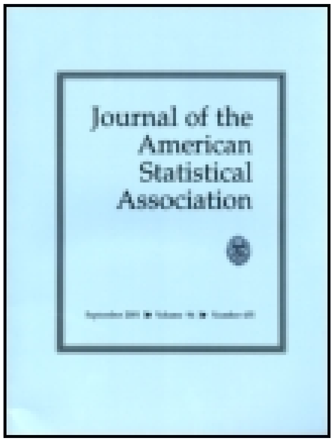

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Local Empirical Likelihood Inference for Varying-Coefficient Density-Ratio Models Based on Case-Control Data

Xu Liu, Hongmei Jiang & Yong Zhou

Accepted author version posted online: 22 Nov 2013.Published online: 13 Jun 2014.

Local Empirical Likelihood Inference for Varying-Coefficient Density-Ratio Models Based on Case-Control Data

Xu LIU, Hongmei JIANG, and Yong ZHOU

In this article, we develop a varying-coefficient density-ratio model for case-control studies. The case and control samples come from two different distributions. Under the model assumption, the ratio of the two densities is related to the linear combination of covariates with varying coefficients through a known function. A special case is the exponential tilt model where the log ratio of the two densities is a linear function of covariates. We propose a local empirical likelihood (EL) approach to estimate the nonparametric coefficient functions. Under some regularity assumptions, the proposed estimators are shown to be consistent and asymptotically normally distributed. The sieve empirical likelihood ratio (SELR) test statistic for detecting whether the varying-coefficients are really constant and other related hypotheses is constructed and it follows approximately a chi-squared distribution. We introduce a modified bootstrap procedure to estimate the null distribution of the SELR when sample size is small. We also examine the performance of proposed method for finite sample sizes through simulation studies and illustrate it with a real dataset. Supplementary materials for this article are available online.

KEY WORDS: Local linear; Logistic regression; SELR statistic; Semiparametric model; Two-sample model.

1. INTRODUCTION

The case-control study is an important method to identify factors associated with disease incidence (Lachin 2000). A cumulative incidence study was discussed by Miettinen (1976) and Prentice and Pyke (1979). Let $y = 1$ denote the development of the disease during the defined accession period, and $y = 0$ the disease-free state at the end of the accession period. Let $\mathbf{x}$ be the associated p-dimensional covariate vector. The standard logistic regression model is

$$P(y = 1|\mathbf{x}) = \frac{\exp\{\tilde{\alpha} + \boldsymbol{\beta}^T\mathbf{x}\}}{1 + \exp\{\tilde{\alpha} + \boldsymbol{\beta}^T\mathbf{x}\}}, \tag{1.1}$$

where $\tilde{\alpha}$ is a scalar parameter and $\boldsymbol{\beta}$ is a p-dimensional vector parameter. Let $g(\mathbf{x}) = P(\mathbf{x}|y = 0)$ and $f(\mathbf{x}) = P(\mathbf{x}|y = 1)$ be the corresponding conditional density distribution functions given the disease status. This logistic model has been shown to be equivalent to the following two-sample density-ratio model or exponential-tilt model (Qin 1998, 1999, 2005):

$$f(\mathbf{x}) = \exp\{\phi(\mathbf{x}, \alpha, \beta)\}g(\mathbf{x}), \tag{1.2}$$

where $\phi(\mathbf{x}, \alpha, \beta) = \alpha + \boldsymbol{\beta}^T\mathbf{x}$, $\alpha = \tilde{\alpha} + \log\{(1 - \pi)/\pi\}$, and $\pi = P(y = 1) = 1 - P(y = 0)$.

Xu Liu is Postdoctoral Fellow (E-mail: xu-liu@northwestern.edu) and Hongmei Jiang is Associate Professor (E-mail: hongmei@northwestern.edu), Department of Statistics, Northwestern University, Evanston, IL 60208. Yong Zhou is Professor, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190 and Department of Statistics, Shanghai University of Finance and Economics, Shanghai 200433 (E-mail: yzhou@amss.ac.cn). The authors thank the editor, the associate editor, the referees, and Dr. T. Severini for helpful suggestions and comments which have led to a substantially improved article; the authors also thank Dr. L. Hou for providing the real data. Liu and Jiang's research was supported in part by NSF DMS-1043080. Liu's research was partially supported by National Science Fund for Distinguished Young Scholars of China (11225103) and National Natural Science Foundation of China (NSFC) (11171293). Zhou's research was partially supported by National Natural Science Foundation of China (NSFC) (71271128), the State Key Program of National Natural Science Foundation of China (71331006), Science Fund for Creative Research Groups (11021161) and NCMIS.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/jasa.

There are many density functions satisfying the relationship in (1.2); for example, (1.2) holds if $g(\mathbf{x})$ and $f(\mathbf{x})$ are either normal densities or exponential densities. Various density-ratio models for some conventional density functions were discussed in Kay and Little (1987). Model (1.2) can also be regarded as a model for a two-sample biased sampling problem (Vardi 1982, 1985). It has been shown recently that the density-ratio model provides a good fit to the observed data in some medical applications (Qin and Zhang 1997; Qin et al. 2002; Zhang 2001), genetic quantitative trait loci analysis (Zou, Fine, and Yandell 2002), and clinical trials with skewed outcomes (White and Thompson 2003).

Here, we generalize the density-ratio model with constant coefficients to a varying-coefficient density-ratio model. In practice, model (1.2) with constant coefficients may not be able to describe the dynamic features contained in the dataset. For instance, the odds risk of disease incidence for some covariates may change with different values of time, age, or other exposure variable. In Section 4, we present a gastric cancer dataset where the association of telomere length and gastric cancer risk, as well as some potential covariates, was studied using about 700 subjects. The effects of telomere length, drinking, and *H. pylori* infection status on gastric cancer risk vary with age. If we simply include an interaction term, that is, the cross product between the covariate and the exposure variable, in the model, the underlying assumption is that the effect of the covariate changes linearly with the exposure variable, which may not reflect the truth. To make the model fit the data better and to capture the dynamic feature of changes, it is natural to consider the varying-coefficient model (Hastie and Tibshirani 1993). Let $\mathbf{x} = (w, \mathbf{z}^T)^T$, where w is a scalar, $\mathbf{z}$ is a d-dimensional vector, and $d = p - 1$. It is not difficult to generalize the scalar w to the setting of a vector. When $\phi(\mathbf{x}, \alpha, \beta) = \phi(w, \mathbf{z}, \alpha, \beta) = \alpha(w) + \boldsymbol{\beta}(w)^T\mathbf{z}$ in model (1.2), we get the varying-coefficient density-ratio model:

$$f(w, \mathbf{z}) = \exp\{\alpha(w) + \boldsymbol{\beta}(w)^T\mathbf{z})\}g(w, \mathbf{z}). \tag{1.3}$$

Journal of the American Statistical Association
June 2014, Vol. 109, No. 506, Theory and Methods
DOI: 10.1080/01621459.2013.858629

In this article, we consider the following general two-sample varying-coefficient density-ratio model:

$$f(w, \mathbf{z}) = \psi\{\alpha(w) + \boldsymbol{\beta}(w)^T \mathbf{z}\} g(w, \mathbf{z}), \tag{1.4}$$

where $\psi\{\cdot\}$ is a nonnegative known function that makes $f(w, \mathbf{z})$ to be a density function, which includes the exponential-tilt model (1.3) as a special case with $\psi(\cdot) = \exp(\cdot)$. In parametric situation, Thomas (1981) and Lustbader, Moolgavkar, Venzon (1984) considered a general relative risk model, a mixture model, $f(\mathbf{z})/g(\mathbf{z}) = (1 + \boldsymbol{\beta}^T \mathbf{z})^\delta [\exp\{\boldsymbol{\beta}^T \mathbf{z}\}]^{1-\delta}$, where δ is a scalar parameter that describes the general shape of the relative risk function. It includes additive relative risk model ($\delta = 1$) and log-linear relative model ($\delta = 0$) as special cases. The mixture model can be easily extended to varying-coefficients model (1.4) with $f(w, \mathbf{z})/g(w, \mathbf{z}) = [1 + \boldsymbol{\beta}^T(w)\mathbf{z}]^\delta [\exp\{\boldsymbol{\beta}^T(w)\mathbf{z}\}]^{1-\delta}$.

Let $g_w(w)$ and $g_{z|w}(\mathbf{z})$ be the marginal density of w and conditional density of $\mathbf{z}$ given w for the control sample, respectively, and let $f_w(w)$ and $f_{z|w}(\mathbf{z})$ be the corresponding densities for the case sample. Assume that both $g_w(w)$ and $f_w(w)$ are bounded away from zero. By (1.4), we have

$$f_{z|w}(\mathbf{z}) = \kappa(w)\psi\{\alpha(w) + \boldsymbol{\beta}(w)^T \mathbf{z}\} g_{z|w}(\mathbf{z}), \tag{1.5}$$

where $\kappa(w) = g_w(w)/f_w(w)$. If $\psi(\cdot) = \exp(\cdot)$, function $\kappa(w)$ can be absorbed into $\alpha(w)$. If the two marginal densities are equal, we have $\kappa(w) = 1$ for all w. These special cases happen frequently in practice, especially in the setting in which both $g(\mathbf{x})$ and $f(\mathbf{x})$ are normal densities; see, for example, the simulation study in Section 3.1. When $\kappa(w)$ is unknown, we can replace it by a consistent estimator, such as a kernel estimator. This varying-coefficient model is still linear in the covariates $\mathbf{z}$, and, hence, it avoids the curse of dimensionality when $\phi(\mathbf{x}, \alpha, \beta)$ is modeled nonparametrically. The coefficients are allowed to vary as smooth functions of w so that the nonlinear interactions between w and $\mathbf{z}$ could be examined. Our model (1.4) is in the same spirit as the varying-coefficient Cox's hazards model proposed by Cai et al. (2007), where the log ratio of the marginal hazard function and the baseline hazard function is a linear function of covariates with varying coefficients.

Here, we consider estimation and inference for the two-sample varying-coefficient density-ratio model (1.4) by constructing the local empirical likelihood (EL) function. In recent decades, the method of EL has taken much of the spotlight in the statistical field since it was first introduced by Owen (1988). Qin (1999) developed an EL method to estimate a mixture proportion using data from two different distributions as well as from a mixture of them. Other related work includes, but is not limited to, Owen (1990, 2001), Chen and Hall (1993), Qin (1993, 1999), and Qin and Lawless (1994). The EL approach is appealing for analyzing the varying-coefficient density-ratio model because the two density functions in (1.4) can be modeled nonparametrically. This nonparametric method of inference has sampling properties similar to the bootstrap. Another advantage of the EL approach is that it takes auxiliary information, such as the density-ratio in (1.4), into account to improve estimation.

When using model (1.4) it is natural to ask whether the coefficient functions are actually constant or whether certain varying-coefficients are statistically different from zero. We consider several related hypothesis test problems using the sieve empirical likelihood ratio (SELR) test statistics proposed by Fan and Zhang (2004). Fan, Zhang, and Zhang (2001) considered the generalized likelihood ratio statistics for several nonparametric models, including nonparametric regression models, varying-coefficient models, and generalized varying-coefficient models, and showed that they follow approximately a chi-squared distribution. Zhang and Gijbels (2003) introduced the SELR for constrained nonparametric regression models with unspecified error distributions. In this article, we show the proposed test statistic asymptotically follows a chi-squared distribution. A modified bootstrap procedure is also given to estimate the null distribution of the test statistic when sample size is small.

The article is organized as follows. In Section 2.1, we formulate the model and describe the local EL procedure and the nonparametric estimates for the coefficient functions. The corresponding asymptotic properties are also presented. A comparison with local maximum likelihood estimation is given for the special case (1.3) in Section 2.2. Section 2.3 discusses estimation of the variances of the estimators. Several hypothesis test problems based on the SELR test statistics are considered in Section 2.4. In Section 3, we present the results of simulation studies that illustrate that our proposed estimates and hypothesis testing procedures work well for small samples. In Section 4, we illustrate the proposed methods with an application for case-control study using a gastric cancer dataset. Some concluding remarks are given in Section 5. The proofs of the main results are collected in the Appendix.

2. LOCAL EL

Let $\mathbf{X}_1, \ldots, \mathbf{X}_{n_1}$ be a sequence of independent and identically distributed random vectors from the control group, each with density $g(\mathbf{x})$, and $\mathbf{X}_{n_1+1}, \ldots, \mathbf{X}_{n_1+n_2}$ be a sequence of independent and identically distributed random vectors from the case group, each with density $f(\mathbf{x})$, where $f(\mathbf{x})$ and $g(\mathbf{x})$ satisfy (1.4), n_1 and n_2 are the number of subjects in the control group and case group, respectively. Let $n = n_1 + n_2$, and $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\} = \{\mathbf{X}_1, \ldots, \mathbf{X}_{n_1}, \mathbf{X}_{n_1+1}, \ldots, \mathbf{X}_{n_1+n_2}\}$ denote the pooled sample. Assume that $n_1/n \to \rho > 0$ as $n \to \infty$. From model (1.4), the EL function derived according to Prentice and Pyke (1979) is

$$\begin{aligned} \ell(\alpha, \boldsymbol{\beta}, G) &= \Pi_{j=1}^{n_1} dG(\mathbf{X}_j) \Pi_{i=n_1+1}^{n} \psi(\mathbf{X}_i) dG(\mathbf{X}_i) \\ &= \Pi_{j=1}^{n} \tilde{p}_j \Pi_{i=n_1+1}^{n} \psi(\mathbf{X}_i), \end{aligned} \tag{2.6}$$

where $\psi(\mathbf{x}) = \psi(w, \mathbf{z}) = \psi(\phi(\mathbf{x})) = \psi\{\alpha(w) + \boldsymbol{\beta}(w)^T \mathbf{z}\}$, $\tilde{p}_i = dG(\mathbf{X}_i)$ and $G(\mathbf{x})$ is the distribution function corresponding to $g(\mathbf{x})$. However, $\ell(\alpha, \boldsymbol{\beta}, G)$ can not be used directly to obtain estimates for $\alpha(\cdot)$ and $\beta(\cdot)$ because $\alpha(\cdot)$ and $\beta(\cdot)$ are infinite-dimensional parameters. Thus, instead of (2.6), we consider the localized conditional EL below.

Assume that all components of $\alpha(\cdot)$ and $\boldsymbol{\beta}(\cdot)$ are smooth so that they admit Taylor's series expansions, that is, for each given w_0 and for w around w_0,

$$\begin{aligned} \boldsymbol{\beta}(w) &\approx \boldsymbol{\beta}(w_0) + \boldsymbol{\beta}'(w_0)(w - w_0), \\ \alpha(w) &\approx \alpha(w_0) + \alpha'(w_0)(w - w_0). \end{aligned} \tag{2.7}$$

Let $\boldsymbol{\xi}(w) = (\alpha(w), \boldsymbol{\beta}(w)^T, \alpha'(w), \boldsymbol{\beta}'(w)^T)^T$, and $\mathbf{X}_i^*(w) = (1, \mathbf{Z}_i^T, W_i - w, \mathbf{Z}_i^T(W_i - w))^T$. For simplicity, denote $\boldsymbol{\xi}(w_0)$ by $\boldsymbol{\xi}$ and $\mathbf{X}_i^*(w_0)$ by $\mathbf{X}_i^*$ for fixed w_0. Then, from model (1.5), the local log empirical likelihood (LEL) function of $\boldsymbol{\xi}$ is

$$\ell(\boldsymbol{\xi}) = \sup\Bigg\{\sum_{i=1}^{n} w_h(W_i, w_0)\log p_i + \sum_{i=n_1+1}^{n} w_h(W_i, w_0)\log(\kappa(W_i)\psi(\boldsymbol{\xi}^T\mathbf{X}_i^*)) \colon p_i \ge 0, 1 \le i \le n, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i(\kappa(W_i)\psi\left(\boldsymbol{\xi}^T\mathbf{X}_i^*\right) - 1) = 0\Bigg\},$$

where $w_h(t_i, t_0) = K_h(t_i - t_0)/\sum_{j=1}^{n} K_h(t_j - t_0)$ is the weight with kernel function $K_h(\cdot) = K(\cdot/h)/h$ and h represents the size of the local neighborhood. The kernel weight is used to give smoother weight to data with w near w_0. The last constraint is the auxiliary information for the EL estimation. It comes from that $F_{z|w}(\mathbf{z}) = \int_{-\infty}^{\mathbf{z}} \kappa(w)\psi(\mathbf{u})dG_{z|w}(\mathbf{u})$ is a conditional cumulative distribution function of $\mathbf{z}$ given w, which results in $E\{\kappa(W)\psi(\mathbf{X})|W = w\} = 1$. By the method of Lagrange multipliers, similar to that used in Owen (1988), we obtain

$$p_i = \frac{w_h(W_i, w_0)}{1 + \lambda(\boldsymbol{\xi})(\kappa(W_i)\psi(\boldsymbol{\xi}^T\mathbf{X}_i^*) - 1)},$$

where $\lambda(\boldsymbol{\xi})$ is determined by the constraint equation

$$0 = \sum_{j=1}^{n} \frac{\kappa(W_j)\psi\left(\boldsymbol{\xi}^T\mathbf{X}_j^*\right) - 1}{1 + \lambda(\boldsymbol{\xi})\left(\kappa(W_j)\psi\left(\boldsymbol{\xi}^T\mathbf{X}_j^*\right) - 1\right)} w_h(W_j, w_0). \quad (2.8)$$

The LEL function, ignoring the constant, can be written as

$$\ell(\boldsymbol{\xi}) = \sum_{i=n_1+1}^{n} \log\{\kappa(W_i)\psi(\boldsymbol{\xi}^T\mathbf{X}_i^*)\} w_h(W_i, w_0) - \sum_{j=1}^{n} \log\{1 + \lambda(\boldsymbol{\xi})(\kappa(W_j)\psi(\boldsymbol{\xi}^T\mathbf{X}_j^*) - 1)\} w_h(W_j, w_0). \quad (2.9)$$

It can be rewritten using the same approach as in Qin (1999). Let $\varrho_n = n_1/n$, $\bar{\gamma}(\mathbf{X}) = \rho + (1-\rho)\psi(\mathbf{X})$, $\bar{m}(\mathbf{X}) = (\kappa(W)\psi(\mathbf{X}) - 1)\bar{\gamma}(\mathbf{X})^{-1}$, and correspondingly

$$\gamma(\boldsymbol{\xi}^T\mathbf{X}_i^*) = ((1-\varrho_n)\kappa(W_i)\psi(\boldsymbol{\xi}^T\mathbf{X}_i^*) + \varrho_n\kappa(W))/(1 - \varrho_n + \varrho_n\kappa(W)),$$
$$m(\boldsymbol{\xi}^T\mathbf{X}_i^*) = (\kappa(W_i)\psi(\boldsymbol{\xi}^T\mathbf{X}_i^*) - 1)\gamma(\boldsymbol{\xi}^T\mathbf{X}_i^*)^{-1}.$$

Then

$$\ell(\boldsymbol{\xi}) = \ell_1(\boldsymbol{\xi}) + \ell_2(\boldsymbol{\xi}), \quad (2.10)$$

where

$$\ell_1(\boldsymbol{\xi}) = -\sum_{j=1}^{n} \log\left\{1 + t(\boldsymbol{\xi})m(\boldsymbol{\xi}^T\mathbf{X}_j^*)\right\} w_h(W_j, w_0),$$

and

$$\ell_2(\boldsymbol{\xi}) = -\sum_{j=1}^{n} \log\{\gamma(\boldsymbol{\xi}^T\mathbf{X}_j^*)\} w_h(W_j, w_0) + \sum_{i=n_1+1}^{n} \log\{\kappa(W_i)\psi(\boldsymbol{\xi}^T\mathbf{X}_i^*)\} w_h(W_i, w_0),$$

in which $t(\boldsymbol{\xi}) = \lambda(\boldsymbol{\xi}) - (1-\varrho_n)/(1 - \varrho_n + \varrho_n\kappa(W))$ is determined by

$$0 = \sum_{i=1}^{n} \frac{m(\boldsymbol{\xi}^T\mathbf{X}_i^*)}{1 + t(\boldsymbol{\xi})m(\boldsymbol{\xi}^T\mathbf{X}_i^*)} w_h(W_j, w_0). \quad (2.11)$$

The advantage of changing variables is a consequence of the fact that $E[m(\boldsymbol{\xi}^T\mathbf{X}^*)w_h(W, w_0)] = O(h^2)$.

By Lemma A.3 in the Appendix $E\partial\ell_2/\partial\boldsymbol{\xi} = O(h^2)$ so that $E\partial\ell/\partial\boldsymbol{\xi} = O(h^2)$ at the value $t(\boldsymbol{\xi}) = 0$; it follows that the expectation of the score function is asymptotically zero. Under certain conditions, the constraint equation determines uniquely an implicit function $t(\boldsymbol{\xi})$ in a neighborhood of $\boldsymbol{\xi}_0$, where $\boldsymbol{\xi}_0$ is the true value of $\boldsymbol{\xi}$. Furthermore, we can prove that $t(\boldsymbol{\xi}) = O_p((nh)^{-1/3})$ by Lemma A.2 and its remarks. In fact, we can prove that $t(\boldsymbol{\xi})$ is of order $O_p((nh)^{-1/2})$.

Let $\hat{\boldsymbol{\xi}}$ be the maximizer of $\ell(\boldsymbol{\xi})$ with known $\kappa(w)$ in (2.10), subject to constraint condition (2.11) at w_0. We replace $\kappa(w)$ in (2.10) by the ratio of two kernel density estimates when it is unknown, and the corresponding estimator will be given in the next section. Let $e_p = (I_p, 0)$ be a $p \times 2p$ matrix, where I_p is p-order identity matrix and 0 is p-order zero matrix. Let $\boldsymbol{\theta}_0(w) = (\alpha_0(w), \boldsymbol{\beta}_0(w)^T)^T$ and $\hat{\boldsymbol{\theta}}(w) = (\hat{\alpha}(w), \hat{\boldsymbol{\beta}}(w)^T)^T$ be the true and estimated function coefficients. Then, $\hat{\boldsymbol{\theta}}(w_0) = e_p\hat{\boldsymbol{\xi}}$ is the LEL estimator desired; this can be done for any w_0.

2.1 Asymptotic Normality

To show conveniently the asymptotic normality of the estimator, we introduce some additional notation. Let $\partial\psi(\phi(\mathbf{x}))/\partial\phi$ denote the derivative of $\psi(\phi)$ with respect to ϕ. Set $c_n = (nh)^{-1/2}(\log h^{-1})^{1/2}$, $\mu_k = \int t^k K(t)dt$, $\nu_k = \int t^k K^2(t)dt$, $\tilde{\mathbf{z}} = (1, \mathbf{z}^T)^T$, $q_0(w) = \rho g_w(w) + (1-\rho) f_w(w)$, $q_0^*(w) = \rho\kappa(w) + 1 - \rho$, $q_1(\mathbf{x}) = \frac{\kappa(w)(\partial\psi(\phi(\mathbf{X}))/\partial\phi)^2}{\psi(\phi(\mathbf{X}))\bar{\gamma}(\mathbf{X})q_0^*(w)}$, $q_{12}(\mathbf{x}) = q_0^*(w)(\partial\psi(\phi(\mathbf{x}))/\partial\phi)/\bar{\gamma}(\mathbf{x})$, and $q_2(\mathbf{x}) = q_0^*(w)(\kappa(w)\psi(\phi(\mathbf{x})) - 1)^2/(\kappa(w)\bar{\gamma}(\mathbf{x}))$. Let $\mathbf{a} = E_G[q_1(\mathbf{X})\tilde{\mathbf{Z}}^{\otimes 2}|W = w_0]$, $b_0 = E_G[q_2(\mathbf{X})|W = w_0]$, $\tilde{\mathbf{a}}_{12} = E_G[q_{12}(\mathbf{X})\tilde{\mathbf{Z}}|W = w_0]$, and $\mathbf{C} = \rho(1-\rho)\mathbf{a} + b_0^{-1}\tilde{\mathbf{a}}_{12}\tilde{\mathbf{a}}_{12}^T$, where $k = 1, 2$, $\mathbf{z}^{\otimes 1} = \mathbf{z}$, and $\mathbf{z}^{\otimes 2}$ denotes $\mathbf{z}\mathbf{z}^T$. Let $B^T = (\alpha_0'', (\boldsymbol{\beta}_0'')^T)$, where α_0'' and $\boldsymbol{\beta}_0''$ denote the second-order derivatives of $\alpha(w)$ and $\boldsymbol{\beta}(w)$ at point w_0, respectively.

Theorem 1. Suppose that assumptions (1)–(7) in the Appendix hold. Let $\mathbf{H}$ be a $2p \times 2p$ diagonal matrix with the first p diagonal elements being 1 and the remaining p diagonal elements being h. Then,

$$\mathbf{H}(\hat{\boldsymbol{\xi}}(w_0) - \boldsymbol{\xi}_0(w_0)) \xrightarrow{P} 0 \quad \text{as} \quad n \to \infty,$$

for any w_0 in the support of w.

Theorem 2. Suppose that assumptions (1)–(7) in the Appendix hold. Then

$$\sqrt{nh}\left\{\left(\mathbf{H}(\hat{\boldsymbol{\xi}}(w_0) - \boldsymbol{\xi}_0(w_0))^T, t(\hat{\boldsymbol{\xi}}(w_0))\right)^T - \frac{1}{2}h^2\mu_2(B^T, \mathbf{0}^T)^T\right\} \xrightarrow{\mathcal{L}} N(0, \Sigma(w_0)/q_0(w_0)) \quad \text{as} \quad n \to \infty,$$

where $\Sigma(w_0)$ is a diagonal matrix with diagonal elements $\nu_0\mathbf{C}^{-1}$, $\frac{1}{\rho(1-\rho)}\nu_2\mu_2^{-2}\mathbf{a}^{-1}$, and $\nu_0\rho(1-\rho)(\rho(1-\rho)b_0 + \tilde{\mathbf{a}}_{12}^T\mathbf{a}^{-1}\tilde{\mathbf{a}}_{12})^{-1}$.

The following corollary is an immediate consequence of the above joint asymptotic normality result for the LEL estimator $\hat{\boldsymbol{\xi}}$.

Corollary 1. Under the conditions of Theorem 2, we have

$$\sqrt{nh}\left\{\hat{\boldsymbol{\theta}}(w_0)-\boldsymbol{\theta}(w_0)-\frac{1}{2}h^2\mu_2 B\right\} \xrightarrow{\mathcal{L}} N(0,\nu_0\mathbf{C}^{-1}/q_0(w_0)) \quad \text{as} \quad n\to\infty,$$

where $\mathbf{C}$ and B are given above.

When $\kappa(w)$ is unknown, we can replace it with a consistent estimator. Because both $g_w(w)$ and $f_w(w)$ are densities, their kernel estimators are $\hat{g}_w(w)=\frac{1}{n_1}\sum_{i=1}^{n_1}K_{h_1}(W_i-w)$ and $\hat{f}_w(w)=\frac{1}{n_2}\sum_{i=n_1+1}^{n}K_{h_1}(W_i-w)$, respectively, where h_1 is a bandwidth that may differ from h. Here, the kernel may also differ from that in (2.9), but for the simplification of technical proofs, we use the same kernel. This gives a consistent estimator for $\kappa(w)$:

$$\hat{\kappa}(w)=\hat{g}_w(w)/\hat{f}_w(w). \tag{2.12}$$

Let $\hat{\boldsymbol{\xi}}^U$ be the maximizer of LEL $\ell(\boldsymbol{\xi})$ in (2.10), subject to constraint condition (2.11) at w_0, with unknown $\kappa(w)$ replaced by $\hat{\kappa}(w)$.

Theorem 3. Suppose that assumptions (1)–(8) in the Appendix hold, then

$$\mathbf{H}(\hat{\boldsymbol{\xi}}^U(w_0)-\boldsymbol{\xi}_0(w_0)) \xrightarrow{P} 0 \quad \text{as} \quad n\to\infty$$

for any w_0 in the support of w, and

$$\sqrt{nh}\left\{\left(\mathbf{H}(\hat{\boldsymbol{\xi}}^U(w_0)-\boldsymbol{\xi}_0(w_0))^T, t(\hat{\boldsymbol{\xi}}^U(w_0))\right)^T-\frac{1}{2}h^2\mu_2(B^T,\mathbf{0}^T)^T\right\} \xrightarrow{\mathcal{L}} N(0,\Sigma(w_0)/q_0(w_0)+\tilde{\Sigma}(w_0)) \quad \text{as} \quad n\longrightarrow\infty,$$

where $\tilde{\Sigma}(w_0)=S^{-1}\tilde{\Gamma}S^{-1}$ and S and $\tilde{\Gamma}$ are defined in Lemma A.1 and Lemma A.4 in the Appendix, respectively.

The LEL estimator is sensitive to the choice of bandwidth h. Bandwidth selection has been intensively studied in the nonparametric estimation field, see Sepanski, Knickerbocker, and Carroll (1994), and Ruppert, Sheathers, and Wand (1995) for good discussions. To avoid the estimation of high order derivatives, we employ a bandwidth selector based on the MSE criterion, called empirical bias bandwidth selection (EBBS) (Ruppert 1997; Carroll, Ruppert, and Welsh 1998). The details of EBBS are provided in the supplementary materials.

2.2 Comparison With the MLE

For the special case (1.3) with $\psi\{\phi(\mathbf{x})\}=\exp\{\phi(\mathbf{x})\}$, $\kappa(w)$ can be absorbed into $\alpha(w)$ in model (1.5), which results in $\phi(\mathbf{x})=\alpha^*(w)+\boldsymbol{\beta}(w)^T\mathbf{z}$, where $\alpha^*(w)=\alpha(w)+\log(\kappa(w))$. Thus, we can avoid estimating $\kappa(w)$ using the LEL procedure, but $\alpha(w)$ is confounded with $\kappa(w)$. In fact, $\hat{\alpha}^*(w_0)-\log(\hat{\kappa}(w_0))$ is a consistent estimator for $\alpha(w_0)$. In this case, the covariance of $(\hat{\alpha}^*(w_0),\hat{\boldsymbol{\beta}}(w_0)^T)^T$ is $\nu_0C^{-1}/q_0(w_0)=\frac{\nu_0}{\rho(1-\rho)q_0(w_0)}((\mathbf{a}^*)^{-1}-M_1)$, where $\mathbf{a}^*=E_G\{\exp(\alpha^*(w_0)+\mathbf{Z}^T\boldsymbol{\beta}(w_0))/\bar{\gamma}(\mathbf{X})\tilde{\mathbf{Z}}^{\otimes 2}|W\}$ and $M_1=\mathrm{diag}(1,\mathbf{0}^T)$ is a p by p diagonal matrix. The local MLE is not feasible for the general model (1.4). However, for the density ratio model (1.3), it can be used to estimate the parameters based on logistic model by reparameterization (Prentice and Pyke 1979),

$$P(\mathbf{x}|y=1)=\frac{\exp\{\alpha^{**}(w)+\beta(w)^T\mathbf{z}\}}{1+\exp\{\alpha^{**}(w)+\beta(w)^T\mathbf{z}\}}\frac{nq(z)}{n_2}, \tag{2.13}$$

where $\alpha^{**}(w)=\alpha^*(w)+\log(n_2/n_1)$, and $nq(z)/n_2$ is considered as a nuisance parameter. The local log-likelihood of logistic model (2.13) for case-control data is

$$\ell_n(\tilde{\boldsymbol{\xi}})=\sum_{i=1}^{n}\ell(\tilde{\boldsymbol{\xi}},W_i,\mathbf{Z}_i)w_h(W_i,w_0), \tag{2.14}$$

where $\ell(\tilde{\boldsymbol{\xi}},W_i,\mathbf{Z}_i)=y\tilde{\boldsymbol{\xi}}^T\mathbf{X}_i^*-\log\{1+\exp(\tilde{\boldsymbol{\xi}}^T\mathbf{X}_i^*)\}$, and $\tilde{\boldsymbol{\xi}}$ is $\boldsymbol{\xi}$ with the first column of $\boldsymbol{\xi}$ being replaced by $\alpha^*(w_0)+\log(n_2/n_1)$. Similarly, the MLE of above log-likelihood (2.14) is $(\tilde{\alpha}(w_0),\tilde{\boldsymbol{\beta}}(w_0))=(\hat{\alpha}^*(w_0)+\log(n_2/n_1),\hat{\boldsymbol{\beta}}(w_0))$ with variance $\frac{\nu_0}{\rho(1-\rho)q_0(w_0)}(\mathbf{a}^*)^{-1}$. It gives the same estimator and covariance as LEL method except for the intercept. The variance of intercept term based on our proposed LEL is smaller than that based on local MLE because the LEL estimator takes into account $E\{\kappa(W)\psi(\mathbf{X})|W=w\}=1$ as an auxiliary information.

2.3 Estimation of Variances

To estimate the covariance matrix $\Sigma(w_0)$, as defined in Theorem 2, we need to estimate S and Γ. The matrix $S=(S_{ij})$, $i,j=1,2$, is defined in Lemma A.1, where $S_{11}=\mathrm{diag}(\rho(1-\rho)\mathbf{a},\rho(1-\rho)\mu_2\mathbf{a})$ is a diagonal matrix, $S_{12}=S_{21}^T=(\tilde{a}_0,\tilde{\mathbf{a}}_1^T,0,\mathbf{0}^T)^T$, and $S_{22}=-b_0$. For $\Gamma=(\Gamma_{ij})$, $i,j=1,2$, as defined in Lemma A.3, $\Gamma_{11}=\mathrm{diag}(\rho(1-\rho)\nu_0\mathbf{a},\rho(1-\rho)\nu_2\mathbf{a})$ is a diagonal matrix, $\Gamma_{12}=\Gamma_{21}^T=\mathbf{0}$, and $\Gamma_{22}=\nu_0 b_0$.

As mentioned previously, $\kappa(w_0)$ can be consistently estimated by (2.12). For convenience, define $\hat{q}_0^*(w_0)=\varrho_n\hat{\kappa}(w_0)+1-\varrho_n$, $\hat{d}_0(\mathbf{x})=\partial\hat{\psi}(\phi(\mathbf{x}))\partial\phi$, $\hat{d}_1(\mathbf{x})=\varrho_n(1-\varrho_n)/(\hat{\psi}(\phi(\mathbf{x}))\hat{\gamma}^2(\mathbf{x}))$, $\hat{d}_{12}(\mathbf{x})=\hat{q}_0^*(w_0)^2/(\hat{\kappa}(w_0)\hat{\gamma}^2(\mathbf{x}))$, and $\hat{d}_2(\mathbf{x})=\hat{q}_0^*(w_0)^2(\hat{\kappa}(w)\hat{\psi}(\phi(\mathbf{x}))-1)^2/(\hat{\kappa}^2(w)\hat{\gamma}^2(\mathbf{x}))$, where $\hat{\psi}(\phi(\mathbf{X}))=\psi(\phi(\hat{\boldsymbol{\xi}}^T\mathbf{X}^*))$, and $\hat{\gamma}(\mathbf{X})=\varrho_n+(1-\varrho_n)\hat{\psi}(\mathbf{X})$. We can estimate S_{11}, S_{12}, and S_{22} by $\hat{S}_{11}=\sum_{j=1}^n\hat{d}_1(\mathbf{X}_j)\hat{d}_0^2(\mathbf{X}_j)(\mathbf{H}^{-1}\mathbf{X}_j^*)^{\otimes 2}w_h(W_j,w_0)$, $\hat{S}_{12}=\sum_{j=1}^n\hat{d}_{12}(\mathbf{X}_j)\hat{d}_0(\mathbf{X}_j)\mathbf{H}^{-1}\mathbf{X}_j^*w_h(W_j,w_0)$, $\hat{S}_{22}=-\sum_{j=1}^n\hat{d}_2(\mathbf{X}_j)w_h(W_j,w_0)$, respectively, and $\hat{S}_{21}=\hat{S}_{12}^T$. Thus, S can be estimated consistently by $\hat{S}_n=(\hat{S}_{ij})$, $i,j=1,2$.

A consistent estimator of $q_0(w_0)$ is $\hat{q}_0(w_0)=\varrho_n\hat{g}_w(w_0)+(1-\varrho_n)\hat{f}_w(w_0)$. For known $\kappa(w)$, the consistent estimates for Γ_{11}, Γ_{12}, and Γ_{22} are $\hat{\Gamma}_n=(\hat{\Gamma}_{ij})/\hat{q}_0(w_0)$, $i,j=1,2$, where $\hat{\Gamma}_{11}=\hat{S}_{11}D$, $\hat{\Gamma}_{22}=-\nu_0\hat{S}_{22}$, $\hat{\Gamma}_{12}=\mathbf{0}$, and $D=\mathrm{diag}(\nu_0 I_p,\nu_2/\mu_2 I_p)$. Therefore, the covariance of the local estimator $(\mathbf{H}(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}),t(\hat{\boldsymbol{\xi}}))^T$ can be estimated consistently by $(nh)^{-1}\hat{S}_n^{-1}\hat{\Gamma}_n\hat{S}_n^{-1}$. By Corollary 1, we can estimate the covariance of $(\hat{\alpha}(w_0),\hat{\boldsymbol{\beta}}^T(w_0))^T$ by

$$(nh)^{-1}e_p\hat{S}_n^{-1}\hat{\Gamma}_n\hat{S}_n^{-1}.$$

For unknown $\kappa(w)$, we can estimate $\tilde{\Gamma}=(\tilde{\Gamma}_{ij})$, $i,j=1,2$, by $\breve{\Gamma}=(\breve{\Gamma}_{ij})$, where $\breve{\Gamma}_{ij}$ is $\tilde{\Gamma}_{ij}$ with $g_w(w_0)$, $\kappa(w_0)$, $q_0^*(w_0)$, $q_0(w_0)$, $\tilde{b}_0(w_0)$, and S_{ij} replaced by their estimators $\hat{g}_w(w_0)$, $\hat{\kappa}(w_0)$, $\hat{q}_0^*(w_0)$, $\hat{q}_0(w_0)=\hat{q}_0^*(w_0)\hat{f}_w(w_0)$, $\breve{b}_0(w_0)=\sum_{i=1}^n\hat{q}_0^*(w_0)\hat{\psi}(\phi(\mathbf{X}_i))w_h(W_i,w_0)/(\hat{\kappa}(w_0)\hat{\gamma}(\mathbf{X}_i))$, and $\hat{S}_{ij}$, respectively. Thus, $(nh)^{-1}\hat{S}_n^{-1}(\hat{\Gamma}_n+\check{\Gamma})\hat{S}_n^{-1}$ is a consistent estimator of covariance of $(\mathbf{H}(\hat{\boldsymbol{\xi}}^U-\boldsymbol{\xi}), t(\hat{\boldsymbol{\xi}}^U))^T$.

2.4 Hypothesis Testing

We first consider a hypothesis test to detect if $\boldsymbol{\theta}(w)$ equals a given function $\boldsymbol{\theta}_0(w)$, that is,

$$H_0:\boldsymbol{\theta}(\cdot)=\boldsymbol{\theta}_0(\cdot) \quad \text{vs.} \quad H_1:\boldsymbol{\theta}(\cdot)\neq\boldsymbol{\theta}_0(\cdot), \tag{2.15}$$

by using the SELR test (Fan and Zhang 2004). A special case is that $\boldsymbol{\theta}_0(\cdot)$ is a given constant number such as $\mathbf{0}$, that is to test if $\boldsymbol{\theta}(\cdot)$ is statistically different from $\mathbf{0}$. We define the log-likelihood function $\ell(\Theta)=\sum_{i=1}^n\{\ell_1(W_i,\hat{\boldsymbol{\xi}}(W_i))+\ell_2(W_i,\hat{\boldsymbol{\xi}}(W_i))\}$, where Θ denotes the space of $\boldsymbol{\theta}(w)$,

$$\ell_1(W_i,\hat{\boldsymbol{\xi}}(W_i))=-\sum_{j=1}^n\log\left\{1+t(\hat{\boldsymbol{\xi}}(W_i))m(\hat{\boldsymbol{\xi}}(W_i)^T\mathbf{X}_j^*(W_i))\right\} w_h(W_j,W_i),$$

and

$$\ell_2(W_i,\hat{\boldsymbol{\xi}}(W_i))=-\sum_{j=1}^n\log\left\{\gamma(\hat{\boldsymbol{\xi}}(W_i)^T\mathbf{X}_j^*(W_i))\right\}w_h(W_j,W_i)+\sum_{k=n_1+1}^n\hat{\boldsymbol{\xi}}(W_i)^T\mathbf{X}_k^*(W_i)w_h(W_k,W_i).$$

Define $\boldsymbol{\xi}_0(w)=(\boldsymbol{\theta}_0(w),\boldsymbol{\theta}_0'(w))$, where $\boldsymbol{\theta}_0'(w)$ is the derivative of $\boldsymbol{\theta}_0(w)$. Alternatively, the log-likelihood under the null hypothesis is $\ell(\boldsymbol{\xi}_0)=\sum_{i=1}^n\{\ell_1(W_i,\boldsymbol{\xi}_0(W_i))+\ell_2(W_i,\boldsymbol{\xi}_0(W_i))\}$. Therefore, the SELR statistic for testing problem (2.15) is constructed as

$$\mathscr{L}(\boldsymbol{\theta}_0)=\ell(\Theta)-\ell(\boldsymbol{\xi}_0). \tag{2.16}$$

We prove that $\mathscr{L}(\boldsymbol{\theta}_0)$ with known $\kappa(w)$ has an asymptotic normal distribution in the following theorem under some regularity conditions. For unknown $\kappa(w)$, similar results can be obtained. We give those conditions and the technical proofs in the supplementary materials. Let $K^*(s)=\int K(t)K(s+t)(1+t(s+t)\mu_2^{-1})dt$ and $\tilde{K}=\int\tilde{K}_1(s)\tilde{K}_2(s)ds$, where $\tilde{K}_1(s)=\int K(t)K(s+t)dt$ and $\tilde{K}_2(s)=\int t(s+t)K(t)K(s+t)dt$. Let $\varsigma_1=\int\kappa(w)/q_0^*(w)dw$, $\varsigma_2=\int\kappa(u)^2/q_0^*(u)^2du$, and $\varsigma_3=\rho(1-\rho)\int\text{tr}(C^{-1}(u)\mathbf{a}(u))du-p\varsigma_2$.

Theorem 4. Assume that some regularity conditions presented in the supplementary materials hold. For known $\kappa(w)$, under H_0 in (2.15), if $\boldsymbol{\theta}_0(w)$ is a linear function of w or $nh^{9/2}\to 0$,

$$\sigma_n^{-1}(2\mathscr{L}(\boldsymbol{\theta}_0)-\mu_n)\xrightarrow{\mathcal{L}}N(0,1),$$

where $\mu_n=\frac{p\varsigma_1}{h}K^*(0)$, $\sigma_n^2=\frac{2a_K}{h}$, and $a_K=p\varsigma_2\int K^*(s)^2ds-2\mu_2^{-1}\varsigma_3\tilde{K}$. Furthermore, the scaled $\mathscr{L}(\boldsymbol{\theta}_0)$ follows approximately a χ^2-distribution with degrees of freedom h^{-1}, that is,

$$2a_K^{-1/2}\mathscr{L}(\boldsymbol{\theta}_0)-b_n\overset{a}{\sim}\chi^2_{h^{-1}},$$

where $b_n=\frac{p\varsigma_1K^*(0)-\sqrt{a_k}}{h\sqrt{a_k}}$.

One may be interested in testing the following composite null hypotheses:

$$H_0:\boldsymbol{\theta}(\cdot)\in\mathcal{A}_0 \quad \text{vs.} \quad H_1:\boldsymbol{\theta}(\cdot)\notin\mathcal{A}_0, \tag{2.17}$$

where $\mathcal{A}_0$ is a subset of functions. For instance, $\mathcal{A}_0$ is the unknown constant function or linear function space. Let $\ell(\mathcal{A}_0)$ be the EL function under the null hypothesis in (2.17). Then, the SELR statistic for the testing problem (2.17) is

$$\mathscr{L}_n=\ell(\Theta)-\ell(\mathcal{A}_0). \tag{2.18}$$

To simplify the test statistic (2.18), we consider the following fabricated testing problem with the simple null hypotheses:

$$H_0':\boldsymbol{\theta}(\cdot)=\boldsymbol{\theta}_0 \quad \text{vs.} \quad H_1':\boldsymbol{\theta}(\cdot)\in\mathcal{A}_0. \tag{2.19}$$

Therefore, the SELR statistic $\mathscr{L}_n$ can be rewritten as

$$\mathscr{L}_n=\mathscr{L}(\boldsymbol{\theta}_0)-\mathscr{L}^*(\boldsymbol{\theta}_0),$$

where $\mathscr{L}(\boldsymbol{\theta}_0)$ is defined in (2.16) and $\mathscr{L}^*(\boldsymbol{\theta}_0)$ is the SELR test statistic for (2.19). Asymptotic properties of $\mathscr{L}^*(\boldsymbol{\theta}_0)$ similar to those in Theorem 4 can be obtained. As a consequence we obtain the asymptotic properties of $\mathscr{L}_n$ for testing (2.17).

Although the asymptotic results for $\mathscr{L}(\boldsymbol{\theta}_0)$ and $\mathscr{L}_n$ are available, they may not perform well when sample size is small. We propose a modified bootstrap method for practical application. Because the intercept $\tilde{\alpha}$ is not estimable in general for the case-control data, the bootstrap samples can be drawn from estimated distributions $\hat{G}(\mathbf{x})$ and $\hat{F}(\mathbf{x})$, where

$$\hat{G}(\mathbf{x})=\sum_{i=1}^n\hat{p}_iI(\mathbf{X}_i<=\mathbf{x}),$$
$$\hat{F}(\mathbf{x})=\sum_{i=1}^n\psi(\hat{\phi}(\mathbf{X}_i))\hat{p}_iI(\mathbf{X}_i<=\mathbf{x}), \tag{2.20}$$

where $\hat{p}_i=n^{-1}[1+\hat{\lambda}(\hat{\boldsymbol{\xi}}_0(W_i))(\psi(\hat{\boldsymbol{\xi}}_0(W_i)^T\mathbf{X}_i^*)-1)]^{-1}$, and $\hat{\boldsymbol{\xi}}_0(\cdot)$ is the EL estimate for the parametric model under the null hypothesis. For example, if $\boldsymbol{\theta}_0(\cdot)$ in (2.15) is constant, then $\hat{p}_i=n^{-1}[1+\hat{\lambda}(\hat{\boldsymbol{\theta}})(\psi(\hat{\phi}(\mathbf{X}_i))-1)]^{-1}$, $\hat{\phi}(\mathbf{X}_i)=\hat{\boldsymbol{\theta}}^T\mathbf{X}_i$, and $\hat{\boldsymbol{\theta}}^T=(\hat{\alpha},\hat{\boldsymbol{\beta}}^T)$ is the EL estimate under the null hypothesis. More specifically, let $\tilde{\mathbf{X}}_1,\ldots,\tilde{\mathbf{X}}_{n_1}$ be drawn independently from $\hat{G}(\mathbf{x})$, and $\tilde{\mathbf{X}}_{n_1+1},\ldots,\tilde{\mathbf{X}}_n$ independently from $\hat{F}(\mathbf{x})$, resulting in the combined bootstrap samples $\tilde{\mathbf{X}}_1,\ldots,\tilde{\mathbf{X}}_n$. Then, we can construct the corresponding bootstrap version of the SELR statistic $\{\tilde{\mathscr{L}}_n^b\}$ for the testing problem (2.15) or (2.17), where $b=1,\ldots,B$, and B is the number of bootstrap samples. Let $\mathscr{L}_n^{\text{lel}}$ be the SELR statistic based on original data, then the p-value is

$$\sum_{b=1}^B I\left\{\mathscr{L}_n^{\text{lel}}\geq\tilde{\mathscr{L}}_n^b\right\}/B.$$

3. SIMULATION STUDIES

3.1 Performance of Estimation

The finite-sample performance of the proposed method is evaluated by simulation studies. We generate $\mathbf{X}_1,\ldots,\mathbf{X}_{n_1}$ and $\mathbf{X}_{n_1+1},\ldots,\mathbf{X}_{n_1+n_2}$ from two densities $f(\mathbf{x})$ and $g(\mathbf{x})$, respectively. We set both densities $f(\mathbf{x})$ and $g(\mathbf{x})$ to be trivariate normal distributions, in which $\mathbf{x}=(w,\mathbf{z}^T)^T$, w is a scalar, $\mathbf{z}^T=(z_1,z_2)$,

Table 1. Simulation results at 5 fixed points (0.4, 0.8, 1.0, 1.3, 1.6) with sample size $n = 400$

			$\varrho_n = 0.3$				$\varrho_n = 0.5$				$\varrho_n = 0.7$			
w_0	Par	True	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
0.4	$\alpha(w)$	−0.130	−0.074	0.132	0.133	94.0	−0.080	0.123	0.126	93.2	−0.098	0.145	0.141	93.0
	$\boldsymbol{\beta}_1(w)$	0.200	0.004	0.228	0.228	94.7	0.018	0.214	0.212	94.7	0.002	0.232	0.227	95.0
	$\boldsymbol{\beta}_2(w)$	0.400	0.007	0.179	0.175	94.4	0.005	0.168	0.161	93.4	0.009	0.176	0.174	94.2
	$\alpha'(w)$	−0.567	−0.035	0.159	0.158	95.9	−0.020	0.143	0.142	95.8	−0.009	0.153	0.149	95.6
	$\boldsymbol{\beta}'_1(w)$	−0.333	−0.112	0.149	0.146	89.0	−0.118	0.143	0.138	88.2	−0.101	0.156	0.146	91.1
	$\boldsymbol{\beta}'_2(w)$	1.000	0.036	0.169	0.157	94.4	0.014	0.144	0.142	94.7	0.002	0.158	0.144	93.0
0.8	$\alpha(w)$	−0.370	−0.079	0.139	0.139	94.9	−0.081	0.138	0.140	94.1	−0.096	0.166	0.162	93.2
	$\boldsymbol{\beta}_1(w)$	0.067	−0.035	0.236	0.235	94.6	−0.024	0.224	0.218	93.7	−0.034	0.236	0.234	95.0
	$\boldsymbol{\beta}_2(w)$	0.800	0.023	0.195	0.191	94.4	0.014	0.186	0.176	93.3	0.015	0.194	0.188	93.6
	$\alpha'(w)$	−0.633	0.018	0.159	0.158	93.7	0.034	0.144	0.142	92.8	0.046	0.153	0.149	92.9
	$\boldsymbol{\beta}'_1(w)$	−0.333	−0.114	0.149	0.145	88.2	−0.121	0.142	0.137	87.8	−0.104	0.156	0.145	90.3
	$\boldsymbol{\beta}'_2(w)$	1.000	0.036	0.170	0.156	94.0	0.014	0.144	0.142	94.3	0.002	0.158	0.144	92.5
1	$\alpha(w)$	−0.500	−0.074	0.152	0.151	95.5	−0.074	0.153	0.153	94.3	−0.089	0.181	0.179	93.2
	$\boldsymbol{\beta}_1(w)$	−0.000	−0.054	0.244	0.243	94.0	−0.046	0.234	0.226	92.7	−0.052	0.243	0.241	94.5
	$\boldsymbol{\beta}_2(w)$	1.000	0.031	0.210	0.205	94.8	0.018	0.201	0.189	93.2	0.017	0.209	0.201	93.1
	$\alpha'(w)$	−0.667	0.046	0.159	0.158	92.4	0.062	0.144	0.142	90.9	0.073	0.153	0.149	89.6
	$\boldsymbol{\beta}'_1(w)$	−0.333	−0.114	0.150	0.145	88.2	−0.123	0.141	0.137	87.6	−0.105	0.156	0.145	90.9
	$\boldsymbol{\beta}'_2(w)$	1.000	0.035	0.171	0.156	93.8	0.014	0.144	0.142	94.3	0.002	0.158	0.144	92.5
1.3	$\alpha(w)$	−0.708	−0.058	0.179	0.176	94.4	−0.055	0.180	0.179	94.3	−0.069	0.212	0.209	93.8
	$\boldsymbol{\beta}_1(w)$	−0.100	−0.082	0.264	0.260	93.1	−0.078	0.253	0.242	92.2	−0.079	0.262	0.259	94.7
	$\boldsymbol{\beta}_2(w)$	1.300	0.042	0.241	0.232	93.9	0.025	0.228	0.214	92.8	0.022	0.239	0.226	92.9
	$\alpha'(w)$	−0.717	0.088	0.160	0.158	89.0	0.104	0.145	0.142	85.7	0.115	0.154	0.149	83.6
	$\boldsymbol{\beta}'_1(w)$	−0.333	−0.114	0.150	0.145	88.0	−0.123	0.142	0.137	87.5	−0.106	0.156	0.145	90.5
	$\boldsymbol{\beta}'_2(w)$	1.000	0.035	0.173	0.157	93.7	0.014	0.145	0.142	94.2	0.003	0.159	0.145	92.4
1.6	$\alpha(w)$	−0.930	−0.032	0.213	0.209	94.2	−0.026	0.213	0.211	93.4	−0.038	0.248	0.244	93.6
	$\boldsymbol{\beta}_1(w)$	−0.200	−0.112	0.290	0.282	92.6	−0.111	0.279	0.264	91.4	−0.107	0.287	0.282	94.1
	$\boldsymbol{\beta}_2(w)$	1.600	0.055	0.280	0.265	93.0	0.031	0.259	0.244	92.6	0.027	0.275	0.256	91.8
	$\alpha'(w)$	−0.767	0.130	0.161	0.158	83.2	0.145	0.146	0.142	78.6	0.156	0.154	0.149	75.9
	$\boldsymbol{\beta}'_1(w)$	−0.333	−0.115	0.152	0.145	87.4	−0.125	0.142	0.137	87.1	−0.107	0.156	0.145	90.4
	$\boldsymbol{\beta}'_2(w)$	1.000	0.036	0.174	0.157	93.7	0.015	0.146	0.142	94.3	0.004	0.161	0.145	92.3

and

$$g(w, \mathbf{z}) = (2\pi)^{-3/2}|\Sigma_g|^{-1/2} \exp\left\{-2^{-1}\left[(w-1, z_1, z_2)\Sigma_g^{-1}(w-1, z_1, z_2)^T\right]\right\},$$

$$f(w, \mathbf{z}) = (2\pi)^{-3/2}|\Sigma_f|^{-1/2} \exp\left\{-2^{-1}[(w-1, z_1, z_2-1)\Sigma_f^{-1}(w-1, z_1, z_2-1)^T]\right\},$$

are trivariate normal densities with means $\mu_g = (1, 0, 0)^T$ and $\mu_f = (1, 0, 1)^T$, and inverses of the covariances

$$\Sigma_g^{-1} = \begin{pmatrix} 1/2 & -2/3 & 1/3 \\ -2/3 & 2 & 0 \\ 1/3 & 0 & 1 \end{pmatrix},$$

$$\Sigma_f^{-1} = \begin{pmatrix} 2/3 & -1/3 & -2/3 \\ -1/3 & 2 & 0 \\ -2/3 & 0 & 1 \end{pmatrix}.$$

Because $f(w, \mathbf{z})/g(w, \mathbf{z}) = \exp\{-z_1(w-1)/3 + z_2 w - (w-1)^2/12 - 2(w-1)/3 - 1/2\}$, we have $\alpha(w) = -(w-1)^2/12 - 2(w-1)/3 - 1/2$, $\boldsymbol{\beta}_1(w) = -(w-1)/3$, and $\boldsymbol{\beta}_2(w) = w$.

We draw 1000 datasets with sample size $n = n_1 + n_2 = 400$ for various values of $\varrho_n = n_1/n$. We choose the Epanechnikov kernel $K(t) = 0.75(1-t^2)_+$ to localize the coefficient functions. The suitable smoothing bandwidth for estimating $\alpha(w)$ and $\boldsymbol{\beta}(w)$ is selected by EBBS as discussed in Section 2.1. In fact, we can also use bandwidth $h^* = c \cdot \hat{\sigma}_w \cdot n^{-1/5}$, where c is a constant and has the value of 1.06 as in Zhou, Wan, and Wang (2008). A simulation study shows the sensitivity disappears when c is not too large, which is omitted here to save space. In Table 1, we summarize the simulation results for $\alpha(w)$, $\boldsymbol{\beta}_1(w)$, $\boldsymbol{\beta}_2(w)$, and their derivatives $\alpha'(w)$, $\boldsymbol{\beta}'_1(w)$, $\boldsymbol{\beta}'_2(w)$, at points $w_0 = 0.4, 0.8, 1.0, 1.3, 1.6$ for $\varrho_n = 0.3, 0.5, 0.7$, respectively, where $w_0 = 1.0$ is the population mean of w. We report the bias of the estimate (Bias) which is estimated by EBBS, the standard deviation of the 1000 estimates (SD), the average of the estimated standard error (SE), and the estimated coverage probability (CP) of the nominal 95% confidence interval for these parameters. The calculation of CP takes the estimated bias into account. The performance of the proposed LEL is satisfactory. The biases are relatively small for all functions. The average SE is very close to SD and this indicates good performance of the proposed variance estimator. For a fixed point w, SE, SD and the absolute value of the bias are smaller for equal sample sizes than for unequal samples sizes. For all functions $\alpha(\cdot)$, $\beta_1(\cdot)$, and $\beta_1(\cdot)$, the empirical coverage levels are close to the nominal level 95%. It is well known that it is hard to estimate the derivatives of the parameters using local linear functions. If one wants to get more accurate estimates of their derivatives,

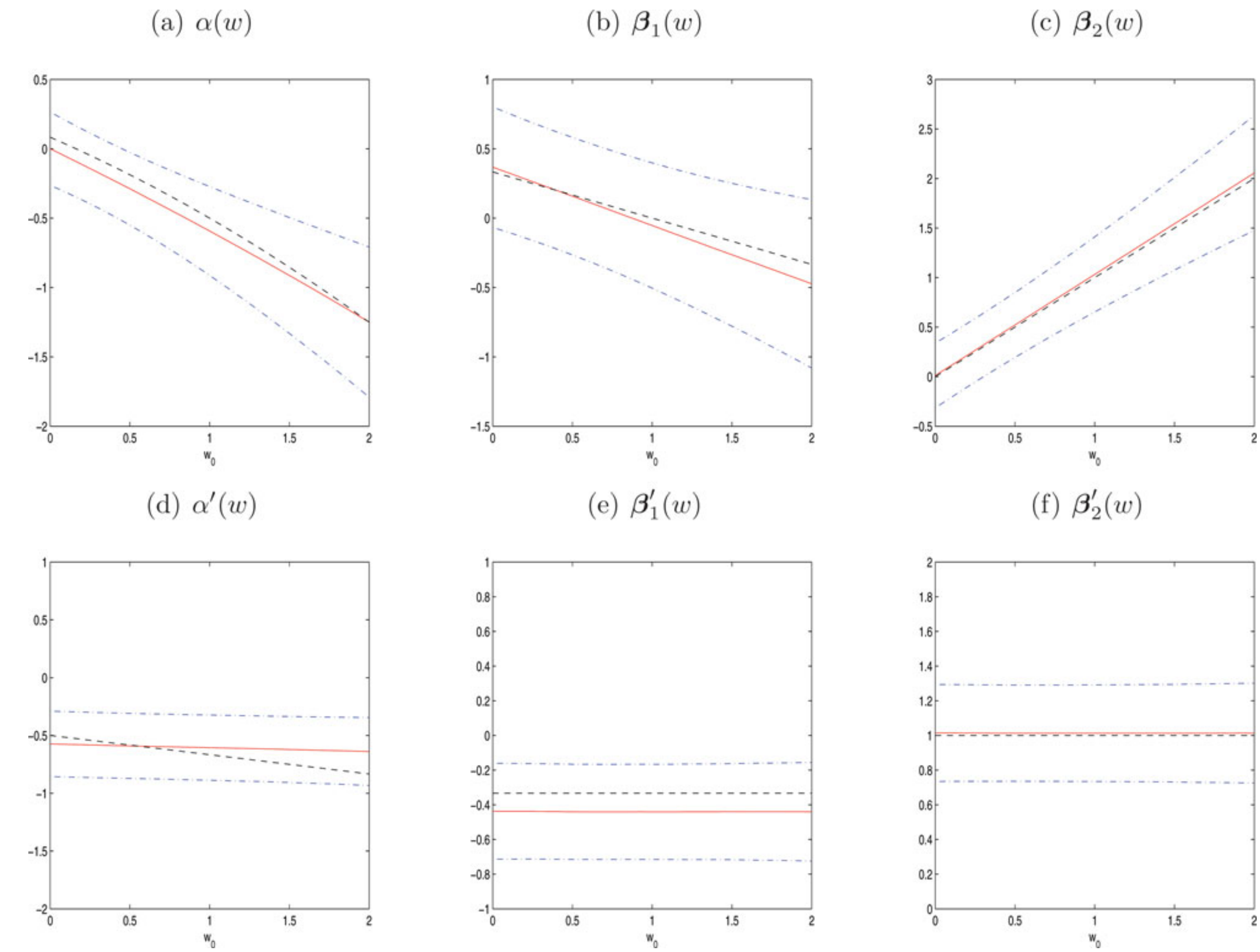

Figure 1. Sample size $n = 400$ and $\varrho_n = 0.6$. Estimator of parameter (solid line); Confidence bands of estimator (dot dashed line); True curve (dashed line).

higher order of local polynomial needs to be used. Fortunately, in case-control study, one usually cares more about the odds ratio which is related to the estimation of parameters instead of their derivatives. We also compare our LEL method with local MLE proposed by Cai, Fan, and Li (2000). To save space, the results are given in the supplementary materials.

In Figure 1, the estimates of $\alpha(w)$, $\boldsymbol{\beta}_1(w)$, $\boldsymbol{\beta}_2(w)$, their derivatives $\alpha'(w)$, $\boldsymbol{\beta}_1'(w)$, $\boldsymbol{\beta}_2'(w)$, and their corresponding confidence bands are plotted for the simulation study with the ratio $\varrho_n = 0.6$, where the estimated curves and confidence bands are the averages over 1000 simulations. We depict 101 points for w in the interval from 0 to 2. All of the estimated curves almost overlap with the corresponding true curves, and the confidence bands are very tight, which implies that the overall mean squared errors (MSE) are all small.

3.2 Performance of Hypothesis Tests

In this section, we explore the numerical performance of proposed hypothesis tests. Consider density ratio model

$$f(w, z) = \psi(\alpha_0(w) + \beta_0(w)z)g(w, z).$$

For simplification, we set $\psi(\cdot) = \exp(\cdot)$ in this simulation. We want to test if the regression coefficients are constant, that is, testing problem (2.17),

$$H_0 : \alpha_0(\cdot) = \alpha_0, \quad \beta_0(\cdot) = \beta_0, \tag{3.21}$$

where α_0 and β_0 are unknown constants. The power of the test is evaluated under a sequence of alternative models indexed by τ,

$$H_1^\tau : \alpha_0(\cdot) = \alpha_0 + \tau(\alpha(\cdot) - \alpha_0), \quad \beta_0(\cdot) = \beta_0 + \tau(\beta(\cdot) - \beta_0),$$

where $\alpha(w) = \exp\{-w^2/2\} + \sin(\pi w)$, $\beta(w) = -3w^2 + \cos(\pi w)$, $\alpha_0 = E\alpha(W)$, and $\beta_0 = E\beta(W)$. Because of equivalence between density-ratio model and logistic model except for the intercept term when $\psi(\cdot)$ is exponential function, we can generate data from logistic model $P(y = 1|(w, z)) = \exp\{\phi(w, z)\}/(1 + \exp\{\phi(w, z)\}$, where $\phi(w, z) = \alpha_0(w) + \beta_0(w)z$. We generate w from uniform $U(-1, 1)$ and Z from standard normal distribution. We first draw m data points $\mathbf{X} = (W, Z)$ and $Y \in \{0, 1\}$ from logistic model, where $m \gg n$, and $n = 400$ is sample size. Denote by D_1 the dataset when $Y = 0$ and by D_2 the dataset when $Y = 1$. Then, draw randomly n_1 data points $\mathbf{X}$ from D_1 as control data, and correspondingly, draw randomly n_2 data points $\mathbf{X}$ from D_2 as case data. Note that the true intercept $\alpha_0(w)$ is unknown which results in the true density functions $g(w, z)$ and $f(w, z)$

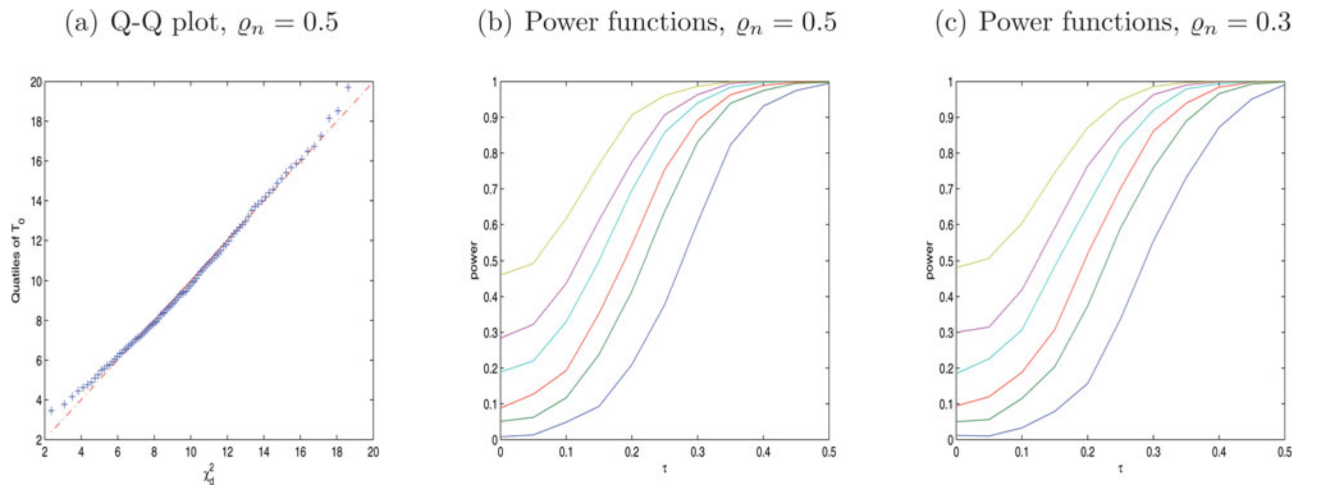

Figure 2. (a) Q-Q plot of scaled $r\mathscr{L}_n$ and χ^2_d distribution with degrees of freedom $d = 9.95$ and scale $r = 2.87$ by 1000 simulations. (b) and (c) power functions.

being unknown, but this has no impact on the hypothesis test (3.21).

To speed up the calculation for estimating $\boldsymbol{\xi}(w_i)$, we employ the one-step method proposed by Cai, Fan, and Li (2000) using 10 grid points $\{w_{(20)}, w_{(60)}, w_{(100)}, \ldots, w_{(340)}, w_{(380)}\}$ where $w_{(i)}$ are the sorted $\{w_i\}_{i=1}^n$. In this simulation, we use the Epanechnikov kernel $K(t) = 0.75(1 - t^2)_+$. The bandwidth is chosen to be $h = 1.8n^{-2/9}$, similar to Fan and Zhang (2004). To check whether the proposed SELR statistic $\mathscr{L}_n$ follows asymptotically the chi-squared distribution, we estimate the mean and variance of $\mathscr{L}_n$ under the null hypothesis based on 1000 Monte Carlo simulations, denoted by $\tilde{\mu}$ and $\tilde{\sigma}^2$. Figure 2(a) depicts the Q-Q plot of the scaled $r\mathscr{L}_n$ versus the χ^2_d distribution with degrees of freedom d, by using 100 points from the 1000 standardized $\mathscr{L}_n$ at $q = 0.01, 0.02, \ldots, 1 \times 100$th quantile, where $r = 2\tilde{\mu}/\tilde{\sigma}^2$ and $d = 2\tilde{\mu}^2/\tilde{\sigma}^2$. It could be seen that the χ^2_d distribution provides good approximation.

Figure 2(b) and (c) depict the power functions for $\varrho_n = 0.5$ and 0.3, and $\tau = (0, 0.05, 0.1, \ldots, 0.5)$ at six different significance levels (0.01, 0.05, 0.1, 0.2, 0.3, 0.5) based on 1000 Monte Carlo simulations and 1000 bootstrap samples for each simulation. Note that when $\tau = 0$, the alternative becomes the null hypothesis, and the Type I error instead of power is calculated. The realized Type I errors at the six significance levels are 0.008, 0.051, 0.088, 0.189, 0.284, 0.460 for $\varrho_n = 0.5$, and 0.012, 0.050, 0.094, 0.185, 0.300, 0.481 for $\varrho_n = 0.3$, which imply that the proposed bootstrap method gives the right levels of test. It is obvious that the power functions increase rapidly as τ increases, which shows that the proposed hypothesis test performs well even when sample size is small.

4. APPLICATION

We now use the proposed method to analyze a dataset from a population-based study of gastric cancer conducted in a high-risk population in Warsaw, Poland, between 1994 and 1996. The association of telomere length with gastric cancer risk as well as potential confounding factors was studied by Hou et al. (2009). The details of the study design were also discussed in Chow et al. (1999). Complete data for several risk factors are available for 701 subjects with 289 (41.23%) cases and 412 controls. The six factors we considered include age in years ("Age"), telomere length ("TL"), pack-years of smoking ("Pak"), drinking status, gender ("Sex"), and *H. pylori* status ("H.p"). Subjects were from 28 to 80 years old with an average age of 62.8 in case and 63.2 in control. Telomere length was measured in blood leukocyte DNA using quantitative real-time PCR. Pack-years of smoking is defined as the number of years smoked multiplied by the number of packages per day. Participants' drinking statuses are divided into three categories: never drinker, former drinker, and current drinker. Two indicator variables are created to indicate the three levels of drinking status: "Drk1" is coded as 1 for former drinker and 0 otherwise, and "Dkr2" is coded as 1 for current drinker and 0 otherwise. *H. pylori* infection is a known risk factor for gastric cancer. An individual's status is defined as negative if tested negative for both *H. pylori* and cagA antibody, and positive if tested positive for at least one of the two markers. More details of characteristics of study subjects are shown in Table S.2 in the supplementary materials.

For this real data, we focus on studying how the effects of risk factors vary with respect to age and which risk factors contribute to gastric cancer. The following varying-coefficient model is considered, as in model (1.4),

$$\begin{aligned}\exp\{\phi(w_i, z_i)\} = \exp\{&\alpha(\text{age}_i) + \beta_1(\text{age}_i) * \text{TL}_i \\ &+ \beta_2(\text{age}_i) * \text{Pak}_i + \beta_3(\text{age}_i) * \text{Drk1}_i \\ &+ \beta_4(\text{age}_i) * \text{Drk2}_i + \beta_5(\text{age}_i) * \text{Sex}_i \\ &+ \beta_6(\text{age}_i) * \text{H.p}_i\}\end{aligned}$$

Before fitting the model, the three variables, "Age," "TL," and "Pak," are standardized by subtracting the respective median and dividing by the standard deviation. To answer the question whether a varying-coefficient model is needed for this dataset, we perform hypothesis test (2.17) using the SELR statistic $\mathscr{L}_n$ (2.18) where w is "Age." We use the Epanechnikov kernel and set bandwidth $h = 5n^{-2/9} = 1.1657$. Based on 1000 bootstrap samples, the p-value is less than 0.005. Therefore, there is a strong evidence that the varying-coefficient model is more appropriate to describe this dataset.

Table 2. Estimates of curves at 4 fixed age points (45, 55, 65, 75)

	Age = 45			Age = 55			Age = 65			Age = 75		
Par	Est	SE	Length	Est	SE	Length	Est	SE	Length	Est	SE	Length
Intercept	−0.283	0.631	2.475	−0.116	0.426	1.669	0.283	0.270	1.059	0.772	0.405	1.587
TL	−0.172	0.189	0.740	−0.166	0.134	0.526	−0.206	0.100	0.391	−0.501	0.190	0.747
Pak	1.012	0.387	1.519	0.737	0.185	0.726	0.435	0.107	0.421	0.068	0.150	0.589
Drk1	1.564	0.726	2.845	1.174	0.412	1.614	0.521	0.277	1.085	−0.579	0.427	1.674
Drk2	−0.749	0.532	2.086	−1.047	0.337	1.320	−0.897	0.253	0.992	−0.904	0.459	1.799
Sex	−0.024	0.518	2.032	0.008	0.330	1.293	0.011	0.238	0.931	0.335	0.395	1.547
H.p	0.504	0.565	2.217	0.355	0.363	1.421	−0.175	0.256	1.003	−0.825	0.404	1.584

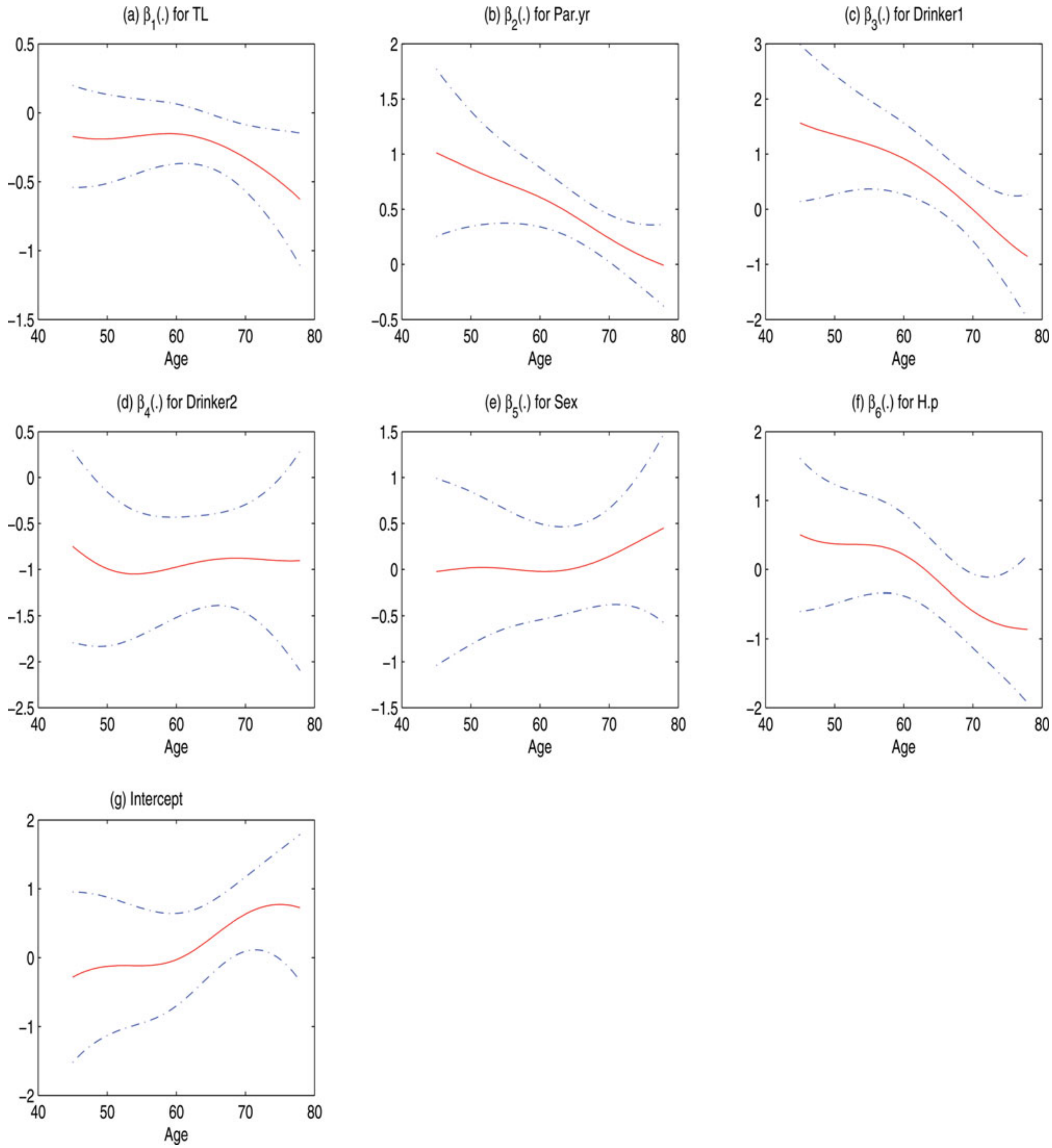

Figure 3. Estimates of parameters (solid line); Confidence band (dashed line).

The results are presented in Table 2 and Figure 3 using bandwidth 0.8004 chosen by EBBS (Section S6.2 in the supplementary materials). In Table 2, for the column labeled with "Par," "Age" denotes the intercept parameter $\alpha(\text{age})$, and others denote the corresponding slope parameters. We report the estimates (Est) of all coefficients at 4 different ages (45, 55, 65, 75), the estimated SE and the length of the 95% confidence interval (Length). The estimates and their corresponding confidence intervals are plotted in Figure 3 as functions of age.

We can see that shorter telomere length is associated with higher risk of getting gastric cancer, and the effect of telomere length increases with age. When other factors are held constant, the odds of getting gastric cancer decrease by about 15% at age 55, 21% at age 65, and 37% at age 75, with each additional unit increase of telomere length. Higher pack-years of smoking is associated with high risk of gastric cancer and the effect of smoking decreases with age. The odds of getting gastric cancer increase about 109% at age 55, 50% at age 65, only 7% at age 75, for each additional unit increase of pack years of smoking. Former drinkers have a higher risk of getting gastric cancer comparing with never drinkers for subjects with age less than 65. The odds for former drinkers are 3.27 times the odds for never drinkers at age 55. The effect of former drinking status disappears for subjects aged 65 and older. It is interesting to observe that the current drinkers have a lower risk of getting gastric cancer with a roughly constant odds ratio of 0.41 across all ages comparing with never drinkers. This phenomenon could be because cancer patients are more likely to stop drinking after the diagnosis or after experiencing stomach pain, which in turn results in smaller proportion of current drinkers in the case group than in the control group. Gender has little effect on the risk because the coefficient is close to zero over the span of age. *H. pylori* positive subjects aged 60 or younger have a higher risk of getting gastric cancer, but *H. pylori* positive subjects aged 65 or older have a lower risk of getting gastric cancer, comparing with *H. pylori* negative subjects. This pattern may be partially explained by the disappearance of *H. pylori* after cancer diagnosis. Another reason could be due to the high *H. pylori* prevalence, 88.32%, in the control group with ages older than 60.

5. DISCUSSION

In this article, we extend the two-sample density-ratio to a more flexible model, a two-sample varying-coefficient density-ratio model, which overcomes the potential misspecification problem of parametric methods. In the general setting, the local MLE is infeasible due to the unspecified densities $g(\mathbf{x})$ and $f(\mathbf{x})$. In the special case with the ratio function $\psi(\cdot) = \exp(\cdot)$, we show that LEL and local MLE are the same except for the intercept term. We localize the two smooth coefficient functions $\alpha(w)$ and $\boldsymbol{\beta}(w)$ using the first order Taylor's expansions. Unlike the ordinary EL method for the standard two-sample density-ratio model (1.2) which was considered by Qin (1999), we estimate the coefficient functions by constructing the LEL function (2.9) while adjusting the constraint condition for the Lagrange multiplier locally at the same time. The asymptotic properties of the LEL estimators are provided. To test if the varying coefficients are constant and other related hypotheses, the SELR statistics are constructed and they follow asymptotically a chi-squared distribution under some regularity conditions. A modified bootstrap procedure is proposed to estimate the SELR's null distribution. Simulation studies and real data analysis illustrate that the proposed model, estimation and hypothesis testing methods implement well with small sample size.

The proposed method can be easily extended to three samples as in Qin (1999) or even more samples, or to studies with more than one kind of disease. Another important extension is to consider the partially linear varying-coefficient density-ratio model

$$f(w, z, u) = \psi\{\alpha(w) + \beta(w)^T z + \gamma^T u\} g(w, z, u),$$

where γ is an unknown parameter. This model is substantially different from the model discussed in the article, and we will explore this semi-varying coefficient model in future work.

APPENDIX: PROOFS

Assumptions:

1. The second derivative of $g(\mathbf{x})$ is continuous and bounded, and the marginal densities $g_w(w)$ and $f_w(w)$ in control and case samples are bounded away from zero.
2. $\alpha(\cdot)$ and $\boldsymbol{\beta}_i(\cdot)$ have continuous second derivatives at all points w_0 in the support of w.
3. $||\bar{m}(\mathbf{x})||^3$, $||\partial\bar{m}(\boldsymbol{\xi}^T\mathbf{X}^*)/\partial\boldsymbol{\xi}||$, and $||\partial^2\bar{m}(\boldsymbol{\xi}^T\mathbf{X}^*)/\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^T||$ are bounded by some integrable function in a neighborhood of the true value $\boldsymbol{\xi}_0$, and $\partial^2\bar{m}(\boldsymbol{\xi}^T\mathbf{X}^*)/\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^T$ is continuous with respect to $\boldsymbol{\xi}$ in this neighborhood, where $||\cdot||$ denotes Euclidean norm, $\mathbf{X}^* = (1, \mathbf{z}^T, w - w_0, \mathbf{z}^T(w - w_0))^T$ is $\mathbf{X}_i^*$ omitting the subscript i, $\bar{m}(\boldsymbol{\xi}^T\mathbf{X}^*) = (\psi(\boldsymbol{\xi}^T\mathbf{X}^*) - 1)\bar{\gamma}(\boldsymbol{\xi}^T\mathbf{X}^*)^{-1}$ and $\bar{\gamma}(\boldsymbol{\xi}^T\mathbf{X}^*) = \rho + (1-\rho)\psi(\boldsymbol{\xi}^T\mathbf{X}^*)$.
4. The matrix $\mathbf{a}$ is nonsingular at the point w_0.
5. The kernel function $K(\cdot)$ is a symmetric density function with compact support.
6. $\varrho_n \to \rho$ as $n = n_1 + n_2 \to \infty$, where $0 < \rho < 1$.
7. $h \to 0$ and $nh \to \infty$ as $n \to \infty$.
8. $h_1 \to 0$, $nh_1 \to \infty$ and $nh_1^4 \to 0$ as $n \to \infty$.

Notation: Define $\boldsymbol{\eta} = \mathbf{H}(\boldsymbol{\xi} - \boldsymbol{\xi}_0)$, and $\boldsymbol{\xi} = \mathbf{H}^{-1}\boldsymbol{\eta} + \boldsymbol{\xi}_0$. Thus, $\ell(\mathbf{H}^{-1}\boldsymbol{\eta} + \boldsymbol{\xi}_0) = \ell_1(\mathbf{H}^{-1}\boldsymbol{\eta} + \boldsymbol{\xi}_0) + \ell_2(\mathbf{H}^{-1}\boldsymbol{\eta} + \boldsymbol{\xi}_0)$. Then, $\hat{\boldsymbol{\eta}}$ maximizes the value of local log-likelihood $\ell(\mathbf{H}^{-1}\boldsymbol{\eta} + \boldsymbol{\xi}_0)$ subject to constraint condition (2.11) in which $\boldsymbol{\xi}$ is replaced by $\mathbf{H}^{-1}\boldsymbol{\eta} + \boldsymbol{\xi}_0$. We write $t(\boldsymbol{\eta}) = t(\mathbf{H}^{-1}\boldsymbol{\eta} + \boldsymbol{\xi}_0)$, $\psi_j(\boldsymbol{\eta}) = \psi(\boldsymbol{\eta}^T\mathbf{H}^{-1}\mathbf{X}_j^* + \boldsymbol{\xi}_0^T\mathbf{X}_j^*)$, $m_j(\boldsymbol{\eta}) = m(\boldsymbol{\eta}^T\mathbf{H}^{-1}\mathbf{X}_j^* + \boldsymbol{\xi}_0^T\mathbf{X}_j^*)$, $\gamma_j(\boldsymbol{\eta}) = \gamma(\boldsymbol{\eta}^T\mathbf{H}^{-1}\mathbf{X}_j^* + \boldsymbol{\xi}_0^T\mathbf{X}_j^*)$, $\ell(\boldsymbol{\eta}) = \ell(\mathbf{H}^{-1}\boldsymbol{\eta} + \boldsymbol{\xi}_0)$, and $\ell_i(\boldsymbol{\eta}) = \ell_i(\mathbf{H}^{-1}\boldsymbol{\eta} + \boldsymbol{\xi}_0)$, where $i = 1, 2$. Let

$$M_{1n}(\boldsymbol{\eta}, t) = -\frac{\partial \ell_2(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} + \sum_{j=1}^{n} \frac{t\,\partial m_j(\boldsymbol{\eta})/\partial \boldsymbol{\eta}}{1 + t m_j(\boldsymbol{\eta})} w_h(W_j, w_0),$$

and

$$M_{2n}(\boldsymbol{\eta}, t) = \sum_{j=1}^{n} \frac{m_j(\boldsymbol{\eta})}{1 + t m_j(\boldsymbol{\eta})} w_h(W_j, w_0).$$

The details of proofs of Lemma A.1–A.4 are given in the supplementary materials.

Lemma A.1. Let assumptions (1)–(7) be satisfied. Then, as $n \to \infty$,

$$-\frac{\partial^2 \ell_2(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T}\bigg|_{\boldsymbol{\eta}=0} \xrightarrow{P} S_{11}, \tag{A.1}$$

$$\sum_{j=1}^{n} \frac{\partial m(\boldsymbol{\xi}_0^T \mathbf{X}_j^*)}{\partial \boldsymbol{\xi}} w_h(W_j, w_0) \xrightarrow{P} S_{12}, \tag{A.2}$$

$$\sum_{j=1}^{n} m^2(\boldsymbol{\xi}_0^T \mathbf{X}_j^*) w_h(W_j, w_0) \xrightarrow{P} S_{22}, \tag{A.3}$$

where $S_{11} = \rho(1-\rho)\mathrm{diag}(\mathbf{a}, \mu_2 \mathbf{a})$, $S_{12} = (\tilde{a}_{12}^T, \mathbf{0}^T)^T$, and $S_{22} = b_0$.

Lemma A.2. Suppose assumptions (1)–(7) are satisfied. Then, at some point $\hat{\boldsymbol{\eta}}$ in the interior of the ball $||\boldsymbol{\eta}|| \le (nh)^{-1/3}$, $\ell(\boldsymbol{\eta})$ attains its local maximum value, and $\hat{t} = t(\hat{\boldsymbol{\eta}})$ satisfies

$$M_{1n}(\hat{\boldsymbol{\eta}}, \hat{t}) = 0 \quad \text{and} \quad M_{2n}(\hat{\boldsymbol{\eta}}, \hat{t}) = 0.$$

Lemma A.2 implies that $\hat{\boldsymbol{\eta}}$ converges to 0 with convergence rate $(nh)^{-1/3}$. More precisely, it can be shown that $\hat{\boldsymbol{\eta}}$ is a $\sqrt{nh}$ consistent estimator of zero, that is, $(\hat{\alpha}(w_0), \hat{\boldsymbol{\beta}}(w_0)^T)^T$ is a $\sqrt{nh}$ consistent estimator of $(\alpha(w_0), \boldsymbol{\beta}(w_0)^T)^T$.

Lemma A.3. Let assumptions (1)–(7) be satisfied. Let $M_n(\mathbf{0}, 0) = (M_{1n}(\mathbf{0}, 0)^T, M_{2n}(\mathbf{0}, 0))^T$, and $Q = (Q_1^T, Q_2)^T$.
Then we have

$$\sqrt{nh}(M_n(\mathbf{0}, 0) - Q) \xrightarrow{\mathcal{L}} N(0, \Gamma),$$

where $Q_1 = -h^2/2\mu_2(\mathbf{a}(\alpha_0'', (\boldsymbol{\beta}_0'')^T), \mathbf{0}^T)^T$, $Q_2 = -h^2/2\mu_2(\alpha_0'', (\boldsymbol{\beta}_0'')^T)\tilde{a}_{12}$, $\Gamma = \mathrm{diag}(\Gamma_{11}, \nu_0 b_0)/q_0(w_0)$, and $\Gamma_{11} = \mathrm{diag}(\rho(1-\rho)\nu_0 \mathbf{a}, \rho(1-\rho)\nu_2 \mathbf{a})$.

Proof of Theorem 1. This theorem is a straightforward consequence of Lemma A.2.

Proof of Theorem 2. Taking derivatives with respect to $\boldsymbol{\eta}$ and t, we have

$$\frac{\partial M_{1n}(\mathbf{0}, 0)}{\partial \boldsymbol{\eta}^T} = -\frac{\partial^2 \ell_2(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T}\bigg|_{\boldsymbol{\eta}=0},$$

$$\frac{\partial M_{2n}(\mathbf{0}, 0)}{\partial t} = \sum_{j=1}^{n} m^2(\boldsymbol{\xi}_0^T \mathbf{X}_j^*) w_h(W_j, w_0),$$

$$\frac{\partial M_{1n}(\mathbf{0}, 0)}{\partial t} = \frac{\partial M_{2n}(\mathbf{0}, 0)}{\partial \boldsymbol{\eta}} = \sum_{j=1}^{n} \frac{\partial m_j(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\bigg|_{\boldsymbol{\eta}=0} w_h(W_j, w_0).$$

Similar to Qin and Lawless (1994), by the assumptions of this Theorem and Lemma A.2, we have

$$0 = M_{1n}(\hat{\boldsymbol{\eta}}, \hat{t}) = M_{1n}(\mathbf{0}, 0) + \frac{\partial M_{1n}(\mathbf{0}, 0)}{\partial \boldsymbol{\eta}^T}\hat{\boldsymbol{\eta}} + \frac{\partial M_{1n}(\mathbf{0}, 0)}{\partial t}\hat{t} + o_p(||\hat{\boldsymbol{\eta}}|| + ||\hat{t}||),$$

$$0 = M_{2n}(\hat{\boldsymbol{\eta}}, \hat{t}) = M_{2n}(\mathbf{0}, 0) + \frac{\partial M_{2n}(\mathbf{0}, 0)}{\partial \boldsymbol{\eta}^T}\hat{\boldsymbol{\eta}} + \frac{\partial M_{2n}(\mathbf{0}, 0)}{\partial t}\hat{t} + o_p(||\hat{\boldsymbol{\eta}}|| + ||\hat{t}||).$$

Then, we have $(\hat{\boldsymbol{\eta}}^T, \hat{t})^T = -S_n^{-1}(M_{1n}(\mathbf{0}, 0)^T, M_{2n}(\mathbf{0}, 0))^T$, where

$$S_n = \begin{pmatrix} \partial M_{1n}(\mathbf{0}, 0)/\partial \boldsymbol{\eta}^T & \partial M_{1n}(\mathbf{0}, 0)/\partial t \\ \partial M_{2n}(\mathbf{0}, 0)/\partial \boldsymbol{\eta}^T & \partial M_{2n}(\mathbf{0}, 0)/\partial t \end{pmatrix}.$$

By Lemma A.1, it is easy to see that $S_n \xrightarrow{P} S$, where $S = (S_{ij})$, $i, j = 1, 2$. From this and $M_n(\mathbf{0}, 0) = O((nh)^{-1/2})$ by Lemma A.3, we have $||\hat{\boldsymbol{\eta}}|| + ||\hat{t}|| = O_p((nh)^{-1/2})$. Then, combining it with Lemma A.3, by Slutsky's theorem we have

$$\sqrt{nh}\left[(\hat{\boldsymbol{\eta}}^T, \hat{t})^T + S^{-1}Q\right] \xrightarrow{\mathcal{L}} N(0, S^{-1}\Gamma S^{-1}).$$

This completes the proof of Theorem 2 noting the definition of $\boldsymbol{\eta}$ and simplification of $S^{-1}Q$ and $S^{-1}\Gamma S^{-1}$.

Let $\hat{M}_{1n}(\boldsymbol{\eta}, t)$ and $\hat{M}_{2n}(\boldsymbol{\eta}, t)$ denote $M_{1n}(\boldsymbol{\eta}, t)$ and $M_{1n}(\boldsymbol{\eta}, t)$, respectively, with $\kappa(W_i)$ replaced by $\hat{\kappa}(W_i)$, $i = 1, \ldots, n$, in (2.12). Define $\tilde{b}_0(w) = E_G\{\psi(\phi(\mathbf{X}))/\bar{\gamma}(\mathbf{X})|W = w\}$.

Lemma A.4. Let assumptions (1) and (8) be satisfied, $\hat{M}_n(\mathbf{0}, 0) = (\hat{M}_{1n}(\mathbf{0}, 0)^T, \hat{M}_{2n}(\mathbf{0}, 0))^T$, and $Q = (Q_1^T, Q_2)^T$. Then, we have

$$\sqrt{nh}(\hat{M}_n(\mathbf{0}, 0) - Q) \xrightarrow{\mathcal{L}} N(0, \Gamma + \tilde{\Gamma}),$$

where Q and Γ is defined in Lemma A.3,

$\tilde{\Gamma} = (\tilde{\Gamma}_{ij})$, $i, j = 1, 2$ and

$$\tilde{\Gamma}_{11} = 3\frac{\rho(1-\rho)\nu_0\kappa(w_0)}{q_0(w_0)q_0^*(w_0)}S_{12}^{\otimes 2},$$

$$\tilde{\Gamma}_{22} = \frac{\nu_0 q_0^*(w_0)^3}{\rho(1-\rho)g_w(w_0)}\tilde{b}_0(w_0)^2 - 2\frac{\nu_0\kappa(w_0)}{g_w(w_0)}b_0\tilde{b}_0(w_0),$$

$$\tilde{\Gamma}_{12} = \tilde{\Gamma}_{21}^T = 2\frac{\nu_0\kappa(w_0)}{g_w(w_0)}\tilde{b}_0(w_0)S_{12} - \frac{\nu_0\rho(1-\rho)\kappa(w_0)^2}{g_w(w_0)q_0^*(w_0)^3}b_0 S_{12}.$$

SUPPLEMENTARY MATERIALS

Proofs of Lemma A.1–A.4 and detailed proofs of Theorem 2–4. The simulation results for the comparison of our LEL method with the local MLE proposed by Cai, Fan, and Li (2000). The details of EBBS for bandwidth selection.

[Received January 2012. Revised October 2013.]

REFERENCES

Cai, J., Fan, J., Zhou, H., and Zhou, Y. (2007), "Hazard Models With Varying Coefficients for Multivariate Failure Time Data," *The Annals of Statistics*, 35, 324–354. [636]

Cai, Z., Fan, J., and Li, R. (2000), "Efficient Estimation and Inferences for Varying-Coefficient Models," *Journal of the American Statistical Association*, 95, 888–902. [641,642,645]

Carroll, R. J., Ruppert, D., and Welsh, A. H. (1998), "Local Estimating Equations," *Journal of the American Statistical Association*, 93, 214–227. [638]

Chen, S. X., and Hall, P. (1993), "Smoothed Empirical Likelihood Confidence Intervals for Quantiles," *The Annals of Statistics*, 21, 1166–1181. [636]

Chow, W. H., Swanson, C. A., Lissowska, J., Grover, F. D., Sobin, L. H., Nasierowska-Guttmejer, A., Radziszewski, J., Regula, J., Hsing, A. W., Jagannatha, S., Zatonski, W., and Blot, W. J. (1999), "Risk of Stomach Cancer in Relation to Consumption of Cigarettes, Alcohol, Tea and Coffee in Warsaw, Poland," *International Journal of Cancer*, 81, 871–876. [642]

Fan, J., Zhang, C., and Zhang, J. (2001), "Generalized Likelihood Ratio Statistics and Wilks Phenomenon," *The Annals of Statistics*, 29, 153–193. [636]

Fan, J., and Zhang, J. (2004), "Sieve Empirical Likelihood Ratio Tests for Nonparametric Functions," *The Annals of Statistics*, 32, 1858–1907. [636,639,642]

Hastie, T., and Tibshirani, R. (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society,* Series B, 55, 757–796. [635]

Hou, L., Savage, S. A., Blaser, M. J., Perez-Perez, G. Hoxha, M. Dioni, L., Pegorano, V., Dong, L. M., Zatonski, W., Lissowska, J., Chow, W. H., and Baccarelli, A. (2009), "Telomere Length in Peripheral Leukocyte DNA and Gastric Cancer Risk," *Cancer Epidemiology Biomarkers Prevention*, 18, 3103–3109. [642]

Kay, R., and Little, S. (1987), "Transformations of the Explanatory Variables in the Logistic Regression Model for Binary Data," *Biometrika*, 74, 495–501. [635]

Lachin, J. M. (2000), *Biostatistical Methods: The Assessment of Relative Risks*, New York: Wiley. [635]

Lustbader, E. D., Moolgavkar, S. H., and Venzon, D. J. (1984), "Tests of the Null Hypothesis in Case-Control Studies," *Biometrics*, 40, 1017–1024. [636]

Miettinen, O. S. (1976), "Estimability and Estimation in Case-Referent Studies," *American Journal of Epidemiology*, 104, 226–235. [635]

Owen, A. B. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249. [636,637]

——— (1990), "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, 18, 90–120. [636]

——— (2001), *Empirical Likelihood*, London: Chapman and Hall. [636]

Prentice, R. L., and Pyke, R. (1979), "Logistic Disease Incidence Models and Case-Control Studies," *Biometrika*, 66, 403–411. [635,636,638]

Qin, J. (1993), "Empirical Likelihood in Biased Sample Problems," *The Annals of Statistics*, 46, 1182–1196. [636]

——— (1998), "Inferences for Case-Control and Semiparametric Two-Sample Density Ratio Models," *Biometrika*, 85, 619–630. [635]

——— (1999), "Empirical Likelihood Ratio Based Confidence Intervals for Mixture Proportions," *The Annals of Statistics*, 27, 1368–1384. [635,636,637,644]

——— (2005), "Marginal Likelihood, Conditional Likelihood and Empirical Likelihood: Connections and Applications," *Biometrika*, 92, 251–270. [635]

Qin, J., Berwick, M., Ashbolt, R. and Dwyer, T. (2002), "Quantifying the Change of Melanoma Incidence by Breslow Thickness," *Biometrics*, 58, 665–670. [635]

Qin, J., and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22, 300–325. [636,645]

Qin, J., and Zhang, B. (1997), "A Goodness-of-Fit Test for Logistic Regression Models Based on Case-Control Data," *Biometrika*, 84, 609–618. [635]

Ruppert, D. (1997), "Empirical-Bias Bandwidths for Lcoal Polynomial Nonparametric Regression and Density Estimation," *Journal of the American Statistical Association*, 92, 1049–1062. [638]

Ruppert, D., Sheathers, S. J., and Wand, M. P. (1995), "An Effective Bandwidth Selector for Local Least Squares Regression," *Journal of the American Statistical Association*, 90, 1257–1270. [638]

Sepanski, J. H., Knickerbocker, R., and Carroll, R. J. (1994), "A Semiparametric Correction for Attenuation," *Journal of the American Statistical Association*, 89, 1366–1373. [638]

Thomas, D. C. (1981), "General Relative-Risk Models for Survival Time and Matched Case-Control Analysis," *Biometrics*, 37, 673–686. [636]

Vardi, Y. (1982), "Nonparametric Estimation in the Presence of Length Bias," *The Annals of Statistics*, 10, 616–620. [635]

——— (1985), "Empirical Distributions in Selection Bias Models," *The Annals of Statistics*, 13, 178–203. [635]

White, I. R., and Thompson, S. G. (2003), "Choice of Test for Comparing Two Groups, With Particular Application to Skewed Outcomes," *Statistics in Medicine*, 22, 1205–1215. [635]

Zhang, B. (2001), "A Information Matrix Test for Logistic Regression Models Based on Case-Control Data," *Biometrika*, 88, 921–932. [635]

Zhang, J., and Gijbels, I. (2003), "Sieve Empirical Likelihood and Extensions of Generalized Least Squares," *Scandinavian Journal of Statistics*, 30, 1–24. [636]

Zhou, Y., Wan, A. T. K., and Wang, X. (2008), "Estimating Equation Inference With Missing Data," *Journal of the American Statistical Association*, 103, 1187–1199. [640]

Zou, F., Fine, J. P., and Yandell, B. S. (2002), "On Empirical Likelihood for a Semiparametric Mixture Model," *Biometrika*, 89, 61–75. [635]