Published in final edited form as:

J Am Stat Assoc. 2016 ; 111(516): 1538–1541. doi:10.1080/01621459.2016.1243479.

Comment

Min Qian

Department of Biostatistics, Columbia University, New York, NY, USA

Abstract

This comment deals with issues related to the article by Chen, Zeng, and Kosorok. We present several potential modifications of the outcome weighted learning approach. Those modifications are based on truncated l_2 loss. One advantage of l_2 loss is that it is differentiable everywhere, which makes it more stable and computationally more tractable.

Keywords

Double robustness; Epanechnikov kernel; Personalized treatment

1. Introduction

We congratulate Chen, Zeng, and Kosorok (hereafter, CZK) for a stimulating and interesting article on the important topic of personalized dose finding. We found the article is enjoyable to read, and we thank the editors for the opportunity to discuss the article. Personalized medicine is an emerging area in medical research. It holds the great potential to improve the quality of patient care. With recent advances in biomedical science, massive amounts of data have been produced on individual patients. How to use high-dimensional data to design personalized treatment is the key to success. Several methods have been proposed to deal with high-dimensional data in the case of limited treatment options (e.g., Qian and Murphy 2011; Zhao et al. 2012; Lu, Zhang, and Zeng 2013). Those methods, however, as discussed by CZK, cannot be directly applied when the number of treatment options is infinite (e.g., in the dose finding problem). CZK developed a novel outcome weighted learning method for personalized dose finding. They substituted the weighted indicator loss in the original optimization problem with a truncated l_1 loss, and used an l_2 penalty to address the overfitting problem. An efficient optimization algorithm was also provided to facilitate computation. In our discussion, we present several modifications of the outcome weighted learning approach. Those modifications are based on truncated l_2 loss. One advantage of l_2 loss is that it is differentiable everywhere, which makes it more stable and computationally more tractable. The proposed modifications are intended as a way to demonstrate the potential of the machine learning framework proposed by CZK.

2. Preliminaries

We adopt the same notations as in CZK. Assume we have n iid trajectories of $(\mathbf{X}, A, R)$, where $\mathbf{X} = (X_1, \dots, X_d)^T \in \mathcal{X}$ is patient-level covariates, A is the assigned treatment dose taking values in a bounded interval $\mathcal{A}$, and R is a scalar “reward,” with large values representing better outcomes. For any individualized dose rule (IDR) $f: \mathcal{X} \rightarrow \mathcal{A}$, the value of f , $\mathcal{V}(f)$, is defined as the expected reward if f is implemented in the study population. The optimal IDR, f^{pt} , is the dose rule that yields maximal expected reward, that is, $f^{\text{pt}} = \arg \max_f \mathcal{V}(f)$. Denote $Q(\mathbf{x}, a) \triangleq E[R|\mathbf{X} = \mathbf{x}, A = a]$. Let $p(a|\mathbf{X})$ be the randomization probability of $A = a$ given $\mathbf{X}$. CZK showed that $\mathcal{V}(f) = E_{\mathbf{X}}[Q(\mathbf{X}, f(\mathbf{X}))] = \lim_{\phi \rightarrow 0^+} \tilde{\mathcal{V}}_{\phi}(f)$, where

$$\tilde{\mathcal{V}}_{\phi}(f) \triangleq E \left(\frac{R}{\phi p(A|\mathbf{X})} \frac{1}{2} I \{ |A - f(\mathbf{X})| \leq \phi \} \right). \quad (1)$$

Note that maximizing $\tilde{\mathcal{V}}_{\phi}(f)$ is computationally intractable due to discontinuity of the 0 – 1 loss. To address this difficulty, CZK proposed to use a surrogate truncated l_1 loss, yielding approximated value function

$$\mathcal{V}_{\phi}(f) \triangleq E \left(\frac{R}{\phi p(A|\mathbf{X})} \max \left\{ 1 - \frac{|A - f(\mathbf{X})|}{\phi}, 0 \right\} \right). \quad (2)$$

Their Theorem 1 showed that $|\mathcal{V}_{\phi}(f) - \mathcal{V}(f)| \leq C\phi$ under mild conditions. Below we extend this result to a general class of loss functions. For any measurable function $g: \mathbb{R} \rightarrow [0, \infty)$, IDR $f: \mathcal{X} \rightarrow \mathcal{A}$ and $\phi > 0$, denote

$$\mathcal{V}_{g,\phi}(f) \triangleq E \left[\frac{R}{\phi p(A|\mathbf{X})} g \left(\frac{A - f(\mathbf{X})}{\phi} \right) \right].$$

We have the following theorem.

Theorem 1

Suppose

$$E_{\mathbf{X}} \left\{ \sup_{a, a' \in \mathcal{A}, a \neq a'} \left| \frac{Q(\mathbf{X}, a') - Q(\mathbf{X}, a)}{a' - a} \right| \right\} = O(1). \quad (3)$$

Assume $g: \mathbb{R} \rightarrow [0, \infty)$ satisfies $\int g(z) dz = 1$ and $\int |z| g(z) dz = O(1)$. Then for any individualized dose rule $f: \mathcal{X} \rightarrow \mathcal{A}$ and $\phi > 0$, there exists a constant $C > 0$ such that $|\mathcal{V}_{g,\phi}(f) - \mathcal{V}(f)| \leq C\phi$.

Proof—First note that

$$\begin{aligned}\mathcal{V}_{g,\phi}(f) &= E \left[\frac{R}{\phi p(A|\mathbf{X})} g \left(\frac{A-f(\mathbf{X})}{\phi} \right) \right] \\ &= E_{\mathbf{X}} \left\{ \frac{1}{\phi} \int Q(\mathbf{X}, a) g \left(\frac{a-f(\mathbf{X})}{\phi} \right) da \right\} \\ &= E_{\mathbf{X}} \left\{ \int Q(\mathbf{X}, \phi z + f(\mathbf{X})) g(z) dz \right\}.\end{aligned}$$

Since $\mathcal{V}(f) = E_{\mathbf{X}}[Q(\mathbf{X}, f(\mathbf{X}))]$, under the condition that $\int g(z) dz = 1$ and $g(z) \geq 0$ for all $z \in \mathbb{R}$, we have

$$\begin{aligned}|\mathcal{V}_{g,\phi}(f) - \mathcal{V}(f)| &= |E_{\mathbf{X}} \{ \int Q(\mathbf{X}, \phi z + f(\mathbf{X})) g(z) dz - \int Q(\mathbf{X}, f(\mathbf{X})) g(z) dz \}| \\ &\leq E_{\mathbf{X}} \left\{ \int |Q(\mathbf{X}, \phi z + f(\mathbf{X})) - Q(\mathbf{X}, f(\mathbf{X}))| g(z) dz \right\} \\ &\leq E_{\mathbf{X}} \left\{ \sup_{a, a' \in \mathcal{A}, a \neq a'} \left| \frac{Q(\mathbf{X}, a') - Q(\mathbf{X}, a)}{a' - a} \right| \int |\phi z| g(z) dz \right\}, \\ &\leq C \phi,\end{aligned}$$

where the last equality follows from conditions (3) and $\int |z| g(z) dz = O(1)$.

Remarks

1. Condition (3) is a mild Lipschitz-type condition. It is easy to verify that this condition holds in the simulation scenarios presented in Section 5.
2. Note that any density function $g(\cdot)$ of a square integrable random variable satisfies $\int g(z) dz = 1$ and $\int |z| g(z) dz = O(1)$. As $\phi \rightarrow 0^+$, $|\mathcal{V}_{\phi}(f) - \mathcal{V}(f)| \rightarrow 0$. To ensure a good approximation of the original indicator function in finite samples, it is natural to consider densities that are symmetric around 0. In another word, $g(\cdot)$ can be viewed as a kernel function. Indeed, the indicator loss in (1) uses the uniform kernel, and the truncated l_1 loss in (2) corresponds to the triangular kernel.

3. Learning IDR with Truncated l_3 Loss

In this section, we consider the truncated l_2 loss, corresponding to the Epanechnikov kernel, that is, $g(u) = \frac{3}{4} \max(1 - u^2, 0)$. Similar to the truncated l_1 loss, the optimization problem can be solved using DC algorithm. In addition, since it is differentiable everywhere in the compact support, an explicit parameter updating formula can be derived. Denote

$$\overline{\mathcal{V}}_{\phi}(f) \triangleq E \left(\frac{3R}{4\phi p(A|\mathbf{X})} \max \left\{ 1 - \frac{[A-f(\mathbf{X})]^2}{\phi^2}, 0 \right\} \right). \quad (4)$$

Note that choosing f to maximize $\overline{\mathcal{V}}_{\phi}(f)$ is equivalent to minimizing

$$\overline{\mathcal{R}}_\phi(f) = E \left(\frac{3R}{4\phi p(A|\mathbf{X})} \min \left\{ \frac{[A - f(\mathbf{X})]^2}{\phi^2}, 1 \right\} \right).$$

For the reason discussed in CZK, we assume $R > 0$ without loss of generality. Consider IDR of the form $f(\mathbf{x}; \boldsymbol{\beta}) = \Phi(\mathbf{x})^\top \boldsymbol{\beta}$, where $\Phi(\mathbf{x})$ is a vector of basis functions of $\mathbf{x}$. For example, $\Phi(\mathbf{x}) = (1, \mathbf{x}^\top)^\top$ represents a linear model of $\mathbf{x}$. Denote $W_i = \frac{R_i}{p(A_i|\mathbf{X}_i)}$, $i = 1, \dots, n$. The parameters $\boldsymbol{\beta}$ can be estimated by minimizing

$$\hat{\mathcal{R}}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n (W_i \min \{ [A_i - \Phi(\mathbf{X}_i)^\top \boldsymbol{\beta}]^2, \phi_n^2 \}) + \lambda_n \|\boldsymbol{\beta}\|^2,$$

where $\phi_n > 0$ and $\lambda_n > 0$ are tuning parameters. ϕ_n measures the closeness of the surrogate loss to the original indicator loss, and λ_n controls the model complexity. It is easy to see that $\hat{\mathcal{R}}(\boldsymbol{\beta})$ can be written as the difference of two convex functions $\hat{\mathcal{R}}(\boldsymbol{\beta}) = \hat{\mathcal{R}}_1(\boldsymbol{\beta}) - \hat{\mathcal{R}}_2(\boldsymbol{\beta})$, where

$$\begin{aligned} \hat{\mathcal{R}}_1(\boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^n (W_i [A_i - \Phi(\mathbf{X}_i)^\top \boldsymbol{\beta}]^2) + \lambda_n \|\boldsymbol{\beta}\|^2 \\ \text{and } \hat{\mathcal{R}}_2(\boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^n W_i ([A_i - \Phi(\mathbf{X}_i)^\top \boldsymbol{\beta}]^2 - \phi_n^2)_+. \end{aligned}$$

Using DC algorithm, we estimate $\boldsymbol{\beta}$ by first initializing $\boldsymbol{\beta}^{(0)}$, then repeatedly updating $\boldsymbol{\beta}$ via

$$\boldsymbol{\beta}^{(t+1)} = \arg \min_{\boldsymbol{\beta}} \left[\hat{\mathcal{R}}_1(\boldsymbol{\beta}) - [\Delta \hat{\mathcal{R}}_2(\boldsymbol{\beta}^{(t)})]^\top (\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)}) \right] \quad (5)$$

until convergence, where $\Delta \hat{\mathcal{R}}_2(\boldsymbol{\beta})$ is the subgradient of $\hat{\mathcal{R}}_2(\boldsymbol{\beta})$. Define the index set

$\Omega_{\phi_n}^{(t)} = \{i=1, \dots, n: |A_i - \Phi(\mathbf{X}_i)^\top \boldsymbol{\beta}^{(t)}| \leq \phi_n\}$. After algebraic simplification, (5) is equivalent to

$$\begin{aligned} \boldsymbol{\beta}^{(t+1)} &= \arg \min_{\boldsymbol{\beta}} \left\{ \sum_{i \in \Omega_{\phi_n}^{(t)}} W_i [A_i - \Phi(\mathbf{X}_i)^\top \boldsymbol{\beta}]^2 + \sum_{i \in \{1, \dots, n\} \setminus \Omega_{\phi_n}^{(t)}} W_i [\Phi(\mathbf{X}_i)^\top (\boldsymbol{\beta} - \boldsymbol{\beta}^{(t)})]^2 + n \lambda_n \|\boldsymbol{\beta}\|^2 \right\} \\ &= \left(n \lambda_n + \sum_{i=1}^n W_i \Phi(\mathbf{X}_i) \Phi(\mathbf{X}_i)^\top \right)^{-1} \times \left(\sum_{i \in \Omega_{\phi_n}^{(t)}} W_i A_i \Phi(\mathbf{X}_i) + \sum_{i \in \{1, \dots, n\} \setminus \Omega_{\phi_n}^{(t)}} W_i \Phi(\mathbf{X}_i) \Phi(\mathbf{X}_i)^\top \boldsymbol{\beta}^{(t)} \right). \end{aligned} \quad (6)$$

4. A Doubly Robust Estimate

The above procedure assumes that the treatment assignment distribution $p(a|\mathbf{X})$ is known or can be estimated consistently. In the case of finite treatment options, an augmented inverse probability weighted estimator of $\mathcal{V}(f)$ has been provided (Zhang et al. 2012). This estimate offers protection against model mis-specification of $p(a|\mathbf{X})$. It is doubly robust in the sense that the resulting estimate is consistent as long as $p(a|\mathbf{X})$ or $Q(\mathbf{x}, a)$ is correctly specified. Below we present a doubly robust estimate of $\mathcal{V}(f)$ in the dose finding setting. Since $E[R - Q(\mathbf{X}, A)|\mathbf{X}, A] = 0$, the value of an IDR f can be written as

$$\mathcal{V}(f) = E[Q(\mathbf{X}, f(\mathbf{X}))] + E \left[\frac{R - Q(\mathbf{X}, A)}{\phi p(A|\mathbf{X})} g_E \left(\frac{A - f(\mathbf{X})}{\phi} \right) \right]$$

for any $\phi > 0$, where $g_E(z)$ denotes the Epanechnikov kernel. Note that the IDR that maximizes $\mathcal{V}(f)$ does not change if R is replaced by $R + c$ for any constant c in the above display. For any $f: \mathcal{X} \rightarrow \mathcal{A}$, $\tilde{Q}: \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$, $\tilde{p}: \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}^+$ and $\phi > 0$, define

$$\mathcal{V}_\phi^D(f; \tilde{Q}, \tilde{p}) \triangleq E[\tilde{Q}(\mathbf{X}, f(\mathbf{X}))] + E \left[\frac{[R - \tilde{Q}(\mathbf{X}, A)]}{\phi \tilde{p}(\mathbf{X}, A)} g_E \left(\frac{A - f(\mathbf{X})}{\phi} \right) \right].$$

Below we show that $\mathcal{V}_\phi^D(f; \tilde{Q}, \tilde{p})$ is a good approximation of $\mathcal{V}(f)$ when $\tilde{Q}(\mathbf{x}, a) = Q(\mathbf{x}, a)$ or $\tilde{p}(\mathbf{x}, a) = p(a|\mathbf{x})$.

Theorem 2

Suppose $\tilde{Q}: \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$ satisfies

$$E_{\mathbf{X}} \left\{ \sup_{a, a' \in \mathcal{A}, a \neq a'} \left| \frac{\tilde{Q}(\mathbf{X}, a') - \tilde{Q}(\mathbf{X}, a)}{a' - a} \right| \right\} = O(1). \quad (7)$$

For any IDR $f: \mathcal{X} \rightarrow \mathcal{A}$ and $\phi > 0$, we have (i) $\mathcal{V}_\phi^D(f; \tilde{Q}, \tilde{p}) = \mathcal{V}(f)$ for any $\tilde{p}: \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}^+$; and (ii) there exists a positive constant C such that $|\mathcal{V}_\phi^D(f; \tilde{Q}, \tilde{p}) - \mathcal{V}(f)| \leq C\phi$.

Proof—(i) follows from the fact that $E[R - Q(\mathbf{X}, A)|\mathbf{X}, A] = 0$. For (ii), note that $\mathcal{V}_\phi^D(f; \tilde{Q}, \tilde{p})$ can be decomposed as

$$E \left(\tilde{Q}(\mathbf{X}, f(\mathbf{X})) \left[1 - \frac{g_E((A - f(\mathbf{X}))/\phi)}{\phi p(A|\mathbf{X})} \right] \right) + E \left[\frac{[\tilde{Q}(\mathbf{X}, f(\mathbf{X})) - \tilde{Q}(\mathbf{X}, A)]}{\phi p(A|\mathbf{X})} g_E \left(\frac{A - f(\mathbf{X})}{\phi} \right) \right] + \overline{\mathcal{V}}_\phi(f),$$

where $\bar{\mathcal{V}}_\phi(f)$ is defined in (4). The first term in the above display is zero since $\int g_E(z) dz = 1$. Thus,

$$\begin{aligned} |\mathcal{V}_\phi^D(f; \tilde{Q}, p) - \mathcal{V}(f)| &= \left| E \left[\frac{[\tilde{Q}(\mathbf{X}, f(\mathbf{X})) - \tilde{Q}(\mathbf{X}, A)]}{\phi p(A|\mathbf{X})} g_E \left(\frac{A - f(\mathbf{X})}{\phi} \right) \right] + \bar{\mathcal{V}}_\phi(f) - \mathcal{V}(f) \right| \\ &\leq \left| E \left[\frac{[\tilde{Q}(\mathbf{X}, f(\mathbf{X})) - \tilde{Q}(\mathbf{X}, A)]}{\phi p(A|\mathbf{X})} g_E \left(\frac{A - f(\mathbf{X})}{\phi} \right) \right] \right| + |\bar{\mathcal{V}}_\phi(f) - \mathcal{V}(f)|. \end{aligned}$$

Using similar techniques as in the proof of Theorem 1, it is easy to verify that the first term in the above display is $O(\phi)$ under condition (7). The result follows from Theorem 1 immediately.

The above theorem suggest that as long as $Q(\mathbf{x}, a)$ or $p(a|\mathbf{x})$ is consistently estimated, maximizing an empirical version of $\mathcal{V}_\phi^D$ will give us a high quality IDR.

Again consider IDR of the form $f(\mathbf{x}, \boldsymbol{\beta}) = \Phi(\mathbf{x})^\top \boldsymbol{\beta}$. We propose to estimate $\boldsymbol{\beta}$ by minimizing

$$\hat{\mathcal{R}}^D(\boldsymbol{\beta}) = -\frac{\phi_n^3}{n} \sum_{i=1}^n \hat{Q}(\mathbf{X}_i, \Phi(\mathbf{X}_i)^\top \boldsymbol{\beta}) + \frac{1}{n} \sum_{i=1}^n [\hat{W}_i \min \{ (A_i - \Phi(\mathbf{X}_i)^\top \boldsymbol{\beta})^2, \phi_n^2 \}] + \lambda_n \|\boldsymbol{\beta}\|^2,$$

where $\hat{Q}(\mathbf{x}, a)$ and $\hat{p}(\mathbf{x}, a)$ are estimates of $Q(\mathbf{x}, a)$ and $p(a|\mathbf{x})$, respectively, $\hat{W}_i = [R_i - \hat{Q}(\mathbf{X}_i, A_i) + c]/\hat{p}(\mathbf{X}_i, A_i)$, and c is a constant so that $W_i \geq 0$ for $i = 1, \dots, n$. ϕ_n and λ_n are tuning parameters. To make the optimization problem computationally tractable, we only consider $\hat{Q}(\mathbf{x}, a)$ that is differentiable and either convex or concave (e.g., linear or quadratic in a). The problem can be solved using DC algorithm as discussed in Section 3.

5. Numerical Studies

In this section, we conduct simulation studies to evaluate the performance of methods proposed in previous sections. In the simulation below the tuning parameter ϕ_n is fixed, and λ_n is selected using cross-validation.

We consider four examples. Scenarios 1 and 2 are the same as those presented in CZK, where the treatment assignment distribution $A \sim U[0, 2]$ is known. Scenarios 3 and 4 are the same as scenarios 1 and 2, respectively, except that $p(a|\mathbf{x})$ is truncated normal ranging from $(0, 2)$ with mean $(-0.5 + 0.5X_1 + 0.5X_2)1_{X_3 < 0} + |0.5 + 1.5X_2|1_{X_3 \geq 0}$ and standard deviation $1_{X_3 < 0}/2 + 1_{X_3 \geq 0}$, and is assumed to be unknown.

Three methods are evaluated: Linear O-learning with l_1 loss as proposed in CZK (LOL- l_1), O-learning with l_2 loss (LOL- l_2), and doubly robust estimate with l_2 loss (DR- l_2). In all methods, we consider linear IDRs of the form $f(\mathbf{X}; \boldsymbol{\beta}) = \beta_0 + \mathbf{X}^\top \boldsymbol{\beta}$. In addition, when $p(a|\mathbf{x})$ is unknown, we estimate it using the boosting algorithm from Zhu, Coffman, and Ghosh (2015) as suggested in CZK. For the DR- l_2 method, we estimate $Q(\mathbf{X}, A)$ using LASSO with model containing covariates $(\mathbf{X}, A, \mathbf{X}^2, \mathbf{X}A, A^2)$. Thus, $Q(\mathbf{X}, A)$ is correctly specified in scenarios 1 and 3.

Table 1 provides mean (SD) for the values of the estimated dose rules over 500 replicated samples, with sample size ranging from 50 to 800. The “true” values were computed from simulated datasets of size $N = 5000$. In scenario 1, LOL- I_1 produces better dose rules when the sample size is small ($n = 50$). As n increases, the l_2 loss-based methods outperform LOL- I_1 . All methods have similar performance in scenario 2. In scenarios 3 and 4 where the propensity score $p(a|\mathbf{x})$ is unknown, the DR- l_2 method clearly works better than the other two methods with moderate to large sample sizes. This evidences the fact that a good estimate of $Q(\mathbf{X}, A)$ helps reducing the potential bias in the doubly robust value estimator, and thus improves the quality of the estimated dose rule.

6. Discussion

Dose finding is of critical importance in drug development. CZK proposed a novel machine learning framework for personalized dose finding using data from a randomized trial. This is an important contribution to the field of dose finding research.

It is worth noting that in practice most dose finding trials are sequential—the dose assigned to subsequent patients depends on previous patients’ outcomes (e.g., Yuan, Nguyen, and Thall 2016). In addition, for treatments of chronic diseases where a drug will be used for a long period of time, the dose is usually repeatedly adjusted within each subject based on his/her ongoing response. We would be interested in hearing CZK’ opinion on how to adapt machine learning techniques to these types of data.

Acknowledgments

Funding

The author gratefully acknowledges support from NIH Grants R21MH108999, R01MH109496, R01GM095722, R01DA039901, and NSF Grant DMS-1307838.

References

- Lu W, Zhang H, Zeng D. Variable Selection for Optimal Treatment Decision. *Statistical Methods in Medical Research*. 2013; 22:493–504. [PubMed: 22116341]
- Qian M, Murphy SA. Performance Guarantees for Individualized Treatment Rules. *Annals of Statistics*. 2011; 39:1180–1210. [PubMed: 21666835]
- Yuan, Y., Nguyen, H., Thall, P. *Bayesian Designs for Phase III Clinical Trials*. Boca Raton, FL: Chapman and Hall; 2016.
- Zhang B, Tsiatis A, Laber E, Davidian M. A Robust Method for Estimating Optimal Treatment Regimes. *Biometrics*. 2012; 68:1010–1018. [PubMed: 22550953]
- Zhao Y, Zeng D, Rush A, Kosorok M. Estimating Individualized Treatment Rules Using Outcome Weighted Learning. *Journal of the American Statistical Association*. 2012; 107:1106–1118. [PubMed: 23630406]
- Zhu Y, Coffman D, Ghosh D. A Boosting Algorithm for Estimating Generalized Propensity Scores with Continuous Treatments. *Journal of Causal Inference*. 2015; 3:25–40. [PubMed: 26877909]

Table 1

Mean (SD) of $\mathcal{V}(\hat{\theta})$ based on 500 replications with n ranging from 50 to 800.

	n	LOL- l_1	LOL- l_2	DR- l_2
Scenario 1	50	5.12 (1.47)	4.67 (0.81)	4.59 (2.26)
	100	3.26 (1.96)	5.48 (0.79)	7.64 (0.35)
	200	6.12 (0.56)	7.01 (0.31)	7.90 (0.05)
	400	7.15 (0.23)	7.58 (0.13)	7.97 (0.02)
	800	7.58 (0.11)	7.80 (0.06)	8.00 (0.01)
Scenario 2	50	7.58 (0.11)	7.80 (0.06)	8.00 (0.01)
	100	1.83 (1.21)	2.04 (0.83)	2.12 (0.73)
	200	2.14 (0.75)	2.32 (0.47)	2.38 (0.41)
	400	2.34 (0.46)	2.50 (0.24)	2.55 (0.19)
	800	2.47 (0.27)	2.59 (0.12)	2.63 (0.09)
Scenario 3	50	3.31 (2.76)	3.25 (1.72)	-0.74 (8.76)
	100	2.58 (2.02)	4.58 (1.14)	5.03 (4.94)
	200	5.35 (0.72)	5.89 (0.54)	7.83 (0.12)
	400	6.02 (0.41)	6.46 (0.36)	7.93 (0.04)
	800	6.20 (0.28)	6.72 (0.25)	7.95 (0.03)
Scenario 4	50	1.29 (0.79)	1.63 (0.59)	1.64 (0.97)
	100	0.85 (0.99)	1.45 (0.67)	1.41 (0.97)
	200	1.13 (0.58)	1.60 (0.49)	1.73 (0.56)
	400	1.20 (0.38)	1.68 (0.34)	2.11 (0.34)
	800	1.25 (0.27)	1.69 (0.23)	2.26 (0.16)