The Annals of Statistics
1993, Vol. 21, No. 2, 625–644

ESTIMATION UP TO A CHANGE-POINT[1]

BY DEAN P. FOSTER AND EDWARD I. GEORGE

University of Pennsylvania and University of Texas at Austin

Consider the problem of estimating μ, based on the observation of $Y_0, Y_1, \dots, Y_n$, where it is assumed only that $Y_0, Y_1, \dots, Y_\kappa$ iid $N(\mu, \sigma^2)$ for some *unknown* κ. Unlike the traditional change-point problem, the focus here is not on estimating κ, which is now a nuisance parameter. When it is known that $\kappa = k$, the sample mean $\bar{Y}_k = \Sigma_0^k Y_i/(k+1)$, provides, in addition to wonderful efficiency properties, safety in the sense that it is minimax under squared error loss. Unfortunately, this safety breaks down when κ is unknown; indeed if $k > \kappa$, the risk of $\bar{Y}_k$ is unbounded. To address this problem, a generalized minimax criterion is considered whereby each estimator is evaluated by its maximum risk under $Y_0, Y_1, \dots, Y_\kappa$ iid $N(\mu, \sigma^2)$ for each possible value of κ. An essentially complete class under this criterion is obtained. Generalizations to other situations such as variance estimation are illustrated.

0. Introduction. Consider the following problem of combining data. Suppose we want to estimate a scalar μ based on $n+1$ observations $Y_0, Y_1, \dots, Y_n$, where we are only willing to assume that $Y_0, Y_1, \dots, Y_\kappa$ iid $N(\mu, \sigma^2)$ for some unknown κ, μ and σ^2. The situation we have in mind is that $Y_0, Y_1, \dots, Y_n$ represents a time series in reverse order, say, $X_t, X_{t-1}, \dots, X_{t-n}$. Thus, Y_0 $(= X_t)$ would be the current observation for which we believe the model $N(\mu, \sigma^2)$ held, and κ might be called the duration of the model. The dilemma is that we would like to obtain many replications from the past to increase estimation precision, while guarding against using unrelated observations which might increase bias.

This problem is similar to the traditional change-point problems where the goal is typically to detect and/or to estimate an abrupt change in the distribution of a sequence of observations. These change-point setups assume that the sequences before and after the change-point are at least exchangeable. Often the before and after distributions belong to the same parametric family and differ by only one or two parameters. The literature on the problems is vast; see, for example, Brown, Durbin and Evans (1975), Chernoff and Zacks (1964), Hinkley (1970), Siegmund (1986) and Smith (1975, 1985). Our problem, however, differs from this literature in two important respects. First, although our setup allows for an abrupt change in the distribution of the observations, no structure at all is imposed after the change. Second, our focus is on

Received August 1989; revised July 1992.
[1]This work was supported by the Graduate School of Business at the University of Chicago.
AMS 1991 *subject classifications*. Primary 62F10; Secondary 62C20, 62L12.
Key words and phrases. Change-point problems, equivariance, Hunt–Stein theorem, minimax procedures, risk, pooling data.

estimating a characteristic of the prechange distribution rather than the time of change, which is now a nuisance parameter. Our problem is much more one of pooling data for efficient estimation. Similar goals are addressed by Mosteller (1948) in a related pooling problem with more structure.

The outline of this paper is as follows. In Section 1, we formalize the problem and define and motivate various risk criteria. These include a generalized minimax criterion and a risk inflation criterion which measures the price of not knowing the change-point κ. Preliminary estimators based on heuristic considerations are examined from this point of view. In Section 2, characterizations of the class of equivariant estimators are obtained along with convenient expressions for our generalized minimax criterion. In Section 3, we obtain a generalization of the Hunt–Stein theorem which shows that any estimator which is generalized minimax (according to our criterion) within the class of equivariant estimators is generalized minimax overall. In Section 4, an essentially complete class with respect to our generalized minimax criterion is obtained. This class is a substantially restrictive subclass of equivariant estimators. In Section 5, we derive a lower bound for the risk inflation of any estimator and describe estimators which obtain this bound. Finally, in Section 6 we describe how our results may be easily extended to other examples of interest, such as where the initial model is a double exponential distribution or a chi-square distribution.

1. Formalizing the problem. We formalize our problem as follows. Let $Y = (Y_0, \dots, Y_n)$ be the observed sequence of observations. We assume that F, the unknown distribution of Y, belongs to at least one of the following families of distributions:

$$\mathscr{F}_k = \{F: Y_0, \dots, Y_k \text{ iid } N(\mu, \sigma^2)\}, \qquad k = 1, \dots, n, \tag{1.1}$$

where μ and σ^2 are unknown. (Because both μ and σ^2 are unknown, at least two "good" observations, Y_0 and Y_1, are needed and so we restrict $k \geq 1$.) Note that these families are nested, $\mathscr{F}_1 \supset \cdots \supset \mathscr{F}_n$, and that any $F \in \mathscr{F}_n$ is identified by μ and σ^2. Defining the change-point

$$\kappa \equiv \kappa(F) = \sup\{k: F \in \mathscr{F}_k\} \tag{1.2}$$

shows how this setup formalizes the situation described in the introduction, where $Y_0, Y_1, \dots, Y_\kappa \sim$ iid $N(\mu, \sigma^2)$ for some unknown κ, μ and σ^2.

Under this setup, a natural criterion for evaluating an estimator $\delta \equiv \delta(Y)$ of μ ($\equiv EY_0$) is the risk criterion of expected scaled squared error loss,

$$R(F, \delta) = E_F\left[(\delta - \mu)^2/\sigma^2\right], \tag{1.3}$$

where $F \in \mathscr{F}_k$ for some $k \in \{1, \dots, n\}$. Unfortunately, because of the vast size of the parameter space (1.1), assessment of estimators by their entire risk functions is an overwhelming task. Instead, we adopt the strategy of using summary risk criteria, which capture properties which any good estimator should possess. In particular, we focus on keeping small the *maximum risk* (MR) under each of the $\mathscr{F}_k$, namely,

$$\mathrm{MR}(\mathscr{F}_k, \delta) \equiv \sup_{F \in \mathscr{F}_k} R(F, \delta) \quad \text{for } k = 1, \dots, n. \tag{1.4}$$

For example, consider the estimator

$$\overline{Y}_k = \frac{1}{k+1} \sum_0^k Y_i, \tag{1.5}$$

the mean of $Y_0, \dots, Y_k$. In this case, $\mathrm{MR}(\mathscr{F}_j, \overline{Y}_k) = 1/(k+1)$ for $j \geq k$ and $\mathrm{MR}(\mathscr{F}_j, \overline{Y}_k) = \infty$ for $j < k$. Note that although $\mathrm{MR}(\mathscr{F}_k, \delta)$ is minimized at $\mathrm{MR}(\mathscr{F}_k, \overline{Y}_k) = 1/(k+1)$, the trade-off between precision and potential bias is extreme for these kth partial means.

In an effort to correct for the deficiencies of $\overline{Y}_k$, one might consider an estimator of the form $\overline{Y}_T$, where T is an estimator of the change-point κ. For simplicity, suppose $\sigma^2 = 1$ were known. Because when $k < \kappa$, $(\overline{Y}_{k+1} - \overline{Y}_k) \sim N(0, a_k^2)$, where $a_k^2 = 1/(k+1)(k+2)$, a reasonable choice for T might be

$$T^* = \inf\left\{k \colon |\overline{Y}_{k+1} - \overline{Y}_k| > ca_k\right\}, \quad \text{or } n \text{ if no such } k, \tag{1.6}$$

where c is a prechosen constant. Note that, equivalently,

$$T^* = \inf\left\{k \colon |Y_{k+1} - \overline{Y}_k| > c\sqrt{(k+2)/(k+1)}\right\},$$

so that T^* is a stopping time based on prediction.

The intuitive appeal of $\overline{Y}_{T^*}$ is that it may capture some of the efficiency of the mean, while guarding against a disastrous change in the underlying process. This trade-off is controlled by c. If c is too small, then $T^* \ll \kappa$ and $\overline{Y}_{T^*}$ will lose efficiency, whereas if c is too large, then $T^* \gg \kappa$ and $\overline{Y}_{T^*}$ may include substantially biased observations. These characteristics are made precise by examining $\mathrm{MR}(\mathscr{F}_k, \overline{Y}_{T^*})$ for $k = 1, \dots, n$. This can be calculated by noting that, for any c and $k < n$, there exists a "malicious" $G_k^* \in \mathscr{F}_k$ with $\kappa(G_k^*) = k$ such that

$$P_{G_k^*}[T^* > k] = P_{G_k^*}[T^* = n],$$

$$P_{G_k^*}\left[\overline{Y}_n = \begin{cases} \overline{Y}_k + c \sum_{k+1}^n a_j, & \text{if } \overline{Y}_k > \mu \\ \overline{Y}_k - c \sum_{k+1}^n a_j, & \text{if } \overline{Y}_k < \mu \end{cases} \middle| \, T^* = n \right] = 1. \tag{1.7}$$

Thus

$$
\begin{aligned}
&\mathrm{MR}(\mathscr{F}_k, \overline{Y}_{T^*}) \\
&\quad = \sup_{F \in \mathscr{F}_k} \sum_{j=1}^{n} E_F\Big[\big(\overline{Y}_j - \mu\big)^2 \Big| T^* = j\Big] P_F[T^* = j] \\
&\quad = \sum_{j=1}^{k} E_{G_k^*}\Big[\big(\overline{Y}_j - \mu\big)^2 \Big| T^* = j\Big] P_{G_k^*}[T^* = j] \\
&\qquad + \left[\frac{1}{k+1} + 2\sqrt{\frac{2}{\pi(k+1)}}\, c \sum_{k+1}^{n} a_j + c^2\left(\sum_{k+1}^{n} a_j\right)^2\right] P_{G_k^*}[T^* = n] \qquad (1.8) \\
&\quad = \sum_{j=1}^{k} \frac{1}{j+1} \pi_c (1 - \pi_c)^{j-1} \\
&\qquad + \left[\frac{1}{k+1} + 2\sqrt{\frac{2}{\pi(k+1)}}\, c \sum_{k+1}^{n} a_j + c^2\left(\sum_{k+1}^{n} a_j\right)^2\right] (1 - \pi_c)^k,
\end{aligned}
$$

where $\pi_c = 2\Phi(-c)$ for Φ the standard normal cdf. Note that the calculation of the expectations and probabilities in the second equality above depends only on the fact that $G_k \in \mathscr{F}_k$. Interpreting the final equality of (1.8), the terms on the left for $j < k$ account for a loss of efficiency, whereas the rightmost expression accounts for potential effect of bias. Although c can be chosen to minimize $\mathrm{MR}(\mathscr{F}_k, \overline{Y}_{T^*})$ in (1.8) for a particular k, no uniformly best choice of c exists which minimizes (1.8) for all k.

As illustrated by $\overline{Y}_k$ and $\overline{Y}_{T^*}$, there is unfortunately no δ which simultaneously minimizes $\mathrm{MR}(\mathscr{F}_1, \delta), \dots, \mathrm{MR}(\mathscr{F}_n, \delta)$. The MR criterion is vector-valued and imposes only a partial ordering on the class of all estimators. Nonetheless, this criterion can be used to rule out many estimators.

DEFINITIONS. An estimator δ is said to be *MR-dominated* by another estimator δ^* if $\mathrm{MR}(\mathscr{F}_k, \delta^*) \le \mathrm{MR}(\mathscr{F}_k, \delta)$, $k = 1, \dots, n$, with strict inequality for some k. An estimator δ is said to be *MR-admissible* if it is not MR-dominated by another δ^*. A class of estimators is said to be essentially complete with respect to MR-admissibility if, given any estimator δ, there exists an estimator δ^* in the class for which $\mathrm{MR}(\mathscr{F}_k, \delta^*) \le \mathrm{MR}(\mathscr{F}_k, \delta)$ for $k = 1, \dots, n$.

MR-admissibility is in fact a generalized minimax criterion. When $n = 1$, MR-admissibility reduces to ordinary minimaxity. Note that MR-admissibility is different than admissibility in terms of risk. Indeed, neither implies the other.

Another approach to selecting an estimator with satisfactory $\mathrm{MR}(\mathscr{F}_k, \delta)$ for all k is to consider a one-dimensional summary criterion such as the following.

The *risk inflation* (RI) of an estimator δ is defined to be

$$\text{(1.9)}\qquad \mathrm{RI}(\delta) \equiv \max_k \sup_{F\in\mathscr{F}_k}\left[\frac{R(F,\delta)}{R(F,\bar{Y}_k)}\right] = \max_k\left[(k+1)\mathrm{MR}(\mathscr{F}_k,\delta)\right],$$

where the second equality follows from the fact that for all $F \in \mathscr{F}_k$, $R(F,\bar{Y}_k) \equiv 1/(k+1)$. The motivation for the risk inflation of δ is based on the fact that $\delta = \bar{Y}_k$ is minimax on $\mathscr{F}_k$, that is, $\mathrm{MR}(\mathscr{F}_k,\bar{Y}_k) = \inf_\delta \mathrm{MR}(\mathscr{F}_k,\delta)$ and so is best in terms of MR. Thus, $\mathrm{RI}(\delta)$ is a measure of the price of not knowing κ. Estimators with small risk inflation are desirable. A similar risk inflation measure is considered in the context of multiple regression by Foster and George (1993).

For $k > 1$, it is easy to see that $\mathrm{RI}(\bar{Y}_k) = \infty$ in accordance with the fact that using $\bar{Y}_k$ when $k > \kappa$ can be extremely dangerous. On the other hand, $\mathrm{RI}(\bar{Y}_1) = (n+1)/2$, in accordance with the fact that for its extreme safety, $\bar{Y}_1$ can pay a very high price in efficiency. It is interesting to consider the risk inflation of the adaptive compromise $\bar{Y}_{T^*}$. It can be shown, using (1.8), that $\mathrm{RI}(\bar{Y}_{T^*})$ is minimized at $c \approx \sqrt{2\log n}$, where $2\,\mathrm{MR}(\mathscr{F}_1,\bar{Y}_{T^*})$ is the dominant term, and $\mathrm{RI}(\bar{Y}_{T^*}) \approx 2(\log n)^3$. Note that as n increases, $\mathrm{RI}(\bar{Y}_{T^*})$ grows much more slowly than $\mathrm{RI}(\bar{Y}_1) = (n+1)/2$.

An even better alternative to $\bar{Y}_{T^*}$ (again assuming $\sigma^2 = 1$) is $\bar{Y}_{T^{**}}$, where

$$\text{(1.10)}\qquad T^{**} = \inf\left\{k\colon \left|\bar{Y}_{k+1} - \bar{Y}_j\right| > c a_{kj} \text{ for some } j \le k\right\}, \quad \text{or } n \text{ if no such } k,$$

$a_{kj} = 1/\sqrt{k+2} + 1/\sqrt{j+1}$ and c is a prechosen constant. The intuitive advantage of this estimator over $\bar{Y}_{T^*}$ is that it does not allow a gradual departure from the initial model. Although it is difficult to obtain an exact expression for $\mathrm{MR}(\mathscr{F}_k,\bar{Y}_{T^{**}})$, an argument similar to (1.9) obtains the bound

$$\text{(1.11)}\qquad \begin{aligned}\mathrm{MR}(\mathscr{F}_k,\bar{Y}_{T^{**}}) \le{}& \sum_{j=1}^{k}\frac{1}{j+1}P_{F_k}[T^{**}=j]\\ &+\left[\frac{1}{k+1} + 2\sqrt{\frac{2}{\pi(k+1)}}\,c a_{k,k-1} + c^2 a_{k,k-1}^2\right],\end{aligned}$$

where $F_k \in \mathscr{F}_k$. Furthermore, it can be shown that (using the same c) T^* is more likely to stop sooner than T^{**} (more precisely, for any $F \in \mathscr{F}_k$, $P_F[T^* = j] \ge P_F[T^{**} = j]$ for $j \le k$). Thus, the left-hand "efficiency loss term" in (1.11) is less than the corresponding term for $\bar{Y}_{T^*}$ in (1.9). It can also be shown, using (1.8), that a bound for $\mathrm{RI}(\bar{Y}_{T^*})$ is obtained when $c \approx \sqrt{2\log n}$, where the bound for $2\,\mathrm{MR}(\mathscr{F}_1,\bar{Y}_{T^{**}})$ is the dominant term. This yields the bound $\mathrm{RI}(\bar{Y}_{T^{**}}) \le 3.3\log n$, a substantial improvement over $\mathrm{RI}(\bar{Y}_{T^*})$. We show in Section 5 that this is close to the best possible risk inflation, which is $O(\log n)$. Although it is difficult to obtain a more complete analytical comparison of $\bar{Y}_{T^{**}}$ with $\bar{Y}_{T^*}$, we show in Section 4 that in terms of MR, estimators similar to $\bar{Y}_{T^{**}}$ are preferable to $\bar{Y}_{T^*}$.

The main thrust of the next three sections is to obtain usefully restrictive classes of estimators of μ which are essentially complete with respect to MR-admissibility. Our principal reduction is obtained by a generalization of the Hunt–Stein theorem which enables us to restrict attention to equivariant estimators. We then obtain an essentially complete subclass of the equivariant estimators which are similar to $\overline{Y}_{T^{**}}$. In Section 5, these results enable us to obtain a lower bound on the risk inflation of any estimator.

2. A class of equivariant estimators. In this section we describe a natural class of equivariant estimators for our problem. Based on the location and scale invariance of the general problem, we consider estimators satisfying

$$\delta(a + bY) = a + b\delta(Y), \tag{2.1}$$

for all real a, b with $b > 0$ [i.e., $\delta(a + bY_0, \dots, a + bY_n) = a + b\delta(Y_0, \dots, Y_n)$]. Such estimators are location and scale equivariant.

DEFINITION. Let $\mathscr{E}$ denote the class of equivariant estimators, that is, those satisfying (2.1).

Investigation of the members of $\mathscr{E}$ is greatly facilitated by making use of the following representations. Based on (2.1), any $\delta \in \mathscr{E}$ may be expressed as

$$\delta(Y) = \overline{Y}_k + V_k \omega_k(S_k, T_k) \quad \text{for } k = 1, \dots, n, \tag{2.2a}$$

$$\overline{Y}_k = \frac{1}{k+1}\sum_0^k Y_i, \qquad V_k = \left[\sum_{i=0}^k \left(Y_i - \overline{Y}_k\right)^2\right]^{1/2}, \qquad Z_{ik} = \frac{Y_i - \overline{Y}_k}{V_k}, \tag{2.2b}$$

$$S_k = (Z_{0k}, \dots, Z_{kk}), \qquad T_k = (Z_{k+1,k}, \dots, Z_{nk}), \qquad T_n \equiv 0, \tag{2.2c}$$

and ω_k is an arbitrary real-valued function. Note that under $F \in \mathscr{F}_k$, $\overline{Y}_k$, V_k and S_k are independent.

In order to treat any $\delta \in \mathscr{E}$ as sequentially determined, it is useful to consider the following sequential bounds. The largest and smallest possible values for δ after only $Y_0, \dots, Y_k$ have been observed are given by

$$\overline{Y}_k + V_k W_k^+(S_k), \quad \text{where } W_k^+(S_k) \equiv \sup_{T_k} \omega_k(S_k, T_k) \tag{2.3a}$$

and

$$\overline{Y}_k + V_k W_k^-(S_k), \quad \text{where } W_k^-(S_k) \equiv \inf_{T_k} \omega_k(S_k, T_k), \tag{2.3b}$$

respectively. The functions $W_k^+(S_k)$ and $W_k^-(S_k)$ are important characteristics of $\omega_k(S_k, T_k)$. For example, the next result shows that for any $\delta \in \mathscr{E}$ these characteristics determine $\mathrm{MR}(\mathscr{F}_k, \delta)$.

LEMMA 2.1. *For any* $\delta \in \mathscr{E}$,

$$\text{(2.4)}\quad \mathrm{MR}(\mathscr{F}_k, \delta) = E_{0,1}\Big[\max\Big\{\big[\bar{Y}_k + V_k W_k^+(S_k)\big]^2, \big[\bar{Y}_k + V_k W_k^-(S_k)\big]^2\Big\}\Big],$$

where $E_{0,1}$ *is expectation with respect to* $Y_0, \dots, Y_k$ *iid* $N(0,1)$.

PROOF. The maximum risk under $\mathscr{F}_k$ of $\delta \in \mathscr{E}$ may be expressed as

$$\begin{aligned}
\mathrm{MR}(\mathscr{F}_k, \delta) &= \sup_{\mathscr{F}_k} E\left[E_F^{\bar{Y}_k, S_k, V_k} \frac{\big[(\bar{Y}_k - \mu) + V_k \omega_k(S_k, T_k)\big]^2}{\sigma^2}\right] \\
&= E\left[\sup_{\mathscr{F}_k} E_F^{\bar{Y}_k, S_k, V_k} \frac{\big[(\bar{Y}_k - \mu) + V_k \omega_k(S_k, T_k)\big]^2}{\sigma^2}\right] \\
&= E\left[\max\left\{\left[\frac{\bar{Y}_k - \mu}{\sigma} + \frac{V_k}{\sigma} W_k^+(S_k)\right]^2, \left[\frac{\bar{Y}_k - \mu}{\sigma} + \frac{V_k}{\sigma} W_k^-(S_k)\right]^2\right\}\right].
\end{aligned}$$

□

Another useful representation of $\delta \in \mathscr{E}$, which is easy to conceptualize, is as a sequence of nested intervals. Define, for $k = 1, \dots, n$, the sequence of intervals

$$\text{(2.5)}\qquad B_k \equiv [B_k^-, B_k^+] \equiv \big[\bar{Y}_k + V_k W_k^-(S_k), \bar{Y}_k + V_k W_k^+(S_k)\big],$$

which by the definition of W_k^+ and W_k^- in (2.3) are nested. Thus, any equivariant δ may be defined by the sequence $B_1, \dots, B_n$ as

$$\text{(2.6)}\qquad B_1 \supset B_2 \supset \cdots \supset B_n \equiv \delta(Y).$$

It can happen that, for some $k < n$, B_k will also be a single point, in which case δ is determined by $Y_0, \dots, Y_k$. Our next result, which follows directly from Lemma 2.1 and (2.5), shows how the maximum risk of $\delta \in \mathscr{E}$ over $\mathscr{F}_k$ may be conveniently expressed in terms of its corresponding interval B_k.

LEMMA 2.2. *For any* $\delta \in \mathscr{E}$, $\mathrm{MR}(\mathscr{F}_k, \delta) = E_F[\sup_{x \in B_k}[(x - \mu)/\sigma]^2]$ *for any* $F \in \mathscr{F}_k$.

In the next two sections we investigate subclasses of $\mathscr{E}$ which contain "good" estimators. In particular we shall focus on the following subclass.

DEFINITION. Let $\mathscr{E}\mathscr{A} \subset \mathscr{E}$ denote the class of estimators which are MR-admissible within $\mathscr{E}$. (Thus $\delta \in \mathscr{E}\mathscr{A}$ iff $\delta \in \mathscr{E}$ and no other $\delta^* \in \mathscr{E}$ MR-dominates δ.)

In Section 3, we show that $\mathscr{E}\mathscr{A}$ is essentially complete with respect to MR-admissibility. Thus, in terms of MR, one can restrict attention to $\mathscr{E}\mathscr{A}$. In Section 4, we show that $\mathscr{E}\mathscr{A}$ consists of estimators δ whose corresponding sequence of intervals $B_1, \dots, B_n$ from (2.5) are as follows. For any Y, first define for $k = 1, \dots, n$, the "t-intervals" around the successive partial means

$$C_k \equiv [C_k^-, C_k^+] \equiv \left[\overline{Y}_k - V_k W_k, \overline{Y}_k + V_k W_k\right], \tag{2.7a}$$

where $W_1, \dots, W_n$ is a sequence of (possibly infinite) predetermined nonnegative constants with $W_n = 0$. Also, let $h_k\colon \mathbb{R} \to \mathbb{R}$, $k = 1, \dots, n$, be a sequence of predetermined functions with $h_n(\cdot) \equiv 0$. Starting with $B_0 \equiv (-\infty, \infty)$, the sequence of intervals $B_1, \dots, B_n$ is defined recursively by the following:

$$\text{if}\quad C_k \subset B_{k-1}^0, \quad \text{then} \quad B_k = C_k;$$

$$\text{if}\quad C_k \not\subset B_{k-1}^0 \quad \text{and} \quad \overline{Y}_k \geq \frac{B_{k-1}^- + B_{k-1}^+}{2},$$

$$\text{then}\quad B_k^+ = B_{k-1}^+ \quad \text{and}$$

$$B_k^- = \max\left[B_{k-1}^-, \min\left[B_{k-1}^+, \overline{Y}_k - V_k h_k\left(\frac{B_{k-1}^+ - \overline{Y}_k}{V_k}\right)\right]\right]; \tag{2.7b}$$

$$\text{if}\quad C_k \not\subset B_{k-1}^0 \quad \text{and} \quad \overline{Y}_k \leq \frac{B_{k-1}^- + B_{k-1}^+}{2},$$

$$\text{then}\quad B_k^- = B_{k-1}^- \quad \text{and}$$

$$B_k^+ = \min\left[B_{k-1}^+, \max\left[B_{k-1}^-, \overline{Y}_k + V_k h_k\left(\frac{\overline{Y}_k - B_{k-1}^-}{V_k}\right)\right]\right],$$

where B_k^0 denotes the interior of B_k. In order to understand this construction better, the reader may find it useful to consider the special case of (2.7) with $h_k \equiv W_k$.

For general δ defined by (2.7), if the successive partial means $\overline{Y}_1, \dots, \overline{Y}_n$ do not vary "too much" so that $C_1 \supset C_2 \supset \cdots \supset C_n$, then $\delta(Y) = C_n \equiv \overline{Y}_n$. However, if $\overline{Y}_k$ is far from the middle of B_{k-1} so that this nesting does not hold, then $\delta(Y)$ will be constrained to lie in B_{k-1}. It is interesting to compare δ (with $V_k \equiv 1$ to account for $\sigma^2 = 1$) and $\overline{Y}_{T**}$ defined by (1.10). Both estimators force the estimate to be contained within the intersection of intervals about previous means $\overline{Y}_k$. However, unlike $\overline{Y}_{T**}$, δ does not necessarily use one of the $\overline{Y}_k$ as the estimate. Instead δ may select an estimate closer than $\overline{Y}_k$ to the first "incompatible" $\overline{Y}_{k+1}$.

We close this section by remaking that, for the case where σ^2 is known, all of the previous results hold by setting $V_k \equiv 1$ throughout. In this situation, the class $\mathscr{E}$ is replaced by translation equivariant estimators of the form

$$\overline{Y}_k + \omega_k(S_k, T_k), \tag{2.8}$$

with $S_k = (Y_0 - \overline{Y}_k, \dots, Y_k - \overline{Y}_k)$ and $T_k = (Y_{k+1} - \overline{Y}_k, \dots, Y_n - \overline{Y}_k)$.

3. The essential completeness of $\mathscr{E}\mathscr{A}$. In this section we show that the class $\mathscr{E}\mathscr{A}$ is essentially complete with respect to MR-admissibility; that is, for any $\delta \notin \mathscr{E}\mathscr{A}$, there is a $\delta^* \in \mathscr{E}\mathscr{A}$ which is at least MR-equivalent to it. Since of course $\mathscr{E}\mathscr{A} \subset \mathscr{E}$, this then shows that $\mathscr{E}$ is essentially complete. This result is obtained by using the basic ideas of the Hunt–Stein theorem [see Berger (1985) and Lehmann (1986)]. The Hunt–Stein theorem, which demonstrates the overall minimaxity of rules which are minimax within the class of invariant rules, holds in general for statistical problems which are invariant under amenable groups [see Bondar and Milnes (1981)]. Although the location-scale group of our problem is amenable, our result extends the Hunt–Stein result to MR-admissibility, a generalization of minimaxity.

Our results here are presented in terms of Lemma 3.1, which shows that any estimator δ can be MR-approximated by some $\delta^* \in \mathscr{E}$, and Theorem 3.1, which concludes with the essential completeness of $\mathscr{E}\mathscr{A}$. For simplicity of notation and argument, the proofs of these results (which may be skipped with no loss of continuity), only consider the case where $\sigma^2 = 1$ is known so that δ is a translation equivariant estimator of the form (2.8). The details of the general case are similar. The proofs are based on the idea that if it were possible to construct $\delta^* \in \mathscr{E}$ from δ via $\delta^*(Y) \equiv \int[\delta(Y+t) - t]\,dt$, then δ^* would be MR-equivalent to or MR-better than δ by Jensen's inequality. Lemma 3.1 approximates this construction to obtain $\delta^* \in \mathscr{E}$ which has MR within ε of δ. Theorem 3.1 then uses a topological argument to show that the limit of such estimators is in $\mathscr{E}\mathscr{A}$ and is MR-equivalent to δ.

LEMMA 3.1. *For any δ and $\varepsilon > 0$, $\exists\ \delta^* \in \mathscr{E}$ such that* $\mathrm{MR}(\mathscr{F}_k, \delta^*) < \mathrm{MR}(\mathscr{F}_k, \delta) + \varepsilon$ *for all k.*

PROOF. For some constant $A > 0$ (to be determined), we will need the following intermediate estimators:

$$\delta^a(Y) = \begin{cases} (\bar{Y}_1 + A), & \text{if } \delta(Y) \geq (\bar{Y}_1 + A), \\ \delta(Y), & \text{otherwise}, \\ (\bar{Y}_1 - A), & \text{if } \delta(Y) \leq (\bar{Y}_1 - A), \end{cases} \tag{3.1}$$

$$\delta^b(Y) = \sum_{i=1}^{M} [\delta^a(Y + 4Ai) - 4Ai]/M, \tag{3.2}$$

$$\delta^c(Y) = 4A\lfloor \bar{Y}_1/4A \rfloor + \delta^b(\bar{Y}_1 \bmod 4A, T_1), \tag{3.3}$$

where $\lfloor\cdot\rfloor$ is the greatest integer part operator, and M is a large integer to be chosen later. Based on these estimators, we will show that

$$\delta^*(Y) = (1/4A)\int_0^{4A} \left[\delta^c(\bar{Y}_1 + a, T_1) - a\right] da \tag{3.4}$$

is the desired estimator. Note that because $\delta^c(Y + 4Ai) = \delta^c(Y) + 4Ai$ for all integer i, it follows that $\delta^*(a + Y) = a + \delta^*(Y)$ for all a, that is, $\delta^* \in \mathscr{E}$. It also follows immediately from this construction, using Jensen's inequality, that for all k, $\mathrm{MR}(\mathscr{F}_k, \delta^*) \le \mathrm{MR}(\mathscr{F}_k, \delta^c)$. Since we may pick A large enough so that, for all k, $\mathrm{MR}(\mathscr{F}_k, \delta^a) \le \mathrm{MR}(\mathscr{F}_k, \delta) + \varepsilon/2$, it suffices to show that we can choose A and M large enough so that $\mathrm{MR}(\mathscr{F}_k, \delta^c) \le \mathrm{MR}(\mathscr{F}_k, \delta^a) + \varepsilon/2$. We consider two cases.

CASE 1 ($A \le \mu \bmod 4A \le 3A$). First choose A large enough so that

$$E\left[|\overline{Y}_1 - \mu| + A\right]^2 I_{[|\overline{Y}_1 - \mu| \ge A]} < \varepsilon/2.$$

[As before, $\overline{Y}_1 \sim N(\mu, 1/2)$.] Now on the set where $|\overline{Y}_1 - \mu| < A$,

$$\delta^c(Y) = 4A\lfloor \mu/4A \rfloor + \delta^b\left(\overline{Y}_1 - 4A\lfloor \mu/4A \rfloor, T_1\right),$$

while if $|\overline{Y}_1 - \mu| \ge A$, $|\delta^c - \overline{Y}_1| \le A$ (this is always true by the definition of δ^a). Thus,

$$\begin{aligned}\mathrm{MR}(\mathscr{F}_k, \delta^c) &\le \mathrm{MR}(\mathscr{F}_k, \delta^b) + E\left[|\overline{Y}_1 - \mu| + A\right]^2 I_{[|\overline{Y}_1 - \mu| \ge A]}\\ &\le \mathrm{MR}(\mathscr{F}_k, \delta^b) + \varepsilon/2 \le \mathrm{MR}(\mathscr{F}_k, \delta^a) + \varepsilon/2\end{aligned}$$

where the last inequality follows by Jensen's inequality.

CASE 2 ($0 \le \mu \bmod 4A \le A$ or $3A \le \mu \bmod 4A \le 4A$). We will consider $0 \le \mu \bmod 4A \le A$. The other case follows similarly. First note that

$$\left(\overline{Y}_1 + 2A\right) \bmod 4A = \begin{cases} \overline{Y}_1 \bmod 4A + 2A, & \text{if } \overline{Y}_1 \bmod 4A \le 2A,\\ \overline{Y}_1 \bmod 4A - 2A, & \text{if } \overline{Y}_1 \bmod 4A > 2A.\end{cases}$$

Thus, if $\overline{Y}_1 \bmod 4A \le 2A$,

$$\delta^c(Y) = 4A\left\lfloor \frac{\overline{Y}_1 + 2A}{4A} \right\rfloor + \sum_{i=1}^{M} \frac{\delta^a\left(4Ai + \left(\overline{Y}_1 + 2A\right) \bmod 4A - 2A, T_1\right) - 4Ai}{M}$$

and if $\overline{Y}_1 \bmod 4A > 2A$,

$$\begin{aligned}\delta^c(Y) = {} & 4A\left\lfloor \frac{\overline{Y}_1 + 2A}{4A} \right\rfloor\\ & + \sum_{i=1}^{M} \frac{\delta^a\left(4A(i+1) + \left(\overline{Y}_1 + 2A\right) \bmod 4A - 2A, T_1\right) - 4A(i+1)}{M}.\end{aligned}$$

Thus,

$$\delta^c(Y) = 4A\left\lfloor \frac{\overline{Y}_1 + 2A}{4A} \right\rfloor + \delta^b\left(\left(\overline{Y}_1 + 2A\right) \bmod 4A - 2A, T_1\right) + \frac{R}{M},$$

where $|R| \le 2A$. Thus, for M large enough and transforming Y to $(Y - 2A)$, the argument for Case 1 may be used to show $\mathrm{MR}(\mathscr{F}_k, \delta^c) \le \mathrm{MR}(\mathscr{F}_k, \delta^a) + \varepsilon/2$. □

THEOREM 3.1. *The class $\mathscr{E}\mathscr{A}$ is essentially complete with respect to MR-admissibility.*

PROOF. We will show that for any δ there exists $\delta^* \in \mathscr{E}\mathscr{A}$ such that $\mathrm{MR}(\mathscr{F}_k, \delta^*) \le \mathrm{MR}(\mathscr{F}_k, \delta)$ for all k. By Lemma 3.1, there exists a sequence $\delta_1, \delta_2, \ldots \in \mathscr{E}$ such that

$$\limsup_{i \to \infty} \mathrm{MR}(\mathscr{F}_k, \delta_i) \le \mathrm{MR}(\mathscr{F}_k, \delta) \quad \text{for all } k.$$

Furthermore, the sequence can be chosen so that no other δ' has the property of MR-dominating the limit, that is, $\mathrm{MR}(\mathscr{F}_k, \delta^*) \le \liminf_{i \to \infty} \mathrm{MR}(\mathscr{F}_k, \delta_i)$ for all k, with strict inequality for some k. Therefore, it suffices to find an estimator $\delta^* \in \mathscr{E}$ such that

$$\mathrm{MR}(\mathscr{F}_k, \delta^*) \le \liminf_{i \to \infty} \mathrm{MR}(\mathscr{F}_k, \delta_i) \tag{3.5}$$

for all k. By (2.2), we may express δ_i as $\delta_i(Y) = \bar{Y}_n + \omega_n^i(S_n)$ (recall $V_n = 1$ since we are assuming $\sigma^2 = 1$). Now from the sequence $\delta_1, \delta_2, \ldots$ we may extract a subsequence $\delta_{i_1}, \delta_{i_2}, \ldots \in \mathscr{E}$ such that $(S_n, \omega_n^{i_j}(S_n))$ converges in distribution as $j \to \infty$. Furthermore, there exists a random vector, say, $(S_n, \omega_n''(S_n))$, which has this limiting distribution. However, then the (possibly randomized) estimator $\delta'' \equiv \bar{Y}_n + \omega_n''(S_n)$ belongs to $\mathscr{E}$ and satisfies (3.5) for $k = n$, by the continuous mapping theorem.

It remains to show that δ'' can be modified (on a set of measure zero) to satisfy (3.5) for all k. For the estimator δ_i, let $W_k^{+,i}, W_k^{-,i}$, $k = 1, \ldots, n$, be the corresponding bounds in (2.3) (recall that $W_n^{+,i} = W_n^{-,i} = \omega_n^i$). From the subsequence $\delta_{i_1}, \delta_{i_2}, \ldots$ extract a further subsequence $\delta_{i'_1}, \delta_{i'_2}, \ldots \in \mathscr{E}$ such that

$$\left(S_1, W_1^{+,i'_j}, W_1^{-,i'_j}, S_2, W_2^{+,i'_j}, W_2^{-,i'_j}, \ldots, S_n, W_n^{+,i'_j}, W_n^{-,i'_j}\right) \tag{3.6}$$

converges in distribution as $j \to \infty$. (Note that the redundancy in this vector causes no problem for convergence in distribution.) Now there exist $W_k^{+,\infty}, W_k^{-,\infty}$, $k = 1, \ldots, n$, with the property that $(S_1, W_1^{+,\infty}, W_1^{-,\infty}, S_2, W_2^{+,\infty}, W_2^{-,\infty}, \ldots, S_n, W_n^{+,\infty}, W_n^{-,\infty})$ has the limiting distribution of (3.6). Since $W_k^{+,\infty}, W_k^{-,\infty}$ are independent of Y_i, $i > k$ for $k = 1, \ldots, n$, we may recursively construct randomized $W_k^{+,\infty}, W_k^{-,\infty}$ which depend only on S_k and $W_i^{+,\infty}, W_i^{-,\infty}$ for $i < k$. Now define

$$\delta^* = \sup\left\{y \colon y \in \bigcap_{i=1}^{k^*} \left[\bar{Y}_i - W_i^{-,\infty}, \bar{Y}_i + W_i^{+,\infty}\right]\right\}, \tag{3.7}$$

where $k^* = \sup\{k\colon \bigcap_{i=1}^{k}[\bar{Y}_i - W_i^{-,\infty}, \bar{Y}_i + W_i^{+,\infty}] \ne \phi\}$. Clearly, the (possibly randomized) estimator δ^* belongs to $\mathscr{E}$. Furthermore, $\bar{Y}_k - W_k^{-,\infty} \le \delta^* \le \bar{Y}_k + W_k^{+,\infty}$ so that by Lemma 2.1 and the continuous mapping theorem, δ^* satisfies (3.5) for all k. □

Note that if δ^* above is a randomized rule, it can be replaced by $\delta^{**} = E(\delta^* | Y) \in \mathscr{E}\mathscr{A}$.

4. A partial characterization of $\mathscr{E}\mathscr{A}$. The purpose of this section is to show that all estimators in $\mathscr{E}\mathscr{A}$ must satisfy (2.7a) and (2.7b). Theorem 3.1 then shows that the class of estimators of the form (2.7a)–(2.7b) is essentially complete. Although the result (2.7a)–(2.7b) stops short of a full characterization of members of $\mathscr{E}\mathscr{A}$, it does eliminate many equivariant estimators which can be MR-dominated. Many of these results are obtained using the following lemma, which allows for a partial "Rao–Blackwellization" of any $\delta \in \mathscr{E}$.

LEMMA 4.1. *For any $\delta \in \mathscr{E}$, $F_n \in \mathscr{F}_n$, g: $\mathbb{R}^{k+1} \to \mathbb{R}$ and equivariant $\tilde{g}$: $\mathbb{R}^{k+1} \to \mathbb{R}$, the following hold:*

(i) $\delta^*(Y) \equiv E_{F_n}[\delta(Y)|g(S_k), \bar{Y}_k, V_k, T_k]$ *has* $\mathrm{MR}(\mathscr{F}_j, \delta^*) \le \mathrm{MR}(\mathscr{F}_j, \delta)$ *for* $j \ge k$.

(ii) $\delta^*(Y) \equiv E_{F_n}[\delta(Y)|\tilde{g}(Y_0, \dots, Y_k), \bar{Y}_k, V_k, T_k]$ *has* $\mathrm{MR}(\mathscr{F}_j, \delta^*) \le \mathrm{MR}(\mathscr{F}_j, \delta)$ *for* $j \ge k$.

PROOF. (i) It suffices to show that for $j \ge k$, for any $F \in \mathscr{F}_j$ there exists $F^* \in \mathscr{F}_j$ such that $R(F, \delta^*) \le R(F^*, \delta)$. From this it will follow that $\mathrm{MR}(\mathscr{F}_j, \delta^*) \le \mathrm{MR}(\mathscr{F}_j, \delta)$.

For $F \in \mathscr{F}_j$, define F^* to be the probability distribution satisfying

$$E_{F^*}[\cdot] = E_F E_{F_n^*}\big[\cdot|g(S_k), \bar{Y}_k, V_k, T_k\big], \tag{4.1}$$

where $F_n^* \in \mathscr{F}_n$ is such that $Y_0, \dots, Y_j$ has the same distribution under F_n^* and F. Note that

$$\delta^* = E_{F_n^*}\big[\delta(Y)|g(S_k), \bar{Y}_k, V_k, T_k\big].$$

First we show that $F^* \in \mathscr{F}_j$. Let A be a cylinder set $A = A^j \times \mathbb{R}^{n-j}$, where $A^j \subset \mathbb{R}^{j+1}$. It suffices to show $P_{F^*}(Y \in A) = P_F(Y \in A)$. Letting $I_A(Y)$ be the indicator function of A,

$$\begin{aligned} P_{F^*}[A] = E_{F^*}[I_A] &= E_F E_{F_n^*}\big[I_A|g(S_k), \bar{Y}_k, V_k, Y_{k+1}, Y_{k+2}, \dots, Y_j, T_j\big] \\ &= E_F E_{F_n^*}\big[I_A|g(S_k), \bar{Y}_k, V_k, Y_{k+1}, Y_{k+2}, \dots, Y_j\big] \\ &= E_F E_F\big[I_A|g(S_k), \bar{Y}_k, V_k, Y_{k+1}, Y_{k+2}, \dots, Y_j\big] = E_F[I_A] = P_F[A]. \end{aligned}$$

Writing $\delta^* = \bar{Y}_k + \omega_k^* V_k$, where $\omega_k^* \equiv E_{F_n^*}[\omega_k|g(S_k), \bar{Y}_k, V_k, T_k]$, it now follows that $R(F, \delta^*) \le R(F^*, \delta)$ since

$$\begin{aligned} \sigma^2 R(F, \delta^*) = E_F(\delta^* - \mu)^2 &= E_F\big(\bar{Y}_k - \mu\big)^2 + 2E_F\big(\bar{Y}_k - \mu\big)V_k\omega_k^* + E_F V_k^2 \omega_k^{*2} \\ &\le E_{F^*}\big(\bar{Y}_k - \mu\big)^2 + 2E_{F^*}\big(\bar{Y}_k - \mu\big)V_k\omega_k + E_{F^*}V_k^2\omega_k^2 = \sigma^2 R(F^*, \delta), \end{aligned}$$

where $E_F(\bar{Y}_k - \mu)^2 = E_{F^*}(\bar{Y}_k - \mu)^2$ because $F, F^* \in \mathscr{F}_j \subset \mathscr{F}_k$, and

$$\begin{aligned} E_{F^*}\big(\bar{Y}_k - \mu\big)V_k\omega_k &= E_F E_{F_n^*}\big[\big(\bar{Y}_k - \mu\big)V_k\omega_k|g(S_k), \bar{Y}_k, V_k, T_k\big] \\ &= E_F\big(\bar{Y}_k - \mu\big)V_k E_{F_n^*}\big[\omega_k|g(S_k), \bar{Y}_k, V_k, T_k\big] \\ &= E_F\big(\bar{Y}_k - \mu\big)V_k\omega_k^* \end{aligned}$$

by (4.1), and

$$E_{F^*}V_k^2\omega_k^2 = E_F E_{F_n^*}\left[V_k^2\omega_k^2 | g(S_k), \bar{Y}_k, V_k, T_k\right] = E_F V_k^2 E_{F_n^*}\left[\omega_k^2 | g(S_k), \bar{Y}_k, V_k, T_k\right]$$
$$\geq E_F V_k^2\left[E_{F_n^*}\left[\omega_k | g(S_k), \bar{Y}_k, V_k, T_k\right]\right]^2 = E_F V_k^2 \omega_k^{*2}$$

by Jensen's inequality.

Statement (ii) follows from statement (i) by letting $\tilde{g} = \bar{Y}_k + V_k g$. □

The following lemma, based on an enhancement of Jensen's inequality, shows when a convex combination of estimators obtains a strict improvement in MR.

LEMMA 4.2. *Let $\delta, \delta^* \in \mathscr{E}$. Suppose for some k, $\mathrm{MR}(\mathscr{F}_k, \delta^*) \leq \mathrm{MR}(\mathscr{F}_k, \delta)$ and that the intervals B_k and B_k^* from* (2.5) *are such that, for any $F \in \mathscr{F}_k$, $P_F[B_k \neq B_k^*] > 0$. Then for any $\rho \in (0,1)$, $\delta^{**} = \rho\delta + (1-\rho)\delta^{**}$ has $\mathrm{MR}(\mathscr{F}_k, \delta^{**}) < \mathrm{MR}(\mathscr{F}_k, \delta)$.*

PROOF. If $\mathrm{MR}(\mathscr{F}_k, \delta^*) < \mathrm{MR}(\mathscr{F}_k, \delta)$, the result follows directly from Jensen's inequality. When $\mathrm{MR}(\mathscr{F}_k, \delta^*) = \mathrm{MR}(\mathscr{F}_k, \delta)$, let B_k^{**} be the interval for δ^{**} from (2.5). For any $F_k \in \mathscr{F}_k$ with $\sigma^2 = 1$, we have by Lemma 2.2 that $\mathrm{MR}(\mathscr{F}_k, \delta^{**}) = E_{F_k}\max[(B_k^{**+} - \mu)^2, (B_k^{**-} - \mu)^2]$.

Because

$$\rho B_k^+ + (1-\rho)B_k^{*+} \geq B_k^{**+} \quad \text{and} \quad B_k^{**-} \geq \rho B_k^- + (1-\rho)B_k^{*-},$$

it follows that

$$\begin{aligned}
&\max\left[(B_k^{**+} - \mu)^2, (B_k^{**-} - \mu)^2\right] \\
&\quad\leq \max\left[\left[\rho(B_k^+ - \mu)^2 + (1-\rho)(B_k^{*+} - \mu)^2\right],\right. \\
&\qquad\qquad \left.\left(\rho(B_k^- - \mu)^2 + (1-\rho)(B_k^{*-} - \mu)^2\right)\right] \qquad (4.2)\\
&\quad\leq \rho\max\left[(B_k^+ - \mu)^2, (B_k^- - \mu)^2\right] \\
&\qquad + (1-\rho)\max\left[(B_k^{*+} - \mu)^2, (B_k^{*-} - \mu)^2\right].
\end{aligned}$$

Without loss of generality, assume that $P_{F_k}[B_k^+ \neq B_k^{*+}] > 0$. Thus, for μ small, there exists a set A with $P_{F_k}[A] > 0$ such that, on A, $B_k^+ \neq B_k^{*+}$ and $(\rho(B_k^+ - \mu)^2 + (1-\rho)(B_k^{*+} - \mu)^2]$ is the larger term in the middle expression of (4.2). It is straightforward to show that on A, the final inequality of (4.2) is strict. By taking expectations (under F_k) of the two sides of (4.2) we then have

$$\mathrm{MR}(\mathscr{F}_k, \delta^{**}) < \rho\,\mathrm{MR}(\mathscr{F}_k, \delta) + (1-\rho)\mathrm{MR}(\mathscr{F}_k, \delta^*) = \mathrm{MR}(\mathscr{F}_k, \delta). \quad □$$

We now proceed to show that the estimators in $\mathscr{E}\mathscr{A}$ satisfy (2.7a)–(2.7b). Our strategy is to impose successively restrictions on $\mathscr{E}$ which leave $\mathscr{E}\mathscr{A}$ intact. This consists of forming a sequence of subclasses $\mathscr{E}_1, \mathscr{E}_2, \ldots, \mathscr{E}_6$ such that $\mathscr{E} \supset \mathscr{E}_1 \supset \mathscr{E}_2 \supset \cdots \supset \mathscr{E}_6 \supset \mathscr{E}\mathscr{A}$. We begin with the definition of $\mathscr{E}_1$, which forces B_k to be a function only of B_{k-1}, $\overline{Y}_k$ and V_k. In what follows, it will be convenient for definitional considerations to let $B_0 \equiv (-\infty, \infty)$.

DEFINITION. Let $\mathscr{E}_1 \subset \mathscr{E}$ consist of those δ for which B_k is a.s. a function only of $B_{k-1}, \overline{Y}_k$, and V_k for $k = 1, \dots, n$.

Before continuing our development, we should clarify our use of almost sure (a.s.). By construction, the interval B_k is measurable with respect to $Y_0, \dots, Y_k$. Therefore, we shall mean any a.s. statement about B_k to be with respect to $F \in \mathscr{F}_k$. In particular, it is convenient to consider the special case $F_n \in \mathscr{F}_n \subset \mathscr{F}_k$ which does not depend on k.

THEOREM 4.1. $\mathscr{E}\mathscr{A} \subset \mathscr{E}_1$.

PROOF. We will prove that for any $\delta \in \mathscr{E} - \mathscr{E}_1$, there exists δ^{**} which MR-dominates δ. This will imply $\delta \notin \mathscr{E}\mathscr{A}$ so that $(\mathscr{E} \cap \mathscr{E}\mathscr{A}) \subset \mathscr{E}_1$.

Pick $\delta \in \mathscr{E} - \mathscr{E}_1$ with its associated $B_1, \dots, B_n$ from (2.5). Because $\delta \notin \mathscr{E}_1$, we can choose k such that B_k is not a function of $B_{k-1}, \overline{Y}_k$ and V_k on a set of positive measure. For any $F_n \in \mathscr{F}_n$, define

$$\delta^*(Y) = E_{F_n}\left[\delta(Y) | B_{k-1}, \overline{Y}_k, V_k, T_k\right]. \tag{4.3}$$

Let $B_1^*, \dots, B_n^*$ be obtained from (2.5) for δ^*. Since $\delta \in B_{k-1}$ and (4.3) is conditioned on B_{k-1}, we have that $\delta^* \in B_{k-1}$. In other words, $B_{k-1}^* \subset B_{k-1}$, which in turn implies $B_j^* \subset B_j$ for $j < k$. By Lemma 2.2, it follows that $\mathrm{MR}(\mathscr{F}_j, \delta^*) \le \mathrm{MR}(\mathscr{F}_j, \delta)$ for $j < k$. Furthermore, since B_{k-1} is an equivariant function of $Y_0, \dots, Y_k$, it follows from (ii) of Lemma 4.1 that $\mathrm{MR}(\mathscr{F}_j, \delta^*) \le \mathrm{MR}(\mathscr{F}_j, \delta)$ for $j \ge k$.

Define $\delta^{**} = (\delta + \delta^*)/2$. By Jensen's inequality, $\mathrm{MR}(\mathscr{F}_j, \delta^{**}) \le \mathrm{MR}(\mathscr{F}_j, \delta)$ for $j = 1, \dots, n$. Finally, $P_{F_n}[B_k \ne B_k^*] \ge 0$ so that, by Lemma 4.2, $\mathrm{MR}(\mathscr{F}_k, \delta^{**}) < \mathrm{MR}(\mathscr{F}_k, \delta)$. □

Our next subclass $\mathscr{E}_2$ restricts δ to be a.s. antisymmetric. As will be seen, this property obtains many of the symmetric aspects of the conditions (2.7a)–(2.7b).

DEFINITION. Let $\mathscr{E}_2 \subset \mathscr{E}_1$ consists of those δ which are a.s. antisymmetric, that is, $B_j(Y) = -B_j(-Y)$ a.s. for $j = 1, \dots, n$.

THEOREM 4.2. $\mathscr{E}\mathscr{A} \subset \mathscr{E}_2$.

PROOF. We will prove that for any $\delta \in \mathscr{E}_2 - \mathscr{E}_1$, there exists δ^* which MR-dominates δ. This will imply $\delta \notin \mathscr{E}\mathscr{A}$ so that $(\mathscr{E}_1 \cap \mathscr{E}\mathscr{A}) \subset \mathscr{E}_2$.

Suppose $\delta \in \mathscr{E}_2 - \mathscr{E}_1$. Define the antisymmetric estimator $\delta^*(Y) = [\delta(Y) - \delta(-Y)]/2$. Note that δ^* also belongs to $\mathscr{E}_1$ and so belongs to $\mathscr{E}_2$. By symmetry, it is obvious that $\mathrm{MR}(\mathscr{F}_j, -\delta(-Y)) = \mathrm{MR}(\mathscr{F}_j, \delta)$ so that by Jensen's inequality, $\mathrm{MR}(\mathscr{F}_j, \delta^*) \le \mathrm{MR}(\mathscr{F}_j, \delta)$ for $j = 1, \dots, n$. Since, for some k, $B_k \ne B_k^*$ on a set of positive measure, $\mathrm{MR}(\mathscr{F}_k, \delta^*) < \mathrm{MR}(\mathscr{F}_k, \delta)$ by Lemma 4.2. □

The next subclass $\mathscr{E}_3$ restricts attention to those δ for which $B_k \subset B_{k-1}^0$ implies $B_k = C_k$ in (2.7a) a.s. (B_{k-1}^0 is the interior of B_{k-1}).

DEFINITION. Let $\mathscr{E}_3 \subset \mathscr{E}_2$ consist of those δ which satisfy the following. Corresponding to δ, there exists a sequence of (possibly infinite) constants $W_1, \dots, W_n$ such that whenever $B_k \subset B_{k-1}^0$, $B_k = [\bar{Y}_k - V_k W_k, \bar{Y}_k + V_k W_k]$ a.s., $k = 1, \dots, n$.

THEOREM 4.3. $\mathscr{E}\mathscr{A} \subset \mathscr{E}_3$.

PROOF. We will prove that, for any $\delta \in \mathscr{E}_2 - \mathscr{E}_3$, there exists δ^{**} which MR-dominates δ. This will imply $\delta \notin \mathscr{E}\mathscr{A}$ so that $(\mathscr{E}_2 \cap \mathscr{E}\mathscr{A}) \subset \mathscr{E}_3$.

Pick $\delta \in \mathscr{E}_2 - \mathscr{E}_3$ with its associated $B_1, \dots, B_n$ from (2.5). Since $\delta \notin \mathscr{E}_3$, we can choose k so that

$$A = \left(Y \colon B_k \subset B_{k-1}^0 \text{ and } B_k \ne \left[\bar{Y}_k - V_k W_k, \bar{Y}_k + V_k W_k\right]\right) \tag{4.4}$$

has positive measure. For $\varepsilon > 0$, define

$$A_\varepsilon = \left(Y \colon [B_k^- - \varepsilon, B_k^+ + \varepsilon] \subset B_{k-1}^0 \cap \left[\bar{Y}_k - 1/\varepsilon, \bar{Y}_k + 1/\varepsilon\right]\right). \tag{4.5}$$

Note that $A \subset \lim_{\varepsilon \to 0} A_\varepsilon$. Now construct

$$\delta^*(Y) = E_{F_n}\left[\delta(Y) | I_{A_\varepsilon}, \left[1 - I_{A_\varepsilon}\right] \cdot B_{k-1}, \bar{Y}_k, V_k, T_k\right]. \tag{4.6}$$

Because I_{A_ε} and $[1 - I_{A_\varepsilon}] \cdot B_{k-1}$ are equivariant functions of $Y_0, \dots, Y_k$, it follows from (ii) of Lemma 4.1 that $\mathrm{MR}(\mathscr{F}_j, \delta^*) \le \mathrm{MR}(\mathscr{F}_j, \delta)$ for $j \ge k$.

Now consider the estimator $\delta^{**} \equiv (1 - \rho)\delta + \rho\delta^*$ with $\rho = \varepsilon^2/2$. By Jensen's inequality, $\mathrm{MR}(\mathscr{F}_j, \delta^{**}) \le \mathrm{MR}(\mathscr{F}_j, \delta)$ for $j \ge k$. Note that by (2.3) and Theorem 4.2, B_k must satisfy $W_k^+(S_k) = -W_k^-(-S_k)$. Thus, $B_k \ne B_k^*$ on $A \cap A_\varepsilon$. Furthermore, because $A \cap A_\varepsilon$ has positive measure for ε small enough, $\mathrm{MR}(\mathscr{F}_k, \delta^{**}) < \mathrm{MR}(\mathscr{F}_k, \delta)$ by Lemma 4.2.

For $j < k$, note that because

$$\sup_{Y \in \mathbb{R}^n} |\delta(Y) - \delta^*(Y)| = \sup_{Y \in A_\varepsilon} |\delta(Y) - \delta^*(Y)| \le 2/\varepsilon$$

we have that $|\delta^{**} - \delta| < \varepsilon$ on A_ε, and $\delta^{**} \equiv \delta$ on A_ε^c, where $\delta^* = \delta$. Since $\delta \pm \varepsilon \in B_{k-1}$ on A_ε (and of course $\delta \in B_{k-1}$ on A_ε^c), $\delta^{**} \in B_{k-1}$. Letting $B_1^{**}, \dots, B_{k-1}^{**}$ be the intervals associated with δ^{**} from (2.5), this implies that $B_j^{**} \subset B_j$ for $j < k$. By Lemma 2.2, $\mathrm{MR}(\mathscr{F}_j, \delta^{**}) \le \mathrm{MR}(\mathscr{F}_j, \delta)$ for $j < k$. □

When $B_k \not\subset B_{k-1}^0$, the behavior of B_k is more complicated. The next subclass $\mathscr{E}_4$ restricts attention to δ for which $B_k \not\subset B_{k-1}^0$ implies that $B_k = B_{k-1} \cap [\bar{Y}_k - U_k, \bar{Y}_k + U_k]$, where $U_k \equiv U_k(B_{k-1}, \bar{Y}_k, V_k)$. This accounts for the endpoint functions h_k in (2.7b).

DEFINITION. Let $\mathscr{E}_4 \subset \mathscr{E}_3$ consist of those δ with associated $B_1, \dots, B_n$ which when $B_k \not\subseteq B_{k-1}^0$ satisfy the following:

$$
\text{(4.7)} \quad
\begin{aligned}
&\text{(i) if } \bar{Y}_k \geq \frac{B_{k-1}^+ + B_{k-1}^-}{2}, \text{ then } B_k^+ = B_{k-1}^+ \quad \text{and} \quad \bar{Y}_k \geq \frac{B_k^+ + B_k^-}{2};\\
&\text{(ii) if } \bar{Y}_k \leq \frac{B_{k-1}^+ + B_{k-1}^-}{2}, \text{ then } B_k^- = B_{k-1}^- \quad \text{and} \quad \bar{Y}_k \leq \frac{B_k^+ + B_k^-}{2}.
\end{aligned}
$$

THEOREM 4.4. $\mathscr{E}\mathscr{A} \subset \mathscr{E}_4$.

PROOF. Follow the proof of Theorem 4.3, replacing A, A_ε and δ^* in (4.4)–(4.6) by

$$
\begin{aligned}
A &= \left(Y\colon B_k \not\subseteq B_{k-1}^0 \text{ but (4.7) violated}\right),\\
A_\varepsilon &= A \cap \left(Y\colon |B_k^+ - B_{k-1}^+| + |B_k^- - B_{k-1}^-| > \varepsilon\right)\\
&\quad \cap \left(Y\colon B_k \subset \left[\bar{Y}_k - \frac{1}{\varepsilon}, \bar{Y}_k + \frac{1}{\varepsilon}\right]\right),\\
\delta^*(Y) &= E_{F_n}\Bigg[\delta(Y) | B_k^+ - B_k^-, \left|\bar{Y}_k - \frac{B_{k-1}^+ + B_{k-1}^-}{2}\right|,\\
&\qquad I_{A_\varepsilon}, \left[1 - I_{A_\varepsilon}\right] \cdot B_{k-1}, \bar{Y}_k, V_k, T_k\Bigg]. \qquad \square
\end{aligned}
$$

The next subclass $\mathscr{E}_5$ restricts attention to those δ for which $C_k \subset B_{k-1}^0$ implies $B_k = C_k$ in (2.7a) a.s.

DEFINITION. For $\delta \in \mathscr{E}_4$, let $W_1, \dots, W_n$ be the sequence of constants associated with $B_1, \dots, B_n$ (via the definition of $\mathscr{E}_3$) with the added stipulation that whenever $B_k \not\subseteq B_{k-1}^0$ a.s., $W_k = \infty$. Let $\mathscr{E}_5 \subset \mathscr{E}_4$ consist of those δ such that whenever $[\bar{Y}_k - V_k W_k, \bar{Y}_k + V_k W_k] \subset B_{k-1}^0$, $B_k = [\bar{Y}_k - V_k W_k, \bar{Y}_k + V_k W_k]$ a.s., $k = 1, \dots, n$.

THEOREM 4.5. $\mathscr{E}\mathscr{A} \subset \mathscr{E}_5$.

PROOF. We will prove that for any $\delta \in \mathscr{E}_4 - \mathscr{E}_5$, there exists δ^{**} which MR-dominates δ. This will imply $\delta \notin \mathscr{E}\mathscr{A}$ so that $(\mathscr{E}_4 \cap \mathscr{E}\mathscr{A}) \subset \mathscr{E}_5$.

Pick $\delta \in \mathscr{E}_4 - \mathscr{E}_5$. Because $\delta \notin \mathscr{E}_5$, we can choose k such that for some $\varepsilon > 0$, $P_{F_n}[A_1] > \varepsilon$ and $P_{F_n}[A_2] > 0$, where

$$
\begin{aligned}
A_1 &\equiv \left(Y\colon [B_k^- - \varepsilon, B_k^+ + \varepsilon] \subset B_{k-1}^0 \text{ and } B_k = \left[\bar{Y}_k - V_k W_k, \bar{Y}_k + V_k W_k\right]\right),\\
A_2 &\equiv \left(Y\colon \left[\bar{Y}_k - V_k W_k - \varepsilon, \bar{Y}_k + V_k W_k + \varepsilon\right] \subset B_{k-1}^0\right).
\end{aligned}
$$

Pick $A_3 \subset A_2$ such that $P_{F_n}[A_3] > 0$, $E_{F_n}[|\bar{Y}_k - (B_{k-1}^+ + B_{k-1}^-)/2| I_{A_3}] \leq \varepsilon^2/2$

and $E_{F_n}[(B_k^+ - B_k^-)I_{A_3}] \leq \varepsilon^2/2$. Define

$$\delta^*(Y) = E_{F_n}\big[\delta(Y)|I_{[A_1 \cup A_3]}, \big[1 - I_{[A_1 \cup A_3]}\big] \cdot B_{k-1}, \bar{Y}_k, V_k, T_k\big].$$

By construction $B_k^* \subset B_{k-1}$ (B_k^* corresponds to δ^*). Thus $B_j^* \subset B_j$ for $j < k$, which in turn implies that $\mathrm{MR}(\mathscr{F}_j, \delta^*) \leq \mathrm{MR}(\mathscr{F}_j, \delta)$ for $j < k$. For $j \geq k$, we appeal to Lemma 4.1, which shows that $\mathrm{MR}(\mathscr{F}_j, \delta^*) \leq \mathrm{MR}(\mathscr{F}_j, \delta)$ for $j \geq k$ by virtue of the fact that $I_{[A_1 \cup A_3]}$ and $[1 - I_{[A_1 \cup A_3]}] \cdot B_{k-1}$ are equivariant functions of $Y_0, \dots, Y_k$.

Define $\delta^{**} = (\delta + \delta^*)/2$. By Jensen's inequality, $\mathrm{MR}(\mathscr{F}_j, \delta^{**}) \leq \mathrm{MR}(\mathscr{F}_j, \delta)$ for $j = 1, \dots, n$. Finally, $P_{F_n}[B_k \neq B_k^*] > 0$ so that, by Lemma 4.2, $\mathrm{MR}(\mathscr{F}_k, \delta^{**}) < \mathrm{MR}(\mathscr{F}_k, \delta)$. □

Finally, the subclass $\mathscr{E}_6$ is defined by (2.7a)–(2.7b). Note that $\mathscr{E}_6 \subset \mathscr{E}_5$ is obtained from $\mathscr{E}_5$ by restricting attention to those δ with h_k depending on at most one of B_{k-1}^+ or B_{k-1}^-.

Definition. Let $\mathscr{E}_6$ consist of those δ with $B_1, \dots, B_n$ satisfying (2.7a)–(2.7b) a.s.

Theorem 4.6. *$\mathscr{E}_6$ is essentially complete.*

Proof. Any $\delta \in \mathscr{E}_5$ can be expressed in the form (2.7a)–(2.7b) except that the endpoint functions h_k may depend on both B_{k-1}^- and B_{k-1}^+. To obtain the final simplification, use that fact that, by (2.3) and Theorem 4.2, B_k for $\delta \in \mathscr{E}_5$ must satisfy $W_k^+(S_k) = -W_k^-(-S_k)$, and apply the argument of Theorem 4.3, replacing A, A_ε and δ^* in (4.4)–(4.6) by

$$A = (Y: h_k \text{ depends on both } B_{k-1}^+ \text{ and } B_{k-1}^-),$$

$$A_\varepsilon = \big(Y: [B_k^- - \varepsilon, B_k^+] \subset B_{k-1} \cap \big[\bar{Y}_k + 1/\varepsilon, \bar{Y}_k + 1/\varepsilon\big]\big),$$

$$\delta^*(Y) = E_{F_n}\big[\delta(Y)|I_{A_\varepsilon}, B_{k-1}^+, \big[1 - I_{A_\varepsilon}\big] \cdot B_{k-1}^-, \bar{Y}_k, V_k, T_k\big]. \qquad \square$$

Unfortunately, the description in (2.7a)–(2.7b) does not fully characterize the members of $\mathscr{E}\mathscr{A}$. The remaining (and very difficult) open question is to find the restrictions which characterize the endpoint functions h_k. Simulation evidence seems to indicate that these functions need not be linear in $\bar{Y}_k$ as we had initially suspected.

Finally, we remark that for the case where σ^2 is known, all of the results of the section hold by setting $V_k \equiv 1$ throughout. In this case, the class $\mathscr{E}$ is replaced by translation equivariant estimators of the form (2.8).

5. A lower bound on the risk inflation. In this section, we obtain lower bounds for the risk inflation of any estimator. Of course, it is immediate that for any δ, $\mathrm{RI}(\delta) \geq 1$. However, we can do much better than this by exploiting Theorem 3.1. For simplicity, we shall restrict attention to the case

$\sigma^2 = 1$ and hence estimators of the form (2.8). The simplification afforded by this restriction makes the main ideas more transparent. Note that in what follows we use the results of the previous sections implicitly assuming they have been modified for the translation equivariant case. We begin with a result which provides a lower bound for the best we might hope for.

THEOREM 5.1. *For* $\sigma^2 = 1$, $\inf_\delta \mathrm{RI}(\delta) \geq M^*$, *where, letting* Φ *be the standard normal cdf,*

$$M^* = \inf_c \left[\max\left[(1 + c^2), 2(n+1) \sup_\alpha \left(\alpha^2 \Phi(-(c+\alpha)) \right) \right] \right]. \tag{5.1}$$

PROOF. By Theorem 3.1, attention may be restricted to δ of the form (2.8). By Lemma 2.2, we may assume that $W_1 < \infty$. Otherwise $\mathrm{MR}(\mathscr{F}_1, \delta) = \infty$. Now note that

$$\mathrm{RI}(\delta) \geq \max[2\,\mathrm{MR}(\mathscr{F}_1, \delta), (n+1)\mathrm{MR}(\mathscr{F}_n, \delta)]. \tag{5.2}$$

By Lemma 2.2, any $\delta \in \mathscr{E}\mathscr{A}$ has $\mathrm{MR}(\mathscr{F}_1, \delta) \geq 1 + W_1^2$. Also, for any $\alpha > 0$,

$$\begin{aligned} \mathrm{MR}(\mathscr{F}_n, \delta) &= E_F(\delta - \mu)^2 \left(I_{[|Y_1 - \mu| > W_1 + \alpha]} + I_{[|Y_1 - \mu| \leq W_1 + \alpha]} \right) \\ &\geq E_F \alpha^2 \left(I_{[|Y_1 - \mu| > W_1 + \alpha]} \right) \geq 2\alpha^2 \Phi(-(W_1 + \alpha)) \end{aligned}$$

since $|Y_1 - \mu| > W_1 + \alpha$ implies $|\delta - \mu| > \alpha$. Inserting both of these bounds into (5.2) yields the desired result. □

Using standard methods to approximate the tail area of Φ in (5.1), the following explicit bound is obtained.

COROLLARY 5.1. *For* $\sigma^2 = 1$ *and large* n, $\mathrm{RI}(\delta) > (\log n)/2$.

Using the fact that $\mathrm{RI}(\overline{Y}_{T^{**}}) \leq 3.3 \log n$ from Section 1, we have the following result, which shows that the bound in Corollary 5.1 is tight (in order of magnitude).

COROLLARY 5.2. *For* $\sigma^2 = 1$ *and large* n, *there exists a* δ *such that* $\mathrm{RI}(\delta) = O(\log n)$.

It appears that one can do slightly better than $\overline{Y}_{T^{**}}$ in terms of risk inflation. To pursue the best lower bound, we obtained Monte Carlo estimates of the risk inflation of various estimators. The version of δ in (2.7a)–(2.7b) with $h_k \equiv W_k = 1/\sqrt{k+1} - 1/\sqrt{n}$ yielded $\mathrm{RI}(\delta) \approx \log n$ for $2 \leq n \leq 40$, just twice the lower bound of Theorem 5.1.

Finally, we remark that the main results of this section can be extended in a natural way to the general case where σ^2 is unknown. However, one must then consider the criterion $\max_{k \geq k_0}[(k+1)\mathrm{MR}(\mathscr{F}_k, \delta)]$ for large k_0.

6. Extensions to other distributions. In this section, we describe how our previous results can be easily extended to other distributional setups.

EXAMPLE 6.1 (Double exponential distribution). Consider the situation where we observe $(Y_1, \dots, Y_n)$ where (1.1) is replaced by

$$\mathscr{F}_k = \{F: Y_1, \dots, Y_k \text{ iid } f(y) = \tfrac{1}{2}\exp(-|y - \mu|)\}, \tag{6.1}$$

where μ is unknown, and we want to estimate μ $(\equiv EY_1)$. Replacing (1.3) by the risk function

$$R(F, \delta) = E_F\big[(\delta - \mu)^2\big], \tag{6.2}$$

the class of equivariant estimators for this problem are those that satisfy

$$\delta(a + Y) = a + \delta(Y). \tag{6.3}$$

Analogous to (2.2) and (2.8), any such translation equivariant δ can here be expressed as

$$\delta(Y) = \tilde{Y}_k + \omega_k(S_k, T_k), \qquad k = 1, \dots, n, \tag{6.4}$$

where $\tilde{Y}_k \equiv \text{median}\{Y_1, \dots, Y_k\}$, $Z_{ik} = (Y_i - \tilde{Y}_k)$, $S_k = (Z_{1k}, \dots, Z_{kk})$ and $T_k = (Z_{k+1,k}, \dots, Z_{nk})$. Note that, under $F \in \mathscr{F}_k$, $\tilde{Y}_k$ and S_k are independent. The decomposition in (2.2) is not appropriate here since $\tilde{Y}_k$ is sufficient here rather than $\bar{Y}_k$ and V_k. Replacing $\bar{Y}_k$ by $\tilde{Y}_k$ and setting $V_k \equiv 1$, straightforward analogies of the previous results are easily seen to hold. In particular, the restriction (2.7a)–(2.7b) with these substitutions yields an essentially complete class.

EXAMPLE 6.2 (Chi-square distribution). Consider the situation where we observe $(Y_1, \dots, Y_n)$ where (1.1) is replaced by

$$\mathscr{F}_k = \{F: Y_1, \dots, Y_k \text{ iid } \sigma^2\chi_1^2\}, \tag{6.5}$$

where σ^2 is unknown, and we want to estimate σ^2 $(\equiv EY_1)$. [An equivalent formulation would have $X_1, \dots, X_k$ iid $N(0, \sigma^2)$, which would come up in our previous formulation if interest focused on estimating σ^2 with μ being the nuisance parameter.] Such a problem might arise when estimating current volatility levels in financial time series, in which case Y_i might be a squared daily return on an asset [see French, Schwert and Stambaugh (1987)]. Replacing (1.3) by the risk function

$$R(F, \delta) = E_F(\delta - \sigma^2)^2/\sigma^4, \tag{6.6}$$

the class of equivariant estimators for this problem are those that satisfy

$$\delta(bY) = b\delta(Y). \tag{6.7}$$

Analogous to (2.2), any scale equivariant δ can here be expressed as

$$\delta(Y) = \bar{Y}_k\omega_k(S_k, T_k), \qquad k = 1, \dots, n, \tag{6.8}$$

where $Z_{ik} = Y_i/\overline{Y}_k$, $S_k = (Z_{1k}, \dots, Z_{kk})$ and $T_k = (Z_{k+1,k}, \dots, Z_{nk})$. Note that under $F \in \mathscr{F}_k$, $\overline{Y}_k$ and S_k are independent.

Although Theorem 4.1 is not applicable because of the asymmetry of the χ_1^2, analogies of the previous results can still be obtained. In particular, the restriction (2.7a)–(2.7b) substituting $C_k = [\overline{Y}_k W_k^-, \overline{Y}_k W_k^+]$ for (2.7a) and the endpoint functions $\overline{Y}_k h_k(B_{k-1}^+/\overline{Y}_k)$ and $\overline{Y}_k h_k(\overline{Y}_k/B_{k-1}^-)$ into (2.7b) yields an essentially complete class.

Acknowledgments. We would like to thank the anonymous referees for many helpful suggestions.

REFERENCES

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BONDAR, J. V. and MILNES, P. (1981). Amenability: A survey for statistical applications of Hunt–Stein and related conditions on groups. *Z. Wahrsch. Verw. Gebiete* **57** 103–128.

BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time (with discussion). *J. Roy. Statist. Soc. Ser. B* **37** 149–192.

CHERNOFF, H. and ZACKS, S. (1964). Estimating the current means of a normal distribution which is subject to changes in time. *Ann. Math. Statist.* **35** 999–1028.

FOSTER, D. P. and GEORGE, E. I. (1993). The risk inflation criterion for multiple regression. Unpublished manuscript.

FRENCH, K. R., SCHWERT, G. W. and STAMBAUGH, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics* **19** 3–29.

HINKLEY, D. V. (1970). Inference about a change-point in a sequence of random variables. *Biometrika* **57** 1–17.

LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. Wiley, New York.

MOSTELLER, F. (1948). On pooling data. *J. Amer. Statist. Assoc.* **43** 231–242.

ROYDEN, H. L. (1968). *Real Analysis*, 2nd ed. Macmillan, London.

SIEGMUND, D. (1986). Boundary crossing probabilities and statistical applications. *Ann. Statist.* **14** 361–404.

SMITH, A. F. M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika* **62** 407–416.

SMITH, A. F. M. (1985). Change-point problems: Approaches and applications. In *Bayesian Statistics* **2** (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith) 83–98. North-Holland, Amsterdam.

DEPARTMENT OF STATISTICS
THE WARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104-6302

CBA 5.202
DEPARTMENT OF MSIS
UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TEXAS 78712-1175