Bias/Variance Trade-off in Estimates of a Process Parameter based on Temporal Data

Dr. Tricia Barfoot

ASQ/ASA Fall Technical Conference

October 5, 2018

Overview

- Motivation: Net Promoter Score
- General Problem
- Weighted Estimating Equations approach
- Results and Impact
- Summary

"The Ultimate Question"

"How likely are you to recommend us to a colleague or friend?"

Detractors (0–6) **Passive** (7–8) **Promoters** (9–10)

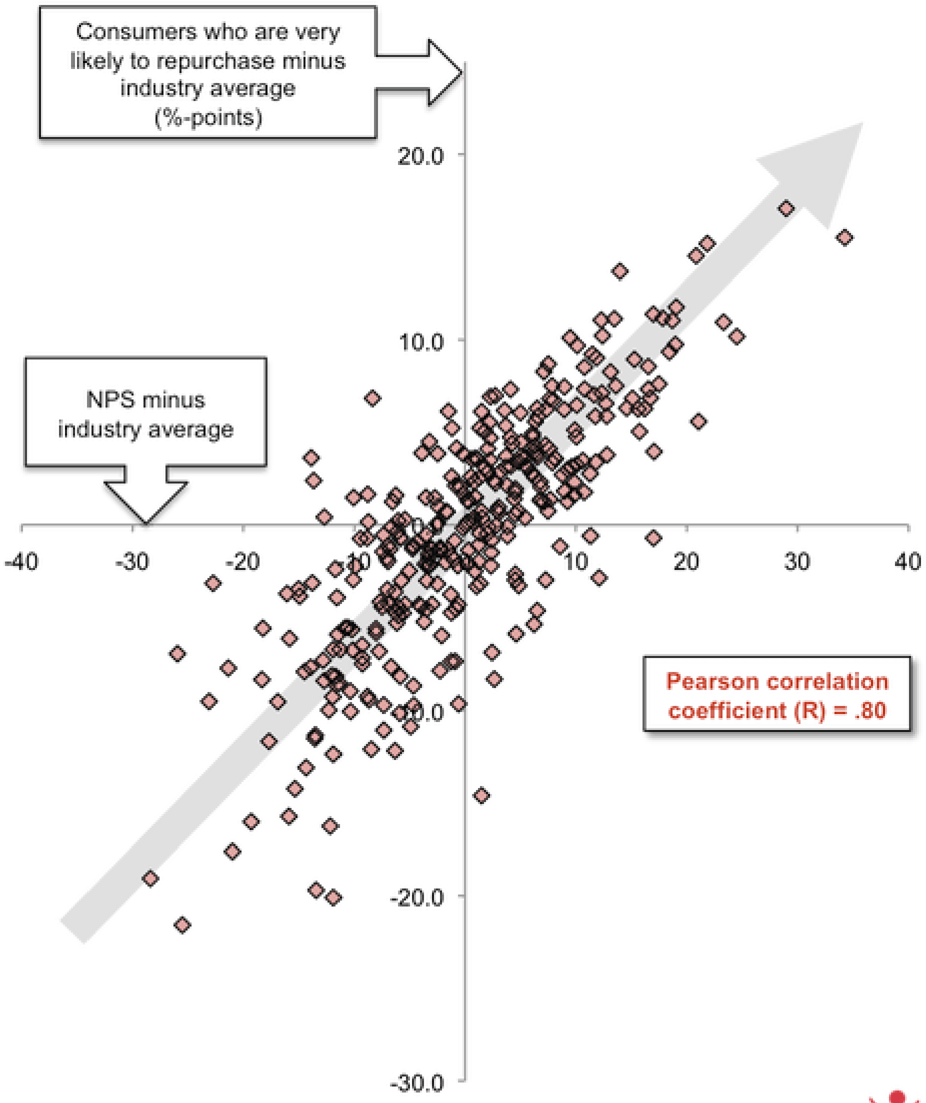
0 1 2 3 4 5 6 7 8 9 10

Not at all likely (0) — **Neutral** (5) — **Extremely likely** (10)

$$\text{Net Promoter Score} = \frac{\#\ \text{promoters} - \#\ \text{detractors}}{\#\ \text{respondents}}$$

NPS Correlates To Future Purchase Intentions

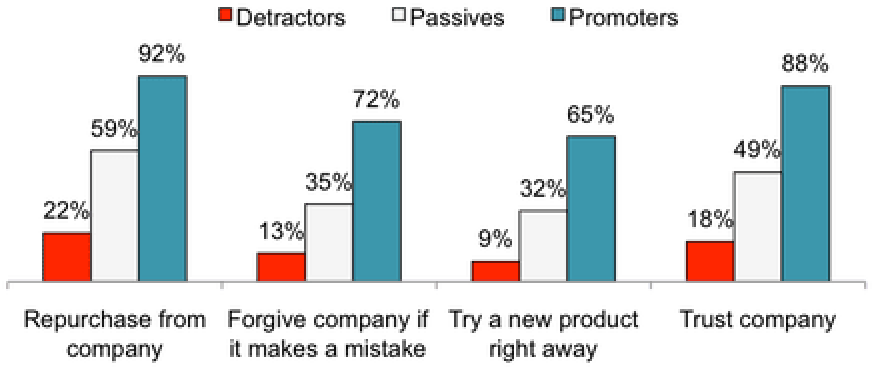
331 Organizations Across 20 Industries

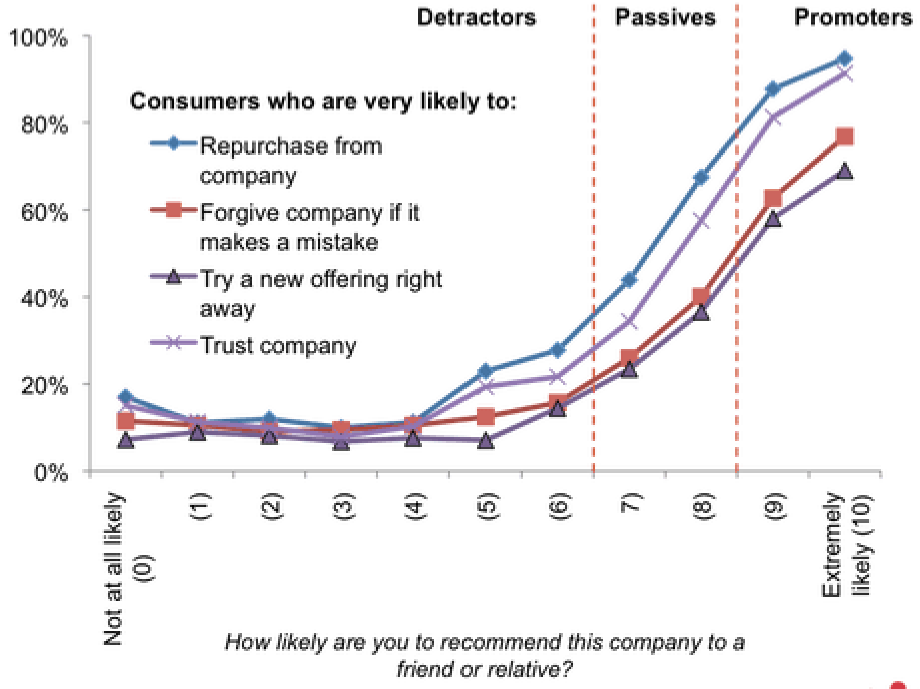


Base: 10,000 U.S. consumers, 331 companies across 20 industries
Source: Temkin Group Q1 2017 Consumer Benchmark Survey

Value of Promoters, Passives, and Detractors Across Industries

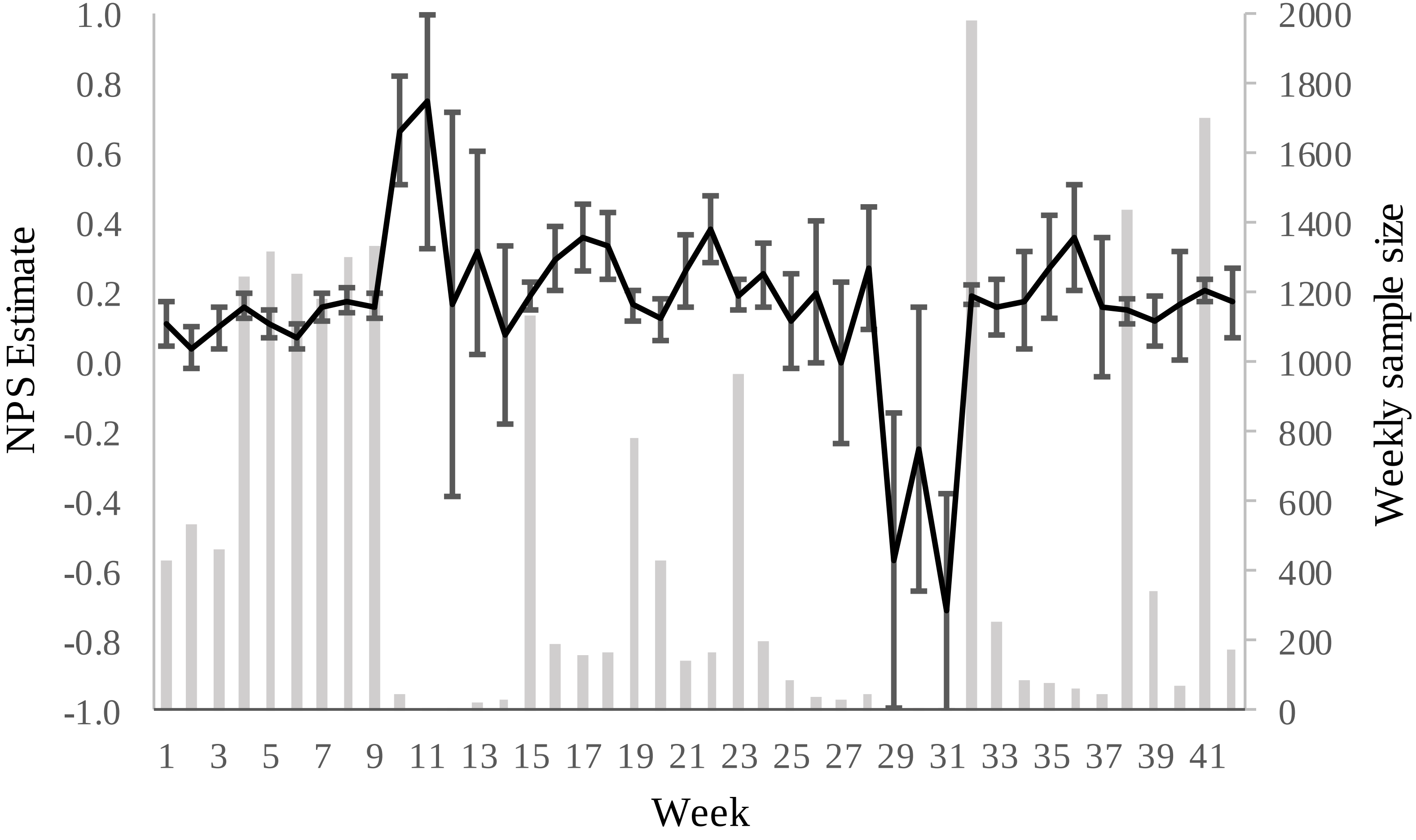
Loyalty of consumers across 20 industries



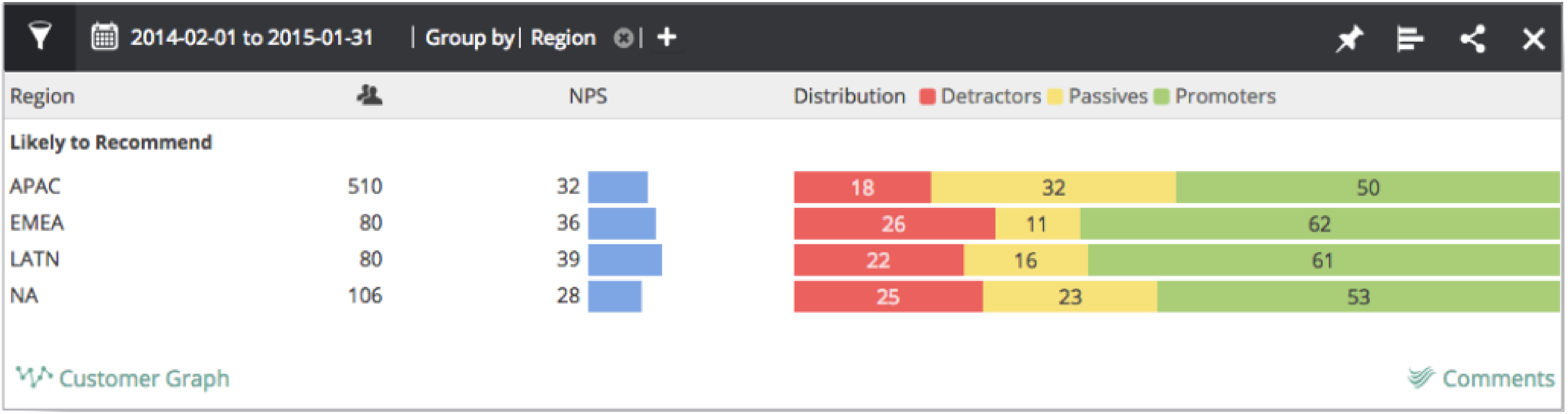


Base: 1,106,750 interactions by U.S. consumers
Source: Temkin Group Q1 2017 Consumer Benchmark Survey

Estimate NPS and monitor over time



Compare NPS across groups



Region		NPS	Distribution: Detractors	Passives	Promoters
Likely to Recommend					
APAC	510	32	18	32	50
EMEA	80	36	26	11	62
LATN	80	39	22	16	61
NA	106	28	25	23	53

General Problem

From a stream of data from a process over time, with subject-level covariates:

- Monitor the estimate of a process parameter over time
- Monitor a test of statistical hypothesis involving present estimates of the parameter multiple streams

Data and Model

- Subject $j = 1, \ldots, n_t$ at time $t = 1, \ldots, T$
- Data d_t: response y_{jt}, independent; covariates x_{jt}
- Likelihood function, $\mathcal{L}_t(d_t; \theta_t)$ involving parameter θ_t

where n_t may be small and θ_t may drift over time, unpredictably

Objectives:

- Estimate parameter at time T: θ_T with an estimate of uncertainty
- Test H_0: $\theta^{(r)} = \theta_0^{(r)}$
- Track estimates $\hat{\theta}_T$ or hypothesis test statistic over time

Bias/Variance Trade-off

Use present time data only	Solve $\frac{\partial l(y_T \mid \theta_T)}{\partial \theta_T} = 0$ for $\hat{\theta}_T$	Low bias High uncertainty
Use all historical data (naïve)	Solve $\sum_{t=1}^{T} \frac{\partial l(y_t \mid \theta_T)}{\partial \theta_T} = 0$ for $\hat{\theta}_T$	High bias Low uncertainty

Weighted Estimating Equations (WEE)

$$Q(\hat{\theta}; y, x, w) = \left[w_1 \frac{\partial \ell_1(y_1; \hat{\theta}, x_1)}{\partial \theta} + w_2 \frac{\partial \ell_2(y_2; \hat{\theta}, x_2)}{\partial \theta} + \cdots + w_T \frac{\partial \ell_T(y_T; \hat{\theta}, x_T)}{\partial \theta} \right] = [0]_{p \times 1}$$

- WEE estimate $\hat{\theta}$ is the solution of this estimating equation for estimating θ_T
- Involves declining weights $w_T, w_{T-1}, \dots, w_1$
- Naïve estimates are special cases at particular values of the weights:
  - Present data only: $w_1, w_2, \dots, w_{T-1} = 0, w_T = 1$
  - All data weighted equally: $w_1 = w_2 =, \dots, = w_T$

Select weights

- Exponentially declining weights reflect slow change in parameter over time

$$w_t = \frac{\lambda(1-\lambda)^{T-t}}{\sum_{t=1}^{T} \lambda(1-\lambda)^{T-t}}$$

- Specify weight parameter $\lambda \in (0,1)$
- Effective sample size: $N_{eff} = \frac{\left(\sum_{t=1}^{T} w_t n_t\right)^2}{\sum_{t=1}^{T} w_t^2 n_t}$
- Depending on the selection of weights, $n_T < N_{eff} < \sum_{t=1}^{T} n_t$

Weighted Information Estimate of Variance

- An **approximate estimate** of the variance of WEE estimate $\hat{\theta}$,
    - assuming no change in θ_t over time
    - based on usual asymptotic properties of score function, $\frac{\partial \ell_t(\theta_t; d_t)}{\partial \theta}$
- Depends on expected information matrix, $I_t(\theta) = -E\left(\frac{\partial^2 \ell_t(\theta_t; \mathcal{D}_t)}{\partial \theta^2}\right)$

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

Distribution of Hypothesis Test Statistic

- For null hypothesis $H_0: \theta^{(r)} = \theta_0^{(r)}$,

  partition $\theta = \left(\theta^{(r)}, \theta^{(u)}\right)$ and estimate θ and $\theta^{(u)}$ by the WEE approach

- **WEE LR test statistic:** $\hat{S} = 2\left(\sum_{t=1}^{T} w_t \ell_t(\hat{\theta}; d_t) - \sum_{t=1}^{T} w_t \ell_t\left(\theta_0^{(r)}, \hat{\theta}^{(u)}; d_t\right)\right)$

- For r-dimensional δ,

$$\frac{\sum_{t=1}^{T} w_t n_t}{\sum_{t=1}^{T} w_t^2 n_t} \hat{S} \overset{\textit{approx}}{\sim} \chi_r^2$$

- Special cases of the weight parameter give usual result, $\hat{S} \overset{approx}{\sim} \chi_r^2$

Exponentially Weighted Moving Average and WEE estimates

- Consider the simple example where $Y_t \sim bin(n_t, \theta_t)$, independent

EWMA Solve, then weight	$\hat{\theta}_{EWMA} = \sum_{t=1}^{T} w_t \hat{\theta}_t$ $= w_1 \bar{y}_1 + w_2 \bar{y}_2 + \cdots + w_T \bar{y}_T$
WEE Weight, then solve	$\hat{\theta}_{WEE} = \frac{w_1 n_1}{\sum_{t=1}^{T} w_t n_t} \bar{y}_1 + \frac{w_2 n_2}{\sum_{t=1}^{T} w_t n_t} \bar{y}_2 + \cdots + \frac{w_T n_T}{\sum_{t=1}^{T} w_t n_t} \bar{y}_T$

- WEE estimates reflect differences in sample size by time period

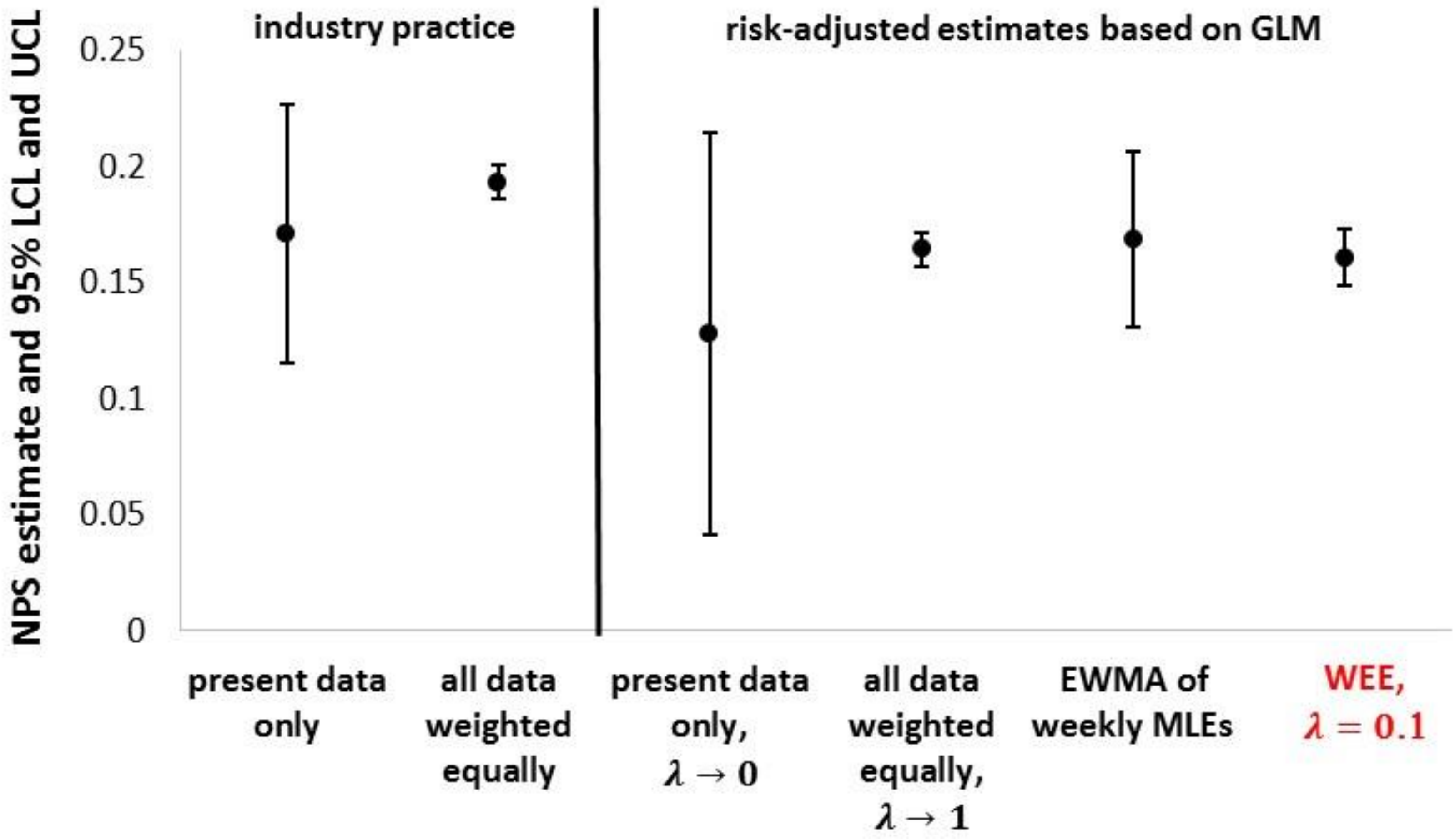
Exponentially Weighted Moving Average and WEE estimates

- Consider the simple example where $Y_t \sim bin(n_t, \theta_t)$, independent

EWMA Solve, then weight	$\widehat{var}_{EWMA} = \left(\frac{w_1^2}{n_1} + \frac{w_2^2}{n_2} + \cdots + \frac{w_T^2}{n_T}\right) \hat{\theta}_{EWMA}(1 - \hat{\theta}_{EWMA})$
WEE Weight, then solve	$\widehat{var}_{WEE} = \frac{w_1^2 n_1 + w_2^2 n_2 + \cdots + w_T^2 n_T}{\left(\sum_{t=1}^{T} w_t n_t\right)^2} \hat{\theta}_{WEE}\left(1 - \hat{\theta}_{WEE}\right)$

- WEE estimates of covariate effects are more precise

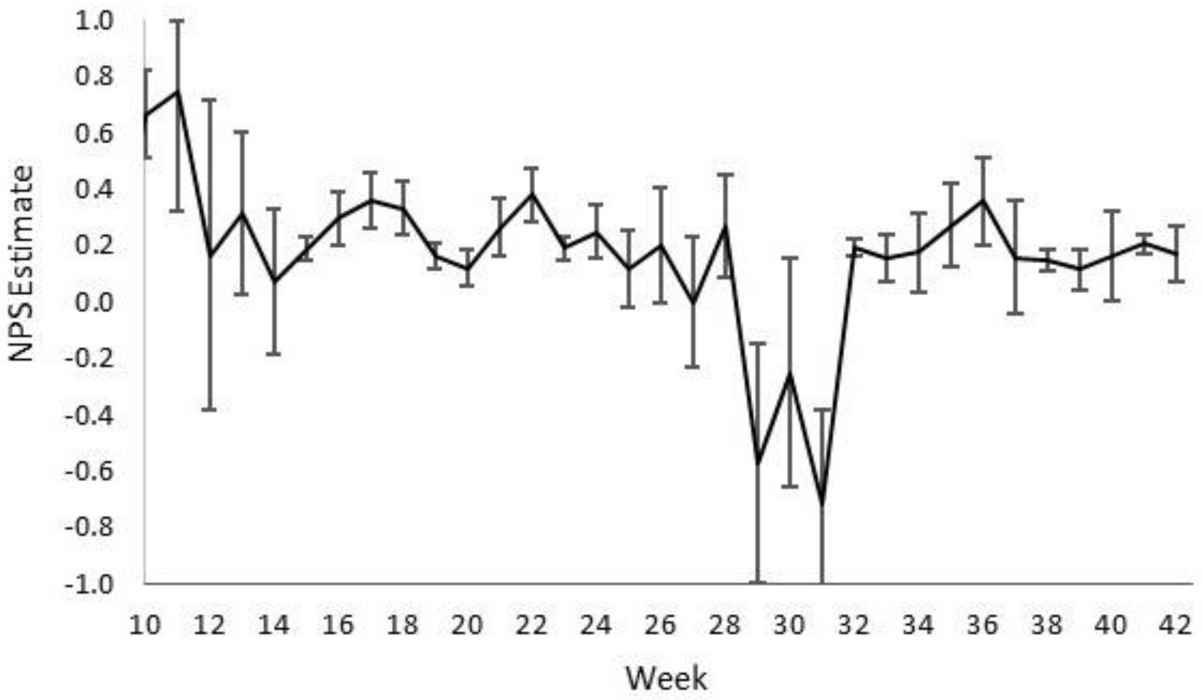
Estimates of Field Population NPS in week 42



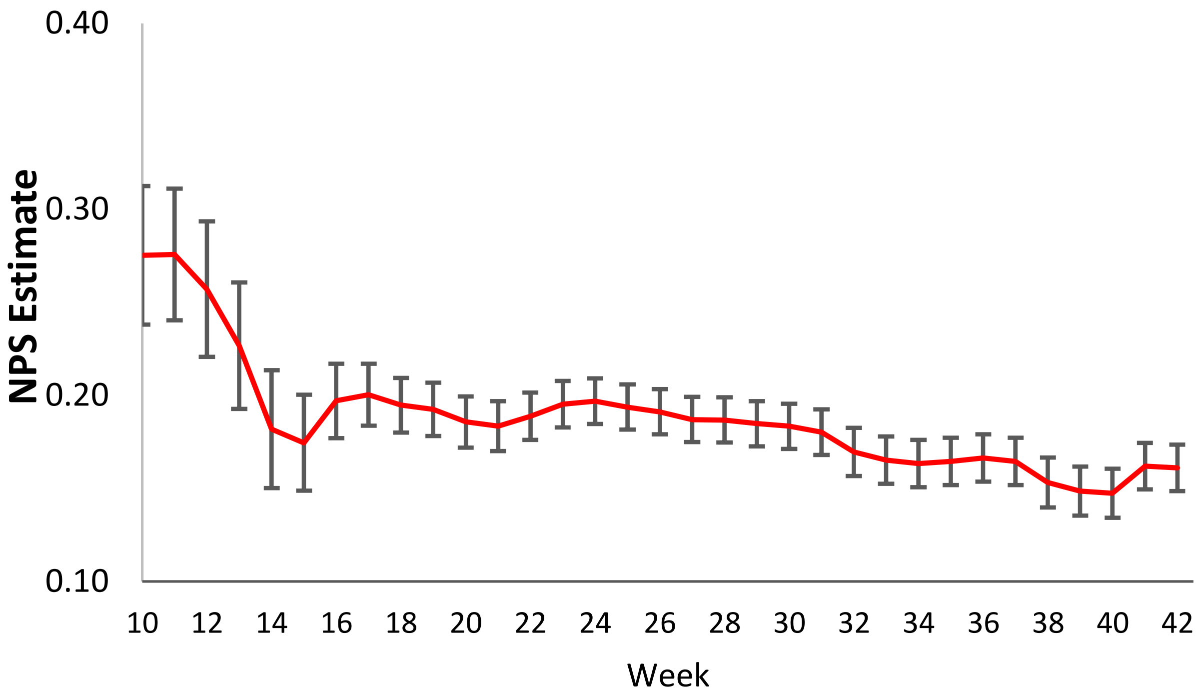
- Realistic dataset
- 2 customer-level covariates
- Multinomial GLM with proportional odds property
- $\lambda = 0.1$
- Standard population is field population

Trends in Field Population NPS

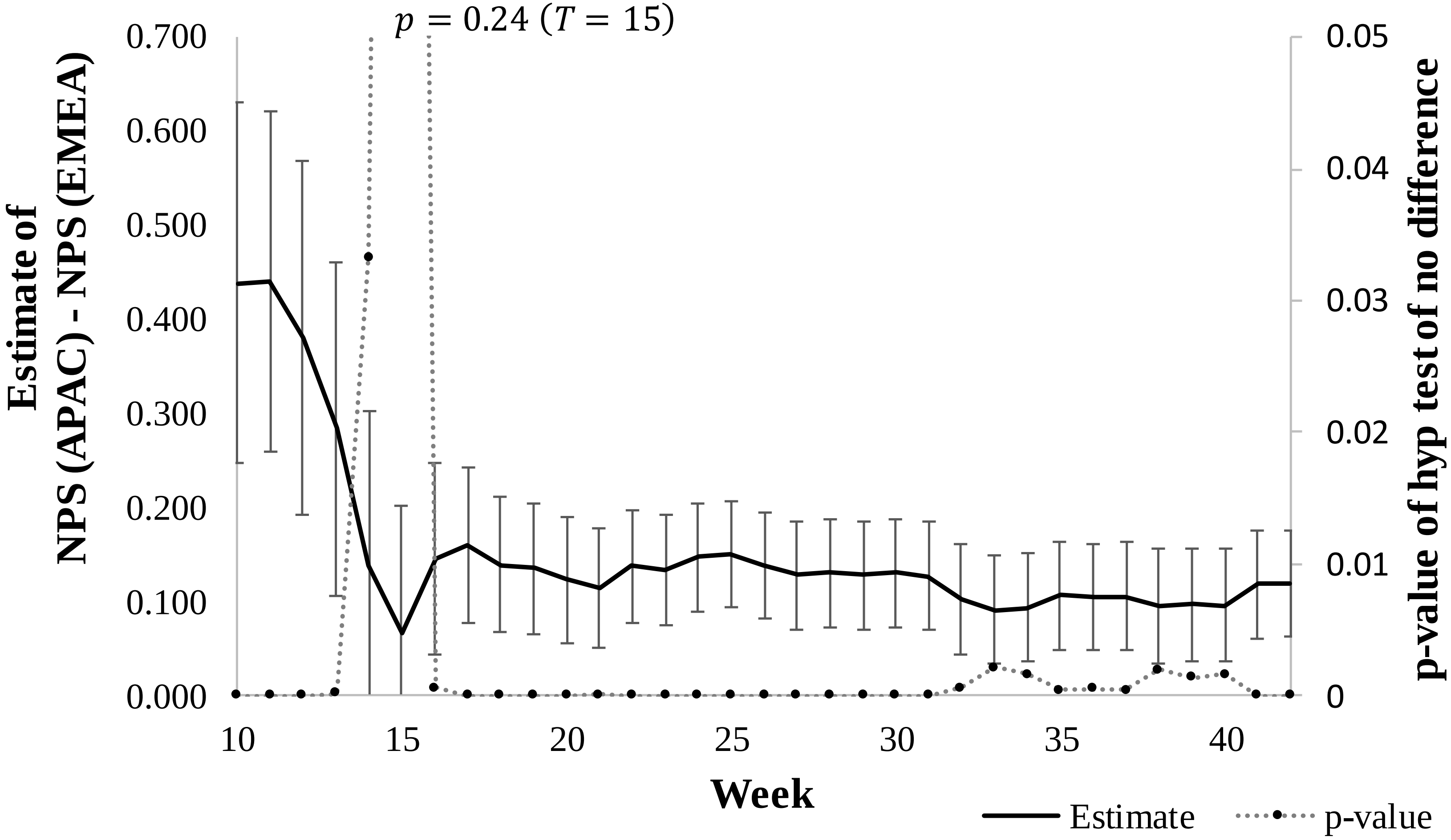
Sample proportion estimates using present data only



risk-adjusted WEE estimates, $\lambda = 0.1$



Trend in comparison between two groups



Implementation

- SAS code to compute WEE estimate, WI estimate of variance, and WEE LR hypothesis test statistic
- Selection of historical time window
- Missing data and sampling zeros
- Selection of time subgroups and weight parameter
- Some large sample sizes
- Known covariate effects or covariate effects that are fixed over time

Summary

- When some sample sizes may be small and the model parameter may change slowly over time in an unpredictable way,

  we need to **regulate a trade-off between bias and variance in the estimate of present performance** based on data over time

> the **Weighted Estimating Equations** approach is an intuitive solution; extends to other problems of importance

- Future work will demonstrate the importance of this work and guide implementation in marketing and healthcare communities