Statistica Sinica **25** (2015), 1045-1064
doi:http://dx.doi.org/10.5705/ss.2013.092

ON REGRESSION FOR SAMPLES WITH ALTERNATING PREDICTORS AND ITS APPLICATION TO PSYCHROMETRIC CHARTS

Mosuk Chow[1], Bing Li[1] and Jackie Q. Xue[2]

[1] *The Pennsylvania State University and*
[2] *Beijing Hold Renhe Culture Investment Ltd*

Abstract: We introduce and study a new type of regression that arises from a kinesiology experiment concerning human's tolerance to temperature and water vapor pressure. In the experiment, a set of pressure and temperature values are collected to construct a psychrometric chart. The problem differs from traditional regression because, for one part of the data, temperature is held fixed while pressure is raised to an equilibrium point; for the other part of the data, pressure is held fixed while temperature is raised to an equilibrium point. The purpose of this peculiar design is to ensure the safety of the participants. Traditional regression is inadequate for modeling this type of data, because the roles of predictor and response alternate. We propose a new regression where the predictor and response alternate while being linked by a bijective function. We study the population and asymptotic properties of this regression, develop test statistics for model selection and analysis of variance, and outline several extensions and refinements. We apply the method to kinesiology data and find, among other things, that the gender difference in psychrometric charts diminishes in old age.

Key words and phrases: Alternating predictor, analysis of covariance, estimating equation, heat-humidity tolerance, SWAP regression.

1. Introduction

Human's tolerance to heat and humidity (as measured by water vapor pressure) is a subject of considerable interest in human physiology — for example in the research on heat-related deaths (Hawkins-Bell (1994); Semenza et al. (1996)), on the physiologic limits to work in heat (Lind (1963, 1970); Belding and Kamon (1972); Kamon and Avellini (1976)), and on the relation between aging and work-heat tolerance (Pandolf (1997); Kenny and Anderson (1988); Drinkwater and Horvath (1979)). A common tool used in such studies is the psychrometric chart, which is based on an alternating design where the response and predictor trade places in one experiment. To our knowledge, there have not been statistical methods available to adequately handle this type of problem. We introduce a new regression that allows the response and the predictor to trade places.

Our inquiry was originated by a data set collected in an experiment in Kinesiology (Zeman (2001)). The study was concerned with epidemics of deaths in heat waves for older people and its purpose was to determine "Upper Limits of the Prescriptive Zone" (ULPZ) on a psychrometric chart of the ambient dry bulb temperature T versus the water vapor pressure P. The study performed a sequence of temperature-pressure tolerance experiments which were age- and sex-specific. Forty healthy subjects, including older men, older women, younger men, and younger women of average fitness were recruited, with each of the 4 groups containing 9 to 11 subjects. The older subjects were aged between 63-80, and the younger subjects between 18-30. For each subject 6 experiments were performed, among which three were under warm and humid conditions, to be called the P_{crit} experiments, and three were under hot and dry conditions, to be called the T_{crit} experiments.

In the three T_{crit} experiments, P was held constant at 12 mmHg, 16 mmHg, or 20 mmHg, and the temperature was increased 1°C every five minutes, starting from 28°C after 30-minute equilibration period. This continued until a tolerance limit T was reached. In the three P_{crit} experiments, T was held constant at 34°C, 36°C, or 38°C while the pressure was increased 1 mmHg every five minutes, starting from 9 mmHg after 30-minute equilibration period. This continued until a pressure tolerance limit P was reached. Thus, experiments always started at regions of pressure and temperature comfortable for the human subjects, and gradually increased one variable. During each experiment, the subjects walked continuously on a treadmill for up to 2.5 hours at a constant speed in an environmental chamber. One point on the ULPZ line was determined as the ambient conditions at which body core temperature was forced out of equilibrium.

To illustrate the data, Figure 1 shows the portion of the data set for the older males. In the upper-left part, temperature acts as the predictor and the pressure acts as the response, whereas in the lower-right part, the pressure acts as the predictor and the temperature acts as the response. The solid curve is the ordinary least squares fit treating pressure as the response, and using a quadratic polynomial of the temperature as the regression function. This analysis is clearly inadequate as, for example, the observations lie almost entirely to the left of the curve at the bottom of the chart.

Our goal is to combine the two parts of the data into a coherent regression analysis, where the regression curve passes through the center of the response variables, whether temperature or pressure. Since the regression is designed for *<u>S</u>amples <u>W</u>ith <u>A</u>lternating <u>P</u>redictors*, we call it SWAP regression.

The rest of the paper is organized as follows. In Section 2 we introduce the SWAP regression estimator and establish its Fisher consistency. In Section 3 we develop the asymptotic theory for SWAP regression, including consistency,

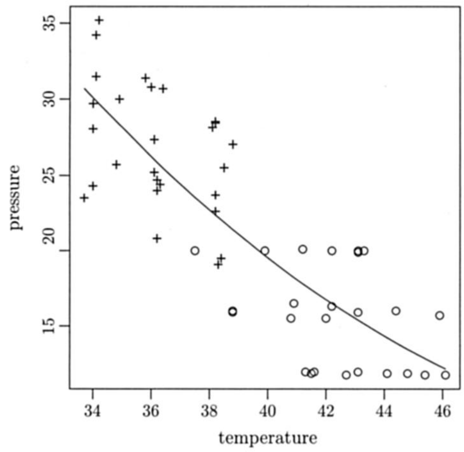

Figure 1. ULPZ data for older males. Plot symbol "+" represents the cases where temperature is held fixed while pressure is gradually increased. Plot symbol "o" represents the cases where pressure is held fixed while temperature is gradually increased. The solid line is the least-squares fit treating temperature as the predictor, and using a quadratic polynomial as the regression function.

asymptotic distribution for the estimator, and a Wilks-type test statistic. In Section 4 we extend SWAP regression to accommodate extra coordinates. This is needed in our data analysis for comparing the psychrometric charts for male and female. In Section 5 we introduce an optimal estimating equation and an adaptively weighted estimating equation for SWAP regression. In Section 6 we apply SWAP regression and related inference methods to the ULPZ data. In Section 7 we compare by simulation the SWAP regression estimators with the conventional regression that ignores the alternating design, and compare three versions of SWAP regression among themselves. Finally, in Section 8, we discuss the general paradigm of SWAP regression and outline how it can be broadened to adapt to a variety of applications. Proofs are in an online Appendix.

2. SWAP Regression

Intuition tells us that in the region where pressure is the predictor, we should regress temperature on pressure, and in the region where temperature is the predictor, we should regress pressure on temperature. We formulate this intuition into a coherent regression system.

Let X and Y denote two random variables, and let Z be a binary variable indicating whether X or Y is the predictor. Thus, when $Z = 0$, $E(Y|X, Z = 0)$ is of interest; when $Z = 1$, $E(X|Y, Z = 1)$ is of interest. Suppose that the sample space of (X, Y, Z) can be represented as the Cartesian product $\Omega_X \times \Omega_Y \times \{0, 1\}$.

For any random variable or vector U, such as $U = X$, $U = (X, Y)$, let P_U denote the distribution of U. Let $L_2(P_U)$ denote the class of functions square-integrable with respect to P_U.

Let $f : \Omega_X \to \Omega_Y$ be a bijection. We assume that

$$E(Y|X, Z = 0) = f(X), \quad E(X|Y, Z = 1) = f^{-1}(Y). \tag{2.1}$$

Thus, although the conditional expectations in the two temperature zones are different, they are related to each other as f and f^{-1}. We define a subfamily of $L^2(P_X)$ by

$$\mathscr{G} = \{g \in L^2(P_X) : g \text{ is an injection from } \Omega_X \text{ to } \Omega_Y,\ g^{-1} \in L^2(P_Y)\}.$$

We propose, at the population level, to minimize the quadratic loss function $Q : \mathscr{G} \to \mathbb{R}$:

$$Q(g) = E[(Y - g(X))^2 I(Z = 0)] + E[(X - g^{-1}(Y))^2 I(Z = 1)] \tag{2.2}$$

over $\mathscr{G}$. We use $\equiv$ to indicate mutually absolute continuity of two measures.

Theorem 1. *If* $\operatorname{var}(Y_1) < \infty$, $\operatorname{var}(X_2) < \infty$, *and* $f \in \mathscr{G}$, *then* f *minimizes* $Q(g)$ *over* $\mathscr{G}$. *If* $P_{X|Z=0} \equiv P_X$ *and* $P_{Y|Z=1} \equiv P_Y$, *then* f *is the (almost surely) unique minimizer of* $Q(g)$ *over* $\mathscr{G}$.

Let $(X_1, Y_1, Z_1), \ldots, (X_n, Y_n, Z_n)$ be an i.i.d. sample of (X, Y, Z) and take $E_n(\cdot)$ to be expectation with respect to the empirical distribution. The sample-level analogue of the objective function $Q(g)$ is

$$Q_n(g) = E_n[(Y - g(X))^2 I(Z = 0) + (X - g^{-1}(Y))^2 I(Z = 1)].$$

The SWAP regression estimator is taken as the minimizer of $Q_n(g)$ over $\mathscr{G}$.

We focus on parametric models. Let $\Theta \subseteq \mathbb{R}^p$, and let $\mathscr{G} = \{g_\theta(\cdot) : \theta \in \Theta\}$. We assume, for each $\theta \in \Theta$, $g_\theta : \Omega_X \to \Omega_Y$ is injective, with $g_\theta \in L_2(P_X)$ and $g_\theta^{-1} \in L_2(P_Y)$. In this context, our parametric SWAP regression estimator is obtained by solving the optimization problem

$$\text{minimize } \ Q_n(g_\theta) \text{ over } \Theta, \quad \text{subject to } \ g_\theta \text{ being injective.}$$

Let $\Theta_0 = \{\theta \in \Theta :\ g_\theta \text{ is injective}\}$. Then the above is equivalent to maximizing $Q_n(g_\theta)$ over Θ_0.

As an example, consider the quadratic polynomials $g_{(a,b,c)}(x) = ax^2 + bx + c$. In order for this function to be injective, we need $-b/(2a)$ to be outside the range of $X_1, \ldots, X_n$.

3. Asymptotic Analysis

3.1. Consistency

The SWAP regression estimator can be viewed as a solution to an estimating equation, whose consistency and asymptotic distribution are standard. See, for example, Crowder (1986), Li (1996, 1997) Heyde (1997, Chap. 12), and van der Vaart (1998, Chap. 5). Traditionally, there are two approaches to consistency of estimating equations, one derived from Wald (1949), another derived from Cramér (1946). We adopt the Cramér's approach because it does not impose global assumptions on the objective function $Q_n(g)$. Suppose that $g_\theta(X)$ is differentiable with respect to θ. Let

$$\begin{aligned} q((X,Y,Z),\theta) = & -2[Y - g_\theta(X)]\Big[\frac{\partial g_\theta(X)}{\partial\theta}\Big] I(Z=0) \\ & -2[X - g_\theta^{-1}(Y)]\Big[\frac{\partial g_\theta^{-1}(Y)}{\partial\theta}\Big] I(Z=1), \end{aligned} \tag{3.1}$$

and consider the estimating equation

$$\frac{\partial Q_n(g_\theta)}{\partial\theta} = E_n q((X,Y,Z),\theta) = 0. \tag{3.2}$$

If there is a θ_0 in the interior of Θ_0 such that $E(Y|X=x, Z=0) = g_{\theta_0}(x)$, then, with probability tending to 1, the minimizer of $Q_n(g_\theta)$ over Θ_0 is a solution to (3.2). Let U abbreviate (X,Y,Z) and $Q_n(\theta)$ abbreviate $Q_n(g_\theta)$.

Theorem 2. *Suppose* (i) *θ_0 belongs to the interior of Θ_0;* (ii) *the function $\theta \mapsto Eq(U,\theta)$ is differentiable at θ_0 under the integral sign, and $E[\partial q(U,\theta_0)/\partial\theta^\intercal]$ is positive definite;* (iii) *for a sufficiently small $\delta > 0$,*

$$E\Big(\sup_{\|\theta-\theta_0\|=\delta} \|q(U,\theta)\|\Big) < \infty;$$

(iv) *for each u, the function $\theta \mapsto q(u,\theta)$ is continuous. Then there is a sequence $\hat\theta_n$ such that $\hat\theta_n \to \theta_0$ almost surely and, with probability 1, $E_n q(U,\hat\theta_n) = 0$ for all but finitely many n.*

3.2. Asymptotic distribution

By standard asymptotic theory for estimating equations, we have

$$\sqrt{n}(\hat\theta - \theta_0) \xrightarrow{\mathscr{D}} N(0, J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)), \tag{3.3}$$

where $J(\theta) = E(\partial q(U,\theta)/\partial\theta^\intercal)$ and $I(\theta) = E(q(U,\theta)q^\intercal(U,\theta))$. See, for example, Heyde (1997, Sec. 12.4). In our context, the matrices $I(\theta)$ and $J(\theta)$ can be written more specifically as

$$\begin{aligned} J(\theta) =& 2E[(\frac{\partial g_\theta(X)}{\partial\theta})(\frac{\partial g_\theta(X)}{\partial\theta^\intercal})I(Z=0)] \\ &+ 2E[(\frac{\partial g_\theta^{-1}(Y)}{\partial\theta})(\frac{\partial g_\theta^{-1}(Y)}{\partial\theta^\intercal})I(Z=1)], \\ I(\theta) =& 4E\left[(Y-g_\theta(X))^2(\frac{\partial g_\theta(X)}{\partial\theta})(\frac{\partial g_\theta(X)}{\partial\theta^\intercal})I(Z=0)\right] \\ &+ 4E\left[(X-g_\theta^{-1}(Y))^2(\frac{\partial g_\theta^{-1}(Y)}{\partial\theta})(\frac{\partial g_\theta^{-1}(Y)}{\partial\theta^\intercal})I(Z=1)\right]. \end{aligned} \tag{3.4}$$

Theorem 3. *Suppose that $\hat\theta$ is a consistent solution to the equation $E_n q(U,\theta) = 0$. Suppose* (i) *$\theta_0 \in \mathrm{int}(\Theta_0)$;* (ii) *$Eq(U,\theta_0) = 0$;* (iii) *the entries of the matrix $E[q(U,\theta_0)q^\intercal(U,\theta_0)]$ are finite and the matrix is positive definite;* (iv) *the function $Eq(U,\theta)$ is differentiable under the integral sign, and the entries of $E[\partial q(U,\theta)/\partial\theta^\intercal]$ are integrable; and* (v) *the sequence of random matrices $\{E_n\partial q(U,\theta)/\partial\theta^\intercal : n \in \mathbb{N}\}$ is stochastically equicontinuous. Then $\sqrt{n}\,(\hat\theta - \theta_0)$ has asymptotic distribution* (3.3) *with $J(\theta_0)$ and $I(\theta_0)$ given by* (3.4).

We can check stochastic equicontinuity as follows. Let $h_n(\theta)$ denote the random function $\vec{[E_n q(U,\theta)/\partial\theta^\intercal]}$. Stochastic equicontinuity of $\{h_n(\theta) : n \in \mathbb{N}\}$ means, for any $\epsilon > 0$ and $\eta > 0$, there is a $\delta > 0$ such that

$$\limsup_{n\to\infty} P\Big(\sup_{\|\theta-\theta_0\|<\delta} \|h_n(\theta) - h_n(\theta_0)\| > \epsilon\Big) < \eta.$$

A sufficient condition for this is, in a neighborhood G of θ_0 ,

$$\sup_{\theta\in G} \|h_n(\theta) - h_n(\theta_0)\| \le M_n(U_1,\ldots,U_n)\|\theta-\theta_0\| \tag{3.5}$$

for some $M_n(U_1,\ldots,U_n) = O_P(1)$. Thus all we need to do is to bound $\|\partial h_n(\theta)/\partial\theta^\intercal\|$ in the vicinity of θ_0 by a function of the order $O_P(1)$ that does not depend on θ.

As an example, if $g_\theta(x) = \theta x$, where the true $\theta_0 > 0$, then

$$h_n(\theta) = 2E_n(X^2|Z=0) - 4\theta^{-3}E_n[(X-\theta^{-1}Y)Y|Z=1] + 2\theta^{-4}E_n(Y^2|Z=1).$$

Hence,

$$\frac{\partial h_n(\theta)}{\partial\theta} = 12\theta^{-4}E_n[(X-\theta^{-1}Y)Y|Z=1] - 12\theta^{-5}E_n(Y^2|Z=1).$$

By the Cauchy-Schwarz inequality, the absolute value of the right-hand side is bounded above by

$$12\theta^{-4}\sqrt{E_n[(X-\theta^{-1}Y)^2|Z=1]}\sqrt{E_n(Y^2|Z=1)} + 12\theta^{-5}E_n(Y^2|Z=1).$$

Because $t \le 1 + t^2$ and $(a+b)^2 \le 3(a^2+b^2)$, the above is no greater than

$$\begin{aligned} &12\theta^{-4}[1 + E_n((X-\theta^{-1}Y)^2|Z=1)](1+E_n(Y^2|Z=1)) + 12\theta^{-5}E_n(Y^2|Z=1) \\ &\le 12\theta^{-4}[1 + 3E_n(X^2|Z=1) + 3\theta^{-2}E_n(Y^2|Z=1)][1+E_n(Y^2|Z=1)] \\ &\qquad + 12\theta^{-5}E_n(Y^2|Z=1). \end{aligned}$$

Let $G = (a,b)$ be a neighborhood of θ_0 with $a > 0$ and $b < \infty$. Then, on G, $|\partial h_n(\theta)/\partial\theta|$ is no greater than

$$\begin{aligned} &12a^{-4}[1 + 3E_n(X^2|Z=1) + 3a^{-2}E_n(Y^2|Z=1)][1+E_n(Y^2|Z=1)] \\ &\qquad + 12a^{-5}E_n(Y^2|Z=1). \end{aligned}$$

Thus, if we assume X and Y have finite fourth moments then the above is of the order $O_p(1)$ and hence (3.5) holds. This type of arguments can be carried out for more complicated functions $g_\theta(x)$.

3.3. Asymptotic distribution for hypothesis testing

We develop a likelihood-ratio type statistic for testing general hypotheses. Let $h : \Theta \to \mathbb{R}^r$ $(r \le p)$, and consider the general hypotheses of the form

$$H_0 : h(\theta) = 0 \quad \text{versus} \quad H_1 : h(\theta) \ne 0. \tag{3.6}$$

Let

$$\hat\theta = \operatorname{argmin}\{Q_n(\theta) : \theta \in \Theta\}, \quad \tilde\theta = \operatorname{argmin}\{Q_n(\theta) : \theta \in \Theta,\ h(\theta) = 0\}.$$

Mimicking the Wilks' statistic (Wilks (1938); Cox and Hinkley (1974, p.92)), we propose the statistic

$$T_n = 2n[Q_n(\tilde\theta) - Q_n(\hat\theta)]. \tag{3.7}$$

Intuitively, if H_0 is correct, then $\hat\theta - \tilde\theta = O_P(n^{-1/2})$ and T_n is at most $O_P(n^{1/2})$ (in fact, it is of the order $O_P(1)$); otherwise, $\hat\theta - \tilde\theta = O_P(1)$ and T_n is of the order $O_P(n)$. In the following, we assume $h(\theta)$ is differentiable and write $H(\theta) = \partial h^\intercal(\theta)/\partial\theta$.

Theorem 4. *If the conditions in Theorem* 3 *hold and h is differentiable, then, under H_0 in* (3.6),

$$T_n \xrightarrow{\mathscr{D}} \sum_{i=1}^{p} \lambda_i K_i, \tag{3.8}$$

where $K_1, \ldots, K_p$ are i.i.d. $\chi^2_{(1)}$ and $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of the matrix

$$\Sigma(\theta) = I^{1/2}(\theta)J^{-1}(\theta)H(\theta)[H^\intercal(\theta)J^{-1}(\theta)H(\theta)]^{-1}H^\intercal(\theta)J^{-1}(\theta)I^{1/2}(\theta). \tag{3.9}$$

Because $H(\theta)$ is of dimension $p\times r$, $\Sigma(\theta)$ has rank r and the last r eigenvalues of $\Sigma(\theta)$ are zero.

In practice, the eigenvalues of $\lambda_1,\ldots,\lambda_p$ can be estimated by those of the matrix (3.9), where θ is substituted by $\hat\theta$ or $\tilde\theta$. Denoting these approximated eigenvalues by $\hat\lambda_1,\ldots,\hat\lambda_p$, the p-value of T_n can be computed by simulation or by an approximation introduced by Bentler and Xie (2000). If

$$\tilde T_n = \frac{\operatorname{tr}(\Sigma_n^2(\hat\theta))\,T_n}{\operatorname{tr}(\Sigma_n(\hat\theta)}, \tag{3.10}$$

where $\Sigma_n(\hat\theta)$ is (3.9) with $E(\cdot)$ replaced by $E_n(\cdot)$ and θ replaced by $\hat\theta$. Then $\tilde T_n$ is approximately distributed as $\chi^2_{(d)}$ where d is the nearest integer to $\operatorname{tr}(\Sigma_n(\hat\theta))^2/\operatorname{tr}(\Sigma_n^2(\hat\theta))$. The first method works well for p-values in the range ≥ 0.01 with 10,000 simulated random numbers. For smaller p-values the second is convenient and works surprisingly well. See also Satterthwaite (1941).

4. Analysis of Covariance

Our data set consists of four groups: younger females, older females, younger males, and older males. After fitting the SWAP regressions to the sub-samples and comparing the psychrometric charts, we observed that the charts of the old-female and the old-male groups are similar, whereas those for young-female and the young-male groups are rather different. Thus it is reasonable to speculate that the gender effect diminishes with age. This can be formulated as testing if two (or more) charts are statistically the same. This is, in essence, an analysis of covariance (ANCOVA) problem, which traditionally refers to the problem of comparing regression curves from several independent samples. We develop a hypothesis test procedure for this problem.

Consider m independent samples

$$\begin{array}{ccc} (X_1^{(1)},Y_1^{(1)},Z_1^{(1)}),\ldots,(X_{n_1}^{(1)},Y_{n_1}^{(1)},Z_{n_1}^{(1)}) & \overset{i.i.d.}{\sim} & (X^{(1)},Y^{(1)},Z^{(1)}), \\ \vdots & & \vdots \\ (X_1^{(m)},Y_1^{(m)},Z_1^{(m)}),\ldots,(X_{n_m}^{(m)},Y_{n_m}^{(m)},Z_{n_m}^{(m)}) & \overset{i.i.d.}{\sim} & (X^{(m)},Y^{(m)},Z^{(m)}). \end{array} \tag{4.1}$$

In our application, $m=2$, representing two gender groups. Suppose, for the kth sample,

$$E_{\theta^{(k)}}(Y^{(k)}|X^{(k)}) = g_{\theta^{(k)}}(X^{(k)}), \quad E_{\theta^{(k)}}(X^{(k)}|Y^{(k)}) = g^{-1}_{\theta^{(k)}}(Y^{(k)}). \tag{4.2}$$

To test whether the m SWAP regression curves are the same, we test

$$H_0 : \theta^{(1)} = \cdots = \theta^{(m)} \quad \text{versus} \quad H_1 : \theta^{(1)},\ldots,\theta^{(m)} \text{ are not all equal.} \tag{4.3}$$

More generally, letting $h : \Theta \times \cdots \times \Theta \to \mathbb{R}^s$ be a differentiable function, where s is a positive integer no greater than mp, we consider the hypotheses

$$H_0 : h(\theta^{(1)}, \ldots, \theta^{(m)}) = 0 \quad \text{versus} \quad H_1 : h(\theta^{(1)}, \ldots, \theta^{(m)}) \neq 0. \tag{4.4}$$

Let $n = n_1 + \cdots n_m$, and take the objective function

$$\mathbb{Q}_n(\theta^{(1)}, \ldots, \theta^{(m)}) = \frac{n_1 Q_{n_1}^{(1)}(\theta^{(1)})}{n} + \cdots + \frac{n_m Q_{n_m}^{(m)}(\theta^{(m)})}{n}, \tag{4.5}$$

where

$$Q_{n_k}^{(k)}(\theta^{(k)}) = E_{n_k}[(Y^{(k)} - g_{\theta^{(k)}}(X^{(k)}))^2 I(Z^{(k)}=0) + (X^{(k)} - g_{\theta^{(k)}}^{-1}(Y^{(k)}))^2 I(Z^{(k)}=1)].$$

Let $(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)})$ be the global maximizer of (4.5) over $\Theta \times \cdots \times \Theta$ and let $\tilde{\theta}$ be the constrained minimizer of (4.5) subject to $h(\theta^{(1)}, \ldots, \theta^{(m)}) = 0$. We test (4.3) using

$$\mathbb{T}_n = 2n[\mathbb{Q}_n(\tilde{\theta}^{(1)}, \ldots, \tilde{\theta}^{(m)}) - \mathbb{Q}_n(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)})].$$

For $k = 1, \ldots, m$, let

$$\begin{aligned} J^{(k)}(\theta^{(k)}) =& 2E_{\theta^{(k)}}[(\frac{\partial g_{\theta^{(k)}}(X^{(k)})}{\partial \theta^{(k)}})(\frac{\partial g_{\theta^{(k)}}(X^{(k)})}{\partial \theta^{(k)}})^\intercal I(Z^{(k)} = 0)] \\ &+ 2E_{\theta^{(k)}}[(\frac{\partial g_{\theta^{(k)}}^{-1}(Y^{(k)})}{\partial \theta^{(k)}})(\frac{\partial g_{\theta^{(k)}}^{-1}(Y^{(k)})}{\partial \theta^{(k)}})^\intercal I(Z^{(k)} = 1)], \\ I^{(k)}(\theta^{(k)}) =& 4E_{\theta^{(k)}}[\sigma_{\theta^{(k)}}^2(X^{(k)})(\frac{\partial g_{\theta^{(k)}}(X^{(k)})}{\partial \theta^{(k)}})(\frac{\partial g_{\theta^{(k)}}(X^{(k)})}{\partial \theta^{(k)}})^\intercal I(Z^{(k)} = 0)] \\ &+ 4E_{\theta^{(k)}}[\tau_{\theta^{(k)}}^2(Y^{(k)})(\frac{\partial g_{\theta^{(k)}}^{-1}(Y^{(k)})}{\partial \theta^{(k)}})(\frac{\partial g_{\theta^{(k)}}^{-1}(Y^{(k)})}{\partial \theta^{(k)}})^\intercal I(Z^{(k)} = 1)], \end{aligned}$$

where $\sigma_{\theta^{(k)}}^2(X^{(k)}) = \operatorname{var}_{\theta^{(k)}}(Y^{(k)}|X^{(k)})$ and $\tau_{\theta^{(k)}}^2(Y^{(k)}) = \operatorname{var}_{\theta^{(k)}}(X^{(k)}|Y^{(k)})$.

Theorem 5. *Suppose* (4.1) *and* (4.2) *hold and the m samples in* (4.1) *are independent, that the assumptions in Theorem* 3 *are satisfied for each of the m subpopulations, and for each $k = 1, \ldots, m$, $\lim_{n\to\infty}(n_k/n) = \alpha_k$ for some $0 < \alpha_k < 1$. Then*

$$\mathbb{T}_n \xrightarrow{\mathscr{D}} \sum_{i=1}^{s} \lambda_i K_i,$$

where $\lambda_1, \ldots, \lambda_s$ are the eigenvalues of the matrix

$$\mathbb{I}^{1/2}\mathbb{J}^{-1}\mathbb{H}(\mathbb{H}^\intercal\mathbb{J}^{-1}\mathbb{H})^{-1}\mathbb{H}^\intercal\mathbb{J}^{-1}\mathbb{I}^{1/2},$$

$$\mathbb{I} = \begin{pmatrix} \alpha_1 I^{(1)}(\theta) & & 0 \\ & \ddots & \\ 0 & & \alpha_m I^{(m)}(\theta) \end{pmatrix}, \ \mathbb{J} = \begin{pmatrix} \alpha_1 J^{(1)}(\theta) & & 0 \\ & \ddots & \\ 0 & & \alpha_m J^{(m)}(\theta) \end{pmatrix}, \tag{4.6}$$

and $\mathbb{H}$ is the mp by s gradient matrix $(\partial h/\partial\theta^{(1)\intercal}, \ldots, \partial h/\partial\theta^{(m)\intercal})^\intercal$.

In practice, the θ in $\mathbb{I}$ and $\mathbb{J}$ is replaced by either $(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)})$ or $(\tilde{\theta}^{(1)}, \ldots, \tilde{\theta}^{(m)})$, the expectations $E(\cdots)$ in $\mathbb{I}$ and $\mathbb{J}$ are replaced by the sample mean $E_n(\cdots)$, and the constants $\alpha_1, \ldots, \alpha_m$ are replaced by $n_1/n, \ldots, n_m/n$. For testing (4.3), we set $h : \Theta \times \cdots \times \Theta \to \mathbb{R}^{m(p-1)}$ as

$$(\theta^{(1)}, \ldots, \theta^{(m)}) \mapsto (\theta^{(1)} - \theta^{(2)}, \ldots, \theta^{(m-1)} - \theta^{(m)}).$$

The gradient $\mathbb{H}$ is the $mp \times m(p-1)$ dimensional matrix

$$\begin{pmatrix} I_p & & 0 \\ -I_p & \ddots & \\ & \ddots & I_p \\ 0 & & -I_p \end{pmatrix}.$$

The conditional variances $\sigma^2_{\theta^{(k)}}(X^{(k)})$ and $\tau^2_{\theta^{(k)}}(Y^{(k)})$ can be replaced by $(Y^{(k)} - g_{\theta^{(k)}}(X^{(k)}))^2$ and $(X^{(k)} - g_{\theta^{(k)}}(Y^{(k)}))^2$, respectively. The rank of the matrix at (4.6) is s, so the test statistic sum extends over only the nonzero eigenvalues.

We can generalize the above ANCOVA model to accommodate arbitrary covariates that might affect the relation between X and Y. Let $V \in \mathbb{R}^k$ be an additional (continuous or discrete) random (or nonrandom) vector that might affect the individual shapes of the relation between X and Y. Suppose X and Y are connected via $E(Y|X = x, Z = 0, V = v) = g_{\theta(v)}(x)$ and $E(X|Y = y, Z = 1, V = v) = g^{-1}_{\theta(v)}(y)$, where θ is a function v. To simplify the problem, we assume that $\theta(\cdot)$ is from a parametric family, say $\theta(v) = f(v, \eta)$, where $\eta \in \mathbb{R}^s$ is the parameter and $u \mapsto f(v, \eta)$ is a known function for each fixed η. The objective function is taken as

$$Q(g) = E_n[Y - g_{f(V,\eta)}(X)I(Z = 0)]^2 + E_n[X - g_{f(V,\eta)}(Y)I(Z = 1)]^2,$$

where, for example, $E_n[Y - g_{f(V,\eta)}(X)I(Z = 0)]^2$ represents the sample average of $[Y_i - g_{f(V_i,\eta)}(X_i)I(Z_i = 0)]^2, \quad i = 1, \ldots, n$. Here the vectors $V_1, \ldots, V_n$ can either be random or fixed.

Our ANCOVA model is a special case with v being an m-dimensional vector that takes only m values: $e_1, \ldots, e_m$ (the standard orthonormal basis of $\mathbb{R}^m$); The function $\theta(v)$ parameterized by the linear relation $\theta(v) = (\theta^{(1)}, \ldots, \theta^{(m)})v$.

5. Optimal and Adaptive Estimation

In this section we introduce an optimal estimating equation for SWAP regression that minimizes the asymptotic variance among the class of linear estimating equations, and an adaptive estimating equation where the optimal weights are estimated from the sample. Consider the estimating equation

$$E_n[a_\theta(X)(Y - g_\theta(X))I(Z = 0) + b_\theta(Y)(X - g_\theta^{-1}(Y))I(Z = 1)] = 0, \qquad (5.1)$$

where $a_\theta : \Omega_X \to \mathbb{R}^+$ and $b_\theta : \Omega_Y \to \mathbb{R}^+$ are weight functions. If $E_\theta(Y|X) = g_\theta(X)$ and $E_\theta(X|Y) = g_\theta^{-1}(Y)$, (5.1) is unbiased (Godambe (1960)):

$$E_\theta[a_\theta(X)(Y - g_\theta(X))I(Z = 0) + b_\theta(Y)(X - g_\theta^{-1}(Y))I(Z = 1)] = 0. \tag{5.2}$$

Here (5.1) is a generalization of (3.1) in which the weight functions are of the special form $a_\theta(X) = -2\partial g_\theta(X)/\partial\theta$ and $b_\theta(Y) = -2\partial g_\theta^{-1}(Y)/\partial\theta$. Denote the function inside $[\cdots]$ in (5.2) as $G_\theta(X,Y,Z)$. By the standard theory of estimating equations, if $\hat\theta$ is a consistent solution to (5.2), then

$$\sqrt{n}(\hat\theta - \theta_0) \xrightarrow{\mathscr{D}} N(0, J^{-1}(a_\theta, b_\theta)I(a_\theta, b_\theta)J^{-\mathsf{T}}(a_\theta, b_\theta)),$$

where

$$\begin{aligned} J(a_\theta, b_\theta) &= E_\theta[\frac{\partial G_\theta(X,Y,Z)}{\partial\theta^\mathsf{T}}], \\ I(a_\theta, b_\theta) &= E_\theta[G_\theta(X,Y,Z)G_\theta^\mathsf{T}(X,Y,Z)]. \end{aligned} \tag{5.3}$$

See, for example, Li (1993, 1996), Li and McCullagh (1994), and Heyde (1997). To emphasize the dependence of I and J on the weight functions, we write them as $I(a_\theta, b_\theta)$ and $J(a_\theta, b_\theta)$. It is then natural to choose the weighting functions $a_\theta(X), b_\theta(Y)$ that minimize the asymptotic variance

$$\mathrm{AV}(a_\theta, b_\theta) = J^{-1}(a_\theta, b_\theta)I(a_\theta, b_\theta)J^{-\mathsf{T}}(a_\theta, b_\theta).$$

This general approach was used in Li (2000, 2001), and Qu, Lindsay, and Li (2000) to derive various optimal estimating equations. It also echoes the construction of quasi likelihood: see Wedderburn (1974), McCullagh (1983), Godambe and Heyde (1987), Godambe and Thompson (1989), and McLeish and Small (1992). We use this approach to derive an optimal estimating equation for SWAP regression.

To simplify notation, let

$$\begin{aligned} \beta_\theta(X,Y) &= (a_\theta(X), b_\theta(Y)), \\ \delta_\theta(X,Y) &= \mathrm{diag}(Y - g_\theta(X), X - g_\theta^{-1}(Y)), \\ c(Z) &= \mathrm{diag}(I(Z=0), I(Z=1)). \end{aligned} \tag{5.4}$$

Then (5.1) becomes $E_n[\beta_\theta(X,Y)c(Z)\delta_\theta(X,Y)] = 0$. Let

$$\gamma_\theta(X,Y) = (\frac{\partial g_\theta(X)}{\partial\theta}, \frac{\partial g_\theta^{-1}(Y)}{\partial\theta}), \quad v_\theta(X,Y) = \mathrm{diag}(\mathrm{var}_\theta(Y|X), \mathrm{var}_\theta(X|Y)).$$

Theorem 6. *Let $\mathscr{W}$ be the family of weighting functions $\beta_\theta(x,y) = (a_\theta(x), b_\theta(y))$ such that the entries of $E_\theta[\beta_\theta(X,Y)c(Z)\gamma_\theta^\mathsf{T}(X,Y)]$ are finite and the matrix is nonsingular; the entries of $E_\theta[\beta_\theta(X,Y)c(Z)v_\theta(X,Y)\beta_\theta^\mathsf{T}(X,Y)]$ are finite and the matrix is nonsingular. If the weighting function $\beta_\theta^*(X,Y) = \gamma_\theta(X,Y)v_\theta^{-1}(X,Y)$ belongs to $\mathscr{W}$, then $\mathrm{AV}(\beta_\theta^*) \le \mathrm{AV}(\beta_\theta)$ for all $\beta_\theta \in \mathscr{W}$ and all $\theta \in \Theta$, where $A \le B$ means $B - A$ is positive semidefinite.*

This result implies that the solution to the estimating equation

$$E_n\left[\frac{\frac{\partial g_\theta(X)}{\partial\theta}(Y-g_\theta(X))I(Z=0)}{\mathrm{var}_\theta(Y|X,Z=0)}+\frac{\frac{\partial g_\theta^{-1}(Y)}{\partial\theta}(X-g_\theta^{-1}(Y))I(Z=1)}{\mathrm{var}_\theta(X|Y,Z=1)}\right]=0, \quad (5.5)$$

has the smallest variance among all estimating equations of the form (5.1). We call this optimal estimator the optimal SWAP, or OSWAP.

To use this optimal estimating equation we need the conditional variance functions $\sigma_\theta^2(X)=\mathrm{var}_\theta(Y|X,Z=0)$ and $\tau_\theta^2(Y)=\mathrm{var}_\theta(X|Y,Z=1)$. We can either choose parametric models for them, as is done in the original quasi likelihood method, or estimate them nonparametrically, as in Chiou and Müller (1998) and Li (2001).

We propose an easy-to-implement parametric estimator of the conditional variance that is quite effective for our purpose. Consider $\mathrm{var}_\theta(Y|X=x,Z=0)=ce^{ax}$, where c and a can depend on θ, but we suppress this dependence for convenience. This function captures the commonly seen heteroscedasticity pattern, with c representing the baseline conditional variance and a controlling whether and to what degree the conditional variance is increasing or decreasing with x. We propose to estimate c and a adaptively from data, as follows. Let $\tilde\theta$ be the SWAP regression estimator. For easy calculation we use a combination of the method of moment combined with L_1 minimization. By the method of moment we have the equation,

$$E_n[(Y-g_{\tilde\theta}(X))^2|Z=0]=E_n(ce^{aX}).$$

Solving this equation for c yields

$$c_n(a)=\frac{E_n[(Y-g_{\tilde\theta}(X))^2|Z=0]}{E_n(e^{aX})}.$$

We estimate a by minimizing $E_n\left(\,|(Y-g_{\tilde\theta}(X))^2-c_n(a)e^{aX}|\,\,|\,Z=0\right)$ over a grid of a, which is easy to compute because a is a scalar. Let $\hat a$ denote the minimizer over the grid. Then we obtain $c_n(\hat a)e^{\hat a x}$ as the estimator of $\mathrm{var}_\theta(Y|X=x,Z=0)$. Let $d_n(\hat b)e^{\hat b x}$ be the parallel estimator of $\mathrm{var}_\theta(X|Y=y,Z=1)$. Substituting these into (5.5), we arrive at the adaptively weighted estimating equation

$$E_n\left[\frac{\frac{\partial g_\theta(X)}{\partial\theta}(Y-g_\theta(X))I(Z=0)}{c_n(\hat a)e^{\hat a X}}+\frac{\frac{\partial g_\theta^{-1}(Y)}{\partial\theta}(X-g_\theta^{-1}(Y))I(Z=1)}{d_n(\hat b)e^{\hat b Y}}\right]=0.$$

We call the solution to this equation the adaptive SWAP regression estimator, or ASWAP. This ASWAP performs well in simulations, even when the true conditional variance is not of the form ce^{ax} (or de^{by}).

6. Data Analysis

We applied SWAP regression to the data described in the Introduction. The full data set contains four sub-samples: younger females (with sample size $n = 51$), younger males ($n = 62$), older females ($n = 52$), older males ($n = 52$). Each subject was repeatedly observed at 5 to 6 design points. For example, the older male group had 52 pairs of observations contributed by 9 subjects. The data originally contain 227 pairs of temperature and pressure values, in which 10 pairs contained missing numbers. To simplify the analysis, here we treated all pairs as independent and deleted the pairs with missing data to focus on the central issue of alternating design.

We use the quadratic model, $g_\theta(x) = ax^2 + bx + c$, where $\theta = (a, b, c)^\intercal$. By inspection and the physical meaning of the data, it is reasonable to restrict g_θ to monotone decreasing functions over the range $[X_{(1)}, X_{(n)}]$. This leads to the constraint

$$\Theta = \{(a, b, c) : \quad 2aX_{(n)} + b < 0\}. \tag{6.1}$$

Under this constraint, $g_\theta : [X_{(1)}, X_{(n)}] \to \mathbb{R}$ is invertible, with inverse

$$g_\theta^{-1}(y) = \frac{-b - \sqrt{b^2 - 4a(c - y)}}{2a},$$

where we have taken the decreasing branch of the two roots.

Let X be the temperature, Y the pressure, with $Z = 0$ when X is the predictor, and $Z = 1$ when Y is the predictor. The SWAP objective function is

$$\begin{aligned} Q_n(g_\theta) = & E_n[(Y - aX^2 - bX - c)^2 I(Z = 0)] \\ & + E_n\Big[X + \Big(b + \frac{\sqrt{b^2 - 4a(c - Y)}}{2a}\Big)^2 I(Z = 1)\Big]. \end{aligned}$$

This is to be minimized subject to (6.1), but in all cases the global minimizers happen to occur within the interior of the set (6.1).

We applied SWAP regression to all four samples, and the fitted curves are presented in Figure 2. As in Figure 1, the plot symbol "+" represents cases for which temperature is the predictor, and the plot symbol "∘" represents cases for which pressure is the predictor. Compared with the least-squares fit (for older males) in Figure 1, where the regression line fits rather poorly the portion of the data where temperature is the response, the SWAP regression lines in Figure 2 go through the centers of the alternating response variables.

The downward-bending quadratic tendencies for older people seem to be stronger than those for younger people. We use the method developed in Section 3.3 to test

$$H_0 : c = 0 \quad \text{versus} \quad H_1 : c \neq 0.$$

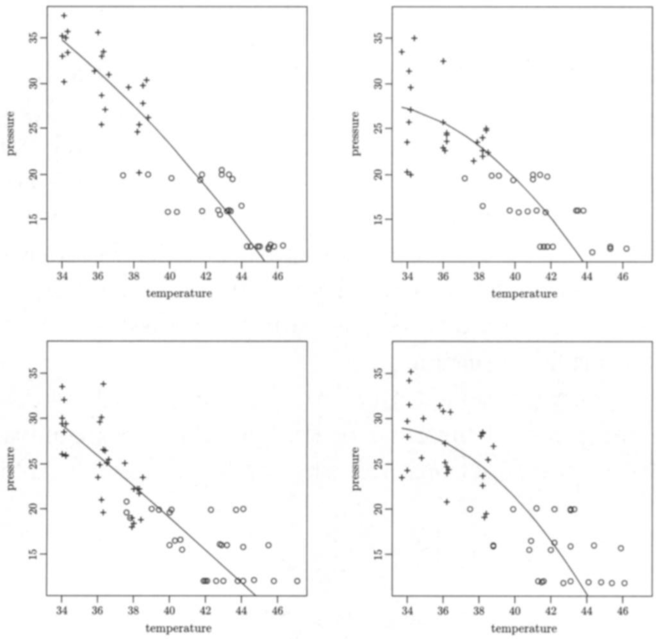

Figure 2. SWAP-regression fits to four ULPZ data sets: upper-left panel for young female, upper-right panel for old female; lower-left panel for young male, and lower-right panel for old male. The four plots are in the same scale.

Table 1. Significance of downward bending tendencies.

group id	$\hat{\lambda}_1$	$\hat{\lambda}_2$	T_n	p-value	p-value*
younger female	13.058	6.785	12.800	0.517	0.525
older female	38.220	7.918	69.130	0.219	0.221
younger male	12.069	9.970	0.722	0.971	0.968
older male	30.752	10.425	119.351	0.062	0.055

for each of the four groups. The results are presented in Table 1, in which columns 2 and 3 are the weights for the weighted χ^2's in Theorem 4, with θ_0 replaced by $\hat{\theta}$ calculated from the quadratic model, and column 4 gives the test statistic T_n at (3.7). The p-values by simulation appear in column 5 and those calculated by (3.10) appear in column 6. In this context the matrix (3.9) has rank 2, and so $\hat{\lambda}_3$ is identically 0.

The quadratic component for older male group is significant at level $\alpha = 0.1$, while the p-values for the older female group are small but not significant.

Comparing the curves in Figure 2 we also observe that the tolerance curves of older men and women are much more similar to each other than those for younger men and women, where women have much higher overall tolerance levels.

Table 2. Differences between genders within age groups.

comparison	$\mathbb{T}_n$	p-value	p-value*
young female vs young male	89.98	0.000	0.000
old female vs old male	812.43	0.260	0.269

To confirm this observation, we test

$$H_0 : \theta^{(1)} = \theta^{(2)} \quad \text{versus} \quad H_1 : \theta^{(1)} \neq \theta^{(2)},$$

where $\theta^{(1)}$ and $\theta^{(2)}$ are the parameter (a, b, c) for the older female and older male groups, respectively. In this case, $\mathbb{H} = (I_3, -I_3)^{\intercal}$. Table 2 presents the result of this test.

The difference between the younger female group and younger male group is very significant, but there is no significant difference between the older female and older male groups, indicating that the gender effect diminishes with age.

7. Simulation Comparisons

We compared the performances of four methods with simulated data for the alternating design: ordinary least squares (OLS) treating X as the predictor and Y as the response, the SWAP regression estimator in Section 2, and OSWAP and ASWAP in Section 5.

We considered two SWAP regression models. The first was a homoscedastic SWAP regression model where the conditional variance $\mathrm{var}(Y|X, Z = 0)$ and $\mathrm{var}(X|Y, Z = 1)$ were constants. Let

$$g_\theta(x) = ax^2 + bx + c, \quad g_\theta^{-1}(y) = \frac{-b - \sqrt{b^2 - 4a(c-y)}}{2a},$$

where $a = -1/2$, $b = -2/5$, and $c = 2$. Here g_θ intersects the axes at $(0, 3)$ and $(3, 0)$ and peaks at $x = -1$. We used the portion of the function with $x > 0$, where function is invertible. For each $i \in \{1, \ldots, n\}$, we first generated Z_i from Bernoulli(0.5). For $Z_i = 0$, we generated X_i from $U(0, 1)$ and then Y_i from the regression model $Y_i = g_\theta(X_i) + 0.5\varepsilon_i$, where $X_i \perp\!\!\!\perp \varepsilon_i$, $\varepsilon_i \sim N(0, 1)$. For $Z_i = 1$, we first generated Y_i from $U(0, 1)$ and then X_i from $X_i = g_\theta^{-1}(Y_i) + 0.5\delta_i$, where $Y_i \perp\!\!\!\perp \delta_i$, $\delta_i \perp\!\!\!\perp \varepsilon_i$, and $\delta_i \sim N(0, 1)$. Thus

$$\text{Model I:} \quad \begin{cases} Z \sim \text{Bernoulli}(0.5), \\ X|Z = 0 \sim U(0,1), \quad Y|Z = 1 \sim U(0,1), \\ Y|X, Z = 0 \sim N(g_\theta(X), 0.5^2), \quad X|Z = 0 \sim U(0,1), \\ X|Y, Z = 1 \sim N(g_\theta^{-1}(Y), 0.5^2), \quad Y|Z = 1 \sim U(0,1). \end{cases} \tag{7.1}$$

Table 3. Comparison of estimation error among four different estimators.

Model	n	OLS		SWAP		OSWAP		ASWAP	
		mean	sd	mean	sd	mean	sd	mean	sd
I	200	0.6058	0.2497	0.0500	0.0629	0.0500	0.0629	0.0493	0.0627
	300	0.5874	0.2167	0.0375	0.0508	0.0375	0.0508	0.0382	0.0516
	400	0.5615	0.1577	0.0268	0.0344	0.0268	0.0344	0.0270	0.0350
II	200	0.5973	0.3799	0.1356	0.1830	0.1066	0.1528	0.1133	0.1637
	300	0.6106	0.3084	0.0941	0.1328	0.0764	0.1101	0.0802	0.1158
	400	0.5647	0.2506	0.0683	0.0942	0.0524	0.0779	0.0534	0.0782

We estimated θ at sample sizes $n = 200, 300, 400$, and simulation sample size $n_{\text{sim}} = 200$. We used mean squared error to assess the accuracy of each estimator, and report the means and standard deviations of $(\hat{\theta}_1 - \theta)^2, \ldots, (\hat{\theta}_{200} - \theta)^2$ for each estimator and each sample size. The results are presented in the upper part of Table 3. We see that the SWAP, OSWAP, and ASWAP perform significantly better than OLS. SWAP and OSWAP are very close because, with conditional variances constant, the optimal estimating equation (5.5) coincides with SWAP in Section 2. ASWAP performs very similarly to SWAP and OSWAP even though the conditional variances are estimated, here the constant conditional variance is of the form ce^{ax} with $a = 0$.

Our second model was a heteroscedastic SWAP regression model where the conditional variances depended on X (or Y). With other settings the same, we had

$$\text{Model II}: \quad \begin{cases} Y|X, Z = 0 \sim N(g_\theta(X), [0.5(1+X)]^2), \\ X|Y, Z = 1 \sim N\Big(g_\theta^{-1}(Y), \Big[0.5\Big(1 + \frac{Y}{2}\Big)\Big]^2\Big). \end{cases}$$

The simulation results are presented in the lower part of Table 3. SWAP, OSWAP, and ASWAP all perform significantly better than OLS. OSWAP is no longer equivalent to SWAP, and OSWAP brings further reduction of estimation error. For ASWAP, we still used the conditional variance model ce^{ax} and de^{bx}, even though the true conditional variances were quadratic polynomials in x or y. Nevertheless, ASWAP brings appreciable reduction of estimation error as compared with SWAP — in fact, ASWAP is closer to OSWAP than to SWAP. This indicates that we only need to capture the ballpark shape of the conditional variances (e.g. increasing or decreasing in x or y) to improve the accuracy of the SWAP estimator.

8. Discussion

We have proposed SWAP regression for a design in which the roles of predictor and response trade places. We developed estimation and inference procedures for unweighted (SWAP) and weighted (OSWAP and ASWAP) estimating equations (or loss functions) for estimation, and a likelihood-ratio type criterion for hypothesis test. Motivated by the structure of a Kinesiology data set, we extended the SWAP estimator to accommodate an extra set of covariates that might affect the relation between X and Y.

This work can be extended in a variety of ways to adapt to other applications and to improve on performance. We outline some possibilities.

8.1. Quantile regression and other robust swap regression

In context of the heat-pressure psychrometric charts, it is natural to consider SWAP median or quantile regression, because it is of practical interest to know the percentage of a certain population (for example older people) that can tolerate certain levels of heat and humidity in a hot summer. Consider the objective function

$$Q(g) = E[\rho(Y, g(X))I(Z = 0) + \rho(X, g^{-1}(Y))I(Z = 1)]$$

among all functions in $\mathscr{G}$, where $\rho : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a general loss function that can be chosen to suit specific purposes. Thus, if $\rho(a, b) = |a - b|$, we have SWAP median regression, where $g(X)$ is the conditional median of Y given X, and $g^{-1}(Y)$ is the conditional median of X given Y. If

$$\rho(a, b) = \begin{cases} c_0(b - a) & \text{if } b > a, \\ c_1(b - a) & \text{if } b < a, \end{cases}$$

we have SWAP quantile regression, where the minimizer of $Q(g)$ is the $c_0/(c_0 + c_1)$th conditional quantile of Y given X, and its inverse the $c_0/(c_0 + c_1)$th conditional quantile of X given Y.

One could choose ρ for robust consideration, such as Huber's loss function (Huber (1964)),

$$\rho(a, b) = \begin{cases} \frac{(b-a)^2}{2} & \text{if } |b - a| \le c, \\ c(|b - a| - \frac{c}{2}) & \text{if } |b - a| > c, \end{cases}$$

to make the SWAP regression robust against outliers.

8.2. Vector-valued X and Y

SWAP regression can be extended to vector-valued X and Y, as follows. Let $X = (X_1, \ldots, X_p)$ and $Y = (Y_1, \ldots, Y_p)$ be p-dimensional random vectors supported on $\Omega_X \subseteq \mathbb{R}^p$ and $\Omega_Y \subseteq \mathbb{R}^p$, respectively. Let $\{g_\theta : \theta \in \Theta\}$ be a parametric family of bijections between Ω_X and Ω_Y. As before, let $Z = 0, 1$ be the design indicator. As in the 1-dimensional case, suppose

$$E_\theta(Y|X, Z = 0) = g_\theta(X), \quad E_\theta(X|Y, Z = 1) = g_\theta^{-1}(Y).$$

We can then construct the objective function as

$$Q(g) = E_n[\,\|Y - g_\theta(X)\|^2 I(Z=0)] + E_n[\,\|X - g_\theta^{-1}(Y)\|^2 I(Z=1)],$$

where $\|\cdot\|$ is the Euclidean norm.

The simplest example of p-dimensional bijection is the marginal bijective mapping $g_\theta(x) = (g_{\theta_1}^{(1)}(x_1), \ldots, g_{\theta_p}^{(p)}(x_p))$, where each $g_{\theta_i}^{(i)}(x_i)$ is a bijection from Ω_{X_i} to Ω_{Y_i}. A slightly more complicated p-dimensional bijection is marginal mappings imposed on linear indices of X and Y. Then we assume there exist nonsingular matrices $A, B \in \mathbb{R}^{p\times p}$ such that $U = AX$ and $V = BY$ and $(v_1, \ldots, v_p) = (g_{\theta_1}^{(1)}(u_1), \ldots, g_{\theta_p}^{(p)}(u_p))$, where, for each $i = 1, \ldots, p$, $g_{\theta_i}^{(i)}$ is a bijection between two subsets of $\mathbb{R}$. Depending on specific applications, we can treat A and B either as known or as parameters to be estimated together with $\theta = (\theta_1, \ldots, \theta_p)$.

8.3. Accounting for dependence

Another direction would extend SWAP regression to take into account the dependence in the data. In our data set, each subject is tested at several design points, which can introduce dependence in the regression error. One can account for the dependence either by introducing random effects, or by explicitly building the dependence into the error covariance structure, as one does in Generalized Estimating Equations (Liang and Zeger (1986)).

Acknowledgements

The authors are grateful to two referees and an associate editor for their many useful suggestions that helped to improve an earlier version of this paper. The research for Mosuk Chow is supported in part by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1RR033184. The research for Bing Li is supported in part Pennsylvania State University CTSI BERD Pilot Award, and a National Science Foundation grant (DMS-1106815). We thank Professor Larry Kenney at the Pennsylvania State University for providing the data that motivated this research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

Belding, H. S. and Kamon, E. (1972). Evaporative coefficients for prediction of safe limits in prolonged exposures to work under hot conditions. *Fed. Proc.* **32**, 1598-1601.

Bentler, P. M. and Xie, J. (2000). Corrections to test statistics in principal hessian directions. *Statist. Probab. Lett.* **47**, 381-389.

Chiou, J. M. and Müller, H.-G. (1998). Quasi-likelihood regression with unknown link and variance functions. *J. Amer. Statist. Assoc.* **93**, 1376-1387.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics.* Chapman & Hall/CRC.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.

Crowder, M. (1986). On consistency and inconsistency of estimating equations. *Econom. Theory* **2**, 305-330.

Drinkwater, B. L. and Horvath, S. M. (1979). Heat tolerance and aging. *Med. Sci. Sports* **11**, 49-55.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31** 1208-1211.

Godambe, V. P. and Heyde, C. C. (1987). Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* **53**, 19-29.

Godambe, V. P. and Thompson, M. E. (1989). An extension of quasi-likelihood estimation. *J. Statist. Plann. Inference* **22**, 137-152.

Hawkins-Bell, L. (1994). Heat-related deaths in Philadelphia and United States, 1993-1994. *Leads from the Morbidity and Mortality Weekly Report, Atlanta, GA* **43**, 453-455.

Heyde, C. C. (1997). *Quasi-Likelihood and its Application: a General Approach to Optimal Parameter Estimation.* Springer.

Huber, P. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73-101.

Kamon, E. and Avellini, B. (1976). Physiological limits to work in the heat and evaporation coefficient for women. *J. Appl. Physiol.* **41**, 71 - 76.

Kenny, W. L. and Anderson, R. K. (1988). Responses of older and younger women to exercise in dry and humid heat without fluid replacement. *Med. Sci. Sports Exerc.* **20**, 155-160.

Li, B. (1993). A deviance function for the quasi-likelihood method. *Biometrika* **80**, 741-753.

Li, B. (1996). A minimax approach to consistency and efficiency for estimating equations. *Ann. Statist.* **24**, 1283-1297.

Li, B. (1997). On the consistency of generalized estimation equations. *Selected Proceedings of the Symposium on Estimating Functions*, 108-129.

Li, B. (2000). Nonparametric estimating equations based on a penalized information criterion. *Canad. J. Statist.* **28**, 621-639.

Li, B. (2001). On quasi likelihood equations with nonparametric weights. *Scand. J. Statist.* **28**, 577-602.

Li, B. and McCullagh, P. (1994). Potential functions and conservative estimating functions. *Ann. Statist.* **22**, 340-356.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

Lind, A. R. (1963). A physiological criterion for setting thermal environmental limits for everyday work. *J. Appl. Physiol.* **18**, 51-56.

Lind, A. R. (1970). Effect of individual variation on upper limit of prescriptive zone climates. *J. Appl. Physiol.* **28**, 57-62.

McCullagh, P. (1983). Quasi-likelihood functions. *Ann. Statist.* **11**, 59-67.

McLeish, D. L. and Small, C. G. (1992). A projected likelihood function for semiparametric models. *Biometrika* **79**, 93-102.

Pandolf, K. H. (1997). Aging and human heat tolerance. *Exp. Aging Res.* **23**, 69-105.

Qu, A. P., Lindsay, B. and Li, B. (2000). Improving generalized estimating equations using quadratic inference functions. *Biometrika* **87**, 823-836.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika* **6**, 309-316.

Semenza, J. C., Rubin, C. H., Falter, K. H., Selanikio, J. D., Flanders, W. D, Howe, H. L. and Wilhelm, J. L. (1996). Heat-related deaths during the July 1995 heat wave in Chicago. *The New England J. Physiol.* **335**, 84-90.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20**, 595-601.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439-447.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9**, 60-62.

Zeman, M. J. (2001). Age and Sex-Specific Psychrometeric Limits and Critical Evaporative Coefficients for unacclimated subjects working in hot environments. A Thesis in Kinesiology at the Pennsylvania State University.

Department of Statistics, The Pennsylvania State University, 315 Thomas, University Park, PA 16802, USA.

E-mail: mchow@stat.psu.edu

Department of Statistics, The Pennsylvania State University, 410 Thomas, University Park, PA 16802, USA.

E-mail: bing@stat.psu.edu

Beijing Exhibition Hotel Office Building 5309, Xicheng District, Beijing, 100044 China.

E-mail: jackiepgh@sina.com

(Received April 2013; accepted May 2014)