Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 68(2) 2014 131-144

Influence of Measures of Significance based Weights in the Weighted Lasso

Tanya P. Garcia[1] and Samuel Müller[2]
[1]*School of Public Health, Texas A&M Health Science Center, College Station, TX 77843-1266*
[2]*School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia*

Received 12 June 2013; Revised 24 November 2013; Accepted 26 November 2013

SUMMARY

When part of the regressors can act on both the response and some of the other explanatory variables, the already challenging problem of selecting variables in a $p > n$ context becomes more difficult. A recent methodology for variable selection in this context links the concept of q-values from multiple testing to the weighted Lasso. In this paper, we show that different informative measures of significance to q-values, such as partial correlation coefficients or Benjamini-Hochberg adjusted p-values, give similarly promising performance as when using q-values.

Keywords: Adjusted p-values, Convex optimization, Partial correlation coefficients, q-values, Variable selection, Weighted Lasso.

1. INTRODUCTION

A common issue in modeling biological and social phenomena is variable selection, where data in such fields often involve more predictor variables than observations. A new difficulty to this already challenging task is selecting predictor variables in a structured way so that an existing hierarchy among the model variables is obeyed. For example, some predictors may be known to act on both the response and other candidate predictors; thus, one must select which candidate variables affect the response after accounting for those predictors known to affect the response. Recently, Garcia *et al*. (2013a) proposed a novel method for handling such structured variable selection problems. Their method involves extracting q-values in multiple hypothesis testing (Storey 2003) and using them as weights in the weighted Lasso (Zou 2006) to appropriately direct the selection procedure. In this paper, we take a closer look at their proposed method and, through various simulation studies, we determine if weights other than the q-values could improve the procedure.

Structured variable selection is needed, for example, in modeling risk factors of childhood obesity. An emerging hypothesis in childhood obesity is that obesity is not only linked to excess caloric intake and inadequate physical activity (Hill and Peters 1998, Anderson *et al*. 1998, Heindel and vom Saal 2009), but also to exposure to endocrine disrupting chemicals which can alter hormones that control weight gain (Baillie-Hamilton 2002, Grun and Blumberg 2006, La Merrill and Birnbaum 2011). There is evidence to suggest that diet is correlated with potential weight gain and with endocrine disrupting chemicals. Thus, a key

Corresponding author: Samuel Müller
E-mail address: samuel.mueller@sydney.edu.au

interest is selecting which endocrine disrupting chemicals affect weight gain (response variable) after accounting for those predictors (*e.g.*, diet) known to affect weight gain.

Structured variable selection is also needed when modeling the relationship between the gut microbiome and features related to body weight regulation. Previous biological experiments warrant that diet is known to regulate body weight (Bray *et al*. 2012) and may alter some microbial groups (Abnous *et al*. 2009 and Li *et al*. 2009). A biological interest is determining which microbes have an effect on body weight regulation after accounting for diet. With such information, scientists can potentially develop targeted therapies to better regulate body weight by altering specific microbes. However, obtaining this information is difficult. One must account for the intricate relationship among microbes, diet, and phenotypes of body weight regulation, and perform the variable selection using data where the number of microbes measured far exceeds the sample size.

Performing structured variable selection is a challenge which extends beyond what earlier selection procedures can handle. These methods include the Lasso (Tibshirani 1996) and its extensions (Yuan and Lin 2006, Zou 2006, Meinshausen and Bühlmann 2010), least angle regression (Efron *et al*. 2004), and selection by controlling the false discovery rates (Benjamini and Hochberg 1995, Storey 2003). To remedy this gap in the literature, Garcia *et al*. (2013a) developed a method that modifies the weights in the weighted Lasso (Zou 2006) such that certain variables are ensured to be in the final model, and that important candidate variables are selected over less important ones. Their method provides a proper multivariate analysis by collectively considering all relevant information in the model variables, and ultimately results in selections with acceptable false positive rates and low false discovery rates. This contrasts from individually assessing which predictors are related to the response through simple measures of correlations or partial correlations.

Our aim in this paper is to explore how the weights in the method of Garcia *et al*. (2013a) could be further improved and generalized. The rest of the paper is organized as follows. Section 2 provides a brief overview of the Garcia *et al*. (2013a) method, along with additional weights that may lead to improved performance. Section 3 evaluates the different weights through various simulation studies. In Section 4, we apply the weighted Lasso with the different proposed weights to a microbiome study. Section 5 concludes the paper. Technical arguments and additional numerical results are provided in the Appendix.

2. MAIN RESULTS

2.1 Motivation for Modified Weighted Lasso

Let the sample size be n, $\mathbf{y} = (y_1, \dots y_n)^T$ be the response variable, and $\mathbf{v}_j$, $j = 1, \dots, m$ denote the $n \times 1$ covariates that are linearly related to $\mathbf{y}$. The covariates are divided into two groups: those that need to be included in the model (*i.e.*, designed covariates), and those that are subject to selection. For ease in presentation, we refer collectively to all covariates as $\mathbf{v}$'s, whereas we denote the designed covariates as $\mathbf{z}$'s and covariates subject to selection as $\mathbf{x}$'s. Specifically, we let the number of designed covariates be m_0, and denote them as $\mathbf{v}_1 := \mathbf{z}_1, \dots, \mathbf{v}_{m_0} := \mathbf{z}_{m_0}$. We also let the number of covariates subject to selection be m_1, and denote them as $\mathbf{v}_{m_0+1} := \mathbf{x}_1, \dots, \mathbf{v}_{m_0+m_1} := \mathbf{x}_{m_1}$. We have that $m = m_0 + m_1$. Without loss of generality, we assume all variables are standardized to have mean zero and sample variance one, so that the intercept is excluded from the regression model. We also suppose that $m_0 + 1 \le n$, but we allow for $m_0 + m_1 = m > n$.

For the $m > n$ variable selection problem for the linear model, a commonly used method is the weighted Lasso (Zou 2006) which minimizes

$$Q(\beta) = \frac{1}{2}\left\| \mathbf{y} - \sum_{k=1}^{m} \mathbf{v}_k \beta_k \right\|^2 + \lambda \sum_{k=1}^{m} w_k |\beta_k|, \quad (1)$$

with respect to $\beta = (\beta_1, \dots, \beta_m)^T$, the vector of regression coefficients. Here, $\lambda > 0$ is a regularization parameter, $w_k > 0$, $k = 1, \dots, m$, are weights, and $\|\cdot\|$ denotes the L_2- norm. We denote the minimizer as $\hat{\beta}$, which will be a function of λ and the weights $\mathbf{w} = (w_1, \dots, w_m)^T$.

To gain insight into the minimizer $\hat{\beta}$, let $\mathbf{r}_{(-k)} = \mathbf{y} - \Sigma_{j \neq k} \mathbf{v}_j \beta_j$, $k = 1, \dots, m$, denote the partial residual after removing the k^{th} covariate. Through a careful derivation involving subgradients (see Appendix A.1), we have that if

$$\left| \mathbf{v}_k^T \mathbf{r}_{(-k)} \right| \le \lambda w_k, \quad (2)$$

then $\hat{\beta}_k = 0$; otherwise,

$$\hat{\beta}_k = \text{sign}(\mathbf{v}_k^T \mathbf{r}_{(-k)})(|\mathbf{v}_k^T \mathbf{r}_{(-k)}| - \lambda w_k) / \mathbf{v}_k^T \mathbf{v}_k, \quad \text{for } k = 1, \ldots, m.$$

From (2), it is apparent that for a fixed λ, variables $\mathbf{v}_k$ with large weights will generally not be included in the model (*i.e*., $\hat{\beta}_k = 0$); whereas, those variables with small weights will generally be included in the model (*i.e*., $\hat{\beta}_k \neq 0$). Using this key property, Garcia *et al*. (2013a) extended the work of Charbonnier *et al*. (2010) and Bergersen *et al*. (2011) by formulating a method so that important variables were included in the model before less important ones.

2.2 Selection of Weights

The method of Garcia *et al*. (2013a) is based on choosing weights so that the designed covariates **z**'s are ensured to be in the model, and that variables subject to selection, **x**'s, are selected according to their significance in the model after accounting for the **z**'s. Candidate weights for **z** and **x** are now described.

2.2.1 Weights for Designed Covariates

To ensure that the **z**'s are included in the model requires that the inequality in (2) does not hold. One way to guarantee this is to set $w_1 = \ldots w_{m_0} = 0$. In our simulations we will set $w_j = 10^{-5}$ on $\mathbf{z}_j$, $j = 1, \ldots, m_0$, to explore to what extent such small weight values lead to the exclusion of the **z**'s.

2.2.2 Weights for Covariates Subject to Selection

Garcia *et al*. (2013a) showed that weights for the **x**'s should measure the significance of each **x** on **y** after accounting for the designed covariates **z**'s. Weights that ignore the effect of the **z**'s are actually inferior to unit weights on the **x**'s; see the simulation study of Garcia *et al*. (2013a). Appropriate weights can thus be based on (i) measures of the partial correlation between $\mathbf{x}_k$, $k = 1, \ldots, m_1$, and **y** after accounting for **z**'s (*i.e*, $\rho_{\mathbf{x}_k, \mathbf{y}|\mathbf{z}_1, \ldots, \mathbf{z}_{m_0}}$), and (ii) measures of the effect of $\mathbf{x}_k$ in the linear regression of **y** on $(\mathbf{z}_1, \ldots, \mathbf{z}_{m_0}, \mathbf{x}_k)$ Specifically, we consider five different types of weights:

1. *Inverted Absolute Partial Correlations:* $w_{m_0+k} = 1/|\rho_{\mathbf{x}_k, \mathbf{y}|\mathbf{z}_1, \ldots, \mathbf{z}_{m_0}}|$ on $\mathbf{x}_k$, $k = 1, \ldots, m_1$, where $\rho_{\mathbf{x}_k, \mathbf{y}|\mathbf{z}_1, \ldots, \mathbf{z}_{m_0}}$ are the partial correlations between $\mathbf{x}_k$ and **y** after controlling for $\mathbf{z}_1, \ldots, \mathbf{z}_{m_0}$. Influential **x**'s will thus have small weights, as desired, since an influential $\mathbf{x}_k$ has large absolute partial correlation and hence small inverted absolute partial correlation. Conversely, less influential **x**'s will have large weights since their inverted absolute partial correlation values will be large.

2. *Inverted Absolute t-Statistics*: $w_{m_0+k} = 1/|t_k|$ on $\mathbf{x}_k$, $k = 1, \ldots, m_1$ where $t_k = \hat{\beta}_k^* / se(\hat{\beta}_k^*)$ are the t-statistics obtained from the individual linear regressions of **y** on $(\mathbf{z}_1, \ldots, \mathbf{z}_{m_0}, \mathbf{x}_k)$, and $\hat{\beta}_k^*$ are the estimated coefficient associated with $\mathbf{x}_k$. Influential **x**'s will thus have small weights, as desired, since an influential $\mathbf{x}_k$ has a large $|t_k|$ value and hence a small $1/|t_k|$ value. Conversely, a less influential $\mathbf{x}_k$ will have a large $1/|t_k|$ value.

3. *p-Values*: $w_{m_0+k} = p_k$ *on* $\mathbf{x}_k$, $k = 1, \ldots, m_1$ where p_k are the p-values obtained from the individual linear regressions of **y** on $(\mathbf{z}_1, \ldots, \mathbf{z}_{m_0}, \mathbf{x}_k)$. A statistically significant $\mathbf{x}_k$ tends to have a small p-value and a non-statistically significant $\mathbf{x}_k$ has a large p-value. Thus, weighing each $\mathbf{x}_k$ with its corresponding p-value will generally lead to including statistically significant **x**'s in the final model.

4. *Benjamini-Hochberg (BH) Adjusted p-Values*: $w_{m_0+k} = p_k^{BH}$ on $\mathbf{x}_k$, $k = 1, \ldots, m_1$, where p_k^{BH} are the Benjamini-Hochberg (Benjamini and Hochberg 1995) adjusted p-values obtained from the individual linear regressions of **y** on $(\mathbf{z}_1, \ldots, \mathbf{z}_{m_0}, \mathbf{x}_k)$. In comparison to p-values, the BH adjusted p-value accounts for the multiplicity of the m_1 tests compared from the m_1 linear regressions. Still, the impact of BH adjusted p-values is similar to that for p-values since a statistically significant $\mathbf{x}_k$ will have a small BH adjusted p-value even after the adjustment, and a statistically non-significant $\mathbf{x}_k$ will have a large BH adjusted p-value.

5. *q-Values*: $w_{m_0+k} = q_k$ on $\mathbf{x}_k$, $k = 1, \ldots, m_1$, where q_k are the q-values (Storey and Tibshirani 2003) obtained from the individual linear regressions of

$\mathbf{y}$ on $(\mathbf{z}_1, ..., \mathbf{z}_{m_0}, \mathbf{x}_k)$. Similarly to adjusted p-values, q-values are a monotone transformation of p-values designed to control the false discovery rate (FDR): the number of false positives found among rejected null hypotheses. As with p-values and BH adjusted p-values, covariates with small q-values are generally inuential to the model, whereas covariates with large q-values are not. We estimate q-values using $\hat{q}(p_{(m_1)}) = \widehat{FDR}(p_{(m_1)})$ and $\hat{q}(p_{(k)}) = \min\{\widehat{FDR}(p_{(k)}), \hat{q}(p_{(k+1)})\}$ for $k = m_1 - 1, m_1 - 2, ..., 1$. Here, $p_{(1)}, ..., p_{(m_1)}$ are the ordered p-values and $\widehat{FDR}(t)$ is the estimated false discovery rate based on rejecting null hypotheses with p-values $\leq t$, $0 \leq t \leq 1$; see Storey and Tibshirani (2003) for the exact form of $\widehat{FDR}(t)$ and the R package "qvalue" (Dabney *et al*. 2011) for an implementation to compute q-values.

The proposed method of Garcia *et al.* (2013a) only focused on q-values as weights for the $\mathbf{x}$'s, but we consider the additional weights listed above. In Section 3, we demonstrate the influence of these weights on the performance of the weighted Lasso.

2.3 Implementation of Weighted Lasso

In practice, the weighted Lasso (*i.e*., minimizing $Q(\beta)$ in (1)) is solved using a least angle regression (LARS) algorithm (Efron *et al*. 2004) which provides the entire sequence of model fits in the Lasso path, along with estimated parameter coefficients. The best descriptive model among all those in the Lasso path is the one that minimizes the penalized loss function

$$M_n(\delta, p^*) = \text{SSE}_{p^*} / \hat{\sigma}^2 - n + \delta p^*. \quad (3)$$

Here, $\delta > 0$, p^* is the number of predictors in the selected model, SSE_{p^*} is the residual sum of squares for the selected model, and $\hat{\sigma}^2$ is an appropriate estimator of the model error variance. For example, when $n > p^*$, $\hat{\sigma}^2$ can be the residual mean square when using all available variables, and when $n < p^*$, $\hat{\sigma}^2$ can be the sample variance of the response vector $\mathbf{y}$ (Hirose *et al*. 2013). Finding the minimizer $\hat{p}^*$ in (3) is equivalent to minimizing $\text{SSE}_{p^*} + \delta\hat{\sigma}^2 p^*$, thus estimating σ^2 well is of little concern.

An important detail of (3) is the choice of δ which controls the penalty on the number of predictors, p^*, in the selected model. Large δ values will inflate the effect of p^* so that minimizing the penalized loss function will require having fewer predictors in the model. Conversely, small δ values will minimize the effect of having many predictors in the model. Consequently, different δ values yield different model fits and observed false discovery rates. Garcia *et al.* (2013a) proposed a modified cross-validation procedure to appropriately select δ; in this paper, we set $\delta = 1$, estimate σ^2 by var($\mathbf{y}$) and focus on the choice of weights instead.

3. SIMULATION STUDIES

3.1 Simulation Design

We evaluated the performance of the different weighting schemes on a simulation study similar to that in Garcia *et al*. (2013a) and one that mimics the real data in Section 4. We supposed there were two treatment groups with 20 subjects in each, and generated $m_1 + 1$ explanatory variables as follows. First, we generated a binary treatment indicator $\mathbf{z}$ where for each subject $i = 1, ..., 40$, $z_i = I(i > 20) - I(i \leq 20)$. Then we generated $\mathbf{x}_k = (x_{1,k}, ..., x_{40,k})^T$, $k = 1, ..., m_1$, such that the first 75% of $\mathbf{x}$'s depend on the treatment indicator $\mathbf{z}$, and the remainder do not. Specifically, for $k = 1, ..., 0.75m_1$, we set $x_{ik} = u_{ik} + z_i s_k$, where u_{ik} were independent uniform (0, 1) random variables and s_k were independent uniform (0.25, 0.5). For $k = 0.75m_1 + 1, ..., m_1$, we generated x_{ik} as independent uniform (0, 1) random variables, with no dependence on the treatment z_i. We generated the response vector as

$$\mathbf{y} = \beta_0\mathbf{z} + \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \beta_3\mathbf{x}_3 + \sum_{k=4}^{m_1-1} \beta_k\mathbf{x}_k + \beta_{m_1}\mathbf{x}_{m_1} + \varepsilon, \quad (4)$$

where ε is normally distributed with mean 0 and covariance $\sigma^2 I$. In summary, $\mathbf{x}_1, ..., \mathbf{x}_{m_1}$ were generated according to four distinct categories:

Group 1. $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ depend on treatment and act on $\mathbf{y}$ even after taking into account treatment;

Group 2. $\mathbf{x}_4, ..., \mathbf{x}_{0.75m_1}$ depend on treatment and do not act on $\mathbf{y}$;

Group 3. $\mathbf{x}_{0.75m_1+1}, ..., \mathbf{x}_{m_1-1}$ neither depend on treatment, nor act on $\mathbf{y}$;

Group 4. $\mathbf{x}_{m_1}$ does not depend on treatment, but acts on $\mathbf{y}$.

To evaluate the performance of the different weights in the weighted Lasso, we considered two different parameter settings. We set $m_1 = 40$, $\sigma^2 = 0.5$ with $\beta = (4.5, 3, -3, -3, \mathbf{0}^T, 3)^T$ and $\beta = (2.5, 1.5, -1.5, -1.5, \mathbf{0}^T, 1.5)^T$ where $\mathbf{0}^T$ is an $(m_1 - 4)$ dimensional vector of zeros. In both settings there are more parameters than observations and 90% of the $\mathbf{x}$'s are redundant. We now present results when $\beta = (4.5, 3, -3, -3, \mathbf{0}^T, 3)^T$, and defer results for $\beta = (2.5, 1.5, -1.5, -1.5, \mathbf{0}^T, 1.5)^T$ to Appendix A.2 because these results are similar except that the variable selection percentages are lower.

Under each parameter setting, we generated 1000 independent data sets, and applied the weighted Lasso with the proposed weights in Section 2. We report the averaged percentages of time variables in each group were selected, and the observed false discovery rate (FDR). The ideal weighting scheme will largely select variables in Groups 1 and 4, while not selecting variables in Groups 2 and 3, and thus have small FDR.

3.2 Simulation Results

From Table 1, we observe that across all weighting schemes, the weighted Lasso has a high rate of true positives and an acceptable false positive rate. Most interestingly, we observed that weights based on the inverted absolute partial correlations and inverted absolute t-statistics equally selected the same percentage of variables in each group. This is because close inspection of many simulation runs showed that the correlation between these two weights was very close to one and corresponding scatterplots showed an almost perfect linear relationship between these two weight vectors. Likewise, weights based on any of the p-values (either with or without adjustment) led to similar results in the variable selection. In fact, both the BH-adjusted p-values and q-values led to exactly the same percentages of selection for each group. This suggests that the weighted Lasso is somewhat robust against monotone transformations of the weights, since q-values and BH adjusted p-values are monotone transformations of p-values. We also note that for each of the four variable groups, the average BH-adjusted p-value is nearly 1.25 times the average q-value (see Table 2). Likewise, the 25%, 50% and 75% quantiles in each group for both the BH-adjusted p-values and q-values have similar proportions.

Table 1. Results from 1000 simulations when $\beta = (4.5, 3, -3, -3, \mathbf{0}^T, 3)^T$. Averaged percentages of time variables in each group were selected and observed false discovery rate (FDR). Ideal weighted Lasso will largely select variables in Groups 1 and 4, and not select variables in Groups 2 and 3.

Average Variables Selection					
Weights	$\lvert\rho_{\mathbf{x},\mathbf{y}\lvert\mathbf{z}}\rvert^{-1}$	$\lvert t\rvert^{-1}$	p	p^{BH}	q
Treatment	100.00	100.00	100.00	100.00	100.00
Group 1	75.53	75.10	73.00	73.03	73.03
Group 2	0.39	0.36	0.27	0.29	0.29
Group 3	1.88	1.60	0.44	0.50	0.50
Group 4	85.70	84.80	75.70	77.80	77.80
FDR	0.08	0.07	0.04	0.04	0.04

Table 2. Summary statistics of weights when $\beta = (4.5, 3, -3, -3, \mathbf{0}^T, 3)^T$. Average weight (Mean), and 25%, 50%, and 75% quantile of weights from each group. Ideal weighting will have small weights in Groups 1 and 4, and large weights in Groups 2 and 3.

Weights		Group 1	Group 2	Group 3	Group 4
$\lvert\rho_{\mathbf{x},\mathbf{y}\lvert\mathbf{z}}\rvert^{-1}$	Mean	2.55	43.97	49.39	2.51
	$Q_{0.25}$	2.01	5.60	6.19	2.51
	$Q_{0.5}$	2.22	9.31	10.25	2.51
	$Q_{0.75}$	2.93	19.79	21.65	2.51
$\lvert t\rvert^{-1}$	Mean	0.38	7.22	8.11	0.37
	$Q_{0.25}$	0.28	0.91	1.00	0.37
	$Q_{0.5}$	0.32	1.52	1.68	0.37
	$Q_{0.75}$	0.45	3.25	3.55	0.37
p	Mean	0.03	0.50	0.50	0.03
	$Q_{0.25}$	0.01	0.27	0.30	0.03
	$Q_{0.5}$	0.01	0.50	0.50	0.03
	$Q_{0.75}$	0.04	0.73	0.70	0.03
p^{BH}	Mean	0.14	0.78	0.78	0.15
	$Q_{0.25}$	0.06	0.71	0.72	0.15
	$Q_{0.5}$	0.09	0.83	0.83	0.15
	$Q_{0.75}$	0.19	0.90	0.89	0.15
q	Mean	0.11	0.63	0.64	0.12
	$Q_{0.25}$	0.05	0.58	0.59	0.12
	$Q_{0.5}$	0.07	0.68	0.68	0.12
	$Q_{0.75}$	0.15	0.74	0.73	0.12

To better visualize the similarities between the different weighting schemes, we generated boxplots of the weights for each group (see Fig. 1). To standardize the weights to similar units of measurement, we divided each of the five sets of weights by their corresponding overall median. Both figures immediately show the striking similarities between the weights based on inverted absolute partial correlations and inverted absolute t-statistics. It also shows striking similarities between the p-values and its adjusted versions. In each case, the variability and spread within each group is practically the same. Though the weights are

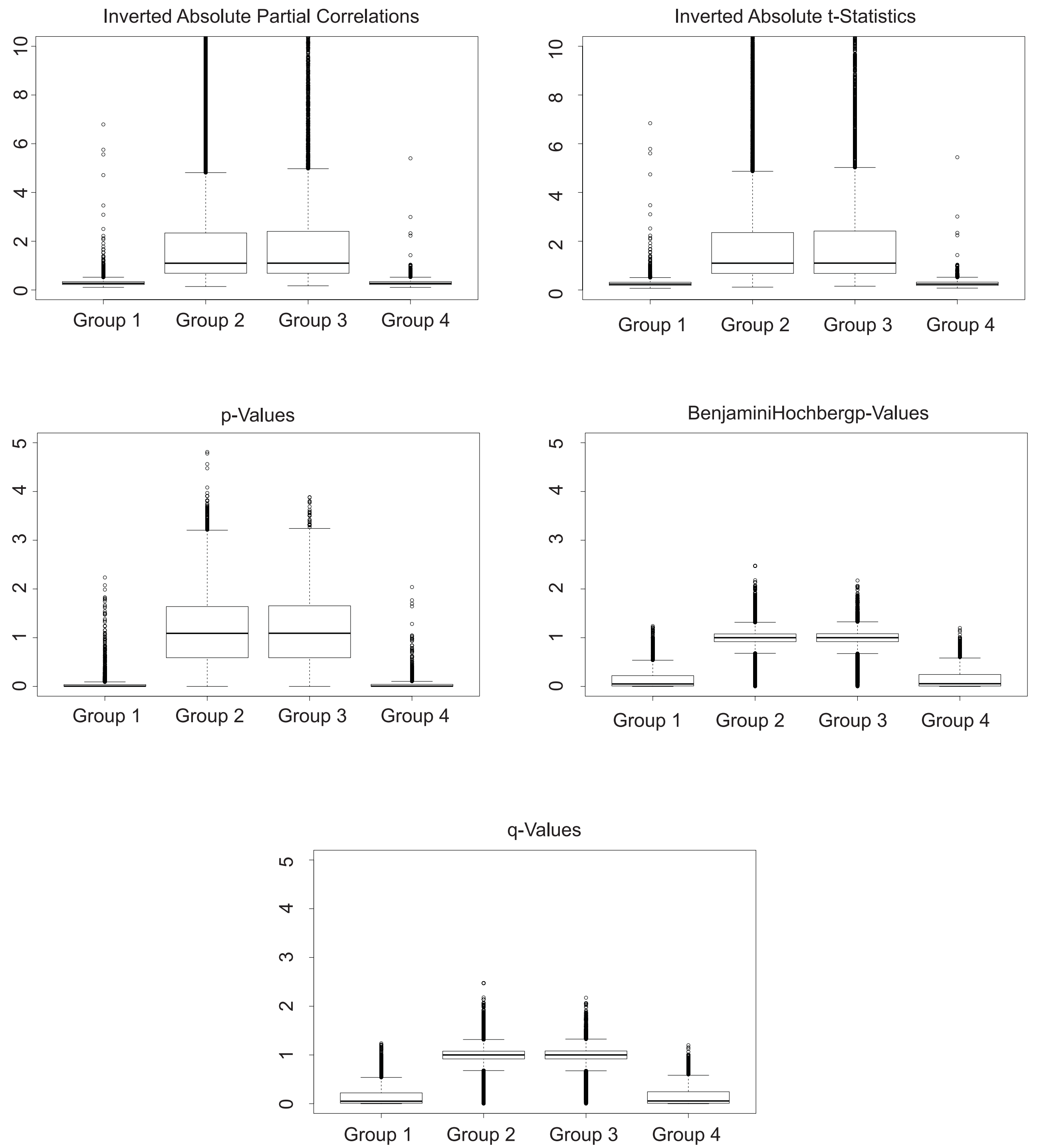

Fig. 1. Boxplots of weights in each group for different weighting types when $\boldsymbol{\beta} = (4.5, 3, -3, -3, \mathbf{0}^T, 3)^T$. All weights have been standardized by the median over all groups.

fundamentally different, their similar features within each group (*i.e.*, variability and spread) is what drives the similar variable selection.

Overall, we observed that weights based on p-values without any adjustment led to the most reasonable results. It selected variables in Groups 1 and 4 most frequently, and variables in Groups 2 and 3 least frequently, and hence, had a low false discovery rate. This suggests that when using the weighted Lasso for variable selection, no transformation of the p-value based weights is necessary. This greatly contrasts from using adjusted p-values to select significant variables where transformation is needed to account for the compounded error in multiple hypothesis testing.

We performed a second simulation study, where in addition to generating variables as previously described, we generated a variable $\mathbf{x}_1^*$ that was correlated with $\mathbf{x}_1$ after adjusting for $\mathbf{z}$, *i.e.*, $\text{corr}(\mathbf{x}_1, \mathbf{x}_1^*) = 0.8$, and such that $\mathbf{x}_1^*$ does not act on $\mathbf{y}$. In this case then, the response variable becomes $\mathbf{y} + \beta_{m_1+2}\mathbf{x}_1^*$ and $\beta_{m_1+2} = 0$. The additional covariate now corresponds to Group 5, defined as

Group 5: $\mathbf{x}_1^*$ is correlated with $\mathbf{x}_1$ after adjusting for treatment, but does not act on $\mathbf{y}$.

Under this setup, we now consider parameter values as $\beta = (4.5, 3, -3, -3, \mathbf{0}^T, 3, 0)^T$ and $\beta = (2.5, 1.5, -1.5, -1.5, \mathbf{0}^T, 1.5, 0)^T$ where $\mathbf{0}^T$ is an $(m_1 - 4)$-dimensional vector of zeros. (Results for the latter β are displayed in Appendix A.2). Here, the ideal weighted Lasso will choose variables in Groups 1 and 4, while not selecting variables in Groups 2, 3, and 5.

From Table 3, we see that the weighted Lasso with weights based on the (adjusted) p-values had a slight edge over weights based on inverted absolute partial correlations or inverted absolute t-statistics. In this case, the weighted Lasso with weights based on the (adjusted) p-values incorrectly chose variables in Group 5 half as often as did the weighted Lasso with weights based on inverted absolute partial correlations or t-statistics. A possible reason for this is that for the weighted Lasso with (adjusted) p-values as weights, the average weight in Group 5 was at least 3.2 times larger than the average weight in Group 1 (see Table 4). In comparison, for the other two weighting schemes, the average weight in Group 5 was at most 2.7 times larger

Table 3. Simulation results from 1000 simulations when $\beta = (4.5, 3, -3, -3, \mathbf{0}^T, 3, 0)^T$. Averaged percentages of time variables in each group were selected and observed false discovery rate (FDR). Ideal weighted Lasso will largely select variables in Groups 1 and 4, and not select variables in remaining groups.

Average Variables Selection					
Weights	$\lvert\rho_{\mathbf{x},\mathbf{y}\mid\mathbf{z}}\rvert^{-1}$	$\lvert t\rvert^{-1}$	p	p^{BH}	q
Treatment	100.00	100.00	100.00	100.00	100.00
Group 1	71.93	71.93	69.93	70.17	70.17
Group 2	0.46	0.43	0.28	0.29	0.29
Group 3	2.00	1.64	0.48	0.57	0.57
Group 4	84.30	83.70	74.30	75.70	75.70
Group 5	8.50	8.10	4.60	4.80	4.80
FDR	0.11	0.10	0.05	0.06	0.06

Table 4. Summary statistics of weights when $(4.5, 3, -3, -3, \mathbf{0}^T, 3, 0)^T$. Average weight (Mean), and 25%, 50%, and 75% quantile of weights from each group. Ideal weighting will have small weights in Groups 1 and 4, and large weights in Groups 2 and 3.

Weights		Group 1	Group 2	Group 3	Group 4	Group 5
$\lvert\rho_{\mathbf{x},\mathbf{y}\mid\mathbf{z}}\rvert^{-1}$	Mean	3.20	41.75	71.45	2.44	8.13
	$Q_{0.25}$	2.02	5.62	6.17	2.44	8.13
	$Q_{0.5}$	2.23	9.46	10.06	2.44	8.13
	$Q_{0.75}$	3.89	20.16	20.68	2.44	8.13
$\lvert t\rvert^{-1}$	Mean	0.49	6.85	11.74	0.36	1.31
	$Q_{0.25}$	0.29	0.91	1.00	0.36	1.31
	$Q_{0.5}$	0.33	1.55	1.64	0.36	1.31
	$Q_{0.75}$	0.60	3.31	3.39	0.36	1.31
p	Mean	0.03	0.50	0.50	0.03	0.18
	$Q_{0.25}$	0.01	0.27	0.30	0.03	0.18
	$Q_{0.5}$	0.01	0.50	0.49	0.03	0.18
	$Q_{0.75}$	0.04	0.74	0.70	0.03	0.18
p^{BH}	Mean	0.13	0.77	0.77	0.12	0.44
	$Q_{0.25}$	0.06	0.70	0.71	0.12	0.44
	$Q_{0.5}$	0.09	0.83	0.82	0.12	0.44
	$Q_{0.75}$	0.19	0.90	0.89	0.12	0.44
q	Mean	0.11	0.63	0.62	0.10	0.36
	$Q_{0.25}$	0.05	0.57	0.57	0.10	0.36
	$Q_{0.5}$	0.07	0.67	0.66	0.10	0.36
	$Q_{0.75}$	0.15	0.73	0.72	0.10	0.36

than the average weight in Group 1. In a sense then, with inverted absolute partial correlations or t-statistics as weights, the weighted Lasso had trouble determining whether Group 5 should be included or not. Consequently the weighted Lasso with weights based on inverted absolute partial correlations or t-statistics had larger FDR than the other weights based on (adjusted) p-values.

As with our first simulation study without Group 5, we see similar behavior between weights based on partial correlations or t-statistics, and between weights based on (adjusted) p-values (see Fig. 2). In both cases, the variability and spread within each group are again similar. In terms of low false discovery rates and overall performance, the weighted Lasso with p-values as weights performs well in this more challenging situation of Group 5 being included. The results of our simulation study suggest that the work in the weighted Lasso of Garcia *et al*. (2013a) can be simplified: one only needs to compute individual p-values, without the additional task of making adjustments.

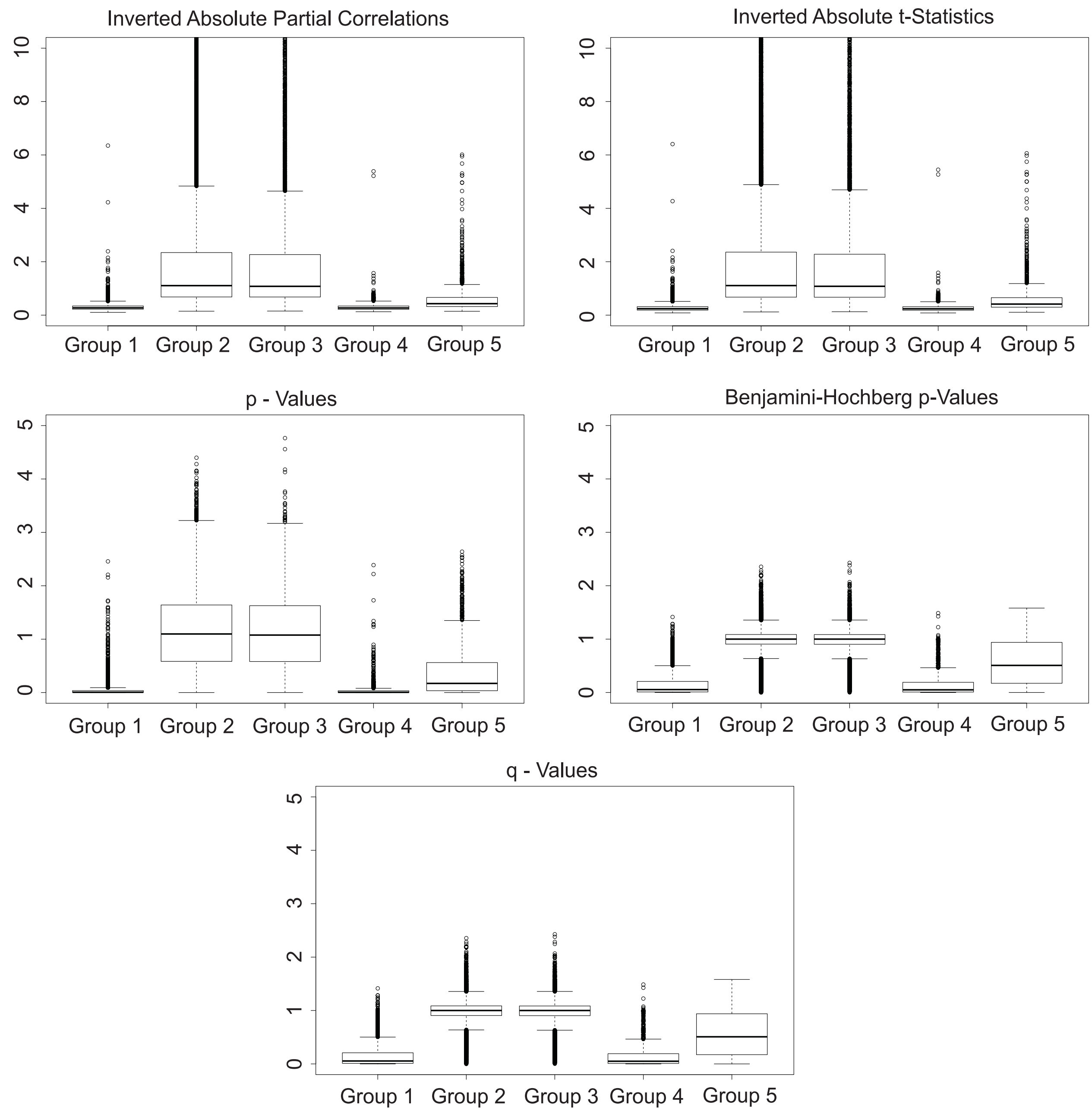

Fig. 2. Boxplots of weights in each group for different weighting types when $\boldsymbol{\beta} = (4.5, 3, -3, -3, \mathbf{0}^T, 3, 0)^T$. All weights have been standardized by the median over all groups.

4. ANALYSIS OF MICROBIAL DATA

Recent studies have indicated a link between body weight regulation and diets rich in dairy (Zemel 2003 and 2005). Other studies demonstrated that diet content highly influences gut microbiome diversity (Abnous *et al*. 2009, Li *et al*. 2009), and, in turn, these gut microbes impact bodyweight regulation components like host energy homeostasis, fat storage, and insulin insensitivity (Musso *et al*. 2011). Motivated by these findings, biologists seek to determine those microbial genera which affect phenotypes related to bodyweight regulation, while incorporating the fact that diet impacts these phenotypes.

To answer this question, a biological study was performed where twenty male, genetically similar mice were randomly assigned to one of two diets. Each diet group contained ten mice and differed as follows: an isolated soy protein diet and a non-fat dry milk. The latter diet is known to result in weight gain, while the former promotes reduced weight gain (Thomas *et al*. 2012). After 12 weeks of feeding, feces from each mouse were collected and analyzed for microbial communities via pyrosequencing. For each mouse, the data available consists of the percentages from 37 different microbes present in the feces, diet and information for plasma insulin concentration in pg insulin/ml plasma. The key interest is to determine which microbes have a relationship with insulin after accounting for diet.

Applying the weighted Lasso under the five proposed weights in Section 2 led to Alistipes and Moryella always being selected. Some of the weights did select additional microbes, but having Alistipes and Moryella selected by all methods suggests that these two microbes have an important impact on insulin. Its biological implications, however, are still not well understood since inter-kingdom signalling and cross-talk between microbes is still a new field (Pacheco and Sperandio 2009). Further biological experiments are needed to truly understand their impact. Still, our method's results are essential in defraying laboratory costs and conserving resources. Future designed experiments can focus entirely on Alistipes and Moryella, rather than the remaining 35 irrelevant microbes.

5. CONCLUSION

We conclude that in the $p > n$ context, when part of the regressors can act on both the response and some of the other explanatory variables, using structural information to construct feature weights in the weighted Lasso greatly aids the variable selection. We have shown that the results from Garcia *et al*. (2013a) extend from using q-values to any other informative measure of significance, such as p-values and adjusted p-values, or partial correlation coefficients and test statistic values. Each such measure of significance has its own merits. By construction, using test statistic values and partial correlation coefficients has the advantage that no distributional assumptions are necessary to construct the weights. In our simulation, using p-values slightly outperformed the other four weights considered. Finally, using BH adjusted p-values and q-values has the advantage that such weights link well to alternative procedures to the weighted Lasso, for example any multiple testing method. In this article we focused on using measures of significance based weights for a structural regression problem. In the future, it will be worthwhile investigating how such weights aid in hierarchical selection problems, for example as in Garcia *et al*. (2013b), where features are classified into groups, subgroups and single components.

ACKNOWLEDGEMENTS

We thank Rosemary Walzem for providing the gut microbiome data set. An open-sourced R version of the proposed procedure is available on the authors' websites. Samuel Müller was supported by the Australian Research Council (DP130100488).

REFERENCES

Abnous, K., Brooks, S.P., Kwan. J., Matias, F., Green-Johnson, J., Selinger, L.B., Thomas, M. and Kalmokoff, M. (2009). Diets enriched in oat bran or wheat bran temporally and differentially alter the composition of the fecal community of rats. *J. Nutrition*, **139**, 2024-2031.

Anderson, R.E., Crespo, C.J., Bartlett, S.J., Cheskin, L.J. and Pratt, M. (1998). Relationship of physical activity and television watching with body weight and level of fatness among children: results from the Third National Health and Nutrition Examination Survey. *J. Amer. Medical Association*, **279**, 938-942.

Baillie-Hamilton, P.F. (2002). Chemical toxins: A hypothesis to explain the global obesity epidemic. *J. Altern Complement Med.*, **8**, 185-192.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc.*, Series **B57**, 289-300.

Bergersen, L.C., Glad, I.K. and Lyng H. (2011). Weighted Lasso with data integration. *Statist. Appl. Genet. Mol. Biol.*, **10**, 1-29.

Bertsekas, D. (1995). *Nonlinear Programming.* Athena Scientific, Belmont, MA.

Bray, G.A., Smith, S.R., de Jonge, L., Xie, H., Rood, J., Martin, C.K., Most, M., Brock, C., Mancuso, S. and Redman, L.M. (2012). Effect of dietary protein content on weight gain, energy expenditure, and body composition during overeating: a randomized controlled trial. *J. Amer. Medical Assoc.,* **307**, 47-55.

Charbonnier, C., Chiquet, J. and Ambroise C. (2010). Weighted-Lasso for structured network inference from time course data. *Statist. Appl. Genet. Mol. Biol.*, 9, Article 15.

Dabney, A.R., Storey, J.D. and with assistance from Gregory R. Warnes (2011). q-value: Q-value estimation for false discovery rate control. R package version 1.26.0. http://CRAN.Rproject. org/package=qvalue.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.*, **32**, 407-499.

Garcia, T.P., Müller, S., Carroll, R.J., Dunn, T.N., Thomas, A.P., Adams, S.H., Pillai, S.D. and Walzem, R.L. (2013a). Structured variable selection with q-values. *Biostatistics*, doi: 10.1093/biostatistics/kxt012.

Garcia, T.P., Müller, S., Carroll, R.J. and Walzem, R.L. (2013b). Identification of important regressor groups, subgroups, and individuals via regularization methods: application to gut microbiome data. *Bioinformatics*, doi: 10.1093/bioinformatics/btt608.

Grun, F. and Blumberg, B. (2006). Environmental obesogens: Organotins and endocrine disruption via nuclear receptor signaling. *Endocrinology*, **147(6)**, S50-55.

Heindel, J.J. and vom Saal, F.S. (2009). Role of nutrition and environmental endocrine disrupting chemicals during the perinatal period on the aetiology of obesity. *Molecular Cellular Endocrinology*, **304**, 90-96.

Hill, J.O. and Peters, J.C. (1998). Environmental contributions to the obesity epidemic. *Science*, **280**, 1371-74.

Hirose, K., Tateishi, S. and Konishi, S. (2013). Tuning parameter selection in sparse regression modeling. *Comput. Statist. Data Anal.*, **59**, 28-40.

La Merrill, M. and Birnbaum, L.S. (2011). Childhood obesity and environmental chemicals. *Mount Sinai J. Medicine*, **78,** 22-48.

Li, F., Hullar, M.A., Schwarz, Y. and Lampe, J.W. (2009). Human gut bacterial communities are altered by addition of cruciferous vegetables to a controlled fruit- and vegetable-free diet. *J. Nutrition* **139,** 1685-1691.

Meinshausen, N. and Bühlmann, P. (2010) Stability selection (with discussion). *J. Roy. Statist. Soc.,* Series **B72**, 417-473.

Musso, G., Gambino, R. and Cassader, M. (2011). Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. *Annual Rev. Medicine,* **62,** 361-380.

Pacheco, A.R. and Sperandio, V. (2009). Inter-kingdom signaling: chemical language between bacteria and host. *Current Opinion Microbiol.,* **12**, 192-198.

Storey, J.D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Statist.,* **31**, 2013-2035.

Storey, J.D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci.,* **100,** 9440-9445.

Thomas, A.P., Dunn, T.N., Drayton, J.B., Oort, P.J. and Adams, S.H. (2012). A high calcium diet containing nonfat dry milk reduces weight gain and associated adipose tissue inammation in diet-induced obese mice when compared to high calcium alone. *Nutrition Metabolism,* **9,** 3.

Tibshirani, R. (1996). Regression shrinkage and variable selection via the Lasso. *J. Roy. Statist. Soc.*, Series **B58**, 267-288.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. Roy. Statist. Soc.*, Series **B68**, 49-67.

Zemel, M.B. (2003). Mechanisms of dairy modulation of adiposity. *J. Nutrition*, **133,** 252S-256S.

Zemel, M.B. (2005). The role of dairy foods in weight management. *J. Amer. College Nutrition*, **24,** 537S-546S.

Zou, H. (2006). The adaptive Lasso and its oracle properties. *J. Amer. Statist. Assoc.*, **101,** 1418-1429.

Appendix: Sketch of Technical Arguments

A.1 Minimizer of Weighted Lasso

A necessary and sufficient condition for $\hat{\beta}$ to be the minimizer of $Q(\beta)$ is that the subdifferential $\partial Q(\beta)$ at $\beta = \hat{\beta}$ is zero (see Bertsekas (1995, p. 736)). Thus, the subgradient of $Q(\beta)$ with respect to β_k must satisfy:

$$0 = -\mathbf{v}_k^T \mathbf{r}_{(-k)} + \mathbf{v}_k^T \mathbf{v}_k \hat{\beta}_k + \lambda w_k u_k, \qquad \text{(A.1)}$$

where

$$u_k = \begin{cases} \dfrac{\hat{\beta}_k}{|\hat{\beta}_k|}, & \hat{\beta}_k \neq 0 \\ \in \{u_k : |u_k| \leq 1\}, & \hat{\beta}_k = 0. \end{cases}$$

Eliminating the k^{th} covariate from the model, *i.e.*, having $\hat{\beta}_k = 0$, is equivalent to (A.1) satisfying:

$$0 = -\mathbf{v}_k^T \mathbf{r}_{(-k)} + \lambda w_k u_k,$$

where $|u_k| \leq 1$. Simplifying, we have that $u_k = \mathbf{v}_k^T \mathbf{r}_{(-k)} / (\lambda w_k)$. But because $|u_k| \leq 1$, it follows that $\hat{\beta}_k = 0$ when $|\mathbf{v}_k^T \mathbf{r}_{(-k)}| / |\lambda w_k| \leq 1$, or equivalently,

$$|\mathbf{v}_k^T \mathbf{r}_{(-k)}| \leq |\lambda w_k|.$$

Because $\lambda > 0$ and $w_k > 0$, the absolute values on the right hand side can be dropped, and thus the result in (2) holds.

On the other hand, when $\hat{\beta}_k \neq 0$, the subgradient equation (A.1) is equivalent to

$$0 = -\mathbf{v}_k^T \mathbf{r}_{(-k)} + \mathbf{v}_k^T \mathbf{v}_k \hat{\beta}_k + \lambda w_k \frac{\hat{\beta}_k}{|\hat{\beta}_k|}$$

$$= -\mathbf{v}_k^T \mathbf{r}_{(-k)} + \left(\mathbf{v}_k^T \mathbf{v}_k + \frac{\lambda w_k}{|\hat{\beta}_k|} \right) \hat{\beta}_k.$$

which implies that

$$\hat{\beta}_k = \frac{\mathbf{v}_k^T \mathbf{r}_{(-k)}}{\mathbf{v}_k^T \mathbf{v}_k + \lambda w_k / |\hat{\beta}_k|}.$$

Taking absolute values of both sides yields

$$|\hat{\beta}_k| = \frac{|\mathbf{v}_k^T \mathbf{r}_{(-k)}|}{|\mathbf{v}_k^T \mathbf{v}_k + \lambda w_k / |\hat{\beta}_k||} = \frac{|\mathbf{v}_k^T \mathbf{r}_{(-k)}|}{\mathbf{v}_k^T \mathbf{v}_k + \lambda w_k / |\hat{\beta}_k|},$$

where the second equality follows from the fact that $\mathbf{v}_k^T \mathbf{v}_k + \lambda w_k / |\hat{\beta}_k| > 0$ given that $\lambda > 0$ and $w_k > 0$. Thus,

$$|\hat{\beta}_k| \left(\mathbf{v}_k^T \mathbf{v}_k + \lambda w_k / |\hat{\beta}_k| \right) = |\mathbf{v}_k^T \mathbf{r}_{(-k)}|,$$

from which we obtain

$$|\hat{\beta}_k| = (|\mathbf{v}_k^T \mathbf{r}_{(-k)}| - \lambda w_k) / \mathbf{v}_k^T \mathbf{v}_k.$$

It is important to note that because $\hat{\beta}_k \neq 0$, condition (2) does not hold; hence, $|\mathbf{v}_k^T \mathbf{r}_{(-k)}| > \lambda w_k$, and so the numerator in $|\hat{\beta}_k|$ above is positive. Finally, using this form of $|\hat{\beta}_k|$ we obtain

$$\hat{\beta}_k = \frac{\mathbf{v}_k^T \mathbf{r}_{(-k)}}{\mathbf{v}_k^T \mathbf{v}_k + \lambda w_k / |\hat{\beta}_k|} = \frac{\mathbf{v}_k^T \mathbf{r}_{(-k)}}{\mathbf{v}_k^T \mathbf{v}_k + \dfrac{\lambda w_k \mathbf{v}_k^T \mathbf{v}_k}{|\mathbf{v}_k^T \mathbf{r}_{(-k)}| - \lambda w_k}}$$

$$= \frac{\mathbf{v}_k^T \mathbf{r}_{(-k)}}{\mathbf{v}_k^T \mathbf{v}_k |\mathbf{v}_k^T \mathbf{r}_{(-k)}| \Big/ \left(|\mathbf{v}_k^T \mathbf{r}_{(-k)}| - \lambda w_k \right)}$$

$$= \frac{\mathbf{v}_k^T \mathbf{r}_{(-k)} \left(|\mathbf{v}_k^T \mathbf{r}_{(-k)}| - \lambda w_k \right)}{\mathbf{v}_k^T \mathbf{v}_k |\mathbf{v}_k^T \mathbf{r}_{(-k)}|}$$

$$= \frac{\mathbf{v}_k^T \mathbf{r}_{(-k)}}{|\mathbf{v}_k^T \mathbf{r}_{(-k)}|} \left(|\mathbf{v}_k^T \mathbf{r}_{(-k)}| - \lambda w_k \right) \Big/ \mathbf{v}_k^T \mathbf{v}_k,$$

which is exactly the form in (3).

A.2 Additional Numerical Results

Results from simulation study when $\beta = (2.5, 1.5, -1.5, -1.5, \mathbf{0}^T, 1.5)^T$ in simulation study 1 and $\beta = (2.5, 1.5, -1.5, -1.5, \mathbf{0}^T, 1.5, 0)^T$ in simulation study 2. The results follow analogously to those in Section 3, except that the variable selection percentages are lower. This is due to the fact that the signal to noise ratio is much lower than for the results presented in Section 3. Still, however, we still observe that the weighted Lasso with p-values as weights generally leads to the most reasonable results in terms of low false discovery rates.

Table A.1. Simulation results from 1000 simulations when $\beta = (2.5, 1.5, -1.5, -1.5, \mathbf{0}^T, 1.5)^T$. Averaged percentages of time variables in each group were selected and observed false discovery rate (FDR). Ideal weighted Lasso will largely select variables in Groups 1 and 4, and not select variables in remaining groups.

Average Variables Selection					
Weights	$\lvert\rho_{\mathbf{x},\mathbf{y}\mid\mathbf{z}}\rvert^{-1}$	$\lvert t\rvert^{-1}$	p	p^{BH}	q
Treatment	100.00	100.00	100.00	100.00	100.00
Group 1	40.57	41.33	46.80	43.53	43.53
Group 2	0.51	0.50	0.57	0.47	0.47
Group 3	2.19	1.96	0.78	0.97	0.97
Group 4	60.20	59.40	52.00	51.50	51.50
FDR	0.16	0.14	0.10	0.10	0.10

Table A.2. Summary statistics of weights when $\beta = (2.5, 1.5, -1.5, -1.5, \mathbf{0}^T, 1.5)^T$. Average weight (Mean), and 25%, 50%, and 75% quantile of weights from each group. Ideal weighting will have small weights in Groups 1 and 4, and large weights in Groups 2 and 3.

Weights		Group 1	Group 2	Group 3	Group 4
$\lvert\rho_{\mathbf{x},\mathbf{y}\mid\mathbf{z}}\rvert^{-1}$	Mean	4.75	146.11	70.32	3.44
	$Q_{0.25}$	2.40	5.58	6.19	3.44
	$Q_{0.5}$	2.73	9.42	10.18	3.44
	$Q_{0.75}$	6.09	19.88	21.45	3.44
$\lvert t\rvert^{-1}$	Mean	0.75	24.01	11.55	0.53
	$Q_{0.25}$	0.36	0.90	1.00	0.53
	$Q_{0.5}$	0.42	1.54	1.66	0.53
	$Q_{0.75}$	0.97	3.26	3.52	0.53
p	Mean	0.07	0.50	0.50	0.07
	$Q_{0.25}$	0.02	0.27	0.30	0.07
	$Q_{0.5}$	0.03	0.50	0.50	0.07
	$Q_{0.75}$	0.10	0.73	0.71	0.07
p^{BH}	Mean	0.28	0.79	0.79	0.29
	$Q_{0.25}$	0.17	0.72	0.73	0.29
	$Q_{0.5}$	0.24	0.84	0.83	0.29
	$Q_{0.75}$	0.37	0.90	0.90	0.29
q	Mean	0.23	0.65	0.65	0.24
	$Q_{0.25}$	0.14	0.59	0.60	0.24
	$Q_{0.5}$	0.20	0.69	0.69	0.24
	$Q_{0.75}$	0.30	0.74	0.74	0.24

Table A.3. Simulation results from 1000 simulations when $\beta = (2.5, 1.5, -1.5, -1.5, \mathbf{0}^T, 1.5)^T$. Averaged percentages of time variables in each group were selected and observed false discovery rate (FDR). Ideal weighted Lasso will argely select variables in Groups 1 and 4, and not select variables in remaining groups.

Average Variables Selection					
Weights	$\lvert\rho_{\mathbf{x},\mathbf{y}\mid\mathbf{z}}\rvert^{-1}$	$\lvert t\rvert^{-1}$	p	p^{BH}	q
Treatment	100.00	100.00	99.90	100.00	100.00
Group 1	37.60	38.37	45.53	42.20	42.20
Group 2	0.67	0.65	0.65	0.62	0.62
Group 3	2.03	1.90	0.94	1.03	1.03
Group 4	57.10	57.10	51.40	51.50	51.50
Group 5	7.30	6.80	4.80	5.70	5.70
FDR	0.20	0.19	0.14	0.15	0.15

Table A.4. Summary statistics of weights when $\beta = (2.5, 1.5, -1.5, -1.5, \mathbf{0}^T, 1.5, 0)^T$. Average weight (Mean), and 25%, 50%, and 75% quantile of weights from each group. Ideal weighting will have small weights in Groups 1 and 4, and large weights in Groups 2 and 3.

Weights		Group 1	Group 2	Group 3	Group 4	Group 5
$\lvert\rho_{\mathbf{x},\mathbf{y}\mid\mathbf{z}}\rvert^{-1}$	Mean	3.69	52.89	51.01	3.62	38.64
	$Q_{0.25}$	2.42	5.60	6.12	3.62	38.64
	$Q_{0.5}$	2.76	9.53	9.99	3.62	38.64
	$Q_{0.75}$	4.49	20.65	20.22	3.62	38.64
$\lvert t\rvert^{-1}$	Mean	0.57	8.68	8.38	0.56	6.33
	$Q_{0.25}$	0.36	0.90	0.99	0.56	6.33
	$Q_{0.5}$	0.42	1.56	1.63	0.56	6.33
	$Q_{0.75}$	0.71	3.39	3.32	0.56	6.33
p	Mean	0.07	0.50	0.50	0.07	0.25
	$Q_{0.25}$	0.02	0.27	0.30	0.07	0.25
	$Q_{0.5}$	0.04	0.51	0.50	0.07	0.25
	$Q_{0.75}$	0.10	0.74	0.69	0.07	0.25
p^{BH}	Mean	0.28	0.78	0.78	0.26	0.55
	$Q_{0.25}$	0.17	0.71	0.72	0.26	0.55
	$Q_{0.5}$	0.24	0.84	0.83	0.26	0.55
	$Q_{0.75}$	0.37	0.91	0.89	0.26	0.55
q	Mean	0.23	0.64	0.64	0.21	0.45
	$Q_{0.25}$	0.14	0.59	0.59	0.21	0.45
	$Q_{0.5}$	0.20	0.69	0.68	0.21	0.45
	$Q_{0.75}$	0.30	0.75	0.73	0.21	0.45

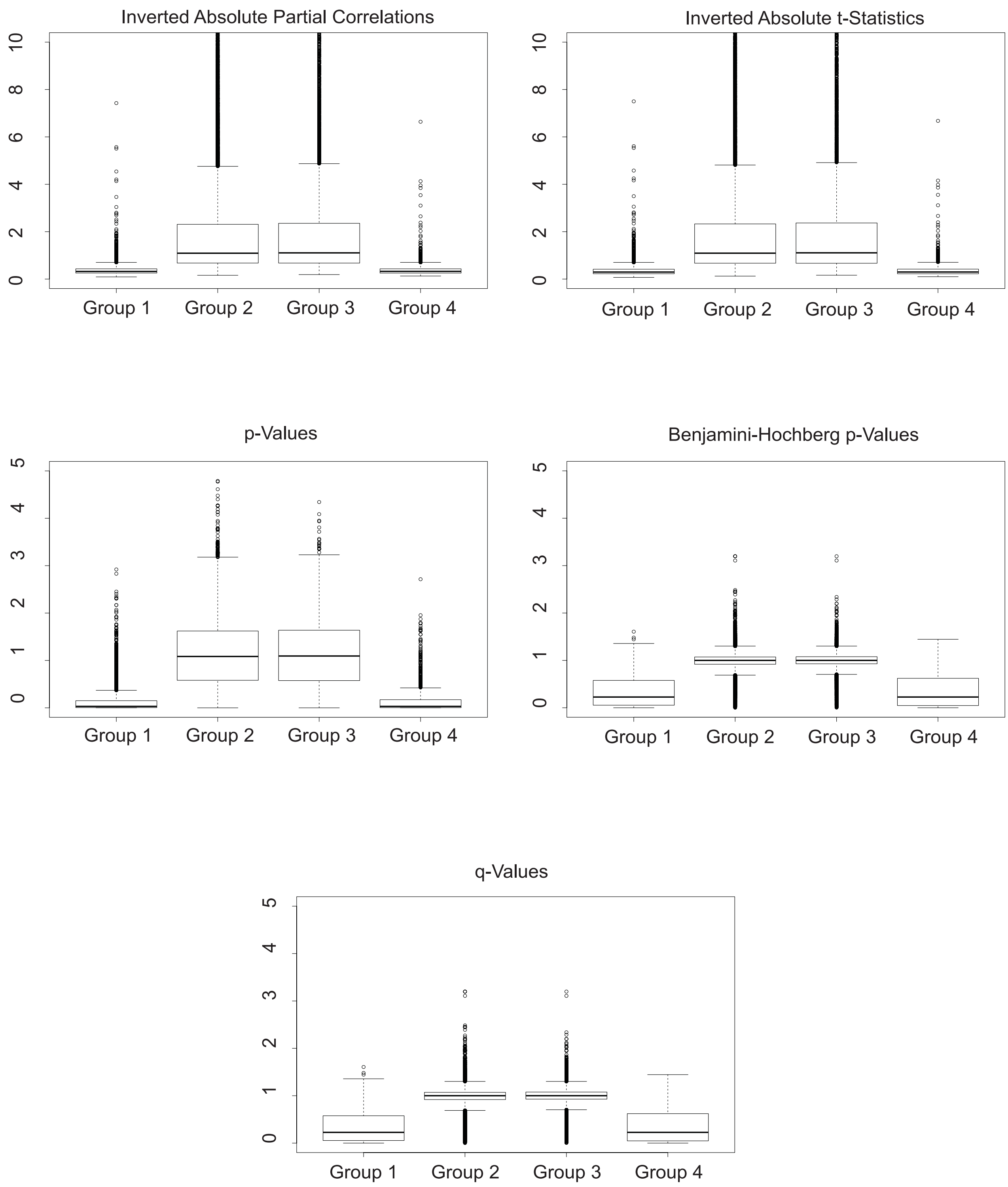

Fig. A.1. Boxplots of weights in each group for different weighting types when $\boldsymbol{\beta} = (2.5, 1.5, -1.5, -1.5, \mathbf{0}^T, 1.5)^T$. All weights have been standardized by the median over all groups.

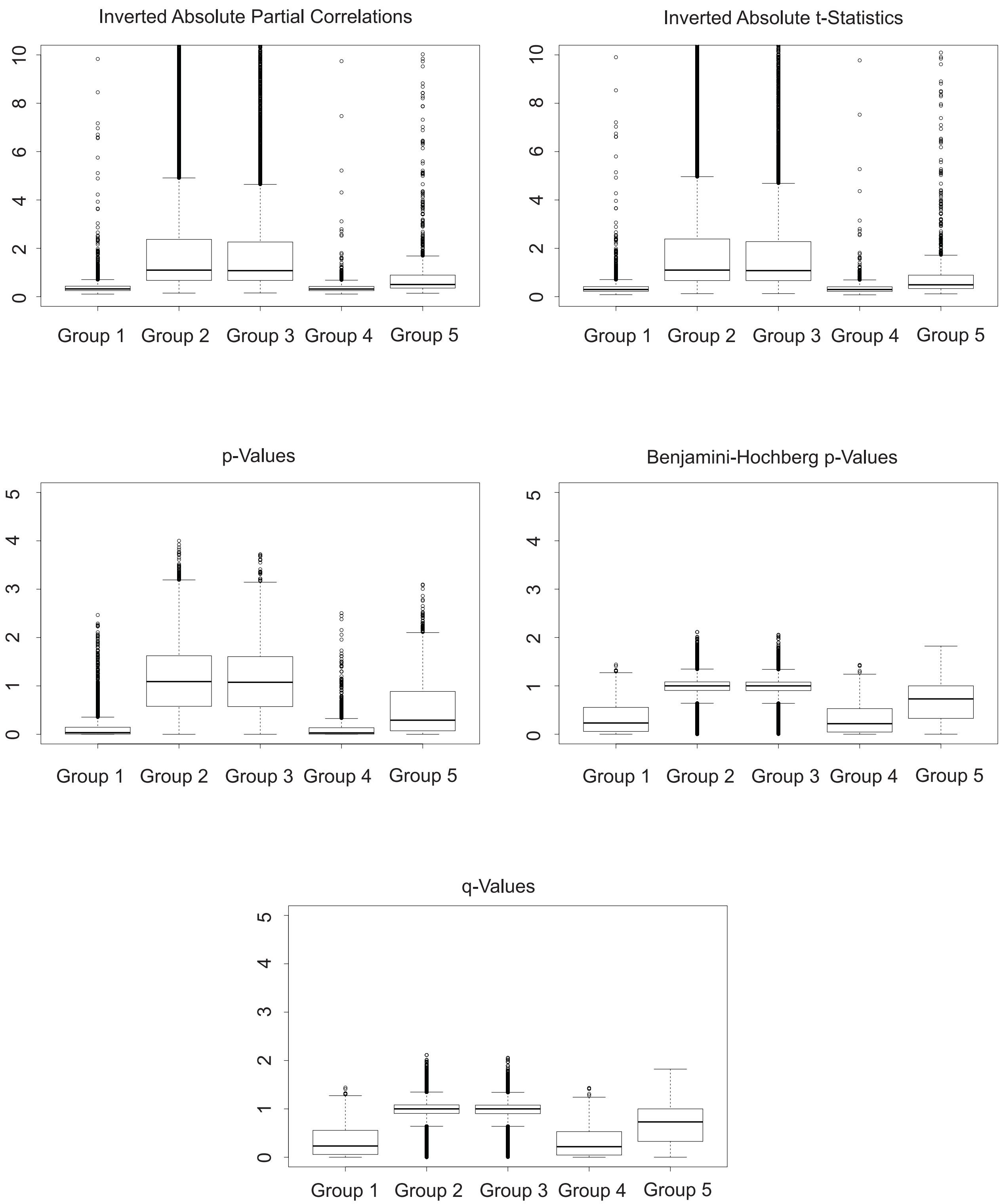

Fig. A.2. Boxplots of weights in each group for different weighting types when $\boldsymbol{\beta} = (2.5, 1.5, -1.5, -1.5, \mathbf{0}^T, 1.5, 0)^T$. All weights have been standardized by the median over all groups.