The Annals of Statistics
2007, Vol. 35, No. 3, 1166–1182
DOI: 10.1214/009053606000001424

ASYMPTOTIC PROPERTIES OF COVARIATE-ADJUSTED RESPONSE-ADAPTIVE DESIGNS

BY LI-XIN ZHANG,[1] FEIFANG HU,[2] SIU HUNG CHEUNG[3] AND WAI SUM CHAN[3]

Zhejiang University, University of Virginia, Chinese University of Hong Kong and Chinese University of Hong Kong

Response-adaptive designs have been extensively studied and used in clinical trials. However, there is a lack of a comprehensive study of response-adaptive designs that include covariates, despite their importance in clinical trials. Because the allocation scheme and the estimation of parameters are affected by both the responses and the covariates, covariate-adjusted response-adaptive (CARA) designs are very complex to formulate. In this paper, we overcome the technical hurdles and lay out a framework for general CARA designs for the allocation of subjects to K (≥ 2) treatments. The asymptotic properties are studied under certain widely satisfied conditions. The proposed CARA designs can be applied to generalized linear models. Two important special cases, the linear model and the logistic regression model, are considered in detail.

1. Preliminaries.

1.1. *Brief history.* In most clinical trials, patients accrue sequentially. Response-adaptive designs provide allocation schemes that assign different treatments to incoming patients based on the previous observed responses of patients. A major objective of response-adaptive designs in clinical trials is to construct a randomized treatment allocation scheme in order to minimize the number of patients that are, assigned to the inferior treatments to a degree that still allows statistical inference with high power. The ethical and other characteristics of response-adaptive designs have been extensively discussed by many authors (e.g., [27]).

Early important work on response-adaptive designs was carried out by Thompson [23] and Robbins [18]. Since then, a steady stream of research [10, 14, 24, 25] in this area has generated various treatment allocation schemes for clinical trials. Some of the advantages of using response-adaptive designs have been recently studied by Hu and Rosenberger [13] and Rosenberger and Hu [20].

Received August 2004; revised August 2006.

[1]Supported by a grant from the National Natural Science Foundation of China (No. 10471126).

[2]Supported by NSF Grant DMS-03-49048.

[3]Supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (Project no. CUHK400204).

AMS 2000 subject classifications. Primary 60F15, 62G10; secondary 60F05, 60F10.

Key words and phrases. Adaptive designs, asymptotic normality, clinical trial, covariate information, generalized linear model, logistic regression.

In many clinical trials, covariate information is available that has a strong influence on the responses of patients. For instance, the efficacy of a hypertensive drug is related to a patient's initial blood pressure and cholesterol level, whereas the effectiveness of a cancer treatment may depend on whether the patient is a smoker or a nonsmoker.

The following notation and definitions are introduced to describe the randomized treatment allocation schemes. Consider a clinical trial with K treatments. Let $\mathbf{X}_1, \mathbf{X}_2, \dots$ be the sequence of random treatment assignments. For the mth subject, $\mathbf{X}_m = (X_{m,1}, \dots, X_{m,K})$ represents the assignment of treatment such that if the mth subject is allocated to treatment k, then all elements in $\mathbf{X}_m$ are 0 except for the kth component, $X_{m,k}$, which is 1. Let $N_{m,k}$ be the number of subjects assigned to treatment k in the first m assignments and write $\mathbf{N}_m = (N_{m,1}, \dots, N_{m,K})$. Then $\mathbf{N}_m = \sum_{i=1}^m \mathbf{X}_i$. Suppose that $\{Y_{m,k},\ k = 1, \dots, K,\ m = 1, 2, \dots\}$ denote the responses such that $Y_{m,k}$ is the response of the mth subject to treatment k, $k = 1, \dots, K$. In practical situations, only $Y_{m,k}$ with $X_{m,k} = 1$ is observed. Denote $\mathbf{Y}_m = (Y_{m,1}, \dots, Y_{m,K})$. Let $\mathcal{X}_m = \sigma(\mathbf{X}_1, \dots, \mathbf{X}_m)$ and $\mathcal{Y}_m = \sigma(\mathbf{Y}_1, \dots, \mathbf{Y}_m)$ be the corresponding sigma fields. A response-adaptive design is defined by

$$\boldsymbol{\psi}_m = E(\mathbf{X}_m | \mathcal{X}_{m-1}, \mathcal{Y}_{m-1}).$$

Now, assume that covariate information is available in the clinical study. Let $\boldsymbol{\xi}_m$ be the covariate of the mth subject and $\mathcal{Z}_m = \sigma(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_m)$ be the corresponding sigma field. In addition, let $\mathcal{F}_m = \sigma(\mathcal{X}_m, \mathcal{Y}_m, \mathcal{Z}_m)$ be the sigma field of the history. A general covariate-adjusted response-adaptive (CARA) design is defined by

$$\boldsymbol{\psi}_m = E(\mathbf{X}_m | \mathcal{F}_{m-1}, \boldsymbol{\xi}_m) = E(\mathbf{X}_m | \mathcal{X}_{m-1}, \mathcal{Y}_{m-1}, \mathcal{Z}_m),$$

the conditional probabilities of assigning treatments $1, \dots, K$ to the mth patient, conditioning on the entire history including the information of all previous $m - 1$ assignments, responses, and covariate vectors, plus the information of the current patient's covariate vector.

A number of attempts have been made to formulate adaptive designs in the presence of covariates. For example, Zelen [26] and Pocock and Simon [17] considered balancing covariates by using the idea of the biased coin design [11]. Atkinson [1–3] tackled this problem by employing the D-optimality criterion with a linear model. The prime concern of these works is to balance allocations over the covariates with treatment assignment probabilities

$$\boldsymbol{\psi}_m = E(\mathbf{X}_m | \mathcal{X}_{m-1}, \mathcal{Z}_m),$$

which differs from the CARA designs. These allocation schemes do not depend on the outcome of the treatment, which is important for adaptive designs that aim to reduce the number of patients that receive the inferior treatment.

The history of incorporating covariates in response-adaptive designs is short. For the randomized play-the-winner rule, Bandyopadhyay and Biswas [8] incorporated polytomous covariates with binary responses. Rosenberger, Vidyashankar and Agarwal [22] considered a CARA design for binary responses that uses a logistic regression model. Their encouraging simulation study indicates that their approach, together with the inclusion of the covariates, significantly reduces the percentage of treatment failures. However, theoretical justification and asymptotic properties have not been given. Further, the applications of their procedure are limited to two treatments with binary responses.

To compare two treatments, Bandyopadhyay and Biswas [9] considered a linear model to utilize covariate information with continuous responses. A limiting allocation proportion was also derived in their design. However, according to their proposed scheme, the conditional assignment probabilities are

$$\boldsymbol{\psi}_m = E(\mathbf{X}_m | \mathcal{F}_{m-1}).$$

The above probabilities do not incorporate the covariates of the incoming patient, which in some cases are crucial. For instance, let the covariate be gender and let there be two treatments. If males and females react very differently to treatments A and B, whether the next patient is male or female is an important element in the assignment of treatment. Recently, Atkinson [4] considered adaptive biased-coin designs for K treatments based on a linear regression model. Atkinson and Biswas ([5] and [6]) proposed adaptive biased-coin designs and Bayesian adaptive biased-coin designs for clinical trials with normal responses. However, none of these articles provided the asymptotic distribution of the estimators and the allocation proportions. Without the asymptotic properties of the estimators, it is often difficult to assess the validity of the statistical inference after using CARA designs.

Instead of working on specific setups, we seek to derive a general framework of CARA designs and provide theoretical foundation for these designs. The technical complexity arises because the assignment of treatment $\mathbf{X}_m$ depends on $\mathcal{F}_{m-1}$ and the covariate information ($\boldsymbol{\xi}_m$) of the incoming patient.

1.2. *Main results and organization of the paper.* The main objectives are (i) to propose a general CARA design that can be applied to cases in which K treatments ($K \geq 2$) are present and to different types of responses (discrete or continuous), and (ii) to study asymptotic properties of the proposed design. These properties provide a solid foundation for both the CARA design and its related statistical inference. Major mathematical techniques, including martingale theory and Gaussian approximation, are employed to develop the asymptotic results.

The rest of the paper is organized as follows. In Section 2 we introduce the general framework of the CARA design. Under this general framework, we are able to derive many new response-adaptive designs. Useful asymptotic results (including strong consistency and asymptotic normality) for both the estimators of the unknown parameters and the allocation proportions are derived. The generalized linear model represents a broad class of applications and is an important tool in the analysis of data that involve covariates. In Section 3 the proposed design is applied to generalized linear models, and two special cases, the linear model and the logistic regression model, are considered in detail. We then conclude our paper with some observations in Section 4. Technical proofs are provided in the Appendix.

2. General CARA design.

2.1. *General framework.* Based on the notation in Section 1, suppose that a patient with a covariate vector $\boldsymbol{\xi}$ is assigned to treatment k, $k = 1, \ldots, K$, and the observed response is Y_k. Assume that the responses and the covariate vector satisfy

$$\mathsf{E}[Y_k|\boldsymbol{\xi}] = p_k(\boldsymbol{\theta}_k, \boldsymbol{\xi}), \qquad \boldsymbol{\theta}_k \in \boldsymbol{\Theta}_k, k = 1, \ldots, K,$$

where $p_k(\cdot,\cdot)$, $k = 1, \ldots, K$, are known functions. Further, $\boldsymbol{\theta}_k$, $k = 1, \ldots, K$, are unknown parameters, and $\boldsymbol{\Theta}_k \subset \mathbb{R}^d$ is the parameter space of $\boldsymbol{\theta}_k$. Write $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$ and $\boldsymbol{\Theta} = \boldsymbol{\Theta}_1 \times \cdots \times \boldsymbol{\Theta}_K$. This model is quite general, and includes the generalized linear models of [16] as special cases. The discussion of the generalized linear models is undertaken in Section 3. We assume that $\{(Y_{m,1}, \ldots, Y_{m,K}, \boldsymbol{\xi}_m),\ m = 1, 2, \ldots\}$ is a sequence of i.i.d. random vectors, the distributions of which are the same as that of $(Y_1, \ldots, Y_K, \boldsymbol{\xi})$.

2.2. *CARA design.* The allocation scheme is as follows. To start, assign m_0 subjects to each treatment by using a restricted randomization. Assume that m ($m \geq Km_0$) subjects have been assigned to treatments. Their responses $\{\mathbf{Y}_j,\ j = 1, \ldots, m\}$ and the corresponding covariates $\{\boldsymbol{\xi}_j,\ j = 1, \ldots, m\}$ are observed. We let $\widehat{\boldsymbol{\theta}}_m = (\widehat{\boldsymbol{\theta}}_{m,1}, \ldots, \widehat{\boldsymbol{\theta}}_{m,K})$ be an estimate of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$. Here, for each $k = 1, \ldots, K$, $\widehat{\boldsymbol{\theta}}_{m,k} = \widehat{\boldsymbol{\theta}}_{m,k}(Y_{j,k}, \boldsymbol{\xi}_j : X_{j,k} = 1, j = 1, \ldots, m)$ is the estimator of $\boldsymbol{\theta}_k$ that is based on the observed $N_{m,k}$-size sample $\{(Y_{j,k}, \boldsymbol{\xi}_j)$: for which $X_{j,k} = 1, j = 1, \ldots, m\}$. Next, when the $(m+1)$st subject is ready for randomization and the corresponding covariate $\boldsymbol{\xi}_{m+1}$ is recorded, we assign the patient to treatment k with probability

$$\psi_k = \mathsf{P}\big(X_{m+1,k} = 1 \big| \mathscr{F}_m, \boldsymbol{\xi}_{m+1}\big) = \pi_k(\widehat{\boldsymbol{\theta}}_m, \boldsymbol{\xi}_{m+1}), \qquad k = 1, \ldots, K, \tag{2.1}$$

where $\mathscr{F}_m = \sigma(\mathbf{X}_1, \ldots, \mathbf{X}_m, \mathbf{Y}_1, \ldots, \mathbf{Y}_m, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)$ is the sigma field of the history and $\pi_k(\cdot,\cdot)$, $k = 1, \ldots, K$, are some given functions. Given $\mathscr{F}_m$ and $\boldsymbol{\xi}_{m+1}$, the response $\mathbf{Y}_{m+1}$ of the $(m+1)$st subject is assumed to be independent of its assignment $\mathbf{X}_{m+1}$. We call the function $\boldsymbol{\pi}(\cdot,\cdot) = (\pi_1(\cdot,\cdot), \ldots, \pi_K(\cdot,\cdot))$ the allocation function that satisfies $\pi_1 + \cdots + \pi_K \equiv 1$. Let $g_k(\boldsymbol{\theta}^*) = \mathsf{E}[\pi_k(\boldsymbol{\theta}^*, \boldsymbol{\xi})]$. From (2.1), it follows that

$$\mathsf{P}(X_{m+1,k} = 1 \big| \mathscr{F}_m) = g_k(\widehat{\boldsymbol{\theta}}_m), \qquad k = 1, \ldots, K. \tag{2.2}$$

Different choices of $\boldsymbol{\pi}(\cdot,\cdot)$ generate different possible classes of useful designs. For example, we can take $\pi_k(\boldsymbol{\theta}, \boldsymbol{\xi}) = R_k(\boldsymbol{\theta}_1\boldsymbol{\xi}^T, \ldots, \boldsymbol{\theta}_K\boldsymbol{\xi}^T)$, $k = 1, \ldots, K$, which includes a large class of applications. Here, $0 < R_k(\mathbf{z}) < 1$, $k = 1, \dots, K$, are real functions that are defined in $\mathbb{R}^K$ with

$$\sum_{k=1}^{K} R_k(\mathbf{z}) = 1 \quad \text{and} \quad R_i(\mathbf{z}) = R_j(\mathbf{z}) \qquad \text{whenever } z_i = z_j. \tag{2.3}$$

For simplicity, it is assumed that $\boldsymbol{\xi}$ and $\boldsymbol{\theta}_k$, $k = 1, \dots, K$, have the same dimensions; otherwise slight modifications are necessary (see Example 3.1 for an illustration). In practice, the functions R_k can be defined as

$$R_k(\mathbf{z}) = \frac{G(z_k)}{G(z_1) + \cdots + G(z_K)}, \qquad k = 1, \dots, K,$$

where G is a smooth real function that is defined in $\mathbb{R}$ and satisfies $0 < G(z) < \infty$. An example is $R_k(\mathbf{z}) = e^{Cz_k}/(e^{Cz_1} + \cdots + e^{Cz_K})$, $k = 1, \dots, K$, for some $C > 0$.

In the two-treatment case, we can let $R_1(z_1, z_2) = G(z_1 - z_2)$ and $R_2(z_1, z_2) = G(z_2 - z_1)$, where G is a real function defined on $\mathbb{R}$ satisfying $G(0) = 1/2$, $G(-z) = 1 - G(z)$ and $0 < G(z) < 1$ for all z. For the logistic regression model, Rosenberger, Vidyashankar and Agarwal [22] suggested using the estimated covariate-adjusted odds ratio to allocate subjects, which is equivalent to defining $R_k(z_1, z_2) = e^{z_k}/(e^{z_1} + e^{z_2})$, $k = 1, 2$. For each fixed covariate $\boldsymbol{\xi}$, we can also choose $\boldsymbol{\pi}(\boldsymbol{\theta}, \boldsymbol{\xi})$ according to [7] and [13]. When $\boldsymbol{\pi}(\boldsymbol{\theta}, \boldsymbol{\xi})$ does not depend on $\boldsymbol{\xi}$, one can use the allocation scheme of [9] for the normal linear regression model. We now introduce some asymptotic properties.

2.3. *Asymptotic properties.* Write $\boldsymbol{\pi}(\boldsymbol{\theta}^*, \mathbf{x}) = (\pi_1(\boldsymbol{\theta}^*, \mathbf{x}), \dots, \pi_K(\boldsymbol{\theta}^*, \mathbf{x}))$, $\mathbf{g}(\boldsymbol{\theta}^*) = (g_1(\boldsymbol{\theta}^*), \dots, g_K(\boldsymbol{\theta}^*))$, $v_k = g_k(\boldsymbol{\theta}) = \mathsf{E}[\pi_k(\boldsymbol{\theta}, \boldsymbol{\xi})]$, $k = 1, \dots, K$, and $\mathbf{v} = (v_1, \dots, v_K)$. We assume that $0 < v_k < 1$, $k = 1, \dots, K$. For the allocation function $\boldsymbol{\pi}(\boldsymbol{\theta}^*, \mathbf{x})$ we assume the following condition.

CONDITION A. Assume that the parameter space $\boldsymbol{\Theta}_k$ is a bounded domain in $\mathbb{R}^d$, and that the true value $\boldsymbol{\theta}_k$ is an interior point of $\boldsymbol{\Theta}_k$, $k = 1, \dots, K$.

1. For each fixed $\mathbf{x}$, $\pi_k(\boldsymbol{\theta}^*, \mathbf{x}) > 0$ is a continuous function of $\boldsymbol{\theta}^*$, $k = 1, \dots, K$.
2. For each $k = 1, \dots, K$, $\pi_k(\boldsymbol{\theta}^*, \boldsymbol{\xi})$ is differentiable with respect to $\boldsymbol{\theta}^*$ under the expectation, and there is a $\delta > 0$ such that

$$g_k(\boldsymbol{\theta}^*) = g_k(\boldsymbol{\theta}) + (\boldsymbol{\theta}^* - \boldsymbol{\theta})\left(\frac{\partial g_k}{\partial \boldsymbol{\theta}^*}\bigg|_{\boldsymbol{\theta}}\right)^T + o(\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|^{1+\delta}),$$

where $\partial g_k/\partial \boldsymbol{\theta}^* = (\partial g/\partial \theta^*_{11}, \dots, \partial g/\partial \theta^*_{Kd})$.

THEOREM 2.1. *Suppose that for $k = 1, \dots, K$,*

$$\widehat{\boldsymbol{\theta}}_{nk} - \boldsymbol{\theta}_k = \frac{1}{n}\sum_{m=1}^{n} X_{m,k}\mathbf{h}_k(Y_{m,k}, \boldsymbol{\xi}_m)\big(1 + o(1)\big) + o(n^{-1/2}) \qquad a.s., \tag{2.4}$$

where $\mathbf{h}_k$ *are* K *functions with* $\mathsf{E}[\mathbf{h}_k(Y_k,\boldsymbol{\xi})|\boldsymbol{\xi}]=\mathbf{0}$. *We also assume that* $\mathsf{E}\|\mathbf{h}_k(Y_k,\boldsymbol{\xi})\|^2<\infty$, $k=1,\dots,K$. *Then under Condition* A, *we have for* $k=1,\dots,K$,

$$\mathsf{P}(X_{n,k}=1)\to v_k,\qquad \mathsf{P}(X_{n,k}=1|\mathscr{F}_{n-1},\boldsymbol{\xi}_n=\mathbf{x})\to\pi_k(\boldsymbol{\theta},\mathbf{x})\qquad a.s. \tag{2.5}$$

and

$$\frac{\mathbf{N}_n}{n}-\mathbf{v}=O\left(\sqrt{\frac{\log\log n}{n}}\right)\qquad a.s.,\qquad \widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}=O\left(\sqrt{\frac{\log\log n}{n}}\right). \tag{2.6}$$

Further, let $\mathbf{V}_k=\mathsf{E}\{\pi_k(\boldsymbol{\theta},\boldsymbol{\xi})(\mathbf{h}_k(Y_k,\boldsymbol{\xi}))^T\mathbf{h}_k(Y_k,\boldsymbol{\xi})\}$, $k=1,\dots,K$, $\mathbf{V}=\mathrm{diag}(\mathbf{V}_1,\dots,\mathbf{V}_K)$, $\boldsymbol{\Sigma}_1=\mathrm{diag}(\mathbf{v})-\mathbf{v}^T\mathbf{v}$, $\boldsymbol{\Sigma}_2=\sum_{k=1}^K\frac{\partial\mathbf{g}}{\partial\boldsymbol{\theta}_k}\mathbf{V}_k(\frac{\partial\mathbf{g}}{\partial\boldsymbol{\theta}_k})^T$ *and* $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_1+2\boldsymbol{\Sigma}_2$. *Then*

$$\sqrt{n}(\mathbf{N}_n/n-\mathbf{v})\xrightarrow{D}N(\mathbf{0},\boldsymbol{\Sigma})\quad and\quad \sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})\xrightarrow{D}N(\mathbf{0},\mathbf{V}). \tag{2.7}$$

REMARK 2.1. Condition (2.4) depends on different estimation methods. In the next section, we show that it is satisfied in many cases.

Theorem 2.1 provides general results on the asymptotic properties of the allocation proportions $N_{n,k}/n$, $k=1,\dots,K$. Sometimes, one may be interested in the proportions for a given covariate (for discrete $\boldsymbol{\xi}$) as discussed in Section 4. Given a covariate $\mathbf{x}$, the proportion of subjects that is assigned to treatment k is

$$\frac{\sum_{m=1}^n X_{m,k}I\{\boldsymbol{\xi}_m=\mathbf{x}\}}{\sum_{m=1}^n I\{\boldsymbol{\xi}_m=\mathbf{x}\}}:=\frac{N_{n,k|\mathbf{x}}}{N_n(\mathbf{x})},$$

where $N_{n,k|\mathbf{x}}$ is the number of subjects with covariate $\mathbf{x}$ that is randomized to treatment k, $k=1,\dots,K$, in the n trials, and $N_n(\mathbf{x})$ is the total number of subjects with covariate $\mathbf{x}$. Write $\mathbf{N}_{n|\mathbf{x}}=(N_{n,1|\mathbf{x}},\dots,N_{n,K|\mathbf{x}})$. The following theorem establishes the asymptotic results of these proportions.

THEOREM 2.2. *Given a covariate* $\mathbf{x}$, *suppose that* $\mathsf{P}(\boldsymbol{\xi}=\mathbf{x})>0$. *Under Condition* A *and* (2.4), *we have*

$$N_{n,k|\mathbf{x}}/N_n(\mathbf{x})\to\pi_k(\boldsymbol{\theta},\mathbf{x})\qquad a.s.\ k=1,\dots,K \tag{2.8}$$

and

$$\sqrt{N_n(\mathbf{x})}\big(\mathbf{N}_{n|\mathbf{x}}/N_n(\mathbf{x})-\boldsymbol{\pi}(\boldsymbol{\theta},\mathbf{x})\big)\xrightarrow{\mathscr{D}}N(\mathbf{0},\boldsymbol{\Sigma}_{|\mathbf{x}}), \tag{2.9}$$

where

$$\boldsymbol{\Sigma}_{|\mathbf{x}}=\mathrm{diag}(\boldsymbol{\pi}(\boldsymbol{\theta},\mathbf{x}))-\boldsymbol{\pi}(\boldsymbol{\theta},\mathbf{x})^T\boldsymbol{\pi}(\boldsymbol{\theta},\mathbf{x})$$
$$+2\sum_{k=1}^K\frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta},\mathbf{x})}{\partial\boldsymbol{\theta}_k}\mathbf{V}_k\left(\frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta},\mathbf{x})}{\partial\boldsymbol{\theta}_k}\right)^T\mathsf{P}(\boldsymbol{\xi}=\mathbf{x}).$$

3. Generalized linear models. In this section, the general results of Section 2 are applied to the generalized linear model (GLM) and its two special cases, the logistic regression model and the linear model (refer to [16] for applications of these models). Suppose, given $\boldsymbol{\xi}$, that the response Y_k of a trial of treatment k has a distribution in the exponential family and takes the form

$$f_k(y_k|\boldsymbol{\xi},\boldsymbol{\theta}_k)=\exp\{(y_k\mu_k-a_k(\mu_k))/\phi_k+b_k(y_k,\phi_k)\} \tag{3.1}$$

with link function $\mu_k=h_k(\boldsymbol{\xi}\boldsymbol{\theta}_k^T)$, where $\boldsymbol{\theta}_k=(\theta_{k1},\dots,\theta_{kd})$, $k=1,\dots,K$, are coefficients. Assume that the scale parameter ϕ_k is fixed. Then $\mathsf{E}[Y_k|\boldsymbol{\xi}]=a_k'(\mu_k)$, $\mathsf{Var}(Y_k|\boldsymbol{\xi})=a_k''(\mu_k)\phi_k$ and

$$\frac{\partial\log f_k(y_k|\boldsymbol{\xi},\boldsymbol{\theta}_k)}{\partial\boldsymbol{\theta}_k}=\frac{1}{\phi_k}\{y_k-a_k'(\mu_k)\}h_k'(\boldsymbol{\xi}\boldsymbol{\theta}_k^T)\boldsymbol{\xi},$$

$$\frac{\partial^2\log f_k(y_k|\boldsymbol{\xi},\boldsymbol{\theta}_k)}{\partial\boldsymbol{\theta}_k^2}=\frac{1}{\phi_k}\{-a_k''(\mu_k)[h_k'(\boldsymbol{\xi}\boldsymbol{\theta}_k^T)]^2+[y_k-a_k'(\mu_k)]h_k''(\boldsymbol{\xi}\boldsymbol{\theta}_k^T)\}\boldsymbol{\xi}^T\boldsymbol{\xi}.$$

Thus, given $\boldsymbol{\xi}$, the conditional Fisher information matrix is

$$\mathbf{I}_k(\boldsymbol{\theta}_k|\boldsymbol{\xi})=-\mathsf{E}\left[\frac{\partial^2\log f_k(Y_k|\boldsymbol{\xi},\boldsymbol{\theta}_k)}{\partial\boldsymbol{\theta}_k^2}\Big|\boldsymbol{\xi}\right]=\frac{1}{\phi_k}a_k''(\mu_k)[h_k'(\boldsymbol{\xi}\boldsymbol{\theta}_k^T)]^2\boldsymbol{\xi}^T\boldsymbol{\xi}.$$

For the observations up to stage m, the likelihood function is

$$L(\boldsymbol{\theta})=\prod_{j=1}^m\prod_{k=1}^K[f_k(Y_{j,k}|\boldsymbol{\xi}_j,\boldsymbol{\theta}_k)]^{X_{j,k}}=\prod_{k=1}^K\prod_{j=1}^m[f_k(Y_{j,k}|\boldsymbol{\xi}_j,\boldsymbol{\theta}_k)]^{X_{j,k}}:=\prod_{k=1}^K L_k(\boldsymbol{\theta}_k)$$

with $\log L_k(\boldsymbol{\theta}_k)\propto\sum_{j=1}^m X_{j,k}\{Y_{j,k}-a_k(\mu_{j,k})\}$, $\mu_{j,k}=h_k(\boldsymbol{\theta}_k^T\boldsymbol{\xi}_j)$, $k=1,2,\dots,K$. The MLE $\widehat{\boldsymbol{\theta}}_m=(\widehat{\boldsymbol{\theta}}_{m,1},\dots,\widehat{\boldsymbol{\theta}}_{m,K})$ of $\boldsymbol{\theta}=(\theta_1,\dots,\theta_K)$ maximizes $L(\boldsymbol{\theta})$ over $\boldsymbol{\theta}\in\boldsymbol{\Theta}_1\times\cdots\times\boldsymbol{\Theta}_K$. Equivalently, $\widehat{\boldsymbol{\theta}}_{m,k}$ maximizes L_k over $\boldsymbol{\theta}_k\in\boldsymbol{\Theta}_k$, $k=1,2,\dots,K$. Rosenberger, Flournoy and Durham [19] established a general result for the asymptotic normality of MLEs from a response-driven design. Rosenberger and Hu [21] gave the asymptotic normality of the regression parameters from a generalized linear model that followed a sequential design with covariate vectors. These two papers neither examined the case of using covariates to adjust the design, nor established the asymptotic properties of the allocation proportions. The next corollary gives results on both the estimators of the parameters and the allocation proportions.

COROLLARY 3.1. *Define*

$$\mathbf{I}_k=\mathbf{I}_k(\boldsymbol{\theta})=\mathsf{E}\{\pi_k(\boldsymbol{\theta},\boldsymbol{\xi})\mathbf{I}_k(\boldsymbol{\theta}_k|\boldsymbol{\xi})\},\qquad k=1,\dots,K. \tag{3.2}$$

Under Condition A, if the matrices $\mathbf{I}_k$, $k=1,2,\dots,K$, are nonsingular and the MLE $\widehat{\boldsymbol{\theta}}_m$ is unique, then under regularity condition (A.13) *in the Appendix, we have* (2.5), (2.6) *and* (2.7) *with $\mathbf{V}_k=\mathbf{I}_k^{-1}$, $k=1,\dots,K$. Moreover, if $\mathsf{P}(\boldsymbol{\xi}=\mathbf{x})>0$ for a given covariate $\mathbf{x}$, then* (2.8) *and* (2.9) *hold.*

The proof is given in the Appendix through the verification of condition (2.4). For both the logistic regression and the linear regression, condition (A.13) is satisfied.

REMARK 3.1. From Corollary 3.1, it follows that

$$\sqrt{N_{n,k}}(\widehat{\boldsymbol{\theta}}_{n,k}-\boldsymbol{\theta}_k)\xrightarrow{D} N(\mathbf{0}, v_k\{\mathsf{E}[\pi_k(\boldsymbol{\theta},\boldsymbol{\xi})\mathbf{I}_k(\boldsymbol{\theta}_k|\boldsymbol{\xi})]\}^{-1}),\qquad k=1,\dots,K. \tag{3.3}$$

It should be noted that the asymptotic variances are different from those of general linear models with a fixed allocation procedure. For the latter, we have

$$\sqrt{N_{n,k}}(\widehat{\boldsymbol{\theta}}_{n,k}-\boldsymbol{\theta}_k)\xrightarrow{D} N(0, \{\mathsf{E}[\mathbf{I}_k(\boldsymbol{\theta}_k|\boldsymbol{\xi})]\}^{-1}),\qquad k=1,\dots,K. \tag{3.4}$$

If the allocation functions $\pi_k(\boldsymbol{\theta},\boldsymbol{\xi})$ do not depend on $\boldsymbol{\xi}$, then $\pi_k(\boldsymbol{\theta},\boldsymbol{\xi})=g_k(\boldsymbol{\theta})=v_k$, and so (3.3) and (3.4) are identical. Our asymptotic variance–covariance matrix of $\widehat{\boldsymbol{\theta}}_n$ is also different from that in Theorem 2 of [7], because the allocation probabilities in their study do not depend on the covariates.

REMARK 3.2. When the distribution of $\boldsymbol{\xi}$ and the true value of $\boldsymbol{\theta}$ are known, the values of $\mathbf{v}=\mathsf{E}[\boldsymbol{\pi}(\boldsymbol{\theta},\boldsymbol{\xi})]$, $\partial\mathbf{g}/\partial\boldsymbol{\theta}_k=\mathsf{E}[\partial\boldsymbol{\pi}(\boldsymbol{\theta},\boldsymbol{\xi})/\partial\boldsymbol{\theta}_k]$ and $\mathbf{I}_k$ in (3.2) can be obtained by computing the expectations, and then the values of the asymptotic variance–covariance matrices $\mathbf{V}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{|\mathbf{x}}$ can be obtained. In practice, we can obtain the estimates as follows.

(a) Estimate $\mathbf{I}_k$ by $\widehat{\mathbf{I}}_{n,k}=\frac{1}{n}\sum_{m=1}^n X_{m,k}\mathbf{I}_k(\widehat{\boldsymbol{\theta}}_{n,k}|\boldsymbol{\xi}_m)$, $k=1,2,\dots,K$; and then the estimator of $\mathbf{V}$ is $\widehat{\mathbf{V}}_n=\operatorname{diag}(\widehat{\mathbf{I}}_{n,1}^{-1},\dots,\widehat{\mathbf{I}}_{n,K}^{-1})$.

(b) Estimate $\boldsymbol{\Sigma}_1$ and $\frac{\partial\mathbf{g}}{\partial\boldsymbol{\theta}_k}$, respectively, by

$$\widehat{\boldsymbol{\Sigma}_1}=\operatorname{diag}\left(\frac{\mathbf{N}_n}{n}\right)-\left(\frac{\mathbf{N}_n}{n}\right)^T\frac{\mathbf{N}_n}{n}\quad\text{and}\quad\widehat{\frac{\partial\mathbf{g}}{\partial\boldsymbol{\theta}_k}}=\frac{1}{n}\sum_{m=1}^n\frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta}^*,\boldsymbol{\xi}_m)}{\partial\boldsymbol{\theta}_k^*}\bigg|_{\boldsymbol{\theta}^*=\widehat{\boldsymbol{\theta}}_n}.$$

(c) Define the estimator $\widehat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$ by $\widehat{\boldsymbol{\Sigma}}=\widehat{\boldsymbol{\Sigma}_1}+2\sum_{k=1}^K\widehat{\frac{\partial\mathbf{g}}{\partial\boldsymbol{\theta}_k}}\widehat{\mathbf{V}}_{n,k}(\widehat{\frac{\partial\mathbf{g}}{\partial\boldsymbol{\theta}_k}})^T$.

(d) For a given covariate $\mathbf{x}$, we can estimate $\boldsymbol{\Sigma}_{|\mathbf{x}}$ by

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}_{|\mathbf{x}}}&=\operatorname{diag}(\boldsymbol{\pi}(\widehat{\boldsymbol{\theta}}_n,\mathbf{x}))-\boldsymbol{\pi}(\widehat{\boldsymbol{\theta}}_n,\mathbf{x})^T\boldsymbol{\pi}(\widehat{\boldsymbol{\theta}}_n,\mathbf{x})\\&\quad+2\sum_{k=1}^K\left(\frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta}^*,\mathbf{x})}{\partial\boldsymbol{\theta}_k^*}\bigg|_{\boldsymbol{\theta}^*=\widehat{\boldsymbol{\theta}}_n}\right)\widehat{\mathbf{V}}_{n,k}\left(\frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta}^*,\mathbf{x})}{\partial\boldsymbol{\theta}_k^*}\bigg|_{\boldsymbol{\theta}^*=\widehat{\boldsymbol{\theta}}_n}\right)^T\frac{\#\{m\le n:\boldsymbol{\xi}_m=\mathbf{x}\}}{n}.\end{aligned}$$

Notice that $\phi_k\mathbf{I}_k(\boldsymbol{\theta}|\boldsymbol{\xi})$ does not depend on ϕ_k. When ϕ_k is unknown, we can estimate $\mathbf{I}_k$ in the same way after replacing ϕ_k with its estimate $\widehat{\phi}_k$.

We now consider two examples, the logistic regression model and the linear model.

EXAMPLE 3.1 (*Logistic regression model*). We consider the case of dichotomous (i.e., success or failure) responses. Let $Y_k = 1$ if a subject being given treatment k is a success and 0 otherwise, $k = 1, \dots, K$. Let $p_k = p_k(\boldsymbol{\theta}_k, \boldsymbol{\xi}) = \mathsf{P}(Y_k = 1|\boldsymbol{\xi})$ be the probability of the success of a trial of treatment k for a given covariate $\boldsymbol{\xi}$, $q_k = 1 - p_k$, $k = 1, \dots, K$. Assume that

$$\text{(3.5)} \qquad \operatorname{logit}(p_k) = \alpha_k + \boldsymbol{\theta}_k \boldsymbol{\xi}^T, \qquad k = 1, \dots, K.$$

Without loss of generality, we assume that $\alpha_k = 0$, $k = 1, 2, \dots, K$, or alternatively, we can redefine the covariate vector to be $(1, \boldsymbol{\xi})$. For each $k = 1, \dots, K$, let $p_{j,k} = p_k(\boldsymbol{\theta}_k, \boldsymbol{\xi}_k)$. With the observations up to stage m, the MLE $\widehat{\boldsymbol{\theta}}_{m,k}$ of $\boldsymbol{\theta}_k$ $(k = 1, \dots, K)$ is that for which $\widehat{\boldsymbol{\theta}}_{m,k}$ maximizes

$$\text{(3.6)} \qquad L_k =: \prod_{j=1}^{m} p_{j,k}^{X_{j,k} Y_{j,k}} (1 - p_{j,k})^{X_{j,k}(1 - Y_{j,k})} \qquad \text{over } \boldsymbol{\theta}_k \in \boldsymbol{\Theta}_k.$$

The logistic regression model is a special case of GLM (3.1) with $\phi_k = 1$, $\mu_k = \log(p_k/q_k)$, $h_k(x) = x$, $b_k(y_k, \phi_k) = 0$ and $a_k(\mu_k) = -\log(1 - p_k) = \log(1 + e^{\mu_k})$. Thus, given $\boldsymbol{\xi}$, the conditional information matrix is $\mathbf{I}_k(\boldsymbol{\theta}_k|\boldsymbol{\xi}) = a_k''(\mu_k)\boldsymbol{\xi}^T\boldsymbol{\xi} = p_k q_k \boldsymbol{\xi}^T \boldsymbol{\xi}$. For Corollary 3.1, we have the following corollary.

COROLLARY 3.2. *Suppose that Condition* A *is satisfied*, $\mathsf{E}\|\boldsymbol{\xi}\|^2 < \infty$ *and the matrix* $\mathsf{E}[\boldsymbol{\xi}^T\boldsymbol{\xi}]$ *is nonsingular. We then have* (2.5), (2.6), (2.7) *with* $\mathbf{V}_k = \mathbf{I}_k^{-1}$ *and* $\mathbf{I}_k = \mathsf{E}\{\pi_k(\boldsymbol{\theta}, \boldsymbol{\xi}) p_k q_k \boldsymbol{\xi}^T \boldsymbol{\xi}\}$, $k = 1, \dots, K$. *Moreover, if* $\mathsf{P}(\boldsymbol{\xi} = \mathbf{x}) > 0$ *for a given covariate* $\mathbf{x}$, *then* (2.8) *and* (2.9) *hold.*

EXAMPLE 3.2 (*Normal linear regression model*). The responses are normally distributed, that is, $Y_k|_{\boldsymbol{\xi}} \sim N(\mu_k, \sigma_k^2)$ with link function $\mu_k = \boldsymbol{\theta}_k \boldsymbol{\xi}^T$. Then the linear model is a special case of GLM (3.1) with $\phi_k = \sigma_k^2$, $a_k(\mu_k) = \mu_k^2/2$ and $h_k(x) = x$. Thus, we have the following corollary.

COROLLARY 3.3. *Suppose that the conditions in Corollary* 3.2 *are satisfied. We then have* (2.5), (2.6), (2.7) *with* $\mathbf{V}_k = \mathbf{I}_k^{-1}$ *and* $\mathbf{I}_k = \mathsf{E}[\pi_k(\boldsymbol{\theta}, \boldsymbol{\xi})\boldsymbol{\xi}^T\boldsymbol{\xi}]/\sigma_k^2$, $k = 1, \dots, K$. *Moreover, if* $\mathsf{P}(\boldsymbol{\xi} = \mathbf{x}) > 0$ *for given* $\mathbf{x}$, *then* (2.8) *and* (2.9) *hold.*

REMARK 3.3. Bandyopadhyay and Biswas [9] considered the normal linear regression model in which $\theta_{11} = \mu_1$, $\theta_{21} = \mu_2$, $\theta_{1j} = \theta_{2j} = \beta_{j-1}$, $j = 2, \dots, d$, and the first component of $\boldsymbol{\xi}$ is 1. Their proposed allocation probabilities are functions of estimates of the unknown parameters that depend only on information of the previous patients, but not on the covariates of the incoming patient. Theorem 1 of [9] gives the consistency property of $N_{n,1}/n$ and $\mathsf{P}(X_{n,1} = 1)$. However, their proof is not correct, since the assignments $\delta_1, \dots, \delta_i$ are functions of the previous responses $Y_1, \dots, Y_{i-1}$ and covariates. In fact, given the assignments $\delta_1, \dots, \delta_i$, the responses $Y_1, \dots, Y_i$ are no longer independent normal variables, which implies that their equation (4) is not valid. Nevertheless, if we let

$$\boldsymbol{\xi} = (1, \widetilde{\boldsymbol{\xi}}), \qquad \mathbf{a} = \mathsf{E}\widetilde{\boldsymbol{\xi}}, \qquad \mathbf{I}_{\widetilde{\xi}} = \mathsf{Var}\{\widetilde{\boldsymbol{\xi}}\},$$

$$v_1 = \Phi\left(\frac{\mu_1 - \mu_2}{T}\right) \quad \text{and} \quad v_2 = 1 - v_1,$$

under our theoretical framework it can be proved that Theorem 1 of [9] is correct. Further, we can show that

$$\begin{aligned}&\sqrt{n}(\widehat{\mu}_{n1} - \mu_1, \widehat{\mu}_{n2} - \mu_2)\\ &\quad\overset{D}{\to} N\left((0,0), \sigma^2\begin{pmatrix} 1/v_1 + \mathbf{a}\mathbf{I}_{\widetilde{\xi}}^{-1}\mathbf{a}^T & \mathbf{a}\mathbf{I}_{\widetilde{\xi}}^{-1}\mathbf{a}^T \\ \mathbf{a}\mathbf{I}_{\widetilde{\xi}}^{-1}\mathbf{a}^T & 1/v_2 + \mathbf{a}\mathbf{I}_{\widetilde{\xi}}^{-1}\mathbf{a}^T\end{pmatrix}\right),\end{aligned}$$

$$\sqrt{n}(\widehat{\beta}_n - \beta) \overset{D}{\to} N(\mathbf{0}, \sigma^2 \mathbf{I}_{\widetilde{\xi}}^{-1})$$

and

$$\sqrt{n}(N_{n1}/n - v_1) \overset{D}{\to} N\left(0, v_1 v_2 + \frac{2\sigma^2}{v_1 v_2}\left(\frac{1}{T}\Phi'\left(\frac{\mu_1 - \mu_2}{T}\right)\right)^2\right),$$

where σ^2 is the variance of the errors in the linear model.

REMARK 3.4. Corollary 3.3 can be generalized to responses not following (3.1). Suppose that the response Y_k of a subject to treatment k, $k = 1, \dots, K$ and its covariate $\boldsymbol{\xi}$ satisfies the linear regression model

$$\mathsf{E}[Y_k|\boldsymbol{\xi}] = p_k(\boldsymbol{\theta}_k, \boldsymbol{\xi}) = \boldsymbol{\theta}_k \boldsymbol{\xi}^T, \qquad k = 1, \dots, K.$$

For the observations up to stage m, let $\widehat{\boldsymbol{\theta}}_{m,k}$ minimize the error sum of squares

$$S_k(\boldsymbol{\theta}_k) = \sum_{j=1}^m X_{j,k}(Y_{j,k} - \boldsymbol{\theta}_k \boldsymbol{\xi}_j^T)^2 \qquad \text{over } \boldsymbol{\theta}_k \in \boldsymbol{\Theta}_k,$$

$k = 1, \dots, K$. Here, $\widehat{\boldsymbol{\theta}}_{m,k}$ is the least-squares estimator (LSE) of $\boldsymbol{\theta}_k$. Then Corollary 3.3 remains true with $\mathbf{V}_k = \mathbf{I}_{\xi k}^{-1}\mathbf{I}_{Yk}\mathbf{I}_{\xi k}^{-1}$, $\mathbf{I}_{\xi k} = \mathsf{E}[\pi_k(\boldsymbol{\theta}, \boldsymbol{\xi})\boldsymbol{\xi}^T\boldsymbol{\xi}]$ and $\mathbf{I}_{Yk} = \mathsf{E}\{\pi_k(\boldsymbol{\theta}, \boldsymbol{\xi})(Y_k - \boldsymbol{\theta}_k\boldsymbol{\xi}^T)^2\boldsymbol{\xi}^T\boldsymbol{\xi}\}$ under the condition $\mathsf{E}\|Y_k\boldsymbol{\xi}\|^2 < \infty$, $k = 1, \dots, K$. This result follows from Theorem 2.1, as condition (2.4) is satisfied with $\mathbf{h}_k = (Y_k - \boldsymbol{\theta}_k\boldsymbol{\xi}^T)\boldsymbol{\xi}\mathbf{I}_{\xi k}^{-1}$.

4. Discussion. This paper makes two major contributions. First, a comprehensive framework of CARA designs is proposed to serve as a paradigm for treatment allocation procedures in clinical trials when covariates are available. It is a very general framework that allows a wide spectrum of applications to very general statistical models, including generalized linear models as special cases. Second, asymptotic properties are obtained to provide a statistical basis for inference after using a CARA design.

When covariate information is not being used in the treatment allocation scheme, an optimal allocation proportion is usually determined with the assistance of some optimality criterion. Jennison and Turnbull [15] described a general procedure to search for an optimal allocation. For CARA designs, how to define and obtain an optimal allocation scheme is still unclear. For a CARA design, we can find optimal allocation for each fixed value of the covariate. Theorem 2.2 provides theoretical support for targeting optimal allocation by using a CARA design for each fixed covariate.

For response-adaptive designs without covariates, Hu and Rosenberger [13] studied the relationship among the power, the target allocation and the variability of the designs. It is important to study the behavior of the power function when a CARA design is used in clinical trials. However, the formulation becomes very different for CARA designs. It is an interesting topic for future research.

APPENDIX

PROOF OF THEOREM 2.1. First, notice that for each $k = 1, \dots, K$, $X_{m+1,k} = X_{m+1,k} - \mathsf{E}[X_{m+1,k}|\mathscr{F}_m] + g_k(\widehat{\boldsymbol{\theta}}_m)$ and then

$$N_{n,k} = \mathsf{E}[X_{1,k}|\mathscr{F}_0] + \sum_{m=1}^{n}(X_{m,k} - \mathsf{E}[X_{m,k}|\mathscr{F}_{m-1}]) + \sum_{m=1}^{n-1} g_k(\widehat{\boldsymbol{\theta}}_m). \tag{A.1}$$

The second term is a martingale. We next show that the third term can be approximated by another martingale. Write $\Delta M_{m,k} = X_{m,k} - \mathsf{E}[X_{m,k}|\mathscr{F}_{m-1}]$, $\Delta \mathbf{T}_{m,k} = X_{m,k}\mathbf{h}_k(Y_{m,k}, \boldsymbol{\xi}_m)$, $k = 1, \dots, K$. Let $\mathbf{M}_n = \sum_{m=1}^n \Delta\mathbf{M}_m$ and $\mathbf{T}_n = \sum_{m=1}^n \Delta\mathbf{T}_m$, where $\Delta\mathbf{M}_m = (\Delta M_{m,1}, \dots, \Delta M_{m,K})$ and $\Delta\mathbf{T}_m = (\Delta\mathbf{T}_{m,1}, \dots, \Delta\mathbf{T}_{m,K})$. Here, the symbol Δ denotes the differencing operand of a sequence $\{z_n\}$, that is, $\Delta z_n = z_n - z_{n-1}$. Then $\{(\mathbf{M}_n, \mathbf{T}_n)\}$ is a multi-dimensional martingale sequence that satisfies

$$|\Delta M_{n,k}| \le 1, \qquad \|\Delta\mathbf{T}_{n,k}\| \le \|\mathbf{h}_k(Y_{n,k}, \boldsymbol{\xi}_n)\| \tag{A.2}$$

and $\mathsf{E}\|\mathbf{h}_k(Y_{n,k}, \boldsymbol{\xi}_n)\|^2 < \infty$, $k = 1, \dots, K$. It follows that

$$\|\mathbf{M}_n\| = O(\sqrt{n}) \quad \text{and} \quad \|\mathbf{T}_n\| = O(\sqrt{n}) \qquad \text{in } L_2. \tag{A.3}$$

Also, according to the law of the iterated logarithm for martingales, we have

$$\mathbf{M}_n = O\big(\sqrt{n \log\log n}\big) \quad \text{a.s.} \quad \text{and} \quad \mathbf{T}_n = O\big(\sqrt{n \log\log n}\big) \quad \text{a.s.} \tag{A.4}$$

From (A.4) and (2.4), it follows that

$$\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = O\left(\sqrt{\frac{\log\log n}{n}}\right) \qquad \text{a.s.} \tag{A.5}$$

From (A.1), (A.3), (A.5) and Condition A, it follows that

$$
\begin{aligned}
N_{n,k} - nv_k &= M_{n,k} + \sum_{m=1}^{n-1}\sum_{j=1}^{K}(\widehat{\boldsymbol{\theta}}_{m,j} - \boldsymbol{\theta}_j)\left(\frac{\partial g_k}{\partial \boldsymbol{\theta}_j}\right)^T + \sum_{m=1}^{n-1} o(\|\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}\|^{1+\delta}) \\
&= M_{n,k} + \sum_{m=1}^{n}\sum_{j=1}^{K}\frac{\mathbf{T}_{m,j}(1+o(1))}{m}\left(\frac{\partial g_k}{\partial \boldsymbol{\theta}_j}\right)^T + o(n^{1/2}) \qquad \text{a.s.} \\
&= M_{n,k} + \sum_{m=1}^{n}\sum_{j=1}^{K}\frac{\mathbf{T}_{m,j}}{m}\left(\frac{\partial g_k}{\partial \boldsymbol{\theta}_j}\right)^T + o(n^{1/2}) \qquad \text{in probability,}
\end{aligned}
$$

that is,

$$
\mathbf{N}_n - n\mathbf{v} = \mathbf{M}_n + \sum_{m=1}^{n}\sum_{j=1}^{K}\frac{\mathbf{T}_{m,j}(1+o(1))}{m}\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}_j}\right)^T + o(n^{1/2}) \qquad \text{a.s.} \tag{A.6}
$$

$$
= \mathbf{G}_n + o(n^{1/2}) \qquad \text{in probability,} \tag{A.7}
$$

where

$$
\mathbf{G}_n = \mathbf{M}_n + \sum_{m=1}^{n}\sum_{j=1}^{K}\frac{\mathbf{T}_{m,j}}{m}\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}_j}\right)^T = \mathbf{M}_n + \sum_{m=1}^{n}\sum_{j=1}^{K}\Delta\mathbf{T}_{m,j}\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}_j}\right)^T \sum_{i=m}^{n}\frac{1}{i}.
$$

The combination of (A.4) and (A.6) yields

$$
\begin{aligned}
\mathbf{N}_n - n\mathbf{v} &= O\big(\sqrt{n\log\log n}\big) + \sum_{m=1}^{n}\sum_{j=1}^{K}\frac{O(\sqrt{m\log\log m})}{m} \\
&= O\big(\sqrt{n\log\log n}\big) \qquad \text{a.s.}
\end{aligned}
$$

(2.5) is obvious by noting (A.5) and the continuity of $\boldsymbol{\pi}(\cdot,\boldsymbol{\xi})$. The proof of consistency is thus obtained. Next, we consider the asymptotic normality. Notice that $\mathbf{M}_n$, $\mathbf{T}_n$ and $\mathbf{G}_n$ are all sums of martingale differences. It is easy to verify that the Lindberg condition is satisfied by (A.2). To complete the proof it suffices to derive the variances. First, the conditional variance–covariance matrices of the martingale difference $\{\Delta\mathbf{M}_n, \Delta\mathbf{T}_n\}$ satisfy

$$
\mathsf{E}[(\Delta\mathbf{M}_n)^T\Delta\mathbf{M}_n|\mathscr{F}_{n-1}] = \operatorname{diag}(\mathbf{g}(\widehat{\boldsymbol{\theta}}_{n-1})) - (\mathbf{g}(\widehat{\boldsymbol{\theta}}_{n-1}))^T\mathbf{g}(\widehat{\boldsymbol{\theta}}_{n-1}) \to \boldsymbol{\Sigma}_1 \qquad \text{in } L_1,
$$

$$
\mathsf{E}[(\Delta\mathbf{T}_{n,k})^T\Delta\mathbf{T}_{n,k}|\mathscr{F}_{n-1}] = \mathsf{E}[\pi_k(\mathbf{x},\boldsymbol{\xi})(\mathbf{h}_k(Y_k,\boldsymbol{\xi}))^T\mathbf{h}_k(Y_k,\boldsymbol{\xi})]|_{\mathbf{x}=\widehat{\boldsymbol{\theta}}_{n-1}} \to \mathbf{V}_k \qquad \text{in } L_1,
$$

$$
\mathsf{E}[(\Delta M_{n,i})^T\Delta\mathbf{T}_{n,j}|\mathscr{F}_{n-1}] = \mathbf{0} \qquad \text{for all } i, j
$$

and

$$\mathsf{E}[(\Delta \mathbf{T}_{n,i})^T \Delta \mathbf{T}_{n,j} | \mathcal{F}_{n-1}] = \mathbf{0}$$

for all $i \neq j$. It follows that $\mathsf{Var}\{\mathbf{T}_n\}/n \to \mathbf{V}$ and

$$\begin{aligned}\mathsf{Var}\{\mathbf{G}_n\} &= \sum_{m=1}^{n} [\boldsymbol{\Sigma}_1 + o(1)] + \sum_{m=1}^{n} \sum_{l=1}^{n} \sum_{j=1}^{K} \frac{l \wedge m}{ml} \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}_j} [\mathbf{V}_j + o(1)] \left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}_j} \right)^T \\ &= n(\boldsymbol{\Sigma}_1 + 2\boldsymbol{\Sigma}_2) + o(n) = n\boldsymbol{\Sigma} + o(n).\end{aligned}$$

By the central limit theorem for martingales [12], it follows that $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = n^{-1/2}\mathbf{T}_n + o(1) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{V})$ and

$$\sqrt{n}(\mathbf{N}_n/n - \mathbf{v}) = n^{-1/2}\mathbf{G}_n + o(1) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Sigma}). \tag{A.8}$$

The proof is now complete. □

PROOF OF THEOREM 2.2. First, according to the law of large numbers, we have

$$\frac{1}{n} \sum_{m=1}^{n} I\{\boldsymbol{\xi}_m = \mathbf{x}\} \to \mathsf{P}(\boldsymbol{\xi} = \mathbf{x}) \qquad \text{a.s.} \tag{A.9}$$

and

$$\begin{aligned}&\frac{1}{n} \sum_{m=1}^{n} X_{m,k} I\{\boldsymbol{\xi}_m = \mathbf{x}\} \\ &\quad = \frac{1}{n} \sum_{m=1}^{n} (X_{m,k} I\{\boldsymbol{\xi}_m = \mathbf{x}\} - \mathsf{E}[X_{m,k} I\{\boldsymbol{\xi}_m = \mathbf{x}\} | \mathcal{F}_{m-1}]) \\ &\qquad + \frac{1}{n} \sum_{m=1}^{n} \pi_k(\widehat{\boldsymbol{\theta}}_{m-1}, \mathbf{x}) \mathsf{P}(\boldsymbol{\xi}_m = \mathbf{x}) \to \pi_k(\boldsymbol{\theta}, \mathbf{x}) \mathsf{P}(\boldsymbol{\xi} = \mathbf{x}) \qquad \text{a.s.,}\end{aligned}$$

and thus (2.8) is proved. We next consider the asymptotic normality. The proof is similar to that of (A.8). The difference lies in the approximation of the process by a new $2K$-dimensional martingale and the calculation of its variance–covariance matrix. Define $\zeta_{n,k}(\mathbf{x}) := \sum_{m=1}^{n} (X_{m,k} - \pi_k(\boldsymbol{\theta}, \mathbf{x})) I\{\boldsymbol{\xi}_m = \mathbf{x}\}$. Then

$$\sqrt{N_n(\mathbf{x})} \left(\frac{\mathbf{N}_{n|\mathbf{x}}}{N_n(\mathbf{x})} - \boldsymbol{\pi}(\boldsymbol{\theta}, \mathbf{x}) \right) = \sqrt{\frac{n}{N_n(\mathbf{x})}} \frac{\zeta_{n,k}(\mathbf{x})}{\sqrt{n}}, \qquad k = 1, \dots, K.$$

Notice (A.9). It is sufficient to prove

$$n^{-1/2}\big(\zeta_{n,1}(\mathbf{x}), \dots, \zeta_{n,K}(\mathbf{x})\big) \xrightarrow{\mathcal{D}} N\big(\mathbf{0}, \boldsymbol{\Sigma}_{|\mathbf{x}} \mathsf{P}(\boldsymbol{\xi} = \mathbf{x})\big). \tag{A.10}$$

With the same argument as is used to derive (A.6), we can obtain

$$\begin{aligned}
\zeta_{n,k}(\mathbf{x}) &= \sum_{m=1}^{n}\big(\Delta\zeta_{n,k}(\mathbf{x}) - \mathsf{E}[\Delta\zeta_{n,k}(\mathbf{x})|\mathcal{F}_{n-1}]\big) \\
&\quad + \sum_{m=1}^{n}\big(\pi_k(\widehat{\boldsymbol{\theta}}_{m-1},\mathbf{x}) - \pi_k(\boldsymbol{\theta},\mathbf{x})\big)\mathsf{P}(\boldsymbol{\xi}=\mathbf{x}) \\
&= \sum_{m=1}^{n}\big(\Delta\zeta_{m,k}(\mathbf{x}) - \mathsf{E}[\Delta\zeta_{m,k}(\mathbf{x})|\mathcal{F}_{m-1}]\big) \\
&\quad + \sum_{j=1}^{K}\sum_{m=1}^{n}\frac{\mathbf{T}_{m,j}}{m}\left(\frac{\partial\pi_k(\boldsymbol{\theta},\mathbf{x})}{\partial\boldsymbol{\theta}_j}\right)^T\mathsf{P}(\boldsymbol{\xi}=\mathbf{x}) + o(n^{1/2}) \qquad \text{in probability} \\
&=: G_{n,k}(\mathbf{x}) + o(n^{1/2}).
\end{aligned}$$

Similar to the proof of (A.8), to complete the proof it suffices to get the variance of $\mathbf{G}_n(\mathbf{x}) = (G_{n,1}(\mathbf{x}),\dots,G_{n,K}(\mathbf{x}))$. Let $\Delta\overline{M}_{n,k}(\mathbf{x}) = \Delta\zeta_{n,k}(\mathbf{x}) - \mathsf{E}[\Delta\zeta_{n,k}(\mathbf{x})|\mathcal{F}_{n-1}]$. The variance–covariance matrix of the martingale difference $\{(\Delta\overline{\mathbf{M}}_n(\mathbf{x}), \Delta\mathbf{T}_n)\}$ then satisfies $\mathsf{E}[(\Delta\overline{M}_{n,k}(\mathbf{x}))^2|\mathcal{F}_{n-1}] \to \pi_k(\boldsymbol{\theta},\mathbf{x})(1-\pi_k(\boldsymbol{\theta},\mathbf{x}))\mathsf{P}(\boldsymbol{\xi}=\mathbf{x})$, $\mathsf{E}[\Delta\overline{M}_{n,k}(\mathbf{x})\Delta\overline{M}_{n,j}(\mathbf{x})|\mathcal{F}_{n-1}] \to -\pi_k(\boldsymbol{\theta},\mathbf{x})\pi_j(\boldsymbol{\theta},\mathbf{x})\mathsf{P}(\boldsymbol{\xi}=\mathbf{x})$ $\forall k\neq j$ and $\mathsf{E}[\Delta\overline{M}_{n,k}(\mathbf{x})\Delta\mathbf{T}_{n,j}|\mathcal{F}_{n-1}] = \mathbf{0}$ $\forall i,j$ in L_1. It follows that

$$\begin{aligned}
\mathsf{Var}\{\mathbf{G}_n(\mathbf{x})\} &= n[\operatorname{diag}(\boldsymbol{\pi}(\boldsymbol{\theta},\mathbf{x})) - \boldsymbol{\pi}(\boldsymbol{\theta},\mathbf{x})^T\boldsymbol{\pi}(\boldsymbol{\theta},\mathbf{x})\mathsf{P}(\boldsymbol{\xi}=\mathbf{x}) + o(1)] \\
&\quad + \sum_{m=1}^{n}\sum_{l=1}^{n}\sum_{j=1}^{K}\frac{l\wedge m}{lm}\frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta},\mathbf{x})}{\partial\boldsymbol{\theta}_j}[\mathbf{V}_j + o(1)]\left(\frac{\partial\boldsymbol{\pi}(\boldsymbol{\theta},\mathbf{x})}{\partial\boldsymbol{\theta}_j}\right)^T\mathsf{P}^2(\boldsymbol{\xi}=\mathbf{x}) \\
&= n\boldsymbol{\Sigma}_{|\mathbf{x}}\mathsf{P}(\boldsymbol{\xi}=\mathbf{x}) + o(n).
\end{aligned}$$

(A.10) is then proved. □

PROOF OF COROLLARY 3.1. By Theorem 2.1, it suffices to verify the condition (2.4). Notice that $\widehat{\boldsymbol{\theta}}_{m,k}$ is a solution to $\partial\log L_k/\partial\boldsymbol{\theta}_k = 0$. The application of Taylor's theorem yields

$$\begin{aligned}
&\frac{\partial\log L_k}{\partial\boldsymbol{\theta}_k}\bigg|_{\boldsymbol{\theta}_k} + (\widehat{\boldsymbol{\theta}}_{m,k}-\boldsymbol{\theta}_k)\left\{\frac{\partial^2\log L_k}{\partial\boldsymbol{\theta}_k^2}\bigg|_{\boldsymbol{\theta}_k} + \int_0^1\left[\frac{\partial^2\log L_k}{\partial\boldsymbol{\theta}_k^2}\bigg|_{\boldsymbol{\theta}_k}^{\boldsymbol{\theta}_k+t(\widehat{\boldsymbol{\theta}}_{m,k}-\boldsymbol{\theta}_k)}\right]dt\right\} \\
&\qquad = \frac{\partial\log L_k}{\partial\boldsymbol{\theta}_k}\bigg|_{\widehat{\boldsymbol{\theta}}_{m,k}} = \mathbf{0},
\end{aligned} \tag{A.11}$$

where $f(x)|_a^b = f(b) - f(a)$. Notice that

$$\frac{\partial\log L_k}{\partial\boldsymbol{\theta}_k} = \sum_{j=1}^{m} X_{j,k}\frac{\partial\log f_k(Y_{j,k}|\boldsymbol{\xi}_j,\boldsymbol{\theta}_k)}{\partial\boldsymbol{\theta}_k} \tag{A.12}$$

and

$$\frac{\partial^2 \log L_k}{\partial \boldsymbol{\theta}_k^2} = \sum_{j=1}^m X_{j,k} \frac{\partial^2 \log f_k(Y_{j,k}|\boldsymbol{\xi}_j, \boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k^2}.$$

We assume the regularity condition

$$H(\delta) =: \mathsf{E}\left[\sup_{\|\mathbf{z}\| \le \delta} \left\| \frac{\partial^2 \log f_k(Y_k|\boldsymbol{\xi}, \boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k^2} \bigg|_{\boldsymbol{\theta}_k}^{\boldsymbol{\theta}_k + \mathbf{z}} \right\| \right] \to 0 \qquad \text{as } \delta \to 0. \tag{A.13}$$

This regularity condition is implied by the simple condition that a_k'', h_k'' are continuous and $\boldsymbol{\xi}$ is bounded. Under (A.13), one can show that

$$\sup_{\|\mathbf{z}\| \le \delta} \left\| \frac{1}{m} \frac{\partial^2 \log L_k}{\partial \boldsymbol{\theta}_k^2} \bigg|_{\boldsymbol{\theta}_k}^{\boldsymbol{\theta}_k + \mathbf{z}} \right\| \le H(\delta) + o(1) \qquad \text{a.s.}$$

However,

$$\sum_{j=1}^m \left\{ X_{j,k} \frac{\partial^2 \log f_k(Y_{j,k}|\boldsymbol{\xi}_j, \boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k^2} - \mathsf{E}\left[X_{j,k} \frac{\partial^2 \log f_k(Y_{j,k}|\boldsymbol{\xi}_j, \boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k^2} \bigg| \mathscr{F}_{j-1} \right] \right\}$$

is a martingale. According to the law of large numbers,

$$\begin{aligned} \frac{\partial^2 \log L_k}{\partial \boldsymbol{\theta}_k^2} &= \sum_{j=1}^m \mathsf{E}\left[X_{j,k} \frac{\partial^2 \log f_k(Y_{j,k}|\boldsymbol{\xi}_j, \boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k^2} \bigg| \mathscr{F}_{j-1} \right] + o(m) \\ &= -\sum_{j=1}^m \{\mathsf{E}[\pi_k(\mathbf{z}, \boldsymbol{\xi}) \mathbf{I}_k(\boldsymbol{\theta}_k|\boldsymbol{\xi})]\}_{\mathbf{z} = \widehat{\boldsymbol{\theta}}_{j-1}} + o(m) \qquad \text{(A.14)} \\ &= -m\mathbf{I}_k + o(m) \qquad \text{a.s.} \end{aligned}$$

The substitution of (A.12) and (A.14) into (A.11) yields

$$m(\widehat{\boldsymbol{\theta}}_{m,k} - \boldsymbol{\theta}_k)\{\mathbf{I}_k + o(1) + O(H(\|\widehat{\boldsymbol{\theta}}_{m,k} - \boldsymbol{\theta}_k\|))\} = \sum_{j=1}^m X_{j,k} \frac{\partial \log f_k(Y_{j,k}|\boldsymbol{\xi}_j, \boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k}.$$

Thus,

$$\widehat{\boldsymbol{\theta}}_{m,k} - \boldsymbol{\theta}_k = \frac{1}{m} \sum_{j=1}^m X_{j,k} \frac{\partial \log f_k(Y_{j,k}|\boldsymbol{\xi}_j, \boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k} \mathbf{I}_k^{-1}(1 + o(1)) \qquad \text{a.s.} \tag{A.15}$$

Notice that

$$\mathsf{E}\left[\frac{\partial \log f_k(Y_{j,k}|\boldsymbol{\xi}_j, \boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k} \bigg| \boldsymbol{\xi}_j \right] = 0$$

and

$$\mathsf{Var}\left\{ \frac{\partial \log f_k(Y_{j,k}|\boldsymbol{\xi}_j, \boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k} \bigg| \boldsymbol{\xi}_j \right\} = \mathbf{I}_k(\boldsymbol{\theta}_k|\boldsymbol{\xi}_j).$$

Hence, condition (2.4) is valid. By Theorem 2.1, the proof is complete. $\square$

Acknowledgments. Special thanks go to the anonymous referees, the Associate Editor and the Editor for their constructive comments, which led to a much improved version of the paper.

REFERENCES

[1] ATKINSON, A. C. (1982). Optimal biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69** 61–67. MR0655670

[2] ATKINSON, A. C. (1999). Optimal biased-coin designs for sequential treatment allocation with covariate information. *Stat. Med.* **18** 1741–1752.

[3] ATKINSON, A. C. (2002). The comparison of designs for sequential clinical trials with covariate information. *J. Roy. Statist. Soc. Ser. A* **165** 349–373. MR1904822

[4] ATKINSON, A. C. (2004). Adaptive biased-coin designs for clinical trials with several treatments. *Discuss. Math. Probab. Statist.* **24** 85–108. MR2118925

[5] ATKINSON, A. C. and BISWAS, A. (2005). Adaptive biased-coin designs for skewing the allocation proportion in clinical trials with normal responses. *Stat. Med.* **24** 2477–2492. MR2112377

[6] ATKINSON, A. C. and BISWAS, A. (2005). Bayesian adaptive biased-coin designs for clinical trials with normal responses. *Biometrics* **61** 118–125. MR2135851

[7] BALDI ANTOGNINI, A. and GIOVAGNOLI, A. (2005). On the large sample optimality of sequential designs for comparing two or more treatments. *Sequential Anal.* **24** 205–217. MR2154938

[8] BANDYOPADHYAY, U. and BISWAS, A. (1999). Allocation by randomized play-the-winner rule in the presence of prognostic factors. *Sankhyā Ser. B* **61** 397–412. MR1745743

[9] BANDYOPADHYAY, U. and BISWAS, A. (2001). Adaptive designs for normal responses with prognostic factors. *Biometrika* **88** 409–419. MR1844841

[10] EISELE, J. and WOODROOFE, M. (1995). Central limit theorems for doubly adaptive biased coin designs. *Ann. Statist.* **23** 234–254. MR1331666

[11] EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403–417. MR0312660

[12] HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York. MR0624435

[13] HU, F. and ROSENBERGER, W. F. (2003). Optimality, variability, power: Evaluating response-adaptive randomization procedures for treatment comparisons. *J. Amer. Statist. Assoc.* **98** 671–678. MR2011680

[14] HU, F. and ZHANG, L.-X. (2004). Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *Ann. Statist.* **32** 268–301. MR2051008

[15] JENNISON, C. and TURNBULL, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman and Hall/CRC, Boca Raton, FL. MR1710781

[16] MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman and Hall, London.

[17] POCOCK, S. J. and SIMON, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* **31** 103–115.

[18] ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58** 527–535. MR0050246

[19] ROSENBERGER, W. F., FLOURNOY, N. and DURHAM, S. D. (1997). Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs. *J. Statist. Plann. Inference* **60** 69–76. MR1453033

[20] ROSENBERGER, W. F. and HU, F. (2004). Maximizing power and minimizing treatment failures in clinical trials. *Clinical Trials* **1** 141–147.

[21] ROSENBERGER, W. F. and HU, M. X. (2002). On the use of generalized linear models following a sequential design. *Statist. Probab. Lett.* **56** 155–161. MR1881169

[22] ROSENBERGER, W. F., VIDYASHANKAR, A. N. and AGARWAL, D. K. (2002). Covariate-adjusted response-adaptive designs for binary response. *J. Biopharm. Statist.* **11** 227–236.

[23] THOMPSON, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* **25** 285–294.

[24] WEI, L. J. and DURHAM, S. (1978). The randomized pay-the-winner rule in medical trials. *J. Amer. Statist. Assoc.* **73** 840–843.

[25] ZELEN, M. (1969). Play the winner rule and the controlled clinical trial. *J. Amer. Statist. Assoc.* **64** 131–146. MR0240938

[26] ZELEN, M. (1974). The randomization and stratification of patients to clinical trials. *J. Chronic Diseases* **27** 365–375.

[27] ZELEN, M. and WEI, L. J. (1995). Foreword. In *Adaptive Designs* (N. Flournoy and W. F. Rosenberger, eds.) IMS, Hayward, CA.

L.-X. ZHANG
DEPARTMENT OF MATHEMATICS
ZHEJIANG UNIVERSITY
HANGZHOU 310028
PEOPLE'S REPUBLIC OF CHINA
E-MAIL: lxzhang@mail.hz.zj.cn

F. HU
DEPARTMENT OF STATISTICS
UNIVERSITY OF VIRGINIA
HALSEY HALL
CHARLOTTESVILLE, VIRGINIA 22904-4135
USA
E-MAIL: fh6e@virginia.edu

S. H. CHEUNG
DEPARTMENT OF STATISTICS
CHINESE UNIVERSITY OF HONG KONG
SHATIN, N.T.
HONG KONG
PEOPLE'S REPUBLIC OF CHINA
E-MAIL: shcheung@sta.cuhk.edu.hk

W. S. CHAN
DEPARTMENT OF FINANCE
CHINESE UNIVERSITY OF HONG KONG
SHATIN, N.T.
HONG KONG
PEOPLE'S REPUBLIC OF CHINA
E-MAIL: chanws@cuhk.edu.hk