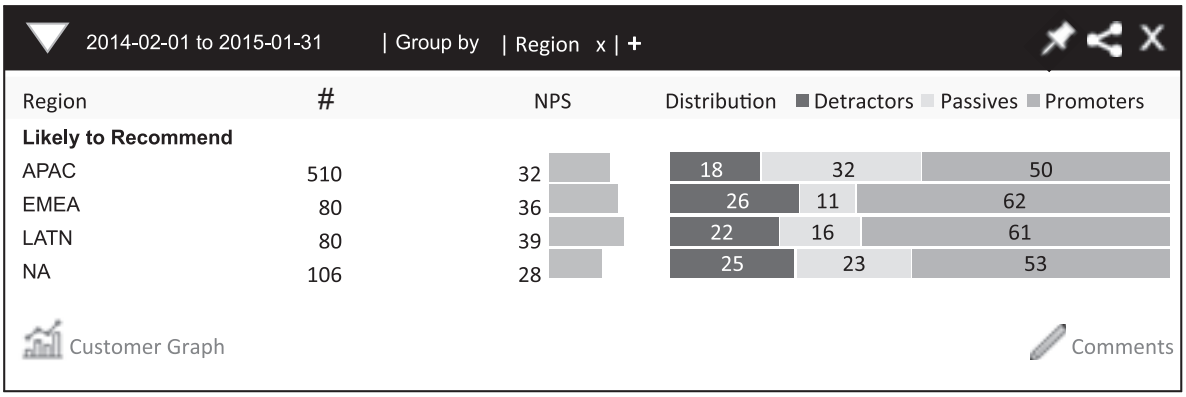

FIGURE 1. Best Practice Customer Loyalty Measure Chart.

peers. Present estimates of the mean performance by lab and a suitable test statistic drive ratings of lab noncompliance, which have important implications for lab operations.

In both examples, the current practice is to select a time period of importance and compare estimates of the parameter or test statistical hypotheses by usual methods based on the observed data from that time period stratified by streams. For example, a consulting company that specializes in best practices in customer loyalty analytics advises companies to estimate present performance through an interface like Figure 1 (Satmetrix (2015)).

Figure 1 shows estimates of the customer loyalty measure known as net promoter score (NPS) based on data observed over one year (2014-02-01 to 2015-01-31) in multiple streams defined by region (APAC: Asia-Pacific, EMEA: Europe/Middle East/Africa, LATN: Latin America, NA: North America). The values under the Distribution heading are the observed proportions of customers who are classified as detractors, passives, and promoters, respectively, based on their response to a customer loyalty survey. The values under the NPS heading are the differences between the observed proportions of promoters and detractors which are the present estimates by region. Sample sizes are given and we see that there are more than six times the number of respondents in APAC relative to either EMEA or LATN. Uncertainty intervals related to size of the samples and bias related to possible changes in performance within the year are not considered. The interface allows the user to change the time period of interest; doing so changes the estimates of the present performance (NPS) and the bias and uncertainty of these estimates. Estimates based on data over a shorter time period have higher uncertainty due to the smaller sample, but less bias when performance drifts over time. There is a bias/variance trade-off in selecting the time period for the data, which is not considered in the current best practice for estimating the performance measure.

These are different problems than usual statistical process monitoring applications where a mean measure or a hypothesis test statistic is monitored relative to an in-control period determined by prior data (Montgomery (2013)). In the problems at hand, decisions are based on an updated estimate of a parameter rather than a comparison to control limits.

We consider the following general data and model. At each time period t, we observe data d_t from a sample of n_t subjects. The complete dataset is $d = \{d_1, \ldots, d_T\}$. Each of the n_t data points in d_t includes a subject-specific outcome response (e.g., the customer's survey response) and may include subject-specific covariates (e.g., the customer's region). We specify a likelihood function $\mathcal{L}_t(d_t; \theta_t)$ to describe the probability of data d_t given the p-dimensional model parameter θ_t. Based on the log-likelihood function $l_t(\theta_t) = \log \mathcal{L}_t(\theta_t)$, we have a p-dimensional vector of score functions for the parameter θ_t, which we denote by

$$\psi_t(\theta_t; d_t) = \frac{\partial l_t(\theta_t; d_t)}{\partial \theta_t}. \tag{1}$$

The elements of the model parameter vector θ_t capture the mean performance and the effects of the covariates. We assume that the elements of the parameter θ_t describe the same attributes of the process across time periods $t = 1, \ldots, T$. We expect that the unknown true value of one or more of the p elements in θ_t may drift slowly over time in an unpredictable

way. For example, the mean customer loyalty measure may drift slowly over time as competitive products are introduced to the market. We purposefully do not attempt to specify how the model parameter might change over time because the change is hard to predict and we want our approach to be flexible.

To effectively monitor the process, we want an accurate and precise estimate of the current (at time T) parameter, θ_T. There are two naïve approaches. We may estimate θ_T using data d_T observed in the most recent time period; however, a small present-time sample size n_T results in an imprecise estimate. Alternatively, we may use all of the data assuming $\theta_t = \theta_T$ for all t; however, we introduce bias if any of the elements of the parameter drift over time periods. We propose an approach to combine present and historical data that sets up a bias/variance trade-off for an estimate of the present value of the parameter.

In Section 2, we propose estimating equations that are better suited for the general problem than either naïve approach. We derive approximations for the variance of the estimator and the distribution of a hypothesis test statistic at time T. We study the suitability of these approximations in a simple analytic example. We illustrate the proposed approach with the problem of monitoring a customer loyalty measure in Section 3.

A state-space approach could be applied to this problem whereby we combine an estimate from present data d_T with another estimate based on previous data $d_1, \ldots, d_{T-1}$. One possibility is the Kalman filter (Grewal and Andrews (2014)). In Section 4, we provide a qualitative comparison of the differences between the proposed approach and the Kalman filter as well as a summary and further discussion.

2. Weighted Estimating Equations

The work of Steiner and MacKay (2014) is the basis of the proposed approach involving weighted estimating equations (WEE). Their motivating surgical performance example involves a binary observation at each time period and a single covariate with known effect. Here, we generalize the approach for the setup described in Section 1. Based on asymptotic theory and the assumption that, in fact, θ_t is constant over time, we derive approximations for the variance of the WEE estimator and the distribution of an appropriate test statistic for inference in a problem requiring a hypothesis test. When θ_t changes slowly over time, we found these approximations are reasonable.

Suppose we have a set of weights $\{w_t;\ t = 1, \ldots, T\}$. We define the p weighted estimating equations by

$$\psi(\hat{\theta}; d, w) = w_1\psi_1(\hat{\theta}; d_1) + \cdots + w_t\psi_t(\hat{\theta}; d_t) + \cdots + w_T\psi_T(\hat{\theta}; d_T) = 0. \quad (2)$$

The terms in $\psi(\hat{\theta}; d, w)$ are the score functions at time $t = 1, \ldots, T$ given in Equation (1) evaluated at the solution of the corresponding weighted estimating equation, which we denote as $\hat{\theta}$. We call $\hat{\theta}$ the WEE estimate of θ_T and let $\tilde{\theta}$ be the corresponding estimator.

The WEE approach is similar to relevance weighted likelihood (Hu and Zidek (2002)) where contributions to likelihood from similar populations are weighted by a relevance measure. These authors suggest data-based methods to select weights based on relevance measures and an optimization criterion. In the general problem of this paper, the score functions by time period have a natural ordering and we expect that one or more of the p elements of model parameter θ_t drifts slowly with time. Accordingly, we use weights that decrease (exponentially) for time periods further in the past. In particular, we propose a weight parameter, λ, $0 \leq \lambda \leq 1$, to define the weights as

$$w_t = \lambda(1-\lambda)^{T-t} \quad (3)$$

for each $t = 1, \ldots, T$. Other definitions of decreasing weights are possible. With Equation (3), the weight for the most recent time period is proportional to λ, the time period before that has weight proportional to $\lambda(1-\lambda)$, the time period before that $\lambda(1-\lambda)^2$, and so on. For convenience, we divide each weight by the same constant $\sum_{t=1}^{T} \lambda(1-\lambda)^{T-t}$ so that $\sum_{t=1}^{T} w_t = 1$. Note that this rescaling does not change the WEE estimate $\hat{\theta}$ or its properties. Under Equation (3), increasing the value of λ increases the relative weight of present data, which reduces bias and increases variance of the estimator. There is subjectivity in the selection of λ, but the value $\lambda = 0.1$ is reasonable in applications we have considered where we assume the parameter of interest is changing slowly relative to the defined time interval.

Note that the two naïve approaches involving either present time data only or the aggregate of historical data weighted equally are particular cases of Equation (2) at the two limiting values of the weight parameter λ.

- As λ approaches 1, w_T approaches 1 and w_t,

$t < T$ approaches 0. The estimating equation involves the present-time data only; i.e., $\psi(\hat{\theta}; d, w) = \psi_T(\hat{\theta}; d_T) = 0$.

- As λ approaches 0, w_t approaches $1/T$ for all $t = 1, \ldots, T$. The estimating equation involves the aggregate of data weighted equally; i.e., $\psi(\hat{\theta}; d, w) = (\sum_{t=1}^{T} \psi_t(\hat{\theta}; d_t))/T = 0$.

Next, we derive approximations for the variance of $\hat{\theta}$ and the distribution of a given test statistic involving $\hat{\theta}$ using the usual asymptotic properties of the information and score functions in the model based on data by time period. For these derivations, we assume that the model parameter θ_t does not change over time $t = 1, \ldots, T$. So there are two sources of error in the approximations; first, the usual error due to the asymptotics and a second error due to the fact that the parameter may have drifted.

2.1. Estimate of Variance

A specific problem of interest may require an estimate of the uncertainty of the WEE estimate $\hat{\theta}$. For example, a manager may want to assess whether the mean performance measure based on $\hat{\theta}$ is significantly different than the competitive benchmark. We assume that the model $\mathcal{L}_t(\tilde{\theta}; \mathcal{D}_t)$ holds for each $t = 1, \ldots, T$ and that the random data $\mathcal{D}_t$ are independent over t. In the case where the model depends on covariates, then we assume that $\mathcal{D}_t$ are independent over t, conditional on the values of the covariates. Note that we do not model changes in the covariates. For θ, the unknown model parameter,

$$I_t(\theta) = -E\left(\frac{\partial^2 \log \mathcal{L}_t(\theta; \mathcal{D}_t)}{\partial \theta^2}\right),$$

is the matrix of expected information about θ at time t, $i_t(\theta) = -l_t''(\theta; d_t)$ is the observed information matrix, and the two are related by $E(i_t(\theta)) = I_t(\theta)$ (Small (2010)). Because the weighted estimating equations combine the usual score functions by time period, we consider an estimate of $\text{var}(\tilde{\theta})$ through the known asymptotic properties of the corresponding information and score functions.

We consider the asymptotic properties of the information and score functions in the case where the total sample size $N = \sum_{t=1}^{T} n_t$ approaches infinity and the number of time periods T remains fixed. In order to preserve the usual asymptotic properties of these functions by time period as $N \to \infty$, we need to preserve some uniformity in the relative distributions of $I_t(\theta)$ by time period $t = 1, \ldots, T$. We require that the relative sample size defined by $c_t = n_t/N$ remains constant for each t so that $n_t \to \infty$ as $N \to \infty$. In the case where the model does not depend on covariates, then each individual subject has the same expected information and so, for fixed c_t, the relative distributions of $I_t(\theta)$, $t = 1, \ldots, T$, stay the same as $N \to \infty$. In the more general problem where the model depends on covariates, then some uniformity in the distribution of samples across the covariate space must be maintained as each $n_t \to \infty$ so that $I_t(\theta))/n_t \to g_t(\theta)$ for some constant matrix $g_t(\theta)$. We derive an approximation for $\text{var}(\tilde{\theta})$ under this asymptotic paradigm.

In Appendix A.1, we sketch a proof to show that $\tilde{\theta}$ is a consistent estimator of the true value θ under usual regularity conditions and under the condition that θ does not change over time. In Appendix A.2, we derive the estimate of the variance of $\tilde{\theta}$ for the case of a model that does not depend on covariates. This extends to the general case where there may be covariates in the model and we refer to the result as the weighted information (WI) estimate of variance,

$$\widehat{\text{var}}_{\text{WI}}(\tilde{\theta}) = \left(\sum_{t=1}^{T} w_t I_t(\hat{\theta})\right)^{-1} \sum_{t=1}^{T} w_t^2 I_t(\hat{\theta}) \times \left(\sum_{t=1}^{T} w_t I_t(\hat{\theta})\right)^{-1}, \quad (4)$$

given weights $\{w_t\}$ and expected information matrices evaluated at the WEE estimate, $\{I_t(\hat{\theta}),\ t = 1, \ldots, T\}$. We use this approximation for the variance of the random variable $\tilde{\theta}$ to estimate the standard error of the WEE estimate $\hat{\theta}$.

Note that Equation (4) at the two special cases of weight values described previously gives the usual estimates of variance. In the case where $w_T = 1$ **and** $w_t = 0$ for all $t < T$, then

$$\begin{aligned}\widehat{\text{var}}_{\text{WI}}(\tilde{\theta}) &= I_T^{-1}(\hat{\theta}) I_T(\hat{\theta}) I_T^{-1}(\hat{\theta}) \\ &= I_T^{-1}(\hat{\theta}).\end{aligned}$$

In the case where $w_t = 1/T$ for all t, then

$$\begin{aligned}\widehat{\text{var}}_{\text{WI}}(\tilde{\theta}) &= \left(\sum_{t=1}^{T} \frac{I_t(\hat{\theta})}{T}\right)^{-1} \sum_{t=1}^{T} \frac{I_t(\hat{\theta})}{T^2} \left(\sum_{t=1}^{T} \frac{I_t\hat{\theta})}{T}\right)^{-1} \\ &= \left(\sum_{t=1}^{T} I_t(\hat{\theta})\right)^{-1}.\end{aligned}$$

We can show that the weighted information estimate of variance is the same result as the sandwich estimate of variance used in Steiner and MacKay

(2014). The sandwich estimate of variance was proposed for maximum-likelihood estimates of a misspecified model or under missing covariate data (White (1982)). The key contribution of the work of this section is that we justify its use as an estimate of the variance of the WEE estimator relative to the specified asymptotic paradigm.

2.2. Distribution of Hypothesis Test Statistic

A specific problem of interest may require a test of hypothesis involving the WEE estimate at the present time T. For example, a process-quality manager responsible for checking the consistency of multiple parallel gauges may want to monitor a test statistic for the hypothesis that the parameters of the model describing the gauge effects are the same. This activity requires an approximation for the distribution of a test statistic involving the WEE estimate under a null hypothesis versus a specified alternative hypothesis.

Here, we consider a test statistic based on a likelihood ratio (LR), though a Wald or score test statistic could also be constructed (Lehmann and Romano (2005)). Consider a partition of the parameter vector θ into $\theta = [\delta, \alpha]$, where δ^T is the vector of parameters of interest for testing and α^T is the vector of unrestricted parameters. Let the number of independent restrictions on parameters in δ^T be r. For example, when monitoring the consistency of M binary gauges, suppose the parameter α represents the pass rate for a baseline gauge and parameters $\delta = (\delta_1, \ldots, \delta_{M-1})^T$ represent the pass rates of the other gauges relative to the baseline. To test for consistency across the M gauges, we use a test of the null hypothesis H_0: $\delta_1 = \cdots = \delta_{M-1} = 0$ versus the alternative H_A: at least one of $\delta_1, \ldots, \delta_{M-1} \neq 0$.

The general null hypothesis of interest is H_0: $\delta = \delta_0$. To construct an LR test statistic, we estimate $\theta = (\delta^T, \alpha^T)^T$ under the unrestricted model and α_0 when δ is restricted to δ_0. The weighted estimating equation in Equation (2) gives WEE estimates $\hat{\theta}$ and $\hat{\alpha}_0$. The WEE approach extends the usual LR test statistic by comparing the weighted log-likelihood contributions by time under the unrestricted and restricted models. The WEE LR test statistic is

$$\hat{S} = 2\left(\sum_{t=1}^{T} w_t l_t(\hat{\theta}; d_t) - \sum_{t=1}^{T} w_t l_t(\delta_0, \hat{\alpha}_0; d_t)\right) \quad (5)$$

at WEE estimates $\hat{\theta}$ and $\hat{\alpha}_0$. We consider the distribution of the corresponding random variable $\tilde{S}$ under the null hypothesis and the asymptotic paradigm discussed in Section 2.1. In Appendix A.3, we derive an approximate distribution for $\tilde{S}$ when $\dim(\theta) = 1$, as is the case where there are no covariates in the model and a single parameter. We show that the result holds when the model has covariates and $I_t(\theta)/n_t \to g(\theta)$; i.e., the average expected information in the limit is the same for all t. We extend this result to the multiparameter case where $\dim(\theta) = p \geq 1$, which gives

$$\frac{\sum_{t=1}^{T} w_t n_t}{\sum_{t=1}^{T} w_t^2 n_t} \tilde{S} \overset{\text{approx}}{\sim} \chi_p^2$$

under the simple null hypothesis. For testing $r < p$ restrictions on θ, then

$$\frac{\sum_{t=1}^{T} w_t n_t}{\sum_{t=1}^{T} w_t^2 n_t} \tilde{S} \overset{\text{approx}}{\sim} \chi_r^2 \quad (6)$$

under the null hypothesis. Note that, at the two special cases of weight values described previously, Equation (6) gives the usual results using present-time data only or all data weighted equally. The extension of Equation (6) to the most general case where $\dim(\theta) \geq 1$ and the average expected information in the limit is not the same for all time periods is not straightforward. This remains as future work.

Note that the estimate of variance of Equation (4) and the weight-adjusted test statistic of Equation (6) do not change if we scale each w_t by the same constant. The argument for consistency and the derivations of the approximate results of Equations (4) and (6) assume that the true value of parameter θ_t is the same across the $t = 1, \ldots, T$ time periods. The general problem of this research expects that the parameter may drift over time and so these results do not hold exactly. Because we restrict our focus to slow changes in the parameter over time, we expect that these results are reasonable approximations. In Section 2.3, we show an example where the parameter changes slowly over time. Here, the WEE approach with an appropriate weight parameter gives an estimate with lower mean-squared error than a naïve approach where no weights are used. This property holds for a wide variety of problems.

2.3. Analytic Example

We look at an example of a simple process with multiple streams to look at properties of the WEE parameter estimate, the WI estimate of variance, and the WEE LR test statistic. The simple process generates binary observations over time from units in two streams. The observations are the quantities

of passed units $y_{1,t}$, $y_{2,t}$ at time t arising from two gauges (streams) performing the same test. The objective is to monitor the difference in the pass rates from the gauges over time. The simplicity of the example is convenient for demonstration purposes. Similar demonstrations can be made over a wide class of models.

We consider random variables

$$Y_{m,t} \sim \text{Binomial}(n_{m,t}, \pi_{m,t})$$

for $m = 1, 2$ that we assume are each independent over $t = 1, \ldots, T$. For $\pi = \pi_2 - \pi_1$, the difference of the mean pass rates at the two streams at the present time, the objective requires us to test the null hypothesis H_0: $\pi = 0$ versus the alternative H_A: $\pi \neq 0$. We expect that one or both of the true values of the elements of $\theta_t = (\pi_{1,t}, \pi_{2,t})$ may change slowly over time.

Replacing the $\pi_{m,t}$ for $t = 1, .., T$ by the common parameter π_m for each $m = 1, 2$, closed-form expressions for $\hat{\pi}$, $\widehat{\text{var}}(\hat{\pi})$, and $\hat{S}$ are possible for this simple example,

$$\hat{\pi} = \hat{\pi}_2 - \hat{\pi}_1$$

$$\widehat{\text{var}}(\hat{\pi}) = \sum_{m=1}^{2} \frac{\hat{\pi}_m(1-\hat{\pi}_m)\sum_{t=1}^{T} w_t^2 n_{m,t}}{\left(\sum_{t=1}^{T} w_t n_{m,t}\right)^2}$$

$$\hat{S} = 2\sum_{m=1}^{2}\sum_{t=1}^{T} w_t \left(\log \frac{\hat{\pi}_m}{\hat{\pi}_0} y_{m,t} + \log\left(\frac{1-\hat{\pi}_m}{1-\hat{\pi}_0}\right)(n_{m,t} - y_{m,t})\right)$$

based on estimates

$$\hat{\pi}_m = \frac{\sum_{t=1}^{T} w_t y_{m,t}}{\sum_{t=1}^{T} w_t n_{m,t}}$$

for $m = 1, 2$ and

$$\hat{\pi}_0 = \frac{\sum_{m=1}^{2}\sum_{t=1}^{T} w_t y_{m,t}}{\sum_{m=1}^{2}\sum_{t=1}^{T} w_t n_{m,t}}.$$

In general, closed-form expressions for the estimates by the WEE approach are possible for those models where closed-form expressions for the MLE estimates are possible. The usual maximum-likelihood estimates (MLEs) involving the historical observations are special cases of these estimates with $w_t = 1$ (or $w_t = 1/T$) for all t.

We further consider this simple example to demonstrate three properties of the WEE approach:

(i) the estimate of $\text{var}(\tilde{\theta})$ in Equation (4) is appropriate.

(ii) the WEE estimator $\tilde{\theta}$ and WI estimate of $\text{var}(\tilde{\theta})$ in Equation (4) are suitable with small changes in θ_t over time periods $t = 1, \ldots, T$.

(iii) the distribution of the weight-adjusted random variable $\tilde{S}$ in Equation (6) is approximately χ^2_r.

The three properties are demonstrated for the simple example in Appendix B.

3. Estimate Net Promoter Score with a Bias/Variance Trade-Off

As discussed in Section 1, the customer loyalty measure is commonly used to focus process and product improvements to drive customer loyalty and achieve business success across many industries. The measure known as net promoter score (NPS) is highly recommended by leading customer experience management firms such as Bain & Company and Satmetrix. The estimate of this measure is based on customer responses to a survey asking the loyalty question, "On a scale of 0–10, how likely is it that you would recommend this company or product to a friend or colleague?" The customer's response classifies them into one of the three categories

- detractors who respond 6 or below,
- passives who respond 7 or 8, and
- promoters who respond 9 or 10.

The quantity NPS is defined as the difference between the proportions of customers who are promoters and detractors. Increasing the proportions of customers who are promoters, decreasing the proportion of detractors, or doing both simultaneously increases the value of NPS. Publicly available information such as NPS Benchmarks (n.d.) shows that many diverse companies report NPS quantities as a measure of business performance. Efficient estimation of NPS is thus a topic of importance.

The current industry practice is a naïve estimate for NPS based on sample proportions of data in streams from an arbitrary time period (Markey et al. (2013)). Estimates by time period are compared with benchmarks and targets and tracked in a trend chart over time. Little or no attention is paid to the impact of sample size, covariate effects, and changing populations over time. Depending on the survey design and fluctuations in response rates, small samples are likely in some time periods. Often, the analysis draws on data from multiple time periods to reduce uncertainty. In the common situation where performance drifts over time, a present-time estimate that

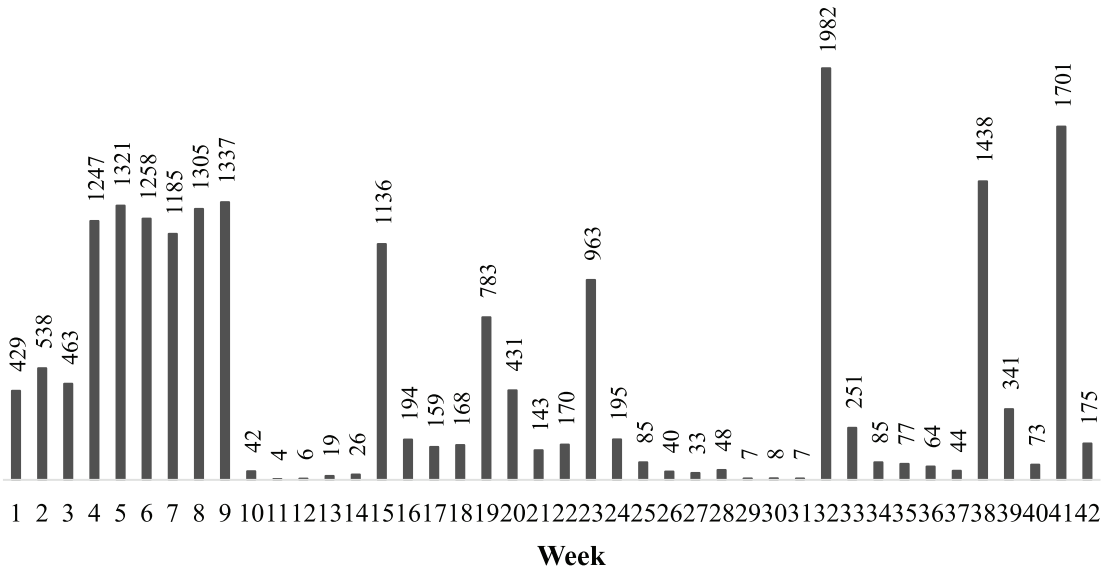

FIGURE 2. Customer Loyalty Dataset: Sample Size by Time Period.

uses present and historical data is biased. We illustrate the application of the WEE approach to set up a bias/variance trade-off in the present-time estimate of NPS with a realistic dataset of observations over time.

We consider data observed on covariates in addition to observations of the process output. These are factors that we expect have an influence an individual's loyalty response. The data are observed from different individuals among a changing customer population. In order to reliably compare estimates of mean NPS across time, we adjust the present-time estimate for the different covariate values among the samples. Further, we illustrate a hypothesis test to compare NPS estimates across levels of a covariate of interest.

The dataset includes customer responses to the survey asking the ultimate question. There are sample responses from 19,981 customers over 42 weeks. The number of customers responding by week over this period, n_t, $t = 1, \ldots, 42$, is given in Figure 2.

Figure 2 shows that the number of customer responses by week varies considerably. There are as few as 4 customer responses and as many as 2000 responses in one week. There are 175 customer responses in the current week. For each sample, we observe the categorized customer response to the ultimate question taking a value from $y = \{1$ (detractor), 2 (passive), 3 (promoter)$\}$. In addition to the response, we also observe two covariate values for each customer: their product variant and the amount of time since their purchase of the product (tenure). The nominal variable $x_1 = \{1, 2, 3, 4\}$ describes the product variant and the interval variable $x_2 = \{0, 2, 6, 12, 17, 24\}$ describes the tenure in months. We define an arbitrary baseline level of the covariates as $x_1 = 1$, $x_2 = 0$. Under the notation introduced in Section 1, at time period t, we observe data $d_t = \{y_{jt}, x_{1,jt}, x_{2,jt};\ j = 1, \ldots, n_t\}$ from the sample of n_t customers.

We estimate NPS at the present time for customers with baseline levels of the covariates (baseline customers) through estimates of two parameters of interest, α_1 and α_2. The parameters α_1 and α_2 represent the mean proportions of baseline customers who are detractors and promoters, respectively. We estimate covariate effects because we are interested in estimating NPS for customers at all possible levels of the covariates. The effects of the other three product variants relative to the baseline are modelled by β_1, β_2, β_3 and product tenure is modelled as a linear change with the term $\beta_4 x_2$. The unknown parameter that we want to estimate is

$$\theta_T = (\alpha_{1,T}, \alpha_{2,T}, \beta_{1,T}, \beta_{2,T}, \beta_{3,T}, \beta_{4,T})^T.$$

For customer j observed at time t having covariate values $x_{1,jt}$ and $x_{2,jt}$, we assume that their response y_{jt} is modelled by $Y_{jt} \sim$ Multinomial$(1, \pi_{1,jt}, 1 - \pi_{1,jt} - \pi_{3,jt}, \pi_{3,jt})$ and that the random variables Y_{jt} are independent across $j = 1, \ldots, n_t$ and $t = 1, \ldots, T$, conditional on the values of the covariates.

TABLE 1. Present-Time Field Population Distribution by Covariate Group at $T = 42$

Product variant	Product tenure [months]					
	0	2	6	12	18	24
1	5	21	95	92	289	1,071
2	20	67	353	490	557	743
3	64	228	1,188	931	522	227
4	524	1,379	1,133	1	0	0

The multinomial rate parameters $\pi_{1,jt}$ and $\pi_{3,jt}$ relate to model parameter θ_t at time t and values of covariates $x_{1,jt}$ and $x_{2,jt}$ through the link functions and linear predictors

$$\begin{aligned}\pi_{1,jt}(\theta_t) = \{&\exp(\alpha_{1,t} + \beta_{1,t} I_{[x_{1,jt}=2]} \\ &+ \beta_{2,t} I_{[x_{1,jt}=3]} + \beta_{3,t} I_{[x_{1,jt}=4]} \\ &+ \beta_{4,t} x_{2,jt})\} \\ \div \{&1 + \exp(\alpha_{1,t} + \beta_{1,t} I_{[x_{1,jt}=2]} \\ &+ \beta_{2,t} I_{[x_{1,jt}=3]} + \beta_{3,t} I_{[x_{1,jt}=4]} \\ &+ \beta_{4,t} x_{2,jt})\} \\ \pi_{3,jt}(\theta_t) = 1 \div \{&1 + \exp(\alpha_{2,t} + \beta_{1,t} I_{[x_{1,jt}=2]} \\ &+ \beta_{2,t} I_{[x_{1,jt}=3]} + \beta_{3,t} I_{[x_{1,jt}=4]} \\ &+ \beta_{4,t} x_{2,jt})\}\end{aligned}$$

for indicator variables $I_{[x_{1,jt}=2]}$, $I_{[x_{1,jt}=3]}$, and $I_{[x_{1,jt}=4]}$. The log-likelihood function describing the probability of data $d_t = \{(x_{1,jt}, x_{2,jt}, y_{jt})$ for $j = 1, \ldots, n_t\}$ including observations from all customers observed at time period t is

$$\begin{aligned}l_t(\theta_t; d_t) = \sum_{j=1}^{n_t} \Big(&I_{[y_{jt}=1]} \log \pi_{1,jt} \\ &+ I_{[y_{jt}=2]} \log(1 - \pi_{1,jt} - \pi_{3,jt}) \\ &+ I_{[y_{jt}=3]} \log \pi_{3,jt} \Big)\end{aligned}$$

for indicator variables $I_{[y_{jt}=1]}$, $I_{[y_{jt}=2]}$, and $I_{[y_{jt}=3]}$.

With an estimate $\hat{\theta}$ for the model parameter θ_T and select covariate values x_1 and x_2, we compute the present estimates for the probabilities that a customer with these covariate values is a detractor or a promoter, which we denote as $\hat{\pi}_1(\hat{\theta}, x_1, x_2)$ and $\hat{\pi}_3(\hat{\theta}, x_1, x_2)$, respectively. An estimate of NPS for a set of customers with these covariate levels is $\widehat{\text{NPS}} = \hat{\pi}_3(\hat{\theta}, x_1, x_2) - \hat{\pi}_1(\hat{\theta}, x_1, x_2)$. An estimate of NPS is possible at any of the possible levels of the two covariates. We define a standard population that is a known, fixed set of the covariate values representing subjects in a population of importance. Then, estimates of NPS are made for each customer in the standard population and we report the mean. The standard population adjusts for differences among covariate levels observed in samples over time. To reliably compare estimates to look for trends across time, it is important that the same standard population be used at each time period. In this application, the sizes of the covariate groups are known in the population of all present customers, which we call the field population. These are given in Table 1.

We summarize the variables, the model, the parameters, and the assumptions for this application in Table 2.

In Table 2, we state the assumption that one or more elements of $\theta_t = (\alpha_{1,t}, \alpha_{2,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \beta_{4,t})^T$ may drift slowly in an unpredictable way over $t = 1, \ldots, T$. In this application, changes to the elements of θ_t over time may occur due to many complex factors; e.g., continuous improvement in the product or manufacturing process, new competitive products in the market, and changing media views of the product. We do not want to assume a stochastic or deterministic model to describe the change in θ_t because it may be difficult to model the contributing factors and the model may only be useful for a short period of time. Instead, we prefer to estimate θ_T assuming that the changes to θ_t over $t = 1, \ldots, T$ are relatively slow and so past data d_t have relevance to estimation of θ_T related to their proximity to the current time. We estimate the single parameter $\hat{\theta}$ through the weighted estimating equation in Equation (2) with weights based on the weight parameter $\lambda = 0.1$ and Equation (3). We know that the estimate $\hat{\theta}$ is a biased estimate of θ_T assuming that $\theta_T \neq \theta_t$ for $t = 1, \ldots, T-1$, but $\hat{\theta}$ has less uncertainty than if we estimate it based on d_T alone. Because the sample size in the current time period is small, reducing uncertainty by incorporating past data becomes important. This is the bias/variance trade-off that is the motivation for using the WEE approach.

Through Equation (4), we calculate the weighted information estimate of variance of $\tilde{\theta}$ involving

$$I_t(\hat{\theta}) = -E\left(\frac{\partial^2 \ell_t(\theta; \mathcal{D}_t)}{\partial \theta^2}\right)\Big|_{\theta=\hat{\theta}},$$

which is the expected information matrix at each time period evaluated at the WEE estimate. Estimates of the variance of the mean values of π_1, π_3, and NPS are computed from $\widehat{\text{var}}_{\text{WI}}(\tilde{\theta})$.

TABLE 2. Model for Net Promoter Score Application

Data, d_t $t = 1, \ldots, T$	$y_{jt} \in \{1, 2, 3\}$: response to ultimate question in one of 3 categories $x_{1,jt} \in \{1, 2, 3, 4\}$: nominal variable representing product variant $x_{2,jt} \in \{0, 2, 6, 12, 17, 24\}$: interval value representing tenure for customer $j = 1, \ldots, n_t$ at week $t = 1, \ldots, 42$
Parameters of interest	$\pi_1(\theta_T)$, $\pi_3(\theta_T)$: proportions of customers who are detractors, promoters at current time T for a fixed standard population of customers, and then NPS $= \pi_3 - \pi_1$.
Distribution of $\mathcal{D}_t$	$Y_{jt} \sim \text{Multinomial}(1, \pi_{1,jt}, 1 - \pi_{1,jt} - \pi_{3,jt}, \pi_{3,jt})$
GLM	$\pi_{1,jt}(\theta_t) = \dfrac{\exp(\alpha_{1,t} + \beta_{1,t} I_{[x_{1,jt}=2]} + \beta_{2,t} I_{[x_{1,jt}=3]} + \beta_{2,t} I_{[x_{1,jt}=4]} + \beta_{4,t} x_{2,jt})}{1 + \exp(\alpha_{1,t} + \beta_{1,t} I_{[x_{1,jt}=2]} + \beta_{2,t} I_{[x_{1,jt}=3]} + \beta_{2,t} I_{[x_{1,jt}=4]} + \beta_{4,t} x_{2,jt})}$ $\pi_{3,jt}(\theta_t) = \dfrac{1}{1 + \exp(\alpha_{2,t} + \beta_{1,t} I_{[x_{1,jt}=2]} + \beta_{2,t} I_{[x_{1,jt}=3]} + \beta_{3,t} I_{[x_{1,jt}=4]} + \beta_{4,t} x_{2,jt})}$
Model parameters	$\theta_t = (\alpha_{1,t}, \alpha_{2,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \beta_{4,t})^T$ with $p = 6$
Objectives	1. Estimate θ_T, the model parameter at the current time 2. Estimate π_1, π_3, and NPS at the current time for a fixed standard population of customers 3. Track estimates of NPS over time
Assumption	One or more elements of $\theta_t = (\alpha_{1,t}, \alpha_{2,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \beta_{4,t})^T$ may drift slowly in an unpredictable way over $t = 1, \ldots, T$.

We compare the WEE estimate for NPS to those by the naïve approaches discussed in Section 2. For the two naïve approaches, estimates $\hat{\pi}_3(\hat{\theta}, x_1, x_2)$ and $\hat{\pi}_1(\hat{\theta}, x_1, x_2)$ and estimates of their variances are calculated through the WEE approach with one of the limiting values of the weight parameter. The naïve approach commonly used in practice (Markey et al. (2013)) estimates NPS based on present-time data only without attention to the values of the covariates among customers in the sample. Here, estimates $\hat{\pi}_3(\hat{\theta}, x_1, x_2)$ and $\hat{\pi}_1(\hat{\theta}, x_1, x_2)$ are sample proportions based on those customers having the particular covariate vector (x_1, x_2) among the present sample. Variances of the sample proportion estimates are estimated by usual methods. The estimate of variance is large when there are few observations for a select combination of covariate levels and the approach is infeasible when no customers are observed at a particular combination. Another naïve approach involves sample proportions estimates based on the aggregate of historical data weighted equally.

Figure 3 gives mean estimates $\widehat{\text{NPS}}$ and the corresponding 95% confidence intervals based on $\widehat{\text{var}}(\widehat{\text{NPS}})$ assuming normality for the standard population in Table 1 by the various approaches.

Figure 3 shows that the estimate by the recommended WEE approach ($\lambda = 0.1$) has less uncertainty than either of the estimates using present-time data only. Its uncertainty is comparable with that of the naïve WEE estimate that uses all historical data. There are some differences between the estimates by the various approaches, but we are unable to assess bias because the true value is unknown. The advantage of the recommended WEE approach over the other approaches depends on the sample sizes and the drift in the parameter over time. Future work will investigate the advantage of WEE over a wide variety of cases through simulation.

Decision makers track the NPS estimates over time to regularly assess and plan improvement activities. In Figure 4, we compare the trends in the current field population estimates between the common naïve approach involving sample proportions based on present-time data only and the WEE approach. Note that there is a difference in the scales of the two vertical axes.

Figure 4 shows a vast difference in the trend of NPS estimates by the two approaches over time. The estimates by the WEE approach are much more precise and show a trend that is not apparent on

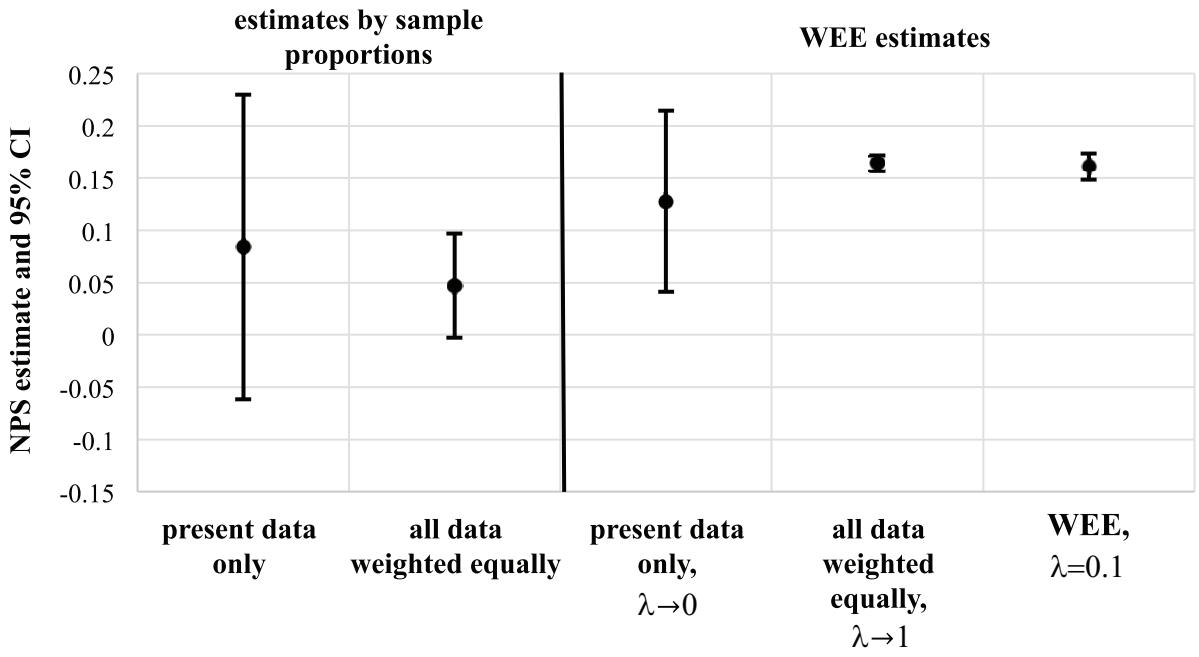

FIGURE 3. Estimates of Field Population NPS by Various Approaches.

the other graph. The WEE approach could have an important impact on the decisions taken by business managers to drive loyalty and growth through a trade-off in bias and variability in population-adjusted NPS estimates and reliable comparisons across time.

In this application, a decision maker may also want to compare NPS estimates across subgroups of the customer population. For example, superior results for a particular product variant may encourage decision makers to target sales of this variant or focus efforts to bring the NPS of other product variants to comparable levels. We consider the test of the hypothesis that NPS for customers with product variant 4 is the same as NPS for customers with product variant 3. In terms of the parameters, we state the null hypothesis for this test as H_0: $\beta_3 - \beta_2 = 0$ versus the alternative H_A: $\beta_3 - \beta_2 \neq 0$. The WEE estimates and relevant quantities to test H_0 versus H_A are given in Table 3.

Table 3 gives evidence to reject the null hypothesis H_0: $\beta_3 - \beta_2 = 0$ in favor of H_A: $\beta_3 - \beta_2 \neq 0$ for a size 0.05 test. The estimates of the proportions are $\hat{\pi}_{1,x_1=3} = 0.30$, $\hat{\pi}_{3,x_1=3} = 0.44$, $\hat{\pi}_{1,x_1=4} = 0.25$, and $\hat{\pi}_{3,x_1=4} = 0.51$. As such, the estimates of NPS for customers at the baseline level of tenure ($x_2 = 0$)

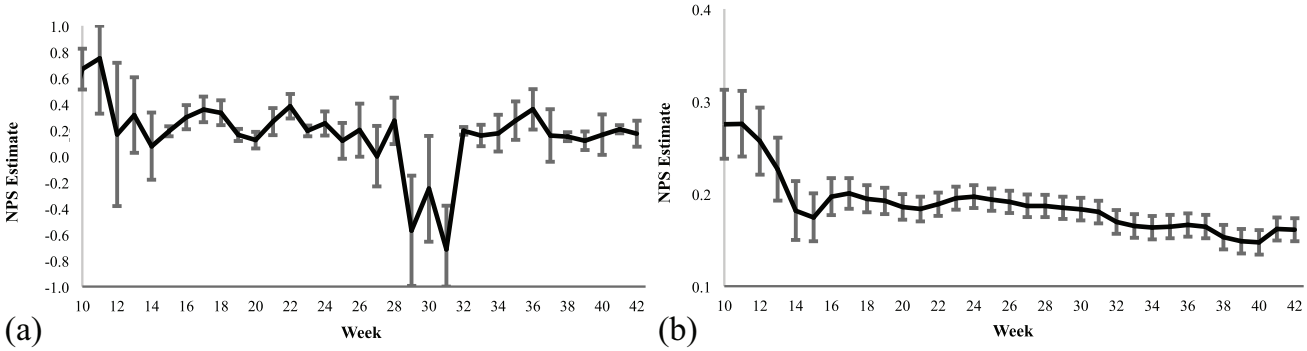

FIGURE 4. Trends in Field Population NPS Estimates. (a) Sample proportion estimates using present data only; (b) WEE approach, $\lambda = 0.1$.

TABLE 3. Hypothesis Test Quantities for H_0: $\beta_3 - \beta_2 = 0$ vs. H_A

Unconstrained model	WEE estimate of θ	$\hat{\theta} = [-0.695, 0.380, -0.0928, -0.147, -0.414, -5.84\text{E-}6]^T$
	Weighted log likelihood	$\sum_{t=1}^{T} w_t l_t(d_t; \hat{\theta}) = -486.218$
Constrained model	WEE estimate of θ	$\hat{\theta}_0 = [-0.899, 0.173, -0.0514, -0.144, -0.144, 0.0100]^T$
	Weighted log likelihood	$\sum_{t=1}^{T} w_t l_t(d_t; \hat{\theta}_0) = -486.754$
WEE LR test statistic (5)		$\hat{S} = 1.072$
Weight-adjusted test statistic		$(\sum_{t=1}^{T} w_t n_t / \sum_{t=1}^{T} w_t^2 n_t)\hat{S} = 17.2$
p-value for H_0 under (6)		$\Pr[\chi_1^2 > (\sum_{t=1}^{T} w_t n_t / \sum_{t=1}^{T} w_t^2 n_t)\hat{S}] < 0.01$

with the two model variants are $\widehat{\text{NPS}}_{x_1=3} = 0.14$ and $\widehat{\text{NPS}}_{x_1=4} = 0.26$. Decision makers have evidence that NPS of product variant 4 is superior to that of product variant 3.

A decision maker may track the estimate of the difference between NPS of the two product variants over time in order to monitor their similarity. The graph of $\widehat{\text{NPS}}_{x_1=4} - \widehat{\text{NPS}}_{x_1=3}$ based on data over the range $T = 10, \ldots, 42$ is given in Figure 5. The 95% confidence interval of each estimate $\widehat{\text{NPS}}_{x_1=4} - \widehat{\text{NPS}}_{x_1=3}$ is based on the WI estimate of variance for $\hat{\theta}$ at that point in time assuming normality. The dotted line shows the p-value for H_0 under Equation (6) at each point in time.

Figure 5 shows that the estimate of NPS for product variant 4 is consistently larger than that of product variant 3 and there is evidence to reject the size 0.05 test of no difference between the two at all points in time except $T = 15$. The earliest customers using product variant 4 are observed in week 10 and so the uncertainty of $\widehat{\text{NPS}}_{x_1=4} - \widehat{\text{NPS}}_{x_1=3}$ decreases after week 10 as more data on this variant are observed. There is a decrease in the estimate of the difference in NPS of the two product variants from week 10 to week 14. The difference between the two product variants is stable from week 16 to week 42. Alternatively, we could monitor the similarity between the mean NPS at the two covariate levels through a graph of the weighted WEE LR test statistic over

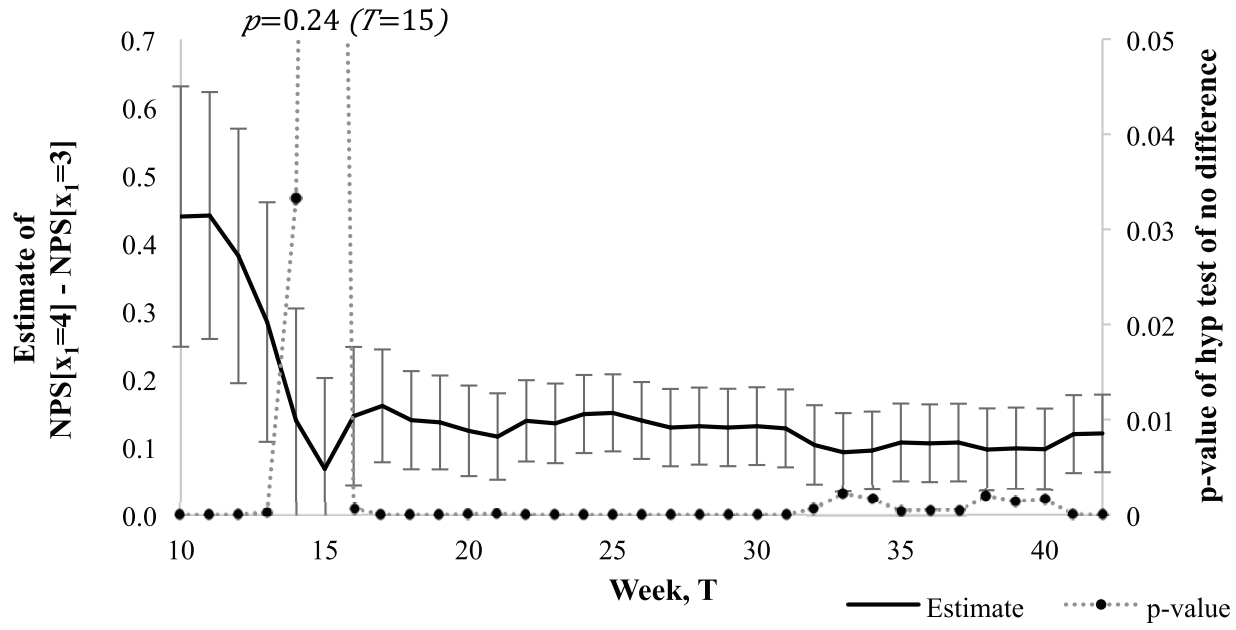

FIGURE 5. WEE Estimates of Difference in NPS for Product Variants 3 and 4.

time. Liu et al. (2008) discuss control limits for a likelihood ratio test statistic that could be adapted in order to formally monitor the weighted WEE LR test statistic.

4. Summary and Discussion

We propose a weighted estimating equation (WEE) approach that offers a trade-off between estimation of a performance measure using present-time data only or historical data over time weighted equally. This trade-off is especially important when sample sizes at some time periods may be small and the parameter in the model for the performance measure may be drifting slowly in an unpredictable way over time. This approach addresses motivating problems with the objective to compare a parameter with a target, over levels of the covariates, across multiple streams, and over time. Specifically, through a realistic dataset, we show the potential of this approach to have an important impact on the ongoing use of the NPS measure based on customer responses to the ultimate question. The WEE approach down weights the contributions of historical data to estimating equations involving scores of the observed data at a common value of the parameter.

We derive an estimate of the variance of the WEE parameter estimator and an approximation for the distribution of a WEE likelihood-ratio test statistic based on asymptotic properties of the score and information functions. For the simple analytic example in Section 2.3, we show the recommended approach has lower mean-squared error than either of the two naïve approaches when there are small changes in the parameter over time and present-time sample size is small. We also show that the given approximation to the distribution of the hypothesis test statistic is suitable for the particular simple example. The WEE approach is straightforward to implement in SAS and the necessary data setup and code are provided in Appendix C.

In this work, we assume that one or more of the elements of the p-dimensional parameter θ_t may drift slowly over time; however, it is possible that the covariate effects may either be known or assumed to be fixed over time. In the case where elements of θ_t are known, we can substitute the known values and reduce the number of estimating functions appropriately. In the case where elements of θ_t are assumed to be fixed over time, an alternative set of estimating functions could be selected without weights for those estimating functions relating to the fixed parameters.

Other methods of analysis are possible. Reviewers of this paper have suggested that we can compare the WEE approach to the Kalman filter (Grewal and Andrews (2014)). Both approaches seek to produce an estimate of θ_T with greater precision by using both current and past data. Each sacrifices unbiasedness for additional precision if one or more parameters change over time.

The usual application of the Kalman filter (KF) assumes an appropriate system model describing the evolution of the state vector (here the parameter) that can be used to estimate the parameter at time T given $\hat{\theta}_{T-1}$. Here, we have no such model so it would be logical to use the current estimate $\hat{\theta}_{T-1}$ to estimate the parameter at time T. Further, to implement the KF, we need to assume a known parameter covariance vector. It is not clear how to select this. We take the weighted average of the two estimates $\hat{\theta}_{T-1}$ and $\hat{\theta}_T$ with dynamic weights based on their precision. If the parameter changes over time, then there is a bias in $\hat{\theta}_{T-1}$.

Unlike the KF, the WEE approach does not combine the current and past estimates. Instead, it creates an estimating function through the weighted average of the likelihood-based score functions across time. The weights are fixed by the choice of λ. Note that the score functions based on d_t, $t = 1, \ldots, T$ are sufficient statistics for the data at each time period and hence contain all of the available information about the parameter. For most models, including the nonlinear model used in our example, the KF estimates $\hat{\theta}_1 \ldots, \hat{\theta}_{T-1}$ are not sufficient statistics and hence information is lost by using $\hat{\theta}_{T-1}$ to summarize the historic data.

In terms of computation for the nonlinear models considered in this paper, both methods require the solution of estimating equations with p unknowns (presuming the KF uses the maximum-likelihood estimate at time T) and similar calculations to find the standard errors. The WEE approach is motivated by applications with small samples in the latest time period. If there are insufficient data at time T to estimate all the parameter components, then a standard implementation of the KF is not applicable. The standard KF implementation could be adapted but it is not obvious how to proceed. With small amounts of data and no system model, present data will have less and less impact on the KF estimate as time goes by and bias in $\hat{\theta}_{T-1}$ is important to consider. In the realistic dataset studied in this paper, there are insufficient data to estimate all of the parameter com-

ponents in 20 of the 42 time periods where data are observed and no obvious system model. Thus, in this case, a standard implementation of the KF is not reliable for updating the NPS estimates over time.

It is not easy to compare quantitatively the performance of the WEE and KF approaches through a simulation study because there are many possible parameter and covariate values and ways that the parameter might drift over time. We suspect that one approach is not uniformly better than others. However, based on the qualitative comparison, we feel the mixed parametric/nonparametric nature of the WEE is a more flexible approach for the estimation problem at hand. Additionally, to implement a change to the usual, naïve method of analysis in practice, decision makers need to be made aware of the reason for the change and the basic premise of the new approach. The WEE approach is an intuitive solution to the bias/variance trade-off problem.

Acknowledgments

We thank the Editor and two referees for their valuable suggestions to improve this paper. This research was supported, in part, by research grant 105240 from the Natural Sciences and Engineering Research Council of Canada.

Appendix A
Asymptotic Results

A.1. Consistency of WEE Estimator

A rigorous proof of consistency of the WEE estimator would follow the method in Wald (1949) for an MLE estimator. Here we outline the main ideas of this proof. We denote θ_0 as the true value of θ that we assume does not change over $t = 1, \ldots, T$ time periods.

Lemma

For any $\theta \neq \theta_0$, we have $E(l_t(X,\theta)) < E(l_t(X,\theta_0))$, where X is a random variable having distribution $f(x,\theta_0)$ and $l_t(x_1,\ldots,x_{n_t}),\theta) = \sum_{j=1}^{n_t} \log f(x_j,\theta)$. See Wald (1949) for proof.

Theorem

Under usual regularity conditions on the family of distributions, the WEE estimate $\hat{\theta}$ is consistent; i.e., $\hat{\theta} \xrightarrow{p} \theta_0$ as $N \to \infty$.

The sketch of the proof is based on the following facts:

- $\hat{\theta}$ is a maximizer of $\sum_{t=1}^{T} w_t l_t(x,\theta)$ by definition.
- θ_0 is the maximizer of $E(l_t(X,\theta))$ by the Lemma. It follows that θ_0 is also the maximizer of $E(\sum_{t=1}^{T} w_t l_t(X,\theta))$.

By the Law of Large Numbers, $\sum_{t=1}^{T} w_t l_t(x,\theta) \xrightarrow{p} E(\sum_{t=1}^{T} w_t l_t(X,\theta))$ for all θ as $N \to \infty$. Because two functions are getting closer, the points of maximum should also get closer, which means that $\hat{\theta} \xrightarrow{p} \theta_0$ as $N \to \infty$.

A.2. Estimate of Asymptotic Variance

We consider the case where the model does not depend on covariates. For $I(\theta)$, the expected information from a single sample, and $I_t(\theta) = n_t I(\theta)$, the expected information from all samples at t, then

$$\mathrm{var}(\psi_t(\theta;\mathcal{D}_t)) = I_t(\theta) = n_t I(\theta) = N c_t I(\theta)$$

because $c_t = n_t/N$ for all t. Then by the Central Limit Theorem,

$$\frac{\psi_t(\theta;\mathcal{D}_t)}{\sqrt{n_t}} \xrightarrow{D} N_p(0, I(\theta)),$$

because ψ_t is the sum of n_t terms each with mean vector 0_p and covariance matrix $I(\theta)$ for each t as $n_t \to \infty$. Because $\mathcal{D}_t$ are assumed to be independent across time $t = 1, \ldots, T$ and w_t and c_t are constants, then

$$\frac{1}{\sqrt{N}} \sum_{t=1}^{T} w_t \psi_t(\theta;\mathcal{D}_t) \xrightarrow{D} N_p\left(0, \sum_{t=1}^{T} w_t^2 c_t I(\theta)\right).$$

We consider the first-order Taylor series approximation of $\psi(\hat{\theta})$ for $\hat{\theta}$ near θ,

$$(\hat{\theta} - \theta) \approx [-\psi'(\theta)]^{-1} \psi(\theta),$$

because $\psi(\hat{\theta}) = 0$. We extend this to an approximation for the corresponding random variable $(\tilde{\theta} - \theta)$ with observed information at time t, $i_t(\theta) = -\psi_t'(\theta)$, so

$$\sqrt{N}(\tilde{\theta} - \theta) \approx \left(\frac{1}{N} \sum_{t=1}^{T} w_t i_t(\theta)\right)^{-1} \frac{1}{\sqrt{N}} \sum_{t=1}^{T} w_t \psi_t(\theta)$$

because $\tilde{\theta}$ is consistent. Then, by Slutsky's theorem,

$$\sqrt{N}(\tilde{\theta} - \theta) \xrightarrow{D} \left(\sum_{t=1}^{T} w_t c_t I(\theta)\right)^{-1} Z$$

as $N \to \infty$, because $E[i_t(\theta)] = I_t(\theta) = N c_t I(\theta)$ and Z is the asymptotic distribution of $(1/\sqrt{N})\times$

$\sum_{t=1}^{T} w_t \psi_t(\theta; \mathcal{D}_t)$. Then, with the previous result for Z,

$$\sqrt{N}(\tilde{\theta} - \theta) \xrightarrow{D} N_p \left(0, \left(\sum_{t=1}^{T} w_t c_t I(\theta)\right)^{-1} \times \sum_{t=1}^{T} w_t^2 c_t I(\theta) \times \left(\sum_{t=1}^{T} w_t c_t I(\theta)\right)^{-1}\right).$$

Then, an estimate for the asymptotic variance of $\tilde{\theta}$ is

$$\widehat{\text{var}}_{\text{WI}}(\tilde{\theta}; \hat{\theta}) = \left(N \sum_{t=1}^{T} w_t c_t I(\hat{\theta})\right)^{-1} N \sum_{t=1}^{T} w_t^2 c_t I(\hat{\theta}) \times \left(N \sum_{t=1}^{T} w_t c_t I(\hat{\theta})\right)^{-1}.$$

More generally, in the case where the model depends on the covariates and $(I_t(\theta))/n_t) \to g_t(\theta)$ as $n_t \to \infty$ for $g_t(\theta)$ a matrix of constants, we extend this estimate as

$$\widehat{\text{var}}_{\text{WI}}(\tilde{\theta}; \hat{\theta}) = \left(\sum_{t=1}^{T} w_t I_t(\hat{\theta})\right)^{-1} \sum_{t=1}^{T} w_t^2 I_t(\hat{\theta}) \times \left(\sum_{t=1}^{T} w_t I_t(\hat{\theta})\right)^{-1}.$$

We refer to this as the weighted information (WI) estimate of variance.

A.3. Distribution of Hypothesis Test Statistic

The likelihood-ratio test of the simple null hypothesis H_0: $\theta = \theta_0$ against the alternative H_A: $\theta \neq \theta_0$ is based on the likelihood-ratio random variable

$$\tilde{S} = 2\left(\sum_{t=1}^{T} w_t l_t(\tilde{\theta}) - \sum_{t=1}^{T} w_t l_t(\theta_0)\right).$$

Consider the second-degree Taylor series approximation of $\sum_{t=1}^{T} w_t l_t(\theta_0)$ for θ_0 near $\hat{\theta}$,

$$\sum_{t=1}^{T} w_t l_t(\theta_0) \approx \sum_{t=1}^{T} w_t l_t(\hat{\theta}) + (\theta_0 - \hat{\theta})^T \sum_{t=1}^{T} w_t l_t'(\hat{\theta}) + \frac{1}{2}(\theta_0 - \hat{\theta})^T \sum_{t=1}^{T} w_t l_t''(\hat{\theta})(\theta_0 - \hat{\theta}).$$

Because $\sum_{t=1}^{T} w_t l_t'(\hat{\theta}) = 0$ and observed information matrix $i_t(\theta) = -l_t''(\theta)$, then

$$\hat{S} = 2\left(\sum_{t=1}^{T} w_t l_t(\hat{\theta}) - \sum_{t=1}^{T} w_t l_t(\theta_0)\right) \approx \sqrt{N}(\hat{\theta} - \theta_0)^T \frac{1}{N} \sum_{t=1}^{T} w_t i_t(\hat{\theta}) \sqrt{N}(\hat{\theta} - \theta_0).$$

We extend this result for $\hat{S}$ to the random variable $\tilde{S}$. We consider the case where the model does not depend on covariates. Then, $\tilde{S}$ has the same asymptotic distribution as

$$\sqrt{N}(\tilde{\theta} - \theta_0)^T \sum_{t=1}^{T} w_t c_t I(\tilde{\theta}) \sqrt{N}(\tilde{\theta} - \theta_0)$$

because $E[i_t(\theta)] = N c_t I(\theta)$. In Appendix A.2, we show that, under regularity conditions and consistency,

$$\sqrt{N}(\tilde{\theta} - \theta_0) \xrightarrow{D} N_p \left(0, \left(\sum_{t=1}^{T} w_t c_t I(\tilde{\theta})\right)^{-1} \times \sum_{t=1}^{T} w_t^2 c_t I(\tilde{\theta}) \times \left(\sum_{t=1}^{T} w_t c_t I(\tilde{\theta})\right)^{-1}\right)$$

as $N \to \infty$. It follows that

$$\sqrt{N}(\tilde{\theta} - \theta_0)^T \left(\sum_{t=1}^{T} w_t c_t I(\tilde{\theta})\right) \left(\sum_{t=1}^{T} w_t^2 c_t I(\tilde{\theta})\right)^{-1} \times \left(\sum_{t=1}^{T} w_t c_t I(\tilde{\theta})\right) \sqrt{N}(\tilde{\theta} - \theta_0) \xrightarrow{D} \chi_p^2$$

as $N \to \infty$. With this asymptotic result, we state an approximation for the distribution of

$$\tilde{S} \sim \sqrt{N}(\tilde{\theta} - \theta_0)^T \sum_{t=1}^{T} w_t c_t I(\tilde{\theta}) \sqrt{N}(\tilde{\theta} - \theta_0)$$

in the case that $\dim(\theta) = 1$. Because $I(\theta)$ is a scalar, then

$$\left(\sum_{t=1}^{T} w_t^2 c_t\right)^{-1} \left(\sum_{t=1}^{T} w_t c_t\right) \tilde{S} \xrightarrow{D} \chi_1^2 \quad \text{as } N \to \infty$$

under the null hypothesis.

More generally, where the model depends on the covariates, we consider the case where $I_t(\theta)/n_t \to g(\theta)$; i.e., the average expected information in the

limit is the same for all t. In the limit as n_t and N get large, then

$$I_t(\theta) \approx n_t g(\theta) \approx N c_t g(\theta)$$

for each $t = 1, \ldots, T$. The previous results in Appendices A.1 and A.2 involving $I(\theta)$ extend to results involving $g(\theta)$. Then, in the case where $g(\theta)$ is a scalar, it follows that

$$\left(\sum_{t=1}^{T} w_t^2 c_t\right)^{-1} \left(\sum_{t=1}^{T} w_t c_t\right) \tilde{S} \xrightarrow{D} \chi_1^2 \quad \text{as } N \to \infty$$

under the null hypothesis. These results extend to the more general case where $\dim(\theta) = p \geq 1$,

$$\left(\sum_{t=1}^{T} w_t^2 c_t\right)^{-1} \left(\sum_{t=1}^{T} w_t c_t\right) \tilde{S} \xrightarrow{D} \chi_p^2 \quad \text{as } N \to \infty$$

under the simple null hypothesis. For testing $r < p$ restrictions on θ, then we can show by a similar argument that

$$\left(\sum_{t=1}^{T} w_t^2 c_t\right)^{-1} \left(\sum_{t=1}^{T} w_t c_t\right) \tilde{S} \xrightarrow{D} \chi_r^2 \quad \text{as } N \to \infty$$

under the null hypothesis. In practice, we replace c_t by n_t/N and use these results to approximate the distribution of the weight-adjusted test statistic

$$\frac{\sum_{t=1}^{T} w_t n_t}{\sum_{t=1}^{T} w_t^2 n_t} \hat{S}.$$

Appendix B
Analytic Example

Based on the random variables

$$Y_{m,t} \sim \text{Binomial}(n_{m,t}, \pi_{m,t}),$$

$m = 1, 2$ with $\tilde{\theta}_t = \{\tilde{\pi}_{1,t}, \tilde{\pi}_{2,t}\}$, the log-likelihood function for observations $y_{1,t}$, $y_{2,t}$ at time t is

$$l_t(y_t \mid \theta_t) = \sum_{m=1}^{2} y_{m,t} \log \pi_{m,t} + (n_{m,t} - y_{m,t}) \log(1 - \pi_{m,t}).$$

Assuming that $\pi_{m,t} = \pi_m$, $m = 1, 2$ for each t, the WEE estimate $\hat{\theta}$ is found by solving $\sum_{t=1}^{T} w(t) \psi_t(y_t \mid \theta) = 0$, which gives

$$\hat{\pi}_m = \frac{\sum_{t=1}^{T} w_t y_{m,t}}{\sum_{t=1}^{T} w_t n_{m,t}}.$$

Because

$$I_t(\theta) = \begin{bmatrix} \dfrac{n_{1,t}}{\pi_1(1-\pi_1)} & 0 \\ 0 & \dfrac{n_{2,t}}{\pi_2(1-\pi_2)} \end{bmatrix},$$

then the estimate of variance of $\tilde{\theta}$ by Equation (4) is

$$\widehat{\text{var}}(\tilde{\pi}_m) = \frac{\sum_{t=1}^{T} w_t^2 n_{m,t} \hat{\pi}_m (1 - \hat{\pi}_m)}{\left(\sum_{t=1}^{T} w_t n_{m,t}\right)^2}, \quad m = 1, 2.$$

The parameter of interest to compare the present pass rates between the two streams is $\pi = \pi_2 - \pi_1$. Based on the preceding estimates,

$$\hat{\pi} = \frac{\sum_{t=1}^{T} w_t y_{2,t}}{\sum_{t=1}^{T} w_t n_{2,t}} - \frac{\sum_{t=1}^{T} w_t y_{1,t}}{\sum_{t=1}^{T} w_t n_{1,t}},$$

$$\widehat{\text{var}}(\tilde{\pi}) = \sum_{m=1}^{2} \frac{\hat{\pi}_m (1 - \hat{\pi}_m) \sum_{t=1}^{T} w_t^2 n_{m,t}}{\left(\sum_{t=1}^{T} w_t n_{m,t}\right)^2}.$$

To test the null hypothesis H_0: $\pi_0 = 0$ versus the alternative H_A: $\theta_0 \neq 0$, the WEE LR test statistic of Equation (5) is

$$\hat{S} = 2 \sum_{m=1}^{2} \sum_{t=1}^{T} w_t \left(\log \frac{\hat{\pi}_m}{\hat{\pi}_0} y_{m,t} + \log \left(\frac{1 - \hat{\pi}_m}{1 - \hat{\pi}_0} \right) (n_{m,t} - y_{m,t}) \right)$$

for $\hat{\pi}_1$ and $\hat{\pi}_2$ as previously stated and

$$\hat{\pi}_0 = \frac{\sum_{m=1}^{2} \sum_{t=1}^{T} w_t y_{m,t}}{\sum_{m=1}^{2} \sum_{t=1}^{T} w_t n_{m,t}}$$

under the null hypothesis. Based on the parameter and test statistic estimates for this simple problem, we consider three properties as follows.

i. the estimate of $\text{var}(\tilde{\theta})$ in Equation (4) is appropriate.

Given the simple model, we estimate $\text{var}(\tilde{\pi})$ directly by the distributions of the random variables $\{Y_{1,t}, Y_{2,t}, t = 1, \ldots, T\}$. The WI estimate of variance by Equation (4) is the same as the closed-form expression of variance derived directly from the distributions of the random variables. Because no asymptotic assumptions are required for the latter formulation, the weighted information estimate of variance is a suitable estimate even when there are small samples for this simple example.

ii. the WEE estimate and WI estimate of $\text{var}(\tilde{\theta})$ in Equation (4) is suitable with small changes in θ_t over time periods $t = 1, \ldots, T$.

We study the effect of a change in true value π on the bias$(\tilde{\pi}) = E(\tilde{\pi}) - \pi$ and estimate of variance $\widehat{\text{var}}(\tilde{\pi})$. For this study, we choose arbitrary values:

- $T = 10$ time periods of data observed.
- sample sizes $n_{1,t} = n_1 = 100$ and $n_{2,t} = n_2 = 60$ for all $t = 1, \ldots, T$.
- stream 2 experiences a positive step change in rate $\pi_{2,t}$ of size Δ at time $t = 6$.

Streams 1 and 2 have the same initial pass rates, so $\pi_{1,1} = \pi_{2,1}$. We vary initial values $\pi_{1,1} = \pi_{2,1}$ and $\Delta = \pi_{2,6} - \pi_{2,5}$ to compare the properties of the WEE estimator over various profiles. Note that, under a change in pass rate at stream 2, the true value of $\pi_{2,t}$ is $\pi_{2,1}$ for $t < 6$ and $\pi_{2,1} + \Delta$ for $t \geq 6$. Then, the quantity $E(Y_{2,t})$ in $E(\tilde{\pi})$ and $I_t(\tilde{\pi})$ depends on t and the size of the change Δ for $t \geq 6$. At the present time $T = 10$, the expected value of estimator $\tilde{\pi}$ is

$$E(\tilde{\pi}; \Delta) = E\left(\frac{\sum_{t=1}^{T} w_t Y_{2,t}}{\sum_{t=1}^{T} w_t n_{2,t}} - \frac{\sum_{t=1}^{T} w_t Y_{1,t}}{\sum_{t=1}^{T} w_t n_{1,t}}\right) = \pi_{2,1} + \Delta \sum_{t=6}^{10} w_t - \pi_{1,1},$$

and its true value is $\pi_{2,1} + \Delta - \pi_{1,1}$. The bias in estimator $\tilde{\pi}$ at time T is

$$\text{bias}(\tilde{\pi}; \Delta) = \Delta\left(\sum_{t=6}^{10} w_t - 1\right).$$

The weighted information estimate of variance of $\tilde{\pi}$ based on Equation (4) is

$$\widehat{\text{var}}_{\text{WI}}(\tilde{\pi}; \Delta) = \frac{a\sum_{t=1}^{10} w_t^2}{n_1} + \frac{b\sum_{t=1}^{5} w_t^2 + c\sum_{t=6}^{10} w_t^2}{n_2\left(b\sum_{t=1}^{5} w_t + c\sum_{t=6}^{10} w_t\right)^2}.$$

with

$$a = \pi_{1,1}(1 - \pi_{1,1}),$$
$$b = \frac{1}{\pi_{2,1}(1 - \pi_{2,1})},$$

and

$$c = \frac{1}{(\pi_{2,1} + \Delta)(1 - \pi_{2,1} - \Delta)}.$$

We study the bias and variance of the WEE estimator through root-mean squared error given by

$$\text{MSE}(\tilde{\pi}, \Delta) = \sqrt{(\text{bias}(\tilde{\pi}, \Delta))^2 + \widehat{\text{var}}(\tilde{\pi}, \Delta)}.$$

We calculate MSE$(\tilde{\pi}, \Delta)$ for values of $\pi_{1,1} = \pi_{2,1}$ in the range of 0.02 to 0.20 and values of Δ in the range of 25% to 100% of each of the starting values $\pi_{1,1} = \pi_{2,1}$. Figure 6 gives contour plots of the relative values of MSE for the values of $\pi_{1,1} = \pi_{2,1}$ and Δ. The relative values compare MSE for the WEE estimator with weight parameter $\lambda = 0.1$ to that of each of the two naïve estimators having limiting values of the weight parameters.

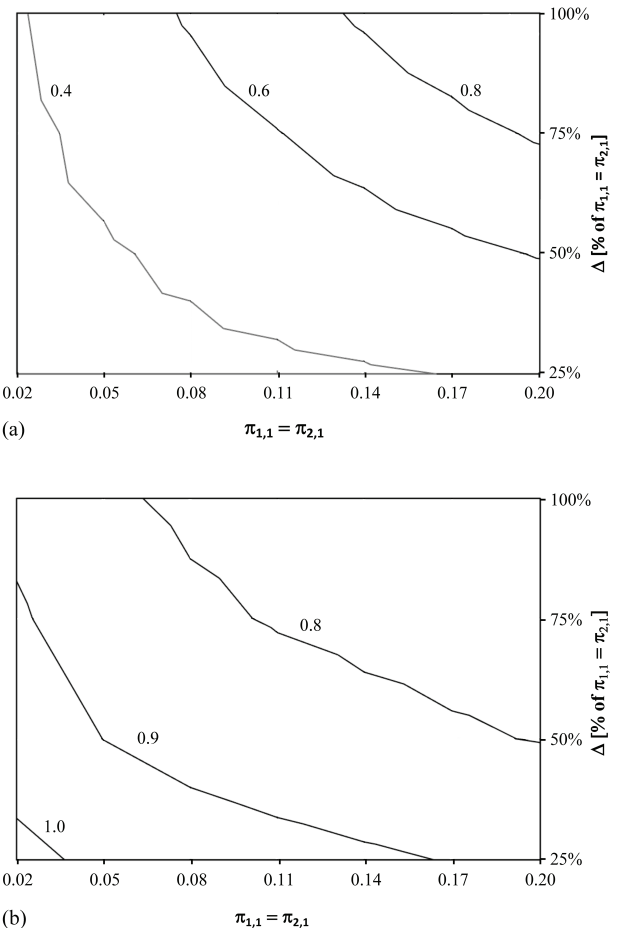

FIGURE 6. Contour Plots of Relative MSE vs. Pass Rates $\pi_{1,1} = \pi_{2,1}$ and Size of Step Change. (a) Relative MSE = $\text{MSE}_{\text{WEE}}/\text{MSE}_{\text{naïve},\lambda\to 1}$; (b) Relative MSE = $\text{MSE}_{\text{WEE}}/\text{MSE}_{\text{naïve},\lambda\to 0}$.

Figure 6 shows that the WEE estimator has lower MSE than either of the naïve estimators for most of the values of $\pi_{1,1} = \pi_{2,1}$ and Δ. The advantage of the WEE estimator over the estimator based on present-time data only is more important when the change in the parameter is small and present-time sample size is small. The advantage of the WEE estimator based on all historical data weighted equally is more pronounced for larger changes in the parameter. We

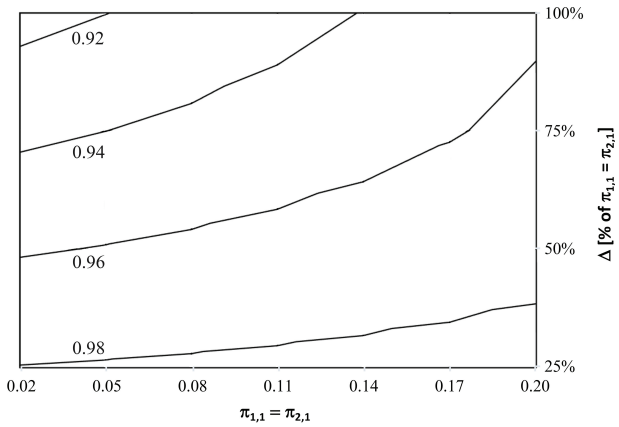

FIGURE 7. Contour Plot of $e(\tilde{\pi}, \Delta)$ by $\pi_{1,1} = \pi_{2,1}$ and Δ.

see that the WEE estimator provides a trade-off between bias and variance relative to the two naïve approaches for this simple example.

For this simple example, we also calculate variance by the distributions of the random variables $\{Y_{1,t}, Y_{2,t}, t = 1, \ldots, T\}$. At the present time $T = 10$, the variance of estimator $\tilde{\pi}$ is

$$\begin{aligned}
&\mathrm{var}_{\mathrm{dist}}(\tilde{\pi}, \Delta) \\
&\quad = \mathrm{var}\left(\frac{\sum_{t=1}^{T} w_t Y_{2,t}}{\sum_{t=1}^{T} w_t n_{2,t}} - \frac{\sum_{t=1}^{T} w_t Y_{1,t}}{\sum_{t=1}^{T} w_t n_{1,t}} \right) \\
&\quad = \frac{\sum_{t=1}^{T} w_t^2 \pi_{1,1}(1-\pi_{1,1})}{n_1} \\
&\qquad + \frac{\sum_{t=1}^{5} w_t^2 \pi_{2,1}(1-\pi_{2,1})}{n_2} \\
&\qquad + \frac{\sum_{t=6}^{10} w_t^2 (\pi_{2,1}+\Delta)(1-\pi_{2,1}-\Delta)}{n_2}.
\end{aligned}$$

We compare the two variances by the ratio of standard deviations, which we denote as $e(\tilde{\pi}, \Delta) = \sqrt{\mathrm{var}_{\mathrm{WI}}(\tilde{\pi}, \Delta)} / \sqrt{\mathrm{var}_{\mathrm{dist}}(\tilde{\pi}, \Delta)}$. Figure 7 gives a contour plot of the values of $e(\tilde{\pi}, \Delta)$ for the WEE estimator with weight parameter $\lambda = 0.1$.

Figure 7 shows that $\mathrm{var}_{\mathrm{WI}}(\tilde{\pi}; \Delta)$ and the variance based on the distributions of $\{Y_{1,t}, Y_{2,t}, t = 1, \ldots, T\}$ are close for these values of $\pi_{1,1} = \pi_{2,1}$ and Δ. We see that the weighted information variance using WEE estimates is a good estimate of variance for this simple example, especially when there is a small change in the parameter.

iii. the distribution of the weight-adjusted random variable $\tilde{S}$ in Equation (6) is approximately χ^2_r.

At time t, consider a test of null hypothesis H_0: $\pi = 0$ versus the alternative H_A: $\pi \neq 0$. To follow, we show by properties of the random variables that

$$E\left(\frac{\sum_{t=1}^{T} w_t n_t}{\sum_{t=1}^{T} w_t^2 n_t} \tilde{S} \right) = 1$$

under the null hypothesis, which agrees with the first moment of the distribution in Equation (6). We validate the second and third moments and 95th percentile of the distribution in Equation (6) through comparison with approximate distributions based on simulated data. Table 4 confirms that the approximate distribution

$$\frac{\sum_{t=1}^{T} w_t n_t}{\sum_{t=1}^{T} w_t^2 n_t} \tilde{S} \overset{\text{approx}}{\rightarrow} \chi^2_1$$

is suitable when N is very large ($N = 1 \times 10^7$) and a useful approximation when N is small ($N = 100$).

We approximate a distribution for $\tilde{S}$ in order to test a hypothesis based on test statistic $\hat{S}$. The random variable $\tilde{S}$ in terms of random variables $Y_{m,t}$, sample sizes $n_{m,t}$, and weights w_t, $t = 1, \ldots, T$, $m = 1, 2$ is

$$\begin{aligned}
\tilde{S} = 2 \sum_{m=1}^{2} \sum_{t=1}^{T} w_t \Bigg(& Y_{m,t} \log \frac{\tilde{\pi}_m}{\tilde{\pi}_{\mathrm{null}}} \\
& + (n_{m,t} - Y_{m,t}) \log \left(\frac{1-\tilde{\pi}_m}{1-\tilde{\pi}_{\mathrm{null}}} \right) \Bigg).
\end{aligned}$$

We approximate $\tilde{S}$ through second-order Taylor series approximations for those terms involving loga-

TABLE 4. Moments of Approximate Distributions of Weight-adjusted Hypothesis Test Statistic

Approximate distribution	Mean	Variance	Skew	95th percentile
$(\sum_{t=1}^{T} w_t n_t / \sum_{t=1}^{T} w_t^2 n_t)\tilde{S} \sim \chi^2_1$	1.000	2.000	2.828	3.841
Simulated distribution with $N = 1 \times 10^7$	1.000	2.008	2.852	3.860
Simulated distribution with $N = 100$	1.019	2.086	2.869	3.914

```
PROC GENMOD data=SAMPLE_DATA order=internal descending;
     class case;
     weight weights;
     MODEL y = x1 x2 / expected dist=binomial;
     repeated subject=case / type=ind ecovb;
     ods output GEEEmpPEst=theta_est GEERCov=covmatrix_est;
RUN;
```

FIGURE 8. SAS Code for WEE Analysis of an Example Dataset.

rithms of the random variables:

- $\log(x)$ for $\sum_{t=1}^{T} w_t Y_{m,t}$ around $\sum_{t=1}^{T} w_t \pi_m n_{m,t}$ for $m = 1, 2$.
- $\log(x)$ for $\sum_{m=1}^{2}\sum_{t=1}^{T} w_t Y_{m,t}$ around $[(\pi_1 + \pi_2)/2]\sum_{m=1}^{2}\sum_{t=1}^{T} w_t n_{m,t}$.

We find the expected value of the approximation for $\tilde{S}$ based on the assumptions

$$Y_{1,t} \sim \text{Binomial}(n_{1,t}, \pi_1),$$
$$Y_{2,t} \sim \text{Binomial}(n_{2,t}, \pi_2),$$

and $Y_{1,t}$, $Y_{2,t}$ independent for each t. Under the null hypothesis with $\pi = \pi_1 - \pi_2 = 0$, then

$$E[\tilde{S}] \approx \frac{\sum_{t=1}^{T} w_t^2 n_t}{\sum_{t=1}^{T} w_t n_t}$$

and

$$E\left[\frac{\sum_{t=1}^{T} w_t n_t}{\sum_{t=1}^{T} w_t^2 n_t}\tilde{S}\right] \approx 1.$$

We validate higher moments of the distribution of $\hat{S}$ through simulation. We consider the empirical distribution of $\hat{S}$ for 100,000 datasets that are generated with $T = 10$, $\pi_1 = \pi_2 = 0.04$, $\lambda = 0.1$, and $n_{1,t} = n_1$, $n_{2,t} = n_2$ for all t. We repeat the simulation study for large $N = 1\times10^7$ and small $N = 100$. Table 4 gives the empirical moments of the distributions of $\hat{S}$.

Table 4 shows that the empirical distributions based on simulation are close to the approximate distribution for this simple example under the select conditions.

Appendix C
SAS Implementation

The weighted estimating equations in Equation (2) can be solved in most regression programs that allow for weights. In SAS, the weighted estimating equations can be solved using PROC GENMOD. Details on this procedure and other resources to use SAS are available at "Resources to help you learn and use SAS" (n.d.). Consider an example dataset called SAMPLE_DATA with one row for each subject that is observed. The dataset contains fields for an index 'case', covariate values 'x_1, x_2', $\{w_t\}$ 'weights', and outcome 'y'. The parameter to estimate includes elements for the mean outcome for a baseline subject and two covariate effects, $\theta_T = (\alpha_T, \beta_{1,T}, \beta_{2,T})$. The SAS statements to estimate $\theta = \theta_T$ by the WEE approach assuming a binomial generalized linear model with a logit link function for SAMPLE_DATA are given in Figure 8. The SAS PROC GENMOD routine also provides the weighted information estimate of the variance of $\hat{\theta}$ given in Equation (3).

The convenience of the existing software functionality for solving the weighted estimating equations makes it convenient to implement the WEE approach and update the estimates over time.

References

Grewal, M. S. and Andrews, A. P. (2014). *Kalman Filtering: Theory and Practice with Matlab*, 4th edition. Hoboken, NJ: Wiley-IEEE Press.

Hu, F. and Zidek, J. V. (2002). "The Weighted Likelihood". *The Canadian Journal of Statistics* 30(4), pp. 347–371.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*, 3rd edition. New York, NY: Springer.

Liu, X.; MacKay, R. J.; and Steiner, S. H. (2008). "Monitoring Multiple Stream Processes". *Quality Engineering* 20, pp. 296–308.

Markey, R.; Reichheld, F. F.; and Dullweber, A. (2013). "A Test of Customer Loyalty". *Bain & Company*, available at www.bain.com/publications/articles/a-test-of-customer-loyalty-smeinfo.aspx.

Montgomery, D. C. (2013). *Introduction to Statistical Quality Control*, 7th edition. Hoboken, NJ: John Wiley & Sons.

NPS Benchmarks (n.d.). *CustomerGauge*, available at www.npsbenchmarks.com.

Reichheld, F. F. and Markey, R. (2011). *The Ultimate Question 2.0: How Net Promoter Companies Thrive in a Customer-Driven World*. Boston, MA: Harvard Business Press.

SAS (n.d.). *Institute for Digital Research and Education UCLA*, available at ats.ucla.edu/stat/sas.

Satmetrix (2015). "Analytics and Reporting". In *Satmetrix Systems, Inc.*, available at http://cdn2.hubspot.net/hub/268441/file-2465864085-pdf/Website_PDF_Downloads/Satmetrix_Analytics_Reporting.pdf.

Small, C. G. (2010). *Expansions and Asymptotics for Statistics*. Boca Raton, FL: Chapman & Hall/CRC.

Steiner, S. H. and MacKay, R. J. (2014). "Monitoring Risk-Adjusted Medical Outcomes Allowing for Changes over Time". *Biostatistics* 15(5), pp. 665–676.

Wald, A. (1949). "Note on the Consistency of the Maximum Likelihood Estimate". *The Annals of Mathematical Statistics* 20(5), pp. 595–601.

White, H. (1982). "Maximum Likelihood Estimation of Misspecified Models". *Econometrica* 50, pp. 1–25.