<table>
<thead>
<tr>
<th>Statistica Sinica Preprint No: SS-13-092R2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Manuscript ID</strong></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
</tr>
</tbody>
</table>
| **Complete List of Authors** | Mosuk Chow  
Bing Li and  
Jackie Q. Xue |
| **Corresponding Author** | Li, Bing |
| **E-mail** | bing@stat.psu.edu |

Notice: Accepted version subject to English editing.
ON REGRESSION FOR SAMPLES WITH ALTERNATING PREDICTORS AND ITS APPLICATION TO PSYCHROMETRIC CHARTS

By Mosuk Chow, Bing Li, and Jackie Q. Xue

The Pennsylvania State University and Express Scripts China

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 is different from traditional regression because, for one part of the data, temperature was held fixed while pressure was raised to an equilibrium point; for the other part of the data, pressure was held fixed while temperature was 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. To handle this type of problems, 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, and develop test statistics for model selection and analysis of variance, and outline several extensions and refinements. We apply the method to the kinesiology data and found, 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). 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 problems. In this paper 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 are 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 are aged between 63-80, and the younger subjects between 18-30. For each subject 6 experiments were performed, among which three are under warm and humid conditions, to be called the $P_{crit}$ experiments, and three are 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 continues until a tolerance limit $T$ is 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 continues until a pressure tolerance limit $P$ is reached. The reason behind this peculiar design is that an experiment always starts at regions of pressure and temperature comfortable for the human subjects, and gradually increases one variable. An ordinary regression design would require some experiments to begin at either high temperature or high pressure, which is clearly undesirable. During each experiment, the subjects walked continu-
ously 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 how the data look like, we present in Figure 1 the portion of the data set for the old males. In the low-temperature zone, the temperature values were fixed and then the critical pressure values were observed correspondingly. In other words, pressure values were random numbers observed from subpopulation with temperature fixed. Therefore, in the upper-left part, the 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 ordinary regression analysis is clearly inadequate: for example, at the bottom of the chart, the observations lie almost entirely to the left of the curve.

Figure 1: ULPZ data for old males. Blue dots represent the cases where temperature is held fixed while pressure is gradually increased. Red dots represent 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 the quadratic polynomial as the regression function.
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 they represent the temperature or the pressure. Since the regression is designed for \textit{Samples With Alternating Predictors}, we will call it the 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, 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 make some further discussions on the general paradigm of SWAP regression and outline how it can be broadened to adapt to a variety of applications. All proofs are relegated to an online Appendix.

2 SWAP regression

Our 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 now 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

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

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

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

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

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

over $\mathcal{G}$. The following theorem shows that the conditional expectation $f(x) = E(Y|X = x, Z = 0)$ is the unique minimizer of this objective function. We use $\equiv$ to indicate mutually absolute continuity of two measures.

**Theorem 1** If $\text{var}(Y_1) < \infty$, $\text{var}(X_2) < \infty$, and $f \in \mathcal{G}$, then $f$ minimizes $Q(g)$ over $\mathcal{G}$. Furthermore, suppose

$$P_{X|Z=0} \equiv P_X, \ P_{Y|Z=1} \equiv P_Y.$$  

Then $f$ is the (almost surely) unique minimizer of $Q(g)$ over $\mathcal{G}$.

Now let $(X_1, Y_1, Z_1), \ldots, (X_n, Y_n, Z_n)$ be an i.i.d. sample of $(X, Y, Z)$. Let $E_n(\cdot)$ denote the sample average (i.e. the 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 defined as the minimizer of $Q_n(g)$ over $\mathcal{G}$.

In this paper we shall focus on parametric models. Let $\Theta \subseteq \mathbb{R}^p$, and let

$$\mathcal{G} = \{g_\theta(\cdot) : \theta \in \Theta\}.$$ 

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 following optimization problem:

minimizing $Q_n(g_\theta)$ over $\Theta$, subject to $g_\theta$ being injective.

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

For example, consider the family of 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 $X_1, \ldots, X_n$; that is, $-b/(2a) \notin [X_{(i)}, X_{(n)}]$, where $X_{(i)}$ is the $i$th order statistic.

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, Chapter 12), and van der Vaart (1998, Chapter 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 $\theta$. Let

$$q((X, Y, Z), \theta) = -2[Y - g_\theta(X)][\partial g_\theta(X)/\partial \theta] I(Z = 0) - 2[X - g_\theta^{-1}(Y)][\partial g_\theta^{-1}(Y)/\partial \theta] I(Z = 1).$$

(3)

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)
Consider the following estimating equation

\[ \frac{\partial Q_n(g_\theta)}{\partial \theta} = E_n q((X, Y, Z), \theta) = 0. \]  \tag{4} 

Assuming that there is a \( \theta_0 \) in the interior of \( \Theta_0 \) 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 \( \Theta_0 \) is a solution to (4). Let \( U \) abbreviate \((X, Y, Z)\) and \( Q_n(\theta) \) abbreviate \( Q_n(g_\theta) \).

**Theorem 2** Suppose (i) \( \theta_0 \) belongs to the interior of \( \Theta_0 \); (ii) the function \( \theta \mapsto Eq(U, \theta) \) is differentiable at \( \theta_0 \) under the integral sign, and \( E[\partial q(U, \theta_0)/\partial \theta^T] \) is positive definite; (iii) for a sufficiently small \( \delta > 0 \),

\[ E \left( \sup_{\|\theta - \theta_0\| = \delta} ||q(U, \theta)|| \right) < \infty; \]

and (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

We now derive the asymptotic distribution of the solution, say \( \hat{\theta} \), to \( E_n q(U, \theta) = 0 \). Again, by standard asymptotic theory for estimating equations, we have

\[ \sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)), \]  \tag{5} 

where

\[ J(\theta) = E(\partial q(U, \theta)/\partial \theta^T), \quad I(\theta) = E(q(U, \theta)q^T(U, \theta)). \]

See, for example, Heyde (1997, Section 12.4). In our context, the matrices \( I(\theta) \) and \( J(\theta) \) can be written more specifically as

\[ J(\theta) = 2E[(\partial g_\theta(X)/\partial \theta)(\partial g_\theta(X)/\partial \theta^T)I(Z = 0)] + 2E[(\partial g_\theta^{-1}(Y)/\partial \theta)(\partial g_\theta^{-1}(Y)/\partial \theta^T)I(Z = 1)], \]

\[ I(\theta) = 4E \left[ (Y - g_\theta(X))^2(\partial g_\theta(X)/\partial \theta)(\partial g_\theta(X)/\partial \theta^T)I(Z = 0) \right] + 4E \left[ (X - g_\theta^{-1}(Y))^2(\partial g_\theta^{-1}(Y)/\partial \theta)(\partial g_\theta^{-1}(Y)/\partial \theta^T)I(Z = 1) \right]. \]  \tag{6} 


We now summarize this result into a theorem, with appropriate regularity conditions.

**Theorem 3** Suppose that \( \hat{\theta} \) is a consistent solution to the equation \( E_n q(U, \theta) = 0 \). Suppose, furthermore:

1. \( \theta_0 \in \text{int}(\Theta_0) \);
2. \( E q(U, \theta_0) = 0 \); the entries of the matrix \( E[q(U, \theta_0)q^T(U, \theta_0)] \) are finite and the matrix itself is positive definite;
3. the function \( E q(U, \theta) \) is differentiable under the integral sign, and the entries of \( E[\partial q(U, \theta)/\partial \theta^T] \) are integrable;
4. the sequence of random matrices \( \{ E_n \partial q(U, \theta)/\partial \theta^T : n \in \mathbb{N} \} \) is stochastically equicontinuous.

Then \( \sqrt{n} (\hat{\theta} - \theta_0) \) has asymptotic distribution (5) with \( J(\theta_0) \) and \( I(\theta_0) \) given by (6).

To use this theorem in practice, we replace \( \theta_0 \) in (6) by \( \hat{\theta} \) and the expectations in \( I(\theta) \) and \( J(\theta) \) by sample averages.

We can check stochastic equicontinuity in Assumption 4 as follows. Let \( h_n(\theta) \) denote the random function \( \text{vec}[E_n q(U, \theta)/\partial \theta^T] \). 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

\[
\limsup_{n \to \infty} P \left( \sup_{\|\theta - \theta_0\| < \delta} \| h_n(\theta) - h_n(\theta_0) \| > \epsilon \right) < \eta.
\]

A sufficient condition for this is, in a neighborhood \( G \) of \( \theta_0 \),

\[
\sup_{\theta \in G} \| h_n(\theta) - h_n(\theta_0) \| \leq M_n(U_1, \ldots, U_n) \| \theta - \theta_0 \| \quad (7)
\]

for some \( M_n(U_1, \ldots, U_n) = O_p(1) \). Thus all we need to do is to bound \( \| \partial h_n(\theta)/\partial \theta^T \| \) in the vicinity of \( \theta_0 \) by a function of the order \( O_p(1) \) that does not depend on \( \theta \).
As an example, consider the linear case where \( g_\theta(x) = \theta x \), where the true \( \theta_0 > 0 \). In this case
\[
h_n(\theta) = 2E_n(X^2|Z = 0) - 4\theta^{-3}E_n[(X - \theta^{-1}Y)Y|Z = 1] + 2\theta^{-4}E_n(Y^2|Z = 1).
\]
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 \( t \leq 1 + t^2 \) and \( (a + b)^2 \leq 3(a^2 + b^2) \), the above is no greater than
\[
12\theta^{-4}[1 + E_n((X - \theta^{-1}Y)^2|Z = 1)](1 + E_n(Y^2|Z = 1)) + 12\theta^{-5}E_n(Y^2|Z = 1)
\]
\[
\leq 12\theta^{-4}[1 + 3E_n(X^2|Z = 1) + 3\theta^{-2}E_n(Y^2|Z = 1)](1 + E_n(Y^2|Z = 1))
\]
\[
+ 12\theta^{-5}E_n(Y^2|Z = 1).
\]
Let \( G = (a, b) \) be a neighborhood of \( \theta_0 \) with \( a > 0 \) and \( b < \infty \). Then, on \( G \),
\[
|\frac{\partial h_n(\theta)}{\partial \theta}| \text{ is no greater than }
\]
\[
12a^{-4}[1 + 3E_n(X^2|Z = 1) + 3a^{-2}E_n(Y^2|Z = 1)](1 + E_n(Y^2|Z = 1))
\]
\[
+ 12a^{-5}E_n(Y^2|Z = 1).
\]
Thus, if we assume \( X \) and \( Y \) have finite fourth moments then the above is of the order \( O_p(1) \) and hence (7) holds. Obviously, this type of arguments can be carried out for more complicated functions \( g_\theta(x) \).

### 3.3 Asymptotic distribution for hypothesis testing

We now develop a likelihood-ratio type statistic for testing general hypotheses. Let \( h : \Theta \to \mathbb{R}^r \) \((r \leq p)\), and consider the general hypotheses of the form
\[
H_0 : h(\theta) = 0 \quad \text{versus} \quad H_1 : h(\theta) \neq 0.
\]

(8)
For example, if \( \theta = (\phi^T, \eta^T)^T \) and we are interested in testing \( \phi = 0 \), then \( h \) reduces to the coordinate projection \( \theta \mapsto \phi \). Let

\[
\hat{\theta} = \text{argmin}\{Q_n(\theta) : \theta \in \Theta\}, \quad \tilde{\theta} = \text{argmin}\{Q_n(\theta) : \theta \in \Theta, \ h(\theta) = 0\}.
\]

Mimicking the Wilks’ statistic (Wilks, 1938; Cox and Hinkley, 1974, page 92), we propose the following statistic

\[
T_n = 2n[Q_n(\tilde{\theta}) - Q_n(\hat{\theta})]. \tag{9}
\]

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, as we will see, 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 denote

\[
H(\theta) = \partial h^T(\theta)/\partial \theta.
\]

**Theorem 4** Suppose the conditions in Theorem 3 hold and \( h \) is differentiable. Then, under \( H_0 \) in (8),

\[
T_n \overset{p}{\to} \sum_{i=1}^p \lambda_i K_i, \tag{10}
\]

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

\[
\Sigma(\theta) = I^{1/2}(\theta)J^{-1}(\theta)H(\theta)[H^T(\theta)J^{-1}(\theta)H(\theta)]^{-1}H^T(\theta)J^{-1}(\theta)I^{1/2}(\theta). \tag{11}
\]

Note that, because \( H(\theta) \) is dimension \( p \times r \), the matrix \( \Sigma(\theta) \) has rank \( r \). Hence the last \( r \) eigenvalues of \( \Sigma(\theta) \), the \( \lambda_{p-r+1}, \ldots, \lambda_p \) in (10), are zero.

In practice, the eigenvalues of \( \lambda_1, \ldots, \lambda_p \) can be estimated by those of the matrix (11), where \( \theta \) 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 using one of the following two methods. The first is by simulating from the distribution \( \hat{\lambda}_1 K_1 + \cdots + \hat{\lambda}_p K_p \), where \( K_1, \ldots, K_p \) are i.i.d. \( \chi^2_{\nu(i)} \), and using the percentage
of random numbers generated from this distribution greater than $T_n$ as the $p$-value. The second is to use the approximation introduced by Bentler and Xie (2000). That is, let

$$
\hat{T}_n = \frac{\text{tr}(\Sigma_n^2(\hat{\theta}))}{\text{tr}(\Sigma_n(\hat{\theta}))},
$$

where $\Sigma_n(\hat{\theta})$ with $E(\cdot)$ replaced by $E_n(\cdot)$ and $\theta$ replaced by $\hat{\theta}$. Then $\hat{T}_n$ is approximately distributed as $\chi^2_d$ where $d$ is the nearest integer to $\frac{\text{tr}(\Sigma_n(\hat{\theta}))^2}{\text{tr}(\Sigma_n^2(\hat{\theta}))}$. The first method works well for $p$-value in the range $\geq 0.01$ with 10000 simulated random numbers. For smaller $p$-values first method requires generating a large number of random numbers, whereas the second sample is very convenient and works surprisingly well. See also Satterthwaite (1941).

### 4 Analysis of covariance

As described in the Introduction, our data set consists of four groups: young females, old females, young males, and old males. After fitting the SWAP regressions to the sub-samples and comparing the psychrometric charts, we observed that the charts of the old-female group and the old-male group are similar, whereas those for young-female group and the young-male group 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 now develop a hypothesis test procedure for this problem.

Consider $m$ independent samples

$$
(X_1^{(1)}, Y_1^{(1)}, Z_1^{(1)}), \ldots, (X_{n_1}^{(1)}, Y_{n_1}^{(1)}, Z_{n_1}^{(1)}) \overset{i.i.d.}{\sim} (X^{(1)}, Y^{(1)}, Z^{(1)}),
$$

$$
\vdots
$$

$$
(X_1^{(m)}, Y_1^{(m)}, Z_1^{(m)}), \ldots, (X_{n_m}^{(m)}, Y_{n_m}^{(m)}, Z_{n_m}^{(m)}) \overset{i.i.d.}{\sim} (X^{(m)}, Y^{(m)}, Z^{(m)}).
$$
In our application, \( m = 2 \), representing two gender groups. Suppose, for the \( k \)th sample,
\[
E_{g(k)}(Y^{(k)}|X^{(k)}) = g_{g(k)}(X^{(k)}), \quad E_{g(k)}(X^{(k)}|Y^{(k)}) = g_{g(k)}^{-1}(Y^{(k)}).
\]  
(14)

We would like to test if the \( m \) SWAP regression curves are the same, which can be formulated as testing
\[
H_0 : \theta^{(1)} = \cdots = \theta^{(m)} \quad \text{versus} \quad H_1 : \theta^{(1)}, \ldots, \theta^{(m)} \text{ are not all equal.} \quad (15)
\]

More generally, letting \( h : \Theta \times \cdots \times \Theta \rightarrow \mathbb{R}^s \) be a differentiable function, where \( s \) is a positive integer no greater than \( mp \), we consider the hypotheses
\[
H_0 : h(\theta^{(1)}, \ldots, \theta^{(m)}) = 0 \quad \text{versus} \quad H_1 : h(\theta^{(1)}, \ldots, \theta^{(m)}) \neq 0. \quad (16)
\]

Let \( n = n_1 + \cdots n_m \), and define the following objective function
\[
\mathcal{Q}_n(\theta^{(1)}, \ldots, \theta^{(m)}) = n_1Q_n^{(1)}(\theta^{(1)})/n + \cdots + n_mQ_n^{(m)}(\theta^{(m)})/n,
\]
(17)

where \( Q_n^{(k)} \) is the objective function defined in Section 3.3 applied to the \( k \)th sample. That is,

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

Let \( (\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)}) \) be the global maximizer of (17) over \( \Theta \times \cdots \times \Theta \) and let \( \check{\theta} \) be the constrained minimizer of (17) subject to \( h(\theta^{(1)}, \ldots, \theta^{(m)}) = 0 \). We propose the statistic
\[
T_n = 2n[\mathcal{Q}_n(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)}) - \mathcal{Q}_n(\check{\theta}^{(1)}, \ldots, \check{\theta}^{(m)})]
\]
to test the hypotheses (15).

For \( k = 1, \ldots, m \), let \( I^{(k)}(\theta^{(k)}) \) and \( J^{(k)}(\theta^{(k)}) \) be the matrices \( I(\theta) \) and \( J(\theta) \) defined in section 3.3 as applied to the \( k \)th subpopulation:
\[
J^{(k)}(\theta^{(k)}) = 2E_{g(k)}[(\partial g_{g(k)}(X^{(k)})/\partial \theta^{(k)})(\partial g_{g(k)}(X^{(k)})/\partial \theta^{(k)})^T I(Z^{(k)} = 0)]
\]
\[
+ 2E_{g(k)}[(\partial g_{g(k)}^{-1}(Y^{(k)})/\partial \theta^{(k)})(\partial g_{g(k)}^{-1}(Y^{(k)})/\partial \theta^{(k)})^T I(Z^{(k)} = 1)],
\]
\[
I^{(k)}(\theta^{(k)}) = 4E_{g(k)}[\sigma^2_{g(k)}(X^{(k)})/(\partial g_{g(k)}(X^{(k)})/\partial \theta^{(k)})(\partial g_{g(k)}(X^{(k)})/\partial \theta^{(k)})^T I(Z^{(k)} = 0)]
\]
\[
+ 4E_{g(k)}[\tau^2_{g(k)}(Y^{(k)})/(\partial g_{g(k)}^{-1}(Y^{(k)})/\partial \theta^{(k)})(\partial g_{g(k)}^{-1}(Y^{(k)})/\partial \theta^{(k)})^T I(Z^{(k)} = 1)],
\]

where \( \sigma^2_{g(k)}(X^{(k)}) = \text{var}_{g(k)}(X^{(k)}|X^{(k)}) \) and \( \tau^2_{g(k)}(Y^{(k)}) = \text{var}_{g(k)}(X^{(k)}|Y^{(k)}) \).
Theorem 5 Suppose

1. the assumptions (13) and (14) hold, and the m samples in (13) are independent;
2. the assumptions in Theorem 3 are satisfied for each of the m subpopulations;
3. for each $k = 1, \ldots, m$, $\lim_{n \to \infty} (n_k/n) = \alpha_k$ for some $0 < \alpha_k < 1$.

Then

$$T_n \xrightarrow{D} \sum_{i=1}^{s} \lambda_i K_i,$$

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

$$I^{1/2}J^{-1}H(H^TJ^{-1}H)^{-1}H^TJ^{-1}I^{1/2}$$

(18)

where

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

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

In practice, the $\theta$ in $I$ and $J$ is replaced by either $(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)})$ or $(\tilde{\theta}^{(1)}, \ldots, \tilde{\theta}^{(m)})$. The expectations $E(\cdots)$ in $I$ and $J$ 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 the analysis of variance hypotheses (15), we set $h : \Theta \times \cdots \times \Theta \to \mathbb{R}^{m(p-1)}$ as

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

Thus the gradient $H$ is the $mp \times m(p-1)$ dimensional matrix

$$\begin{pmatrix} I_p & 0 \\ -I_p & \ddots \\ \vdots & \ddots & I_p \\ 0 & \cdots & -I_p \end{pmatrix}.$$
The conditional variances $\sigma^2_{\theta(k)}(X^{(k)})$ and $\tau^2_{\theta(k)}(Y^{(k)})$, since they appear within (unconditional) expectations, can be replaced by

$$(Y^{(k)} - g_{\theta(k)}(X^{(k)}))^2, \quad (X^{(k)} - g_{\theta(k)}(Y^{(k)}))^2,$$

respectively. Also note that the rank of the matrix (18) is $s$, so that $\lambda_1, \ldots, \lambda_s$ therein are the only nonzero eigenvalues.

We can further generalize the above ANCOVA model to accommodate arbitrary covariates (not necessarily categorical as in the above ANCOVA model) that might affect the relation between $X$ and $Y$. In addition to $U = (X, Y, Z)$, let $V \in \mathbb{R}^k$ be a (continuous or discrete) random (or nonrandom) vector that may affect the individual shapes of the relation between $X$ and $Y$. Suppose $X$ and $Y$ are connected by a family of bijections depending on $v$:

$$E(Y|X = x, Z = 0, V = v) = g_{\theta(v)}(x),$$
$$E(X|Y = y, Z = 1, V = v) = g_{\theta(v)}^{-1}(y),$$

where $\theta$ 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 $\eta$. The objective function is now generalized 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

$$[Y_i - g_{f(V, \eta)}(X_i)I(Z_i = 0)]^2, \quad i = 1, \ldots, n.$$

The vectors $V_1, \ldots, V_n$ can either be random or fixed. If $V$ is fixed, conditional expectations such as $E(Y|X = x, Z = 0, V = v)$ should be interpreted as $E(Y|X = x, Z = 0)$ for a fixed $v$; the notation $E_n(\cdot)$ still mean sample average in this case.
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, which 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 following estimating equation

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

where $a_\theta : \Omega_X \to \mathbb{R}^+$ and $b_\theta : \Omega_Y \to \mathbb{R}^+$ are weight functions. Under the assumption

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

the estimating equation (19) is unbiased (Godambe, 1960). That is,

$$E_\theta[a_\theta(X)(Y - g_\theta(X))I(Z = 0) + b_\theta(Y)(X - g_\theta^{-1}(Y))I(Z = 1)] = 0.$$ (20)

Also note that (19) is a generalization of the estimating equation (3), in which the weight functions are of the special forms

$$a_\theta(X) = -2\partial g_\theta(X)/\partial \theta, \quad b_\theta(Y) = -2\partial g_\theta^{-1}(Y)/\partial \theta.$$

Denote the function inside $[\cdots]$ in (20) as $G_\theta(X, Y, Z)$. By the standard theory of estimating equations, if $\hat{\theta}$ is a consistent solution to (20), then

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{p} N(0, J^{-1}(a_\theta, b_\theta)I(a_\theta, b_\theta)J^{-T}(a_\theta, b_\theta)),$$
where
\[
J(a, b) = E_\theta [\partial G_\theta(X, Y, Z)/\partial \theta^T],
\]
\[
I(a, b) = E_\theta [G_\theta(X, Y, Z)G_\theta^T(X, Y, Z)].
\] (21)

See, for example, Li (1993, 1996), and Li and McCullagh (1994), and Heyde (1997). Here, to emphasize the dependence of \( I \) and \( J \) on the weight functions, we write them as \( I(a, b) \) and \( J(a, b) \) instead of \( I(\theta) \) and \( J(\theta) \), as were used previously. It is then natural to choose the weighting functions \( a_\theta(X), b_\theta(Y) \) that minimize the asymptotic variance
\[
AV(a, b) = J^{-1}(a, b)I(a, b)J^{-T}(a, b).
\]

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 (1973), McCullagh (1983), Godambe and Heyde (1987), Godambe and Thompson (1989), and Small and McLeish (1993). We now use this approach to derive an optimal estimating equation for SWAP regression.

To simplify notation, let
\[
\beta_\theta(X, Y) = (a_\theta(X), b_\theta(Y)),
\]
\[
\delta_\theta(X, Y) = \text{diag}(Y - g_\theta(X), X - g_\theta^{-1}(Y)),
\]
\[
c(Z) = \text{diag}(I(Z = 0), I(Z = 1)).
\] (22)

Then the estimating equation (19) becomes
\[
E_n[\beta_\theta(X, Y)c(Z)\delta_\theta(X, Y)] = 0.
\]

Let
\[
\gamma_\theta(X, Y) = (\partial g_\theta(X)/\partial \theta, \partial g_\theta^{-1}(Y)/\partial \theta), \quad v_\theta(X, Y) = \text{diag}(\text{var}_\theta(Y|X), \text{var}_\theta(X|Y)).
\]

**Theorem 6** Let \( \mathcal{W} \) be the family of weighting functions \( \beta_\theta(x, y) = (a_\theta(x), b_\theta(y)) \) such that
1. the entries of matrix $E_0[\beta_0(X, Y)c(Z)\gamma_0^T(X, Y)]$ are finite and the matrix is nonsingular;

2. the entries of matrix $E_0[\beta_0(X, Y)c(Z)v_0(X, Y)\beta_0^T(X, Y)]$ are finite and the matrix is nonsingular.

If the weighting function $\beta_0^*(X, Y) = \gamma_0(X, Y)v_0^{-1}(X, Y)$ belongs to $\mathcal{W}$, then

$$AV(\beta_0^*) \leq AV(\beta_0).$$

for all $\beta_0 \in \mathcal{W}$ and all $\theta \in \Theta$, where $A \leq B$ means $B - A$ is positive semidefinite.

This theorem implies that the solution to the estimating equation

$$E_n[(\partial g_0(X)/\partial \theta)(Y - g_0(X))I(Z = 0)/\text{var}_n(Y|X, Z = 0) + (\partial g_0^{-1}(Y)/\partial \theta)(X - g_0^{-1}(Y))I(Z = 1)/\text{var}_n(X|Y, Z = 1)] = 0,$$

has the smallest variance among all estimating equations of the form (19). We call this optimal estimator the optimal SWAP, or OSWAP.

To use this optimal estimating equation we need the conditional variance functions $\sigma_n^2(X) = \text{var}_n(Y|X, Z = 0)$ and $\tau_n^2(Y) = \text{var}_n(X|Y, Z = 1)$. We can either choose parametric models for them, as is done in the original quasi likelihood method, or to estimate them nonparametrically, as in Chiou and Müller (1999) and Li (2001). Here, we propose an easy-to-implement parametric estimator of the conditional variance that is quite effective for our purpose. Consider

$$\text{var}_n(Y|X = x, Z = 0) = ce^{ax}.$$

Note that $c$ and $a$ here generally depend on $\theta$, but we suppress this dependence from the notation 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 in Section 2. 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 the $L_1$ criterion:

$$E_n(||(Y - g_{\tilde{\theta}}(X))^2 - c_n(a)e^{ax}|| | Z = 0)$$

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 $c_n(\hat{a})e^{ax}$ as the estimator of $\text{var}_Y(Y|X = x, Z = 0)$. Let $d_n(\hat{b})e^{by}$ be the parallel estimator of $\text{var}_Y(X|Y = y, Z = 1)$. Substituting these into (23), we arrive at the following adaptively weighted estimating equation

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

We call the solution to this equation the adaptive SWAP regression estimator, or ASWAP.

The reason for choosing an $L_1$- instead of an $L_2$-criterion is that the latter involves the fourth moments and tends to be unstable. As shown in the simulation studies in Section 7, this ASWAP performs very well, even when the true conditional variance is not of the form $ce^{ax}$ (or $de^{by}$).

6 Data Analysis

We now apply SWAP regression to analyze the data described in the Introduction. The full data set contains four sub-samples: young females (with
sample size \( n = 51 \), young males \((n = 62)\), old females \((n = 52)\), old males \((n = 52)\). As we have explained, each subject is repeatedly observed at 5 to 6 design points. For example, for the old male group, the 52 pairs of observations are contributed by 9 subjects. The data originally contain 227 pairs of temperature and pressure values, in which 10 pairs contain missing numbers. To simplify analysis we treat all pairs as independent and delete the pairs with missing data. A more careful analysis should take care of possible dependence caused by repeated observations and possible bias caused by discarding missing cases, but in this paper we ignore these issues to focus on the central issue of alternating design. This simplified treatment does not seem to seriously impede the clarity of analysis for this particular data.

We use the quadratic model

\[ g_\theta(x) = ax^2 + bx + c, \]

where \( \theta = (a, b, c)^T \). By an inspection and the physical meaning of the data, it is reasonable to restrict \( g_\theta \) to monotone decreasing functions over the range \([X_{(1)}, X_{(n)}]\). This leads to constraint

\[ \Theta = \{(a, b, c) : 2aX_{(n)} + b < 0 \}. \] (24)

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

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

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

Let \( X \) represent the temperature, \( Y \) represent the pressure, \( Z = 0 \) represents the cases where \( X \) is the predictor, and \( Z = 1 \) represents the cases where \( Y \) is the predictor. The SWAP objective function takes the form

\[
Q_n(g_\theta) = E_n[(Y - aX^2 - bX - c)^2I(Z = 0)] \\
+ E_n[X + (b + \sqrt{b^2 - 4a(c - Y)}/(2a)^2)I(Z = 1)].
\]

This is to be minimized subject to the constraint (24), but in all cases the global minimizers occur within the interior of the set (24), so that we never needed to actually evoke constrained minimization.
We applied SWAP regression to all four samples, and the fitted curves are presented in Figure 2, with young females in the upper-left panel, old females in the upper-right panel, young males in the lower-left panel, and old males in the lower-right panel. As in Figure 1, the blue dots represent the cases for which temperature is the predictor, and the red dots represent the cases for which pressure is the predictor. Comparing with the least-squares fit (for old males) in Figure 1, where the regression line rather poorly fits 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.

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.
It is interesting to note that the downward-bending quadratic tendencies for old people seem to be stronger than those for younger people. We now use the method developed in Section 3.3 to test the hypothesis

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

for each of the 4 groups. The results are presented in the Table 1, in which the columns 2 and 3 are the weights for the weighted \( \chi^2 \)'s in Theorem 4, with \( \theta_0 \) is replaced by \( \hat{\theta} \) calculated from the quadratic model, and column 4 shows the values of the test statistic \( T_n \) defined in (9). The \( p \)-values are computed using two methods described in the last paragraph of section 3.3, with the \( p \)-values calculated by simulation appearing column 5 and those calculated by the approximation (12) appearing in column 6. Note that in our context the matrix (11) has rank 2, and so \( \hat{\lambda}_3 \) is identically 0.

Table 1: Significance of downward bending tendencies

<table>
<thead>
<tr>
<th>group id</th>
<th>( \hat{\lambda}_1 )</th>
<th>( \hat{\lambda}_2 )</th>
<th>( T_n )</th>
<th>( p )-value</th>
<th>( p )-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>young female</td>
<td>13.058</td>
<td>6.785</td>
<td>12.800</td>
<td>0.517</td>
<td>0.525</td>
</tr>
<tr>
<td>old female</td>
<td>38.220</td>
<td>7.918</td>
<td>69.130</td>
<td>0.219</td>
<td>0.221</td>
</tr>
<tr>
<td>young male</td>
<td>12.069</td>
<td>9.970</td>
<td>0.722</td>
<td>0.971</td>
<td>0.968</td>
</tr>
<tr>
<td>old male</td>
<td>30.752</td>
<td>10.425</td>
<td>119.351</td>
<td>0.062</td>
<td>0.055</td>
</tr>
</tbody>
</table>

We see that the quadratic component (downward-bending tendency) for old male group is significant at level \( \alpha = 0.1 \). The \( p \)-values for the old female group is nonsignificant, but is much smaller than those from the two young groups. Significantly, the downward-bending tendency for the old male group cannot be observed from the traditional quadratic regression fit (Figure 1).

Comparing the curves in Figure 2 we also observe that the tolerance curves of old men and women are much more similar to each other than those for young men and women, where women have much higher overall tolerance levels. To confirm this observation, we now use the procedure developed in Section 4 to test the hypotheses

\[ H_0 : \theta^{(1)} = \theta^{(2)} \quad \text{versus} \quad H_1 : \theta^{(1)} \neq \theta^{(2)}, \]
where $\theta^{(1)}$ and $\theta^{(2)}$ are the parameter $(a, b, c)$ for the old female and old male groups, respectively. In this case, $H = (I_3, -I_3)^T$. Table 2 presents the result of this test.

<table>
<thead>
<tr>
<th>comparison</th>
<th>$T_n$</th>
<th>$p$-value</th>
<th>$p$-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>young female vs young male</td>
<td>89.98</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>old female vs old male</td>
<td>812.43</td>
<td>0.260</td>
<td>0.269</td>
</tr>
</tbody>
</table>

The difference between young female group and young male group is very significant; but there is no significant difference between old female group and old male group, indicating that the gender effect diminishes with age.

7 Simulation comparisons

In this section compare the performances of four methods for analyzing simulated data with the alternating design: the 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 consider two SWAP regression models. The first one is a homoscedastic SWAP regression model where the conditional variance $\text{var}(Y|X, Z = 0)$ and $\text{var}(X|Y, Z = 1)$ are constants. Let

$$g_a(x) = ax^2 + bx + c, \quad g_a^{-1}(y) = \left[-b - \sqrt{b^2 - 4a(c - y)}\right]/(2a),$$

where $a = -1/2$, $b = -2/5$, and $c = 2$. This is a downward-facing parabola that intersects to two axes at $(0, 3)$ and $(3, 0)$ and peaks at $x = -1$. We use the portion of the function with $x > 0$. In this region the function is invertible with its inverse shown above. For each $i \in \{1, \ldots, n\}$, we first generate $Z_i$ from Bernoulli(0.5). If $Z_i = 0$, we generate $X_i$ from $U(0, 1)$ and then $Y_i$ from the regression model

$$Y_i = g_a(X_i) + 0.5\epsilon_i,$$
where $X_i \perp \varepsilon_i$, $\varepsilon_i \sim N(0, 1)$. If $Z_i = 1$, we first generate $Y_i$ from $U(0, 1)$ and then $X_i$ from

$$X_i = g_\theta^{-1}(Y_i) + 0.5\delta_i,$$

where $Y_i \perp \delta_i$, $\delta_i \perp \varepsilon_i$, and $\delta_i \sim N(0, 1)$. To summarize, the first SWAP regression model is specified by the following distributions:

\begin{align*}
\text{Model I:} & \\
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{align*}

(25)

We estimate $\theta$ using the four methods for sample sizes $n = 200, 300, 400$, and simulation sample size $n_{\text{sim}} = 200$. We use mean squared error to assess the accuracy of each estimator. Specifically, let $\hat{\theta}_1, \ldots, \hat{\theta}_{200}$ be the estimates by one of the four procedures for the 200 simulation samples, we report the mean (mean in Table 3) and standard deviation (sd in Table 3) 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. This is because OLS is not consistent for the alternating design. We note that SWAP and OSWAP are numerically identical. This is because when the conditional variances are constant and optimal estimating equation (23) coincides with SWAP in Section 2. Also note that ASWAP performs very similarly to SWAP and OSWAP even though the conditional variances are estimated. In this case, the constant conditional variance belongs to the parametric family $ae^{ax}$ with $a = 0$.

Our second model (Model II) is a heteroscedastic SWAP regression model where the conditional variances depends on $X$ (or $Y$). That is, we make the
following change to (25):

Model II: \[
\begin{align*}
Y | X, Z = 0 & \sim N(g_0(X), [0.5(1 + X)]^2), \\
X | Y, Z = 1 & \sim N(g_1^{-1}(Y), [0.5(1 + Y/2)]^2).
\end{align*}
\]

The simulation results are presented in the lower part of Table 3. As in Model I, SWAP, OSWAP, and ASWAP all perform significantly better than OLS. Because of heteroscedasticity, OSWAP is no longer equivalent to SWAP, and OSWAP brings further reduction of estimation error. For ASWAP, we still use the conditional variance model \(ce^{-x}x\) and \(de^{-y}y\), even though they are no longer the true conditional variance models, which are 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.

### Table 3: Comparison of estimation error among four different estimators

<table>
<thead>
<tr>
<th>Model</th>
<th>(n)</th>
<th>OLS</th>
<th>SWAP</th>
<th>OSWAP</th>
<th>ASWAP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>sd</td>
<td>mean</td>
<td>sd</td>
<td>mean</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>200</td>
<td>.6058</td>
<td>.2497</td>
<td>.0500</td>
<td>.0629</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>.5874</td>
<td>.2167</td>
<td>.0375</td>
<td>.0508</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>.5615</td>
<td>.1577</td>
<td>.0268</td>
<td>.0344</td>
</tr>
<tr>
<td>II</td>
<td>200</td>
<td>.5973</td>
<td>.3799</td>
<td>.1356</td>
<td>.1830</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>.6106</td>
<td>.3084</td>
<td>.0941</td>
<td>.1328</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>.5647</td>
<td>.2506</td>
<td>.0683</td>
<td>.0942</td>
</tr>
</tbody>
</table>

### 8 Discussions

In this paper we propose a new type of regression, the SWAP regression, for a special design where the roles of predictor and response trade places. We developed estimation and inference procedures for this type of problems. We
focussed on the 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 special structure of a Kinesiology data set, we also extended the SWAP estimator to accommodate an extra set of covariates that may affect the relation between $X$ and $Y$.

This new type of regression can be further extended in a variety of ways to adapt to other applications and to improve its performance. We now outline several such 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 old people) that can tolerate certain levels of heat and humidity in a hot summer. More generally, we consider the following objective function

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

among all functions in $\mathcal{G}$, where $\rho: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is a general loss function that can be chosen to suit specific purposes. For example, if we choose $\rho(a, b) = |a - b|$, then 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 we choose

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

then 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 can also choose $\rho$ for robust consideration, such as Huber’s loss func-
tion (Huber, 1964):
\[
\rho(a, b) = \begin{cases} 
(1/2)(b - a)^2 & \text{if } |b - a| \leq c \\
 c(|b - a| - c/2) & \text{if } |b - a| > c
\end{cases}
\]
to make the SWAP regression robust against outliers.

8.2 Vector-valued X and Y

SWAP regression can also 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 \(\Omega_X\) and \(\Omega_Y\). As before, let \(Z = 0, 1\) be the design indicator such that, when \(Z = 0\), \(X\) is the predictor, and when \(Z = 1\), \(Y\) is the predictor. 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 by mimicking the one-dimensional case, as follows:
\[
Q(g) = E_\theta[\|Y - g_\theta(X)\|^2I(Z = 0)] + E_\theta[\|X - g_\theta^{-1}(Y)\|^2I(Z = 1)],
\]
where \(\| \cdot \|\) is the Euclidean norm.

The simplest example of \(p\)-dimensional bijection is the marginal bijective mapping; that is,
\[
g_\theta(x) = (g_{\theta_1}^{(1)}(x_1), \ldots, g_{\theta_p}^{(p)}(x_p)),
\]
where each \(g_{\theta_i}^{(i)}(x_i)\) is a bijection from \(\Omega_{X_i}\) to \(\Omega_{Y_i}\). A slightly more complicated \(p\)-dimensional bijection is marginal mappings imposed on linear indices of \(X\) and \(Y\). That is, 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 for further development, which is especially relevant to the data we considered, is to 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 may 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. In particular, they led to the extensions to multivariate case (Section 8.2) and to the situations where a set of arbitrary covariates affect individual SWAP regression curves (second part of Section 4). 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 would like to thank Professor Larry Kenney at the Pennsylvania State University for providing the data that motivate this research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References


