The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1880–1906
DOI: 10.1214/16-AOAS947
© Institute of Mathematical Statistics, 2016

LINKING LUNG AIRWAY STRUCTURE TO PULMONARY FUNCTION VIA COMPOSITE BRIDGE REGRESSION

BY KUN CHEN[1,2,*], ERIC A. HOFFMAN[1,†], INDU SEETHARAMAN[‡], FEIRAN JIAO[†], CHING-LONG LIN[1,†] AND KUNG-SIK CHAN[1,†]

University of Connecticut, University of Iowa† and Kansas State University ‡*

The human lung airway is a complex inverted tree-like structure. Detailed airway measurements can be extracted from MDCT-scanned lung images, such as segmental wall thickness, airway diameter, parent-child branch angles, etc. The wealth of lung airway data provides a unique opportunity for advancing our understanding of the fundamental structure-function relationships within the lung. An important problem is to construct and identify important lung airway features in normal subjects and connect these to standardized pulmonary function test results such as FEV1%. Among other things, the problem is complicated by the fact that a particular airway feature may be an important (relevant) predictor only when it pertains to segments of certain generations. Thus, the key is an efficient, consistent method for simultaneously conducting group selection (lung airway feature types) and within-group variable selection (airway generations), i.e., bi-level selection. Here we streamline a comprehensive procedure to process the lung airway data via imputation, normalization, transformation and groupwise principal component analysis, and then adopt a new composite penalized regression approach for conducting bi-level feature selection. As a prototype of composite penalization, the proposed composite bridge regression method is shown to admit an efficient algorithm, enjoy bi-level oracle properties and outperform several existing methods. We analyze the MDCT lung image data from a cohort of 132 subjects with normal lung function. Our results show that lung function in terms of FEV1% is promoted by having a less dense and more homogeneous lung comprising an airway whose segments enjoy more heterogeneity in wall thicknesses, larger mean diameters, lumen areas and branch angles. These data hold the potential of defining more accurately the "normal" subject population with borderline atypical lung functions that are clearly influenced by many genetic and environmental factors.

1. Introduction. The human lung airway is a complex, fractal-like [Weibel (2015)], inverted tree-like structure up to 28 segmental generations starting from the trachea, resulting from repeated branch bifurcations/trifurcations. Figure 1 shows two human lung airway trees, each of which consists of the trachea and

Received January 2015; revised April 2016.
[1]Supported in part by NIH U01 HL114494 and NIH R01 HL089897.
[2]Supported in part by NSF DMS-16-13295.
Key words and phrases. Bi-level variable selection, composite penalization, feature extraction, lung airway data, pulmonary function tests.

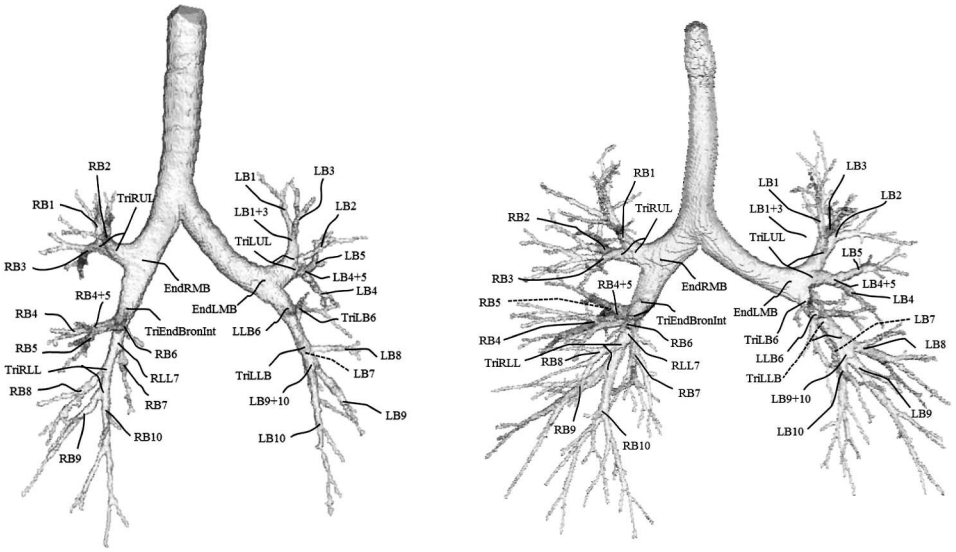


FIG. 1. *Airway trees of two normal subjects*; *subject of the left* (*right*) *diagram has the lowest* (*highest*) *standardized FEV*1%.

other labeled segments of different generations [Gao (2010)]. Using multi-detector row computed tomography (MDCT), detailed in vivo measurements of individual human lung airways have been increasingly collected [Palagyi et al. (2006), Tschirren et al. (2005a, 2005b)]. From each MDCT image, thousands of measurements of lung airway segments up to the 7th generation can be extracted. These highly detailed, regional views of the lung airway geometry enable detailed exploration of the integrated lung structure and its association with pulmonary functions. Besides these geometric features, the MDCT scan also provides other important lung features such as the lung parenchymal radio-density. Precisely, radio-density from MDCT is measured in Hounsfield Units (HU), which range from -1000 to $+1000$. By definition, air has -1000 HU, water has 0 HU, and bone is the most dense, which has $+1000$ HU. Lung parenchymal radio-density, that is, the radio-density of the portion of the lung involved in gas exchange, can be indicative of certain lung disease and lung function deterioration.

Lung function can be assessed by spirometry, representing global measures, using various pulmonary function tests that provide measures of airflow, lung volumes, diffusing capacity and more along with the response of these properties to bronchodilators. In particular, the forced expiratory volume in 1 second (FEV1) measures the maximum air volume that can forcibly be blown out by a subject in one second after a full inspiration effort. It is known that FEV1 is related to both sex and age: men generally have higher FEV1 values than women of the same age, and FEV1 generally increases with age until around 20 years old after which it declines. Thus, the FEV1 of a subject is often calibrated by the ratio of

the FEV1 to the predicted FEV1, which is the mean FEV1 of the subpopulation with age, sex and body composition matching those of the subject; see Hankinson, Odencrantz and Fedan (1999), which provided the spirometric reference values from a sample of the general U.S. population which have been commonly used in the literature. The preceding ratio is denoted as FEV1%, and it quantifies lung obstruction. Having FEV1% greater than 80% is generally classified as normal, while having FEV1% less then 40% is considered a symptom of severe lung obstruction. However, this value provides little to no information in regards to the regional distribution of lung pathologies serving to effect this single measure, and regional disease can be undetected by this single metric.

Even within the normal population, there is substantial variation in FEV1%, and little is known about the mechanism underlying the variation in that some normal subjects have exceptionally high FEV1% while others have borderline lung obstruction. As the first step toward understanding the variation in FEV1% among normal subjects, we examine how the local lung airway structures and the global parenchymal radio-density measures affect FEV1% using a collection of lung-image data from a cohort of normal subjects gathered under an NIH-sponsored Biomedical Engineering Research Partnership (BERP, NIH-HL-064368). The question regarding the reliability of parenchymal measurements from MDCT has received much attention in the medical literature, as the quantification of the parenchymal pathology can be affected by variations in inspiratory and expiratory efforts, scanner type, radiation dose and image reconstruction algorithms; see Iyer et al. (2014) for a review on the repeatability of the BERP data. The data on the normal subjects we use here consist of volumetric MDCT images gathered at spirometrically controlled [Fuld et al. (2012)] full inspiration (total lung capacity) and 20% vital capacity (functional residual capacity) using a scanning protocol outlined in Iyer et al. (2014) so that the levels of inspiratory and expiratory efforts were both controlled when performing MDCT.

The preceding medical problem may be approached using a high-dimensional regression analysis, with the FEV1% being the response variable and the lung airway variables the predictors. However, in order to fully embrace the many unique characteristics of the lung airway data, a customized, comprehensive dimension reduction approach is required. For each lung, the MDCT measurements comprise a large number of measurements per airway segment, for example, wall thickness, airway diameter, lumen area, segment angle, rate of tapering, etc. The airway variables of the same type and from the same generation/segment are generally highly positively correlated. Besides, the distortion of the airway structure due to lung disease as reflected in each segmental measure could be weak and unstable, and such local distortions need to be amplified by properly summarizing information across airway segments. A comprehensive data-processing procedure is thus needed to extract an interpretable and yet parsimonious set of candidate airway features. More importantly, the problem is complicated by the fact that a particular airway feature may be an important predictor only when it pertains to segments of

certain generations. It is desirable to determine which lung features at a given generation are significant predictors of FEV1%. We tackle this problem by grouping the airway variables by feature type and then simultaneously doing group selection and within-group feature selection. Following Huang et al. (2009), we refer to this general methodology as bi-level variable selection. Using our established framework, other pulmonary function metrics and their association with lung airway structure could be explored similarly.

With the properly constructed lung airway features, the bi-level selection methodology is the key ingredient to carry out the lung airway analysis. The need for bi-level selection is also motivated by the fact that, in the study of the association between FEV1% and lung airway features, we do not possess an ideal grouping configuration such that variables in any group are either all important (relevant) variables (i.e., with nonzero true coefficients) or all irrelevant variables (with zero coefficients). In real applications including our study, certain groups may be *mixed* in the sense that they contain both relevant and irrelevant variables. Thus, it is desirable to develop a variable selection approach that allows flexible incorporation of the prior grouping information that is also robust to the presence of mixed groups.

While our research is motivated by the lung study, the bi-level variable selection methodology is widely applicable. For example, a multi-level categorical variable can be coded as a group of dummy variables. When a group of such dummy variables is selected, it is also desirable to determine which categories truly matter and which shall be collapsed to the baseline. In genetic studies where the genes can be grouped based on the pathways, it is of importance to both select relevant pathways and identify a few useful genes along each selected pathway. Huang et al. (2009) developed the group bridge method for bi-level variable selection. Due to the use of the nonconvex bridge penalty at the group level, the method enjoys group selection consistency. However, at the individual level, the method exhibits similar behaviors as Lasso [Tibshirani (1996)], which often leads to within-group overselection. Zhao, Rocha and Yu (2009) proposed a composite absolute penalty, which combines the properties of norm penalties at the across-group and within-group levels to facilitate hierarchical variable selection. Breheny and Huang (2009, 2011) proposed a general form of composite penalty for bi-level selection. Bi-level selection techniques are also critical in the integrative analysis of multiple data sets, especially in high-throughput genomic studies [Liu, Ma and Huang (2014), Ma et al. (2011)]; see Huang, Breheny and Ma (2012) for a recent review of the group and bi-level selection methods.

To the best of our knowledge, however, little progress has been made to rigorously investigate the bi-level selection methodology and theory via composite penalization scheme since the pioneer work by Huang et al. (2009). Motivated by the lung airway study, we propose a composite bridge method for bi-level selection. Unlike the group bridge method in which an ℓ_1 penalty is used to induce within-group sparsity, our approach adopts another nonconvex bridge penalty for

within-group regularization, and hence the name composite bridge. Our approach is a prototype in the nonconvex composite penalization scheme, and thus our analysis can be readily extended to the general composite and hierarchical penalization methods.

Section 2 presents our comprehensive approach of linking the lung airway data to the pulmonary function measured by FEV1% via bi-level feature selection. Specifically, we describe the lung airway data and the related scientific problems in Section 2.1, and a streamlined procedure to derive crafted lung airway features from lung imaging data in Section 2.2. The problem is then reduced to a bi-level variable selection task as described in Section 2.3, for which we develop the new composite bridge regression approach in Section 2.4. Since the proposed method is applicable in many problems, the methodology and the algorithm are presented in general terms. In Section 2.5, our results show that lung function in terms of FEV1% is promoted by a number of expected characteristics of the airway tree structure and the lung parenchyma. These findings certainly guide and prompt further clinical and data investigations for reaching the goal of devising more informative, personalized calibration of FEV1% and other pulmonary function tests by taking into account important airway features derived from MDCT lung imaging data. Section 3 provides both theoretical and empirical justifications of the composite bridge method. In particular, we show that it enjoys the oracle properties for both group selection and within-group selection, and superior empirical performance over several existing methods. All the technical statements of our findings are given in the Appendix, with their proofs and some additional simulation examples given in the Supplementary Material [Chen et al. (2016)]. We provide our conclusions in Section 4.

2. Linking FEV1% and lung airway features.

2.1. *Lung airway and lung function data*. Our study is based on data from 132 subjects with normal lung function [Hoffman, Simon and McLennan (2006)]. From the MDCT-image of each human lung, numerous detailed in vivo measurements were obtained to quantify the airway tree structure, including luminal and wall characteristics along with branching patterns. Along each airway segment, there were hundreds of measurements taken at various locations along the luminal centerlines [Palagyi et al. (2006), Tschirren et al. (2005a, 2005b, 2005c)]; these measurements were processed by the Pulmonary Analysis Software Suite (PASS) [Guo et al. (2008)], an image reconstruction software, into 49 summary statistics per airway segment, which comprise part of the covariates for our analysis. These variables can be categorized into several feature types, for example, *wall thickness*, *inner/outer airway diameter*, *inner/outer airway perimeter*, *inner/outer lumen area* and *segment angle*. For instance, 13 summary statistics are related to wall thickness. As an example, we illustrate how PASS calculates the so-called "average–average wall thickness." At every voxel position along the centerline

of an airway segment, the wall thickness is measured along rays cast from the center of gravity of the airway lumen. Rays are cast every half degree, resulting in a total of 720 measurements. These 720 wall thickness measurements are averaged into one mean wall thickness. The mean measurements for all centerline voxel positions along the middle third part of the segment are then averaged to obtain the average–average wall thickness. Other wall thickness measures, including average–minimum wall thickness (the average of the minimum measurements for all centerline voxel positions) and maximum–minimum wall thickness (the maximum of the minimum measurements for all centerline voxel positions), are similarly constructed. The extents of variation in wall thickness are also calculated, for example, standard-deviation-average wall thickness, standard-deviation-minimum wall thickness, etc.

Higher generational image data are affected both by limitations of scanner resolution and by motion artifacts, and so they are less reliable and also admit substantial missing values. Therefore, quantitation of the airway tree has been limited to the segmental airways plus two generations beyond along 5 major pathways by manual verification, namely, RB1, RB4, RB10, LB1 and LB10. For the details of the naming convention of the airway segment, see, for example, Gao (2010). In our analysis, the raw airway data are first preprocessed via imputation, normalization and transformation. Specifically, about 1.1% of the airway measurements are missing, and the missing values for each subject are imputed by the corresponding average measurements from the cohort of subjects with the same gender. The measurements are normalized to adjust for variations in body size using the subject's *height*, that is, the length, area and volume measurements are normalized by the height, the square of the height and the cubic of the height, respectively. Geometric features related to segment angles are not adjusted. The original variables tend to have skewed marginal distributions, and we thus work with log-transformed data.

Our interest is to identify which lung airway features from MDCT-scanned lung images are associated with FEV1%, but it is essential to adjust for potential confounding effects due to parenchymal (lung tissue) anomalies, as measured by mean parenchymal (radio-)density, standard deviation of parenchymal density (measured at total lung capacity) and the percent of lung voxels with attenuation below -856 HU. In the regression analysis to be presented in Section 2.3, we have used log-transformed FEV1% as the response variable; for convenience, we still refer to it as the FEV1% in the sequel. By definition, the FEV1% is a gender- and age-adjusted pulmonary function measure, and so we do not include gender and age in the analysis.

2.2. *Airway feature construction and extraction.* To improve data reliability, reduce data dimensions and enhance the biological signals, it is pivotal to construct meaningful and interpretable airways features from the segmental lung airway data. This is done in the following two steps.

First, we aggregate the segmental airway measurements to create generation-level summaries. Specifically, we compute generational average measures for generations 0–6, for example, for average–average wall thickness of generation 3, we computed its average value over all segments in generation 3 to get the generational average–average wall thickness. Similarly, we compute the generational average within-segment variation measures, for example, within-segment standard deviation of the average–average wall thickness of generation 3. Besides, the variation among all segments of the same generation can also be very informative. We thus compute various between-segment variation measures of generations 1 to 4 for which we have at least 4 segments per generation, for example, between-segment standard deviation of the average–average wall thickness of generation 3.

Second, as the airway variables of the same type and in the same generation/segment are generally highly positively correlated, we conduct groupwise principal component analysis (PCA) to further reduce the dimensionality and alleviate the collinearity problem. For instance, the correlation coefficients among the wall thickness variables in any particular generation are generally above 0.90, and so are the variables measuring the variation in thickness. The variables of each particular type and generation are summarized (replaced) by their leading principal components (PC). Thus, the thickness variables per generation are summarized by their leading principal components, and so are the variables measuring the extents of variation in thickness, etc. Due to high positive correlation among the variables, all PC loadings are of the same sign, and hence set to be positive without loss of generality. Therefore, each principal component roughly corresponds to the average of the correlated lung variables. This approach accounts for on average 89% of the within-group variation, yet the leading principal components largely preserve the interpretation of the original features. The number of variables is reduced by 70%, resulting in 87 lung airway feature variables for the subsequent regression analysis to be elaborated below.

2.3. *Grouping of the airway features and the need for bi-level selection.* To identify the important airway features associated with FEV1%, it is natural to consider a penalized regression approach, with the FEV1% serving as the response and the lung airway variables and parenchymal density variables as the linear predictors. The problem is challenging for several reasons. The number of candidate variables ($d = 90$) is comparable to the sample size ($n = 132$). Moreover, there exist various sources of variations affecting the spirometric measurements [Becklake (1985)]. For example, the FEV1 measurements can be affected by inspiratory and expiratory efforts. Another fact that may be less well known is that the FEV1 measurements of each individual may even vary substantially within a day, and the highest values usually occur around noon. These "unwanted variations" may lead to low signal-to-noise ratio in the regression analysis of FEV1% on the lung airway structure. Despite these difficulties, some characteristics of lung airways and lung disease mechanisms can be potentially utilized to boost the performance of feature

selection. In particular, any lung airway feature may be an important predictor of FEV1% only when it pertains to segments of certain generations. This important prior scientific knowledge may hold the key of conducting a successful penalized regression analysis.

We thus group the constructed lung airway variables by feature types, and develop a new penalized regression method to simultaneously conduct group selection (lung airway feature types) and within-group variable selection (airway generations), that is, bi-level selection. Specifically, each group consists of the generational mean leading principal component for a particular feature type (wall thickness, lumen area, circularity, etc.), and corresponding groups of the between-segment (within-segment) variation of the features. The between-segment standard deviations will be abbreviated as BSSD and the within-segment counterparts WSSD. The mean and standard deviation of the parenchymal density form a group by their own. The process consolidates the $d = 90$ variables into $J = 17$ groups, with the group sizes varying from 2 to 7.

Below we list the lung features and their grouping structure. We develop an efficient penalized regression approach to conduct bi-level selection in Section 2.4 below:

- Mean parenchymal density, standard deviation of parenchymal density and percentage of voxels below -856 HU form a group.
- Forty-one generational average segmental characteristics including wall thickness, diameter, perimeter, lumen area and circularity, for generations 0–6, plus segmental angle for generations 1–6. These variables form 6 feature groups.
- Twenty-eight generational within-segment variation variables (standard deviation) of wall thickness, diameter, perimeter, and lumen area, for generations 0–6. These variables form 4 groups.
- Eighteen generational between-segment variation variables (standard deviation) of wall thickness, diameter, perimeter, lumen area, angle and circularity, for generations 3–5. These variables form 6 groups.

2.4. *Composite bridge regression.* Since the bi-level feature selection approach is applicable in many applications, we shall present the new methodology in general terms, but in the mean time we refer back to the lung airway analysis whenever necessary.

Consider the multiple linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon} = \sum_{k=1}^{d} \mathbf{x}_k \beta_{0k} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$ is the response vector, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_d)$ the design matrix, $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0d})'$ a vector of regression coefficients, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$ an error vector consisting of uncorrelated random errors with mean 0 and variance

σ^2. We assume the response is centered and the predictors are standardized so that there is no intercept term in the model. Throughout the paper, we use $\|\cdot\|_q$ to denote the ℓ_q norm for any $q \geq 0$.

Let $A_1, \ldots, A_J$ be subsets of $\{1, \ldots, d\}$ representing prior grouping structure of the predictors, and $\boldsymbol{\beta}_{A_j} = (\beta_k, k \in A_j)'$ be the vector of regression coefficients in the jth group, with $\boldsymbol{\beta}_{0A_j}$ being the corresponding true values. Without loss of generality, suppose only the first J_1 groups are relevant, that is, $\boldsymbol{\beta}_{0A_j} \neq \mathbf{0}$ for $j = 1, \ldots, J_1$ and $\boldsymbol{\beta}_{0A_j} = \mathbf{0}$ for $j = J_1 + 1, \ldots, J$. We further assume that, in each of the first J_1 groups, only a subset of the predictors is important. For each A_j, $j = 1, \ldots, J_1$, let $A_j^1 = \{k; \beta_{0k} \neq 0, k \in A_j\}$ and $A_j^2 = \{k; \beta_{0k} = 0, k \in A_j\}$. As in Huang et al. (2009), the J groups may overlap with each other and their union is allowed to be a proper subset of all the predictors. Note that the jth group among the first J_1 groups is a *mixed group* if $A_j^2 \neq \varnothing$.

Consider estimating the vector of coefficients $\boldsymbol{\beta}_0$ in model (2.1) by minimizing the following objective function:

$$L_n(\boldsymbol{\beta}) = \Big\| \mathbf{y} - \sum_{k=1}^{d} \mathbf{x}_k \beta_k \Big\|_2^2 + \lambda_n \sum_{j=1}^{J} c_j \Big(\sum_{k \in A_j} |\beta_k|^\mu \Big)^\gamma. \tag{2.2}$$

The first term is the sum of squared errors, which is appropriate in the case of Gaussian errors. The second term is a penalty function incorporating the known grouping structure, where $\mu \geq 0$, $\gamma \geq 0$, the c_j's are group-level weights adjusting for the dimensions or magnitudes of each group of coefficients, and λ_n is a tuning parameter controlling the degrees of penalization.

The choice of μ and γ in the above penalty form holds the key for controlling the sparsity-inducing behaviors in the individual and the group levels. In fact, with certain choices of $\mu \geq 0$ and $\gamma \geq 0$, (2.2) subsumes many existing penalized estimation approaches or formulations, but may not lead to satisfactory bi-level selection. When $\gamma = \mu = 1$, it reduces to Lasso, and when $\gamma = 1$ and $0 < \mu < 1$, it reduces to the bridge regression. The group Lasso method corresponds to $\mu = 2$ and $\gamma = 1/2$, in which c_j is commonly chosen as $\sqrt{|A_j|}$, the square root of the number of predictors in group j. The group Lasso penalty induces group sparsity due to the ℓ_1 norm penalization at the group level, which cannot achieve selection consistency in general and tends to overselect groups of variables. Zhao, Rocha and Yu (2009) proposed the composite absolute penalty, and the focus was on the case $\mu > 1$ and $\gamma = 1$ and the design of the overlapping pattern to achieve hierarchical selection. Another interesting case is that, when $\mu = 1$ and $0 < \gamma < 1$, the above criterion becomes the group bridge method [Huang et al. (2009)]. The method penalizes the ℓ_1 norms of the groups of coefficients using a bridge penalty, hence inducing bi-level sparsity. Huang et al. (2009) showed that the group bridge enjoys group selection consistency. Yet since its selection performance at the individual level is analogous to that of the Lasso, the group bridge cannot achieve the bi-level selection consistency in general.

Motivated by the aforementioned works and the lung airway data structure, we propose to minimize the above objective function (2.2) with

$$\mu \in (0, 1) \quad \text{and} \quad \gamma \in (0, 1).$$

Unless otherwise noted, we set $c_j = |A_j|^{1-\gamma}$, accounting for varying group sizes. Unlike the group bridge method, we adopt another bridge penalty to induce within-group sparsity, and hence we refer to the method as the *composite bridge penalized regression*. As we will show in Section 3 and the Appendix, this intuitive extension is simple yet consequential, leading to variable selection consistency at both the group and individual levels simultaneously for any particular choices of $\gamma \in (0, 1)$ and $\eta \in (0, 1)$. For simplicity and to streamline the idea, we fix $\gamma = \mu = 0.5$ in all numerical studies. The choice of μ or γ can be made data adaptive, which may further boost the empirical performance of the proposed approach at the cost of increased computational efforts.

The minimization of the objective function (2.2) is challenging, as the composite bridge penalty is nonconvex for $\mu \in (0, 1)$, $\gamma \in (0, 1)$. Motivated by Huang et al. (2009), we show that an equivalent minimization problem can be formulated through an augmented variable approach, and an efficient iterative algorithm is then developed for solving (2.2). Define

$$\begin{aligned} S_{ln}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\delta}) = {} & \Big\| \mathbf{y} - \sum_{k=1}^{d} \mathbf{x}_k \beta_k \Big\|_2^2 \\ & + \sum_{j=1}^{J} \theta_j^{1-\frac{1}{\gamma}} c_j^{\frac{1}{\gamma}} \Big(\sum_{k \in A_j} \delta_k^{1-\frac{1}{\mu}} |\beta_k| + \psi \sum_{k \in A_j} \delta_k \Big) + \tau_n \sum_{j=1}^{J} \theta_j, \end{aligned} \tag{2.3}$$

where $\tau_n = \lambda_n^{1/(1-\gamma)} \gamma^{\gamma/(1-\gamma)} (1 - \gamma)$ and $\psi = \mu^{\mu/(1-\mu)}(1 - \mu)$. The following proposition shows the equivalence between the minimizers of (2.3) and (2.2).

PROPOSITION 2.1. *The composite bridge estimator $\hat{\boldsymbol{\beta}}_n$ minimizes (2.2) if and only if*

$$(\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\delta}}) = \arg \min_{(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\delta})} S_{ln}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\delta}) \quad \textit{subject to} \quad \boldsymbol{\theta} \geq 0, \boldsymbol{\delta} \geq 0$$

for some $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\delta}}$, where $S_{ln}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\delta})$ is given in (2.3).

To see the main idea behind our augmented variable approach, denote

$$\hat{\boldsymbol{\delta}}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \arg \min_{\boldsymbol{\delta} \geq \mathbf{0}} \{ S_{ln}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\delta}) \}, \qquad \hat{\boldsymbol{\theta}}(\boldsymbol{\beta}) = \arg \min_{\boldsymbol{\theta} \geq \mathbf{0}} \{ S_{ln}(\boldsymbol{\beta}, \boldsymbol{\theta}, \hat{\boldsymbol{\delta}}(\boldsymbol{\beta}, \boldsymbol{\theta})) \},$$

that is, $\hat{\boldsymbol{\delta}}(\boldsymbol{\beta}, \boldsymbol{\theta})$ is obtained by minimizing $S_{ln}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$ with $(\boldsymbol{\beta}, \boldsymbol{\theta})$ held fixed; the profile objective function $S_{ln}(\boldsymbol{\beta}, \boldsymbol{\theta}, \hat{\boldsymbol{\delta}}(\boldsymbol{\beta}, \boldsymbol{\theta}))$ becomes a function of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ only, and then $\hat{\boldsymbol{\theta}}(\boldsymbol{\beta})$ is obtained by minimizing $S_{ln}(\boldsymbol{\beta}, \boldsymbol{\theta}, \hat{\boldsymbol{\delta}}(\boldsymbol{\beta}, \boldsymbol{\theta}))$

with respect to $\boldsymbol{\theta}$ with $\boldsymbol{\beta}$ held fixed. It turns out that both $\hat{\boldsymbol{\delta}}(\boldsymbol{\beta}, \boldsymbol{\theta})$ and $\hat{\boldsymbol{\theta}}(\boldsymbol{\beta})$ admit closed-form solutions that are easy to compute, and, more importantly, once they are plugged into S_{ln}, the profile function $S_{ln}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}(\boldsymbol{\beta}), \hat{\boldsymbol{\delta}}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}(\boldsymbol{\beta})))$ recovers the original objective function $L_n(\boldsymbol{\beta})$! This suggests an iterative algorithm for minimizing $L_n(\boldsymbol{\beta})$. Denote the solution at the $(s-1)$th iteration as $\boldsymbol{\beta}^{(s-1)}$. We first compute $\hat{\boldsymbol{\theta}}(\boldsymbol{\beta}^{(s-1)})$ and $\hat{\boldsymbol{\delta}}(\boldsymbol{\beta}^{(s-1)}, \hat{\boldsymbol{\theta}}(\boldsymbol{\beta}^{(s-1)}))$, and then minimize $S_{ln}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}(\boldsymbol{\beta}^{(s-1)}), \hat{\boldsymbol{\delta}}(\boldsymbol{\beta}^{(s-1)}, \hat{\boldsymbol{\theta}}(\boldsymbol{\beta}^{(s-1)})))$ to obtain $\boldsymbol{\beta}^{(s)}$. The latter problem turns out to be an adaptive Lasso [Zou (2006)] problem in $\boldsymbol{\beta}$, which could be solved efficiently by many methods. By construction,

$$\begin{aligned} L_n(\boldsymbol{\beta}^{(s-1)}) &= S_{ln}(\boldsymbol{\beta}^{(s-1)}, \hat{\boldsymbol{\theta}}(\boldsymbol{\beta}^{(s-1)}), \hat{\boldsymbol{\delta}}(\boldsymbol{\beta}^{(s-1)}, \hat{\boldsymbol{\theta}}(\boldsymbol{\beta}^{(s-1)}))) \\ &\geq S_{ln}(\boldsymbol{\beta}^{(s)}, \hat{\boldsymbol{\theta}}(\boldsymbol{\beta}^{(s-1)}), \hat{\boldsymbol{\delta}}(\boldsymbol{\beta}^{(s-1)}, \hat{\boldsymbol{\theta}}(\boldsymbol{\beta}^{(s-1)}))) \\ &\geq S_{ln}(\boldsymbol{\beta}^{(s)}, \hat{\boldsymbol{\theta}}(\boldsymbol{\beta}^{(s)}), \hat{\boldsymbol{\delta}}(\boldsymbol{\beta}^{(s)}, \hat{\boldsymbol{\theta}}(\boldsymbol{\beta}^{(s)}))) \\ &= L_n(\boldsymbol{\beta}^{(s)}). \end{aligned}$$

Therefore, the objective function decreases monotonically along the iterations. The convergence of the algorithm is thus guaranteed, although not necessarily to the global minimum as the objective function is nonconvex. Based on our limited experience, the proposed method is stable and fast in practice. The detailed proof of Proposition 2.1 is given in the Supplementary Material [Chen et al. (2016)], and the resulting iterative algorithm for solving (2.3) is presented in Algorithm 1.

The proposed algorithm is then an iteratively reweighted adaptive Lasso procedure. At each iteration, the penalty level for a regression coefficient is determined by (2.4), combining information from both its current value and the coefficient values of the groups it belongs to. Another way to reveal this connection between our algorithm and the adaptive Lasso is to minimize the composite bridge penalized least squares problem (2.2) via local linear approximation [Zou and Li (2008)]. Up to a constant, the first-order approximation of the penalty for any coefficient, say, β_k, with other βs held fixed, yields an adaptive Lasso penalty for β_k, with the adaptive weight taking the same form as in (2.4). We note that, with some local approximation of the penalty term, we can alternatively implement an element-wise coordinate descent algorithm [Friedman, Hastie and Tibshirani (2010)] for solving the problem; see Breheny and Huang (2011). Although we mainly focus on Gaussian data and the squared error loss, our method can be readily extended to generalized linear models and models with more general loss functions, for example, Bregman divergence loss [Zhang, Jiang and Chai (2010)].

2.5. *Results of the lung airway analysis.* We use the composite bridge (CoB) method developed in Section 2.4 to select important lung airway variables jointly associated with FEV1%. For comparison, we have also tested several other methods, including group bridge (GrB) [Huang et al. (2009)], composite mimimax concave penalization (CoMCP) [Zhang (2010)], group Lasso (GrLasso) [Yuan and Lin

Algorithm 1 Composite Bridge Penalized Regression Algorithm (CoBRA)

Initialization: start with an initial estimator $\boldsymbol{\beta}^{(0)}$.

repeat

Step 1: calculate

$$\theta_j^{(s)} = c_j(\lambda_n\gamma)^{\frac{\gamma}{\gamma-1}}\Big(\sum_{k\in A_j}|\beta_k^{(s-1)}|^{\mu}\Big)^{\gamma}, \qquad j=1,2\ldots,J,$$

$$\delta_k^{(s)} = \mu^{\frac{\mu}{\mu-1}}|\beta_k^{(s-1)}|^{\mu}, \qquad k=1,\ldots,d.$$

Step 2: solve the adaptive Lasso problem,

$$\begin{aligned}\boldsymbol{\beta}^{(s)} &= \arg\min_{\boldsymbol{\beta}}\Big\|\mathbf{y}-\sum_{k=1}^{d}\mathbf{x}_k\beta_k\Big\|_2^2 + \sum_{j=1}^{J}\{\theta_j^{(s)}\}^{1-\frac{1}{\gamma}}c_j^{\frac{1}{\gamma}}\sum_{k\in A_j}\{\delta_k^{(s)}\}^{1-\frac{1}{\mu}}|\beta_k| \\ &= \arg\min_{\boldsymbol{\beta}}\Big\|\mathbf{y}-\sum_{k=1}^{d}\mathbf{x}_k\beta_k\Big\|_2^2 + \lambda_n\sum_{k=1}^{d}w_{1k}^{(s)}|\beta_k|,\end{aligned}$$

where

$$w_{1k}^{(s)} = \gamma\mu\sum_{j:k\in A_j}c_j\|\boldsymbol{\beta}_{A_j}^{(s-1)}\|_{\mu}^{\mu(\gamma-1)}|\beta_k^{(s-1)}|^{\mu-1}. \tag{2.4}$$

$k \leftarrow k+1$.

until convergence, i.e., $\|\boldsymbol{\beta}^{(s)}-\boldsymbol{\beta}^{(s-1)}\|_2/\|\boldsymbol{\beta}^{(s-1)}\|_2 < \epsilon$, where ϵ is some tolerance level, e.g., $\epsilon = 10^{-4}$.

(2006)] and group minimax concave penalization (GrMCP) [Breheny and Huang (2009)]. While our main focus is on bi-level selection and group selection to take advantage of the prior grouping information, the Lasso method [Tibshirani (1996)] is also included as a benchmark. For each method, we consider a grid of 200 values of its tuning parameter (equally spaced on the log scale) that produces a whole spectrum of candidate models, and select the best model via tenfold cross-validation, based on the predictive performance of the models.

Figure 2 displays a heatmap of FEV1% and the lung features used in the analysis. Each subject is represented by a column, and the columns are sorted by their observed FEV1% values. The topmost row gives the FEV1% values, followed by a block of variables selected by either CoB, CoMCP or GrB. For better visualization, each row is centered, divided by the maximum absolute value of its entries, and then multiplied by the sign of its sample correlation coefficient with the topmost row. As such, the entries of each adjusted row are always within -1 and 1, and if the corresponding variable has high predictive power, the row will show similar color change as the topmost row. There does not appear to be any single strong

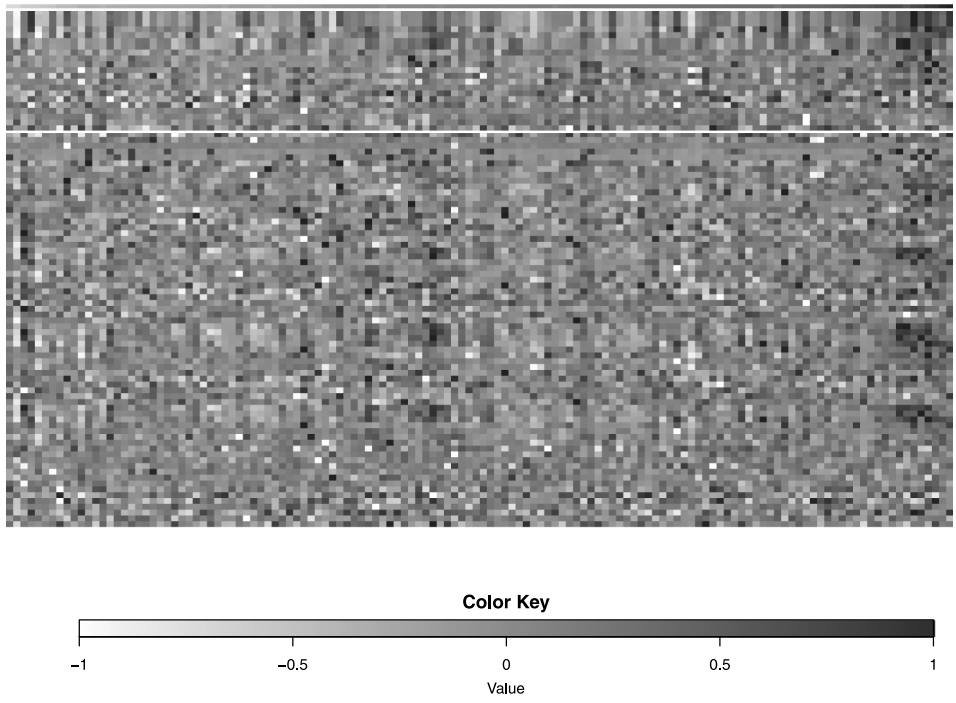


FIG. 2. *Heatmap of FEV*1% *and the lung features used in the analysis. Each subject is represented by a column, and the columns are sorted by their FEV*1% *value. The topmost row gives the FEV*1% *values, followed by a block of variables selected by either CoB, CoMCP or GrB, and another block of unselected variables.*

predictor for FEV1%. Indeed, the impact of the lung airway structure on FEV1% can be properly assessed only by a joint regression analysis.

Figure 3 displays a diagram showing the lung feature selection results by various methods. Table 1 tabulates the coefficient estimates of the selected models by the three bi-level selection methods, namely, CoB, CoMCP and GrB. The reported estimation and inference results are based on the least-squares refitting of the selected models [Efron (2004), Meinshausen and Bühlmann (2006)]. Both CoB and CoMCP selected seven feature groups, four of which are common. CoMCP additionally selected mean thickness, circularity and BSSD of area, all of which, however, do not contain significant variables as judged by the p-values from the least squares refitting. CoB additionally selected BSSD of wall thickness, WSSD of diameter and mean area, two of which contain significant variables. The GrB method tends to select fewer feature types attributed to within-group overselection, but the coefficients of the selected features have the same signs as those of the other methods. Out of the 9 nonzero coefficient estimates based on the CoB method, 6 of them are significant at the 5% level, while only 4 (5) out of 9 (11) nonzero GrB (CoMCP) estimates are significant. This suggests that the proposed

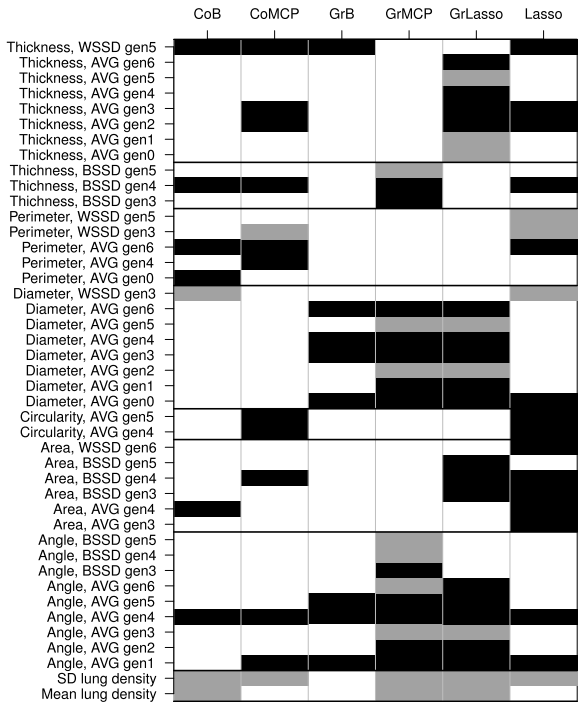


FIG. 3. *A diagram showing the selected features by various methods. Dark gray means the coefficient estimate of a selected variable is positive, light gray means the coefficient estimate of a selected variable is negative, and white means a variable is not selected.*

CoB method does a better job of picking the component variables in the selected features. These results are consistent with the findings from the simulation studies to be reported in the next section.

Figure 4 shows a good fit of the composite bridge regression model, and the residuals are approximately normally distributed. The following inference can be drawn based on the composite bridge regression fit:

1. FEV1% decreases significantly with the mean and the standard deviation of the parenchymal density at the 5% level.

2. Everything else being equal, higher FEV1% is associated with (i) higher within-segment variation in wall thickness (*fifth generation*, where a bold-faced generation indicates significance at the 5% level), (ii) higher between-segment variation in wall thickness (fourth generation), (iii) smaller within-segment variation in diameter (*third generation*), (iv) larger lumen area (*fourth generation*), (v) larger perimeter (zeroth and sixth generations), (vi) larger angle between parent segment and daughter segment (*fifth generation*).

Many of these findings are consistent with the current understandings regarding the association between lung function and airway distortion. The normal airways

TABLE 1

Variable selection and coefficient estimation results using composite bridge, composite MCP and group bridge. The reported estimation and inference results are based on refitting the selected models using the least squares method. BSSD (WSSD) stands for between-segment (within-segment) standard deviation

			Composite bridge			**Composite MCP**			**Group bridge**		
Feature	**Type**	**Gen#**	**Coef**	**std**	***p*-value**	**Coef**	**std**	***p*-value**	**Coef**	**std**	***p*-value**
Lung density	Mean		−0.037	0.012	2.50E−03						
	SD		−0.045	0.011	1.00E−04	−0.035	0.011	2.80E−03			
Wall thickness	Mean	2				0.011	0.010	2.66E−01			
	Mean	3				0.007	0.011	5.27E−01			
	WSSD	5	0.026	0.011	2.10E−02	0.035	0.011	1.20E−03	0.036	0.011	1.00E−03
	BSSD	4	0.019	0.010	6.86E−02						
Diameter	Mean	0							0.009	0.016	5.85E−01
	Mean	1							−0.003	0.017	8.74E−01
	Mean	3							0.010	0.016	5.12E−01
	Mean	4							0.048	0.015	2.10E−03
	Mean	6							0.012	0.010	2.52E−01
	WSSD	3	−0.026	0.010	7.00E−03						
Lumen area	Mean	4	0.044	0.012	3.00E−04						
	BSSD	4				0.009	0.012	4.82E−01			
Perimeter	Mean	0	0.020	0.011	6.56E−02						
	Mean	4				0.036	0.013	7.60E−03			
	Mean	6	0.016	0.010	1.13E−01	0.013	0.010	1.85E−01			
Angle	Mean	4				0.030	0.010	2.30E−03	0.034	0.010	1.30E−03
	Mean	5	0.026	0.010	7.60E−03	0.021	0.010	3.52E−02	0.024	0.010	1.76E−02
	Mean	5							0.008	0.010	4.38E−01
Circularity	Mean	4				0.018	0.012	1.29E−01			
	Mean	5				0.019	0.011	9.73E−02			
		#variables	9			11			9		
		Adjusted R^2	0.387			0.392			0.323		

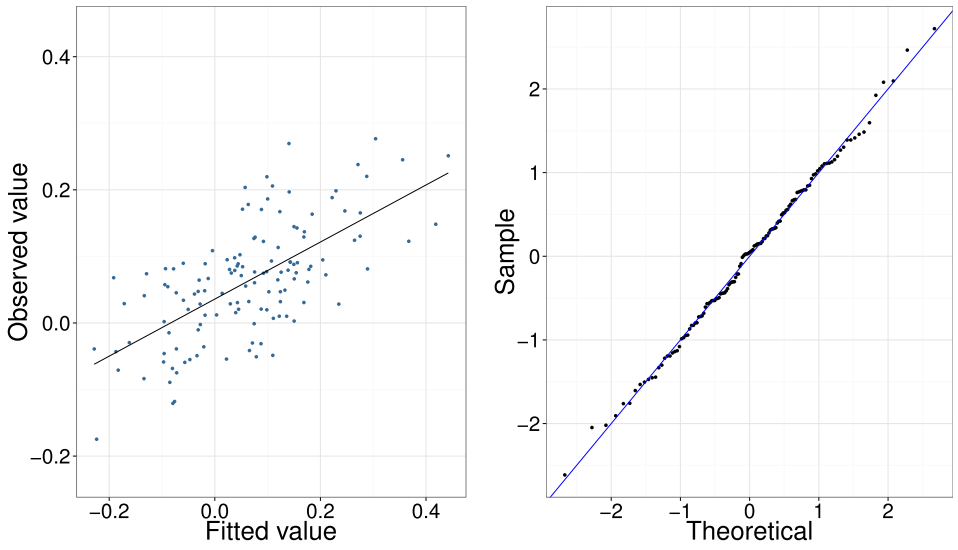


FIG. 4. *Left diagram*: *scatter plot of the observed FEV*1% *values vs the fitted values*. *Right Diagram*: *the normal Q–Q plot for the residuals*.

would have segments with a more spacious lumen (items iv and v), larger within-segment homogeneity in diameter (item iii) and larger branch angles (item vi); see Nakano et al. (2009) and Montesantos et al. (2013). Moreover, these effects are revealed after adjusting for those of the mean and standard deviation of parenchymal densities, both of which have negative impacts on the FEV1%, which is expected as normal lung parenchyma would be neither too dense nor too heterogeneous.

The finding on the positive association between the variation in wall thickness with FEV1% may seem somewhat surprising. But it has recently been demonstrated that airway segments of subjects with chronic obstructive pulmonary disease (COPD) have significantly thinner walls [Smith et al. (2014)], and so our finding may indicate that some variation in wall thickness is a symptom of a normal lung airway.

3. Theoretical properties and empirical performance of composite bridge.

3.1. *Theoretical properties*. We have explored the theoretical properties of the composite bridge estimator following the framework developed by Huang, Horowitz and Ma (2008) and Huang et al. (2009). All the technical statements of our findings are given in the Appendix, and the proofs are provided in the Supplementary Material [Chen et al. (2016)]. Here we briefly discuss the rationales of the theoretical framework and the implications of our main findings.

The technical conditions given in A1–A5 concern the error distribution, the degree of variable overlapping in group specification, the rate of the tuning parameter and the model size, which are all standard. Theorem A.1 provides an estimation error bound of composite bridge, which quantifies how the estimation performance is affected by the noise level, the number of predictors, the sample size, etc. Theorems A.2 and A.3 establish the variable selection properties of the composite bridge method, that is, for any predetermined $0 < \gamma, \mu < 1$, the composite bridge estimator can identify the correct groups and the correct nonzero elements within each selected group with probability converging to one. In Theorem A.4, we further show that the composite bridge estimators of the nonzero coefficients are asymptotically normal with the limiting distribution the same as the case when the true sparsity pattern is known a priori. Therefore, Theorems A.2–A.4 together imply that the composite bridge achieves the powerful *bi-level oracle property*. To the best of our knowledge, this property of the composite penalization approach is established for the first time.

In our analysis, we require a full rank design so that $d_n \leq n$, where d_n denotes the number of covariates, but d_n is allowed to grow at a certain rate as $n \to \infty$. To ensure group selection consistency, we allow $d_n = o(1)n^{(1-\mu\gamma)/(2-\mu\gamma)}$, which is faster than $o(1)n^{(1-\gamma)/(2-\gamma)}$, the rate allowed by the group bridge method [Huang et al. (2009)]. This shows that choosing the nonconvex bridge penalty as the inner penalty further improves group selection. To achieve bi-level selection consistency, our method allows $d_n = o(1)n^{(1-\mu)/(2-\mu)}$, which is essentially the same as the rate allowed by bridge regression [Huang, Horowitz and Ma (2008)] under similar conditions. From the group selection perspective, the established bi-level oracle properties show that our method is *adaptive* to various scenarios of potential group configurations. In particular, our method can achieve a faster rate (comparing to group bridge) in identifying irrelevant groups, that is, groups containing only variables with zero coefficients. For selecting relevant predictors from the mixed groups, that is, groups containing both relevant predictors (with nonzero coefficients) and irrelevant predictors (with zero coefficients), our method is as good as the individual variable selection method (comparing to bridge regression). Therefore, the composite bridge method flexibly incorporates the prior grouping information and is adaptive to the potential presence of mixed groups.

3.2. *Simulation*. We compare the empirical performances of the proposed composite bridge (CoB) method, group bridge (GrB), composite mimimax concave penalization (CoMCP) [Zhang (2010)], group Lasso (GrLasso), group minimax concave penalization (GrMCP) and Lasso [Tibshirani (1996)]. The GrMCP and CoMCP estimators are computed using the *grpreg* package [Breheny and Huang (2009)] in R [R Development Core Team (2015)]; all other methods are implemented in R.

We consider several examples covering various practical scenarios, for example, bi-level sparsity, group sparsity, varying group sizes, correlation within/among

groups, etc. In examples 1–3 to be elaborated below, the number of predictors is around 40 and the sample size is $n = 200$, adopting similar setups as the examples in Huang et al. (2009). The setup of example 4 mimics the lung airway application. In the Supplementary Material [Chen et al. (2016)], we provide another example for investigating the performance of the proposed method under the high-dimensional scenario and the impact of group size. Under each setting, the simulation experiment is replicated 400 times.

EXAMPLE 1 ($n = 200$, $d = 42$). There are $J = 6$ groups of variables, with $|A_1| = |A_2| = |A_3| = 10$ and $|A_4| = |A_5| = |A_6| = 4$. To generate $d = 42$ covariates, we first form $n \times 1$ vectors $\mathbf{r}_1, \ldots, \mathbf{r}_d$ and $\mathbf{z}_1, \ldots, \mathbf{z}_J$; all the entries in these vectors are independently generated from $N(0, 1)$. The covariates $\mathbf{x}_1, \ldots, \mathbf{x}_d$ are generated as

$$\mathbf{x}_k = (\mathbf{z}_{g_k} + \mathbf{r}_k)/\sqrt{2}, \qquad 1 \le k \le d,$$

where $(g_1, \ldots, g_d) = (\text{rep}(1, 10), \text{rep}(2, 10), \text{rep}(3, 10), \text{rep}(4, 4), \text{rep}(5, 4), \text{rep}(6, 4))$, indicating the group membership structure, where $\text{rep}(a, l)$ denotes a vector of length l whose entries are all equal to a. Therefore, the covariates within each group are correlated, while the covariates from different groups are uncorrelated. The response $\mathbf{y}$ is then generated using model (2.1), where

$$\begin{aligned}
\boldsymbol{\beta}_{0A_1} &= (1, -2, 1.25, 1, -1, 1, 3, -1.5, 2, -2)', \\
\boldsymbol{\beta}_{0A_2} &= (-1.5, 3, 1, -2, 1.5, 0, 0, 0, 0, 0)', \\
\boldsymbol{\beta}_{0A_3} &= (0, \ldots, 0)', \qquad \boldsymbol{\beta}_{0A_4} = (2, -2, 1, 1.5)', \\
\boldsymbol{\beta}_{0A_5} &= (-1.5, 1.5, 0, 0)', \qquad \boldsymbol{\beta}_{0A_6} = (0, \ldots, 0)',
\end{aligned}$$

and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, 4\mathbf{I})$.

EXAMPLE 2 ($n = 200$, $d = 42$). The model is the same as in Example 1, except that $\boldsymbol{\beta}_{0A_2} = (-1.5, 3, 0, \ldots, 0)'$ and $\boldsymbol{\beta}_{0A_4} = (2, 0, 0, 0)'$, and so there are several very sparse non-null groups.

EXAMPLE 3 ($n = 200$, $d = 40$). In this example, all the coefficients in a nonzero group are nonzero, and so this is an ideal group selection setup. We set $J = 5$, $|A_1| = \cdots = |A_5| = 8$ and $d = 40$. We first simulate $\mathbf{r}_1, \ldots, \mathbf{r}_{40}$ independently from $N(\mathbf{0}, \mathbf{I})$. Next, to generate $\mathbf{z}_j$ vectors ($j = 1, \ldots, J$), we simulate n independent samples from a J-dimensional Gaussian distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$, where the (h, l)th entry of $\boldsymbol{\Sigma}$ equals $\sigma_{hl} = 0.4^{|h-l|}$. Then the covariates $\mathbf{x}_1, \ldots, \mathbf{x}_{40}$ are generated as

$$\mathbf{x}_{8(j-1)+k} = \{\mathbf{z}_j + \mathbf{r}_{8(j-1)+k}\}/\sqrt{2}. \qquad 1 \le j \le 5, 1 \le k \le 8.$$

In this way, the AR(1) correlation structure of the $\mathbf{z}_j$s induces correlation across different groups of covariates. The response vector is computed using model (2.1), where

$$\boldsymbol{\beta}_{0A_1} = (1, 1, 1.5, 2, 2.5, 3, 3.5, 4)', \qquad \boldsymbol{\beta}_{0A_2} = (2, 2, 2, 2, 2, 2, 2, 2)',$$
$$\boldsymbol{\beta}_{0A_3} = \boldsymbol{\beta}_{0A_4} = \boldsymbol{\beta}_{0A_5} = \mathbf{0}', \quad \text{and} \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, 4\mathbf{I}).$$

EXAMPLE 4 ($n = 132, d = 90$). In this example, the dimensions and the group configurations are exactly the same as the lung airway data application. Based on the feature types of the lung airway variables, there are $J = 17$ variable groups, with group size varying from 2 to 7. The true model structure is the same as the one selected by the composite bridge method, and the true nonzero coefficients are set to be ten times of the estimates in Table 2 so that the values range from 1.5 to 4.5 in magnitude. The rest of the setup is the same as that of example 3, hence, both within-group and between-group correlations exist among the generated predictors.

In our simulation study, to alleviate inaccuracy in the empirical tuning for ensuring a fair comparison, we tune each method based on its predictive accuracy evaluated with an independently generated validation data set of sample size 500. For each method, we compute the solutions over a grid of 200 values of its tuning parameter (equally spaced on the log scale) that produces a whole spectrum of candidate models, and then select the model with the smallest prediction error. We have also tried using BIC [Fan and Tang (2013), Schwarz (1978)] for tuning, and, as expected, in general the variable selection performance improves while the estimation accuracy becomes slightly worse. The results are consistent with our findings reported below, and hence are omitted for brevity.

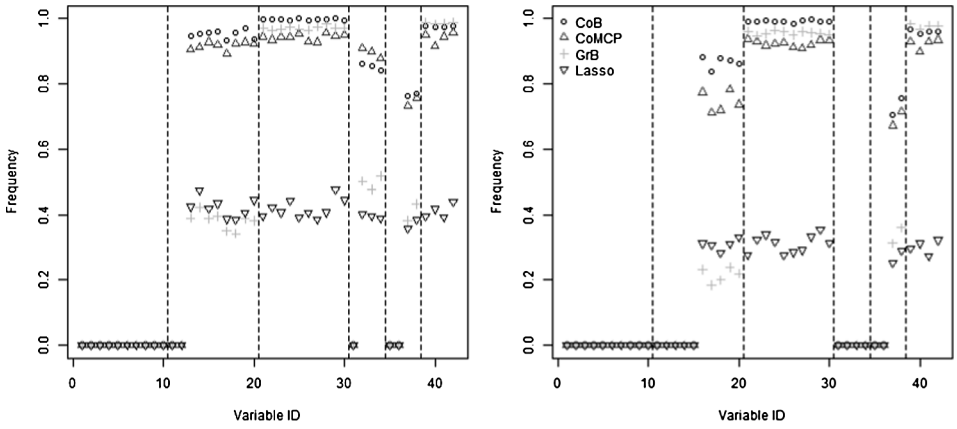
For each method, the model accuracy is measured by the average of the mean squared errors from all runs (Model Error), that is, $\|\mathbf{X}\boldsymbol{\beta}_0 - \mathbf{X}\hat{\boldsymbol{\beta}})\|^2/n$. To evaluate the group-level selection accuracy, we compute the average number of selected nonzero groups (No. of Groups) and the frequency of correct identification of the group sparsity structure (Correct Groups). To evaluate the individual variable selection performance, we compute the average number of nonzero coefficients (No. of Var.), the false negative rate (FNR) and the false discovery rate (FDR).

Tables 2 summarizes the simulation results for examples 1–4. Examples 1–2 are in favor of the bi-level selection methods. The CoB and CoMCP methods have comparable performance in terms of estimation accuracy and variable selection, and they both outperform GrB. CoMCP tends to select slightly more groups than needed, and yet it has the best estimation performance and maintains a low false discovery rate comparable to that of CoB. This indicates that CoMCP may falsely select a few variables from the null groups, behaving more toward an individual variable selection method [Breheny (2015)]. In contrast, GrB yields many more

TABLE 2

Simulation results for examples 1–4. *CoB*, *composite bridge*; *GrB*, *group bridge*; *CoMCP*, *composite MCP*; *GrMCP*, *group MCP*; *GrLasso*, *group Lasso*. *The average values* (*avg*) *are based on* 400 *simulation runs*, *and the standard deviations* (*sd*) *are also reported*

		CoB	**CoMCP**	**GrB**	**GrMCP**	**GrLasso**	**Lasso**
				Example 1			
Model Error	avg	0.54	0.52	0.60	0.63	0.82	0.82
	sd	0.19	0.19	0.19	0.20	0.23	0.23
No. of Groups (4)	avg	4.16	4.77	4.17	4.34	5.92	5.99
	sd	0.41	0.70	0.42	0.62	0.27	0.05
Correct Group (%)		86.3%	38.8%	73.8%	85.3%	0.0%	0.0%
No. of Vars (21)	avg	22.45	23.95	26.79	30.49	41.68	35.59
	sd	1.60	2.12	2.26	4.56	1.16	2.95
FDR (%)		6.0%	11.7%	21.1%	29.8%	49.6%	40.6%
FNR (%)		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
				Example 2			
Model Error	avg	0.41	0.36	0.50	0.64	0.82	0.70
	sd	0.16	0.14	0.18	0.20	0.23	0.21
No. of Groups (4)	avg	4.08	4.63	4.11	4.38	5.93	5.97
	sd	0.28	0.67	0.34	0.65	0.26	0.17
Correct Group (%)		91.8%	47.5%	72.0%	90.5%	0.0%	0.0%
No. of Vars (16)	avg	16.42	17.30	22.98	30.75	41.71	30.87
	sd	1.47	1.94	2.81	4.85	1.12	3.67
FDR (%)		8.0%	12.3%	33.8%	50.2%	64.0%	50.7%
FNR (%)		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
				Example 3			
Model Error	avg	0.36	0.36	0.36	0.35	0.53	0.46
	sd	0.14	0.14	0.14	0.14	0.19	0.17
No. of Groups (2)	avg	2.16	2.75	2.13	2.47	4.59	4.58
	sd	0.42	0.94	0.39	0.78	0.60	0.64
Correct Group (%)		85.8%	52.8%	67.5%	88.5%	0.25%	0.5%
No. of Vars (16)	avg	16.24	17.03	16.35	19.78	36.74	21.59
	sd	0.71	1.54	1.26	6.25	4.78	2.77
FDR (%)		1.3%	5.2%	2.0%	14.1%	55.7%	24.8%
FNR (%)		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
				Example 4			
Model Error	avg	2.77	2.78	3.75	7.73	9.49	4.08
	sd	1.31	1.22	1.43	2.13	2.48	1.49
No. of Groups (7)	avg	6.87	8.85	7.60	8.28	14.75	13.87
	sd	0.95	1.75	1.27	1.69	1.43	1.89
Correct Group (%)		46.3%	12.0%	32.5%	6.0%	0.0%	0.0%
No. of Vars (9)	avg	9.94	13.66	20.90	46.34	80.47	27.90
	sd	2.13	3.56	5.61	9.26	6.84	5.84
FDR (%)		13.5%	33.7%	56.0%	81.3%	88.8%	66.9%
FNR (%)		7.2%	4.6%	3.7%	6.6%	0.8%	1.7%



(a) Example 1 (b) Example 2

FIG. 5. *Relative frequency plot of each covariate not being selected in examples* 1–2. *The vertical lines indicate the group structure. As neither of the methods yields a false negative in the two examples, the four types of symbols at the bottom of each figure are overlaid.*

false positives due to its inner ℓ_1 penalty, but it behaves better than CoMCP at the group-level selection. For each covariate in a simulated model, we plot in Figures 5 the relative frequency that it is not selected. It is clear that the GrB method often yields false positives in non-null groups, and the other two bi-level selection methods perform much better due to their enhanced individual selection property. Not surprisingly, all three bi-level methods outperform GrMCP, GrLasso and Lasso, and more so in example 2 when several non-null groups are quite sparse. In example 3, the model is only sparse at the group level, which has exactly the structure the group selection methods aim to recover. While the performance of GrMCP and GrLasso improves, the bi-level selection methods, especially CoB, may still outperform them by a considerable margin. This shows that individual-level section consistency may reinforce the selection at the group level, which agrees with our theoretical analysis. We note that both CoB and GrB can handle overlapping of predictors in group assignments. We thus have examined the effect of overlapping by modifying example 1 to allow a few important or irrelevant predictors to be grouped into more than one group. Both CoB and GrB remain to work well, as the simulation results (unreported) are very similar to those of the reported nonoverlapping case. Example 4 uses the same group structure and model dimensions as in the real application, and allows for both within-group and between-group correlations. Here the sample size is comparable to the number of predictors, which makes the problem more challenging. The performances of the CoB and CoMCP methods are comparable, except that CoMCP tends to select a few more variables than needed and CoB has slightly larger false negatives, although they were both tuned based on prediction error. Again, in this bi-level selection scenario, these two

nonconvex composite penalization methods substantially outperform other methods.

4. Discussion. We have developed a comprehensive procedure for extracting parsimonious and interpretable lung airway features from noisy and high-dimensional lung airway tree data, and then study their association with lung function variables using a new composite bridge regression method. In particular, our findings indicate the importance of the between-segment and within-segment heterogeneities in studying lung function. The found relationships between lung function and airway characteristics indicate the potential for future work to collect more data in order to derive a systematic approach for assessing FEV1% (and other pulmonary function tests) after adjusting for the subject's parenchymal density and airway tree features, which has the potential of more accurately identifying "normal" subjects with borderline atypical lung functions.

We have established the bi-level oracle properties for the composition of the bridge penalty in penalized regression, and demonstrated that the method can take full advantage of prior grouping information and be adaptive to the presence of mixed groups. In the theoretical analysis we allow a diverging number of predictors but still require $n > d_n$. There are many directions for future research. Our results may shed light on developing a unified theory for bi-level selection via the general nonconvex composite penalty forms under high-dimensional setups [Huang, Breheny and Ma (2012)]. Bi-level and hierarchical variable selection techniques are critical in the integrative analysis of multiple data sets, especially in high-throughput genomic studies [Liu, Ma and Huang (2014), Ma et al. (2011)]. It is thus pressing to develop efficient algorithms and investigating properties of multiple-level penalization. In time series analysis, the model terms in an ARMA model form an AR group, an MA group and groups associated with exogenous variables, and they may also be grouped based on the lag orders. Motivated by Chen and Chan (2011), it would be interesting to explore the use of bi-level penalization methods to identify a parsimonious subset ARMA model. In multivariate regression, there are naturally grouped coefficients associated with different responses, predictors or latent structures in which bi-level selection can be very effective, for example, it can be used to identify the relevant predictors as well as the subset of selected predictors associated with each response variable [Chen, Chan and Stenseth (2012), Chen and Huang (2012)].

APPENDIX: THEORETICAL PROPERTIES OF THE COMPOSITE BRIDGE METHOD

Recall that, without loss of generality, we have assumed that

$$\begin{aligned} \boldsymbol{\beta}_{0A_j} &\neq \mathbf{0}, \qquad 1 \le j \le J_1, \\ \boldsymbol{\beta}_{0A_j} &= \mathbf{0}, \qquad J_1 + 1 \le j \le J. \end{aligned}$$

For each A_j, $j = 1, \ldots, J_1$, $A_j^1 = \{k; \beta_{0k} \neq 0, k \in A_j\}$ and $A_j^2 = \{k; \beta_{0k} = 0, k \in A_j\}$. Let $B_2 = \bigcup_{j=J_1+1}^{J} A_j$ be the union of the groups with zero coefficients. Let $B_1 = B_2^c$, $B_1^1 = \{k; \beta_{0k} \neq 0, k \in B_1\}$ and $B_1^2 = \{k; \beta_{0k} = 0, k \in B_1\}$. Note that A_j may include relevant predictors or irrelevant predictors either in B_1 or B_2. Denote $\boldsymbol{\beta}_{0B_j} = (\beta_{0k}, k \in B_j)'$ for $j = 1, 2$, and define other subvectors of $\boldsymbol{\beta}_0$ similarly. Assume the variables are arranged so that $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}'_{0B_1^1}, \boldsymbol{\beta}'_{0B_1^2}, \boldsymbol{\beta}'_{0B_2})'$. Since $\boldsymbol{\beta}_{0B_1^2} = 0$ and $\boldsymbol{\beta}_{0B_2} = 0$, the response variable is fully explained by the relevant variables belonging to B_1^1 within the first J_1 groups. In this notation, $\hat{\boldsymbol{\beta}}_{nB_1^1}$, $\hat{\boldsymbol{\beta}}_{nB_1^2}$ and $\hat{\boldsymbol{\beta}}_{nB_2}$ are respectively the estimates of $\boldsymbol{\beta}_{0B_1^1}$, $\boldsymbol{\beta}_{0B_1^2}$ and $\boldsymbol{\beta}_{0B_2}$ from the composite bridge estimator $\hat{\boldsymbol{\beta}}_n$.

Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d)$, $\mathbf{X}_1 = (\mathbf{x}_k, k \in B_1)$, $\mathbf{X}_{11} = (\mathbf{x}_k, k \in B_1^1)$ and $\mathbf{X}_{12} = (\mathbf{x}_k, k \in B_1^2)$. Define

$$\boldsymbol{\Sigma}_n = \frac{1}{n}\mathbf{X}'\mathbf{X}, \qquad \boldsymbol{\Sigma}_{1n} = \frac{1}{n}\mathbf{X}_1'\mathbf{X}_1, \qquad \boldsymbol{\Sigma}_{11n} = \frac{1}{n}\mathbf{X}_{11}'\mathbf{X}_{11}. \tag{A.1}$$

Let ρ_n and ρ_n^* be the smallest and largest eigenvalues of $\boldsymbol{\Sigma}_n$, and let τ_{1n} and τ_{1n}^* be the smallest and largest eigenvalues of $\boldsymbol{\Sigma}_{11n}$.

We consider the following conditions:

A1. The errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are uncorrelated with mean zero and finite variance σ^2.

A2. The $c_n^* = \max_k \sum_{j=1}^{J} I(k \in A_j)$ is bounded, and

$$\frac{\lambda_n^2 \eta_n^2}{n\rho_n \sigma^2 d} = M_n = O(1),$$

where $\eta_n = \{\sum_{j=1}^{J_1} c_j^2 \|\boldsymbol{\beta}_{0A_j^1}\|_{2\mu-2}^{2\mu-2} \|\boldsymbol{\beta}_{0A_j^1}\|_{\mu}^{2\mu(\gamma-1)}\}^{1/2}$.

A3. The constants c_j's are scaled to satisfy $\min_{j \leq J} c_j \geq 1$, and

$$\frac{\lambda_n \rho_n^{1-\mu\gamma/2}}{d^{1-\mu\gamma/2} \rho_n^* n^{\mu\gamma/2}} \to \infty.$$

A3*. The constants c_j's are scaled to satisfy $\min_{j \leq J} c_j \geq 1$, and

$$\frac{\lambda_n \rho_n^{1-\mu/2}}{d^{1-\mu/2} \rho_n^* n^{\mu/2}} \to \infty.$$

A4. There exists constant $\tau_1^* < \infty$ such that $\tau_{1n}^* \leq \tau_1^*$ for all n.

Assumption A1 is standard about the error distribution. Assumptions A2 and A3 are about the degree of overlapping, the growth rate of the tuning parameter and the growth rate of the model size; they imply a full rank design with $\operatorname{rank}(\mathbf{X}) = d \leq n$, $\rho_n > 0$, and $\tau_{1n} > 0$. The first three assumptions are used to establish the group-level selection consistency. To establish the individual-level selection consistency,

however, A3 shall be replaced by a stronger version A3*. A4 ensures that the largest eigenvalue of $\mathbf{X}_{11}$ is bounded.

In general, the selection consistency at the individual level is stronger than that at the group level. This fact is reflected in the above required assumptions. Note that A3* implies A3, as individual-level selection consistency implies group-level consistency. Moreover, the choice of the group-level penalty does not have a direct impact on the within-group variable selection. On the other hand, A3 involves both μ and γ, and it is evident that the choice of the within-group or inner penalty determines the behavior of individual variable selection, and hence also influences the group selection performance. Similar to Huang et al. (2009), for $[B_1^1, \boldsymbol{\beta}_{0B_1^1}, J_1]$ fixed but unknown, Assumptions A2 and A3 hold when

$$\text{(A.2)}\qquad \begin{aligned} &\text{(a) } (1/\rho_n) + \rho_n^* + \sum_{j=1}^{J_1} c_j^2 = O(1), \qquad \text{(b) } \frac{\lambda_n}{n^{1/2}} \to \lambda_0 < \infty, \\ &\text{(c) } \frac{\lambda_n d^{\mu\gamma/2}}{dn^{\mu\gamma/2}} \to \infty, \end{aligned}$$

provided that $c_j \geq 1$ and $c_n^* = O(1)$. For Assumptions A2 and A3*, (c) is strengthened to

$$\text{(A.3)}\qquad \text{(c*) } \frac{\lambda_n d^{\mu/2}}{dn^{\mu/2}} \to \infty.$$

Our main results about the properties of the proposed composite bridge estimator are summarized in the following theorems. We remark that Theorems 1 and 2 also apply to the group bridge method when $\mu = 1$ [Huang et al. (2009)]. All proofs are provided in the Supplementary Material [Chen et al. (2016)].

THEOREM A.1 (Estimation Error Bound). *Suppose that* $0 < \mu \leq 1, 0 < \gamma < 1$ *and Assumptions* A1–A2 *hold. Then*

$$E\big(\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|_2^2\big) \leq \frac{\sigma^2 d}{n\rho_n}\big(8 + 64 c_n^* M_n\big).$$

THEOREM A.2 (Group Selection Consistency). *Suppose that* $0 < \mu \leq 1, 0 < \gamma < 1$ *and Assumptions* A1–A3 *hold. Then*

$$\text{(A.4)}\qquad P(\hat{\boldsymbol{\beta}}_{nA_j} = 0, j > J_1) \to 1,$$

as $n \to \infty$.

THEOREM A.3 (Individual Selection Consistency). *Suppose that* $0 < \mu < 1$, $0 < \gamma \leq 1$ *and Assumptions* A1, A2, A3* *and* A4 *hold. Then* (A.4) *holds, and*

$$P(\hat{\boldsymbol{\beta}}_{nA_j^2} = 0, j \leq J_1) \to 1,$$

as $n \to \infty$.

THEOREM A.4 (Asymptotic Distribution). *Suppose* $\{B_1^1, \boldsymbol{\beta}_{0B_1^1}, J_1\}$ *are fixed unknowns and* (A.2) *holds. Suppose further that* $\boldsymbol{\Sigma}_{1n} \to \boldsymbol{\Sigma}_1$ *and* $n^{-1/2}\mathbf{X}_1'\boldsymbol{\varepsilon}_1 \to_d \mathbf{W}_1 \sim N(0, \sigma^2\boldsymbol{\Sigma}_1)$, *and, consequently,* $\boldsymbol{\Sigma}_{11n} \to \boldsymbol{\Sigma}_{11}$ *and* $n^{-1/2}\mathbf{X}_{11}'\boldsymbol{\varepsilon}_{11} \to_d \mathbf{W}_{11} \sim N(0, \sigma^2\boldsymbol{\Sigma}_{11})$. *Then*

$$\sqrt{n}\hat{\boldsymbol{\beta}}_{nB_2} \underset{d}{\to} \mathbf{0}, \qquad \sqrt{n}\hat{\boldsymbol{\beta}}_{nB_1^2} \underset{d}{\to} \mathbf{0},$$

and

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{nB_1^1} - \boldsymbol{\beta}_{0B_1^1}) \underset{d}{\to} \arg\min V_{11}(\mathbf{u}), \qquad \mathbf{u} \in \mathbb{R}^{|B_1^1|},$$

where

$$V_{11}(\mathbf{u}) = -2\mathbf{u}'\mathbf{W}_{11} + \mathbf{u}'\boldsymbol{\Sigma}_{11}\mathbf{u} + \mu\gamma\lambda_0 \sum_{j=1}^{J_1} c_j \|\boldsymbol{\beta}_{0A_j}\|_\mu^{\mu(\gamma-1)} \sum_{k\in A_j^1} u_k |\beta_{0k}|^{\mu-1} \operatorname{sgn}(\beta_{0k}).$$

Acknowledgment. The authors thank Mr. Dakai Jin for drawing and labeling the airway trees. The authors are grateful to the referee, the Associate Editor and the Editor for their valuable comments and suggestions.

SUPPLEMENTARY MATERIAL

Supplement to"Linking lung airway structure to pulmonary function via composite bridge regression" (DOI: 10.1214/16-AOAS947SUPP; .pdf). We provide the technical details in the theoretical investigation of the proposed method and an additional simulation example to investigate the impact of group size.

REFERENCES

BECKLAKE, M. R. (1985). Concepts of normality applied to the measurement of lung function. *Am. J. Med.* **80** 1158–1164.

BREHENY, P. (2015). The group exponential lasso for bi-level variable selection. *Biometrics* **71** 731–740. MR3402609

BREHENY, P. and HUANG, J. (2009). Penalized methods for bi-level variable selection. *Stat. Interface* **2** 369–380. MR2540094

BREHENY, P. and HUANG, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Stat.* **5** 232–253. MR2810396

CHEN, K. and CHAN, K.-S. (2011). Subset ARMA selection via the adaptive Lasso. *Stat. Interface* **4** 197–205. MR2812815

CHEN, K., CHAN, K.-S. and STENSETH, N. CHR. (2012). Reduced rank stochastic regression with a sparse singular value decomposition. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 203–221. MR2899860

CHEN, L. and HUANG, J. Z. (2012). Sparse reduced-rank regression for simultaneous dimension reduction and variable selection. *J. Amer. Statist. Assoc.* **107** 1533–1545. MR3036414

CHEN, K., HOFFMAN, E. A., SEETHARAMAN, I., JIAO, F., LIN, C.-L. and CHAN, K.-S. (2016). Supplement to "Linking lung airway structure to pulmonary function via composite bridge regression." DOI:10.1214/16-AOAS947SUPP.

EFRON, B. (2004). The estimation of prediction error: Covariance penalties and cross-validation. *J. Amer. Statist. Assoc.* **99** 619–642. MR2090899

FAN, Y. and TANG, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 531–552. MR3065478

FRIEDMAN, J. H., HASTIE, T. J. and TIBSHIRANI, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1–22.

FULD, M. K., GROUT, R. W., GUO, J., MORGAN, J. H. and HOFFMAN, E. (2012). Systems for lung volume standardization during static and dynamic MDCT-based quantitative assessment of pulmonary structure and function. *Acad. Radiol.* **19** 930–940.

GAO, W. (2010). Development of human lung query atlas. Dissertation, Univ. Iowa.

GUO, J., FULD, M. K., ALFORD, S. K., REINHARDT, J. M. and HOFFMAN, E. A. (2008). Pulmonary Analysis Software Suite 9.0: Integrating quantitative measures of function with structural analyses. In *First International Workshop on Pulmonary Image Analysis* 283–292.

HANKINSON, J. L., ODENCRANTZ, J. R. and FEDAN, K. B. (1999). Spirometric reference values from a sample of the general U.S. population. *Am. J. Respir. Crit. Care Med.* **159** 179–187.

HOFFMAN, E. A., SIMON, B. A. and MCLENNAN, G. (2006). State of the art. A structural and functional assessment of the lung via multidetector-row computed tomography: Phenotyping chronic obstructive pulmonary disease. *Proc. Am. Thorac. Soc.* **3** 519–532.

HUANG, J., BREHENY, P. and MA, S. (2012). A selective review of group selection in high-dimensional models. *Statist. Sci.* **27** 481–499. MR3025130

HUANG, J., HOROWITZ, J. L. and MA, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Ann. Statist.* **36** 587–613. MR2396808

HUANG, J., MA, S., XIE, H. and ZHANG, C.-H. (2009). A group bridge approach for variable selection. *Biometrika* **96** 339–355. MR2507147

IYER, K. S., GRANT, R. W., ZAMBA, G. K. and HOFFMAN, E. A. (2014). Repeatability and sample size assessment associated with computed tomography-based lung density metrics. *Journal of the COPD Foundation* **1** 97–104.

LIU, J., MA, S. and HUANG, J. (2014). Integrative analysis of cancer diagnosis studies with composite penalization. *Scand. Stat. Theory Appl.* **41** 87–103. MR3181134

MA, S., HUANG, J., WEI, F., XIE, Y. and FANG, K. (2011). Integrative analysis of multiple cancer prognosis studies with gene expression measurements. *Stat. Med.* **30** 3361–3371. MR2861619

MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363

MONTESANTOS, S., KATZ, I., FLEMING, J., MAJORAL, C., PICHELIN, M., DUBAU, C., PIEDNOIR, B., CONWAY, J., TEXEREAU, J. and CAILLIBOTTE, G. (2013). Airway morphology from high resolution computed tomography in healthy subjects and patients with moderate persistent asthma. *Anat Rec* (*Hoboken*) **296** 852–866.

NAKANO, Y., THO, N. V., YAMADA, H., OSAWA, M. and NAGAO, T. (2009). Radiological approach to asthma and COPD—the role of computed tomography. *Allergol. Intern.* **58** 323–331.

PALAGYI, K., TSCHIRREN, J., HOFFMAN, E. A. and SONKA, M. (2006). Quantitative analysis of pulmonary airway tree structuress. *Comput. Biol. Med.* **36** 974–976.

R DEVELOPMENT CORE TEAM (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014

SMITH, B. M., HOFFMAN, E. A., RABINOWITZ, D., BLEECKER, E., CHRISTENSON, S., COUPER, D., DONOHUE, K. M., HAN, M. K., HANSEL, N. N., KANNER, R. E. et al. (2014).

Comparison of spatially matched airways reveals thinner airway walls in COPD. The Multi-Ethnic Study of Atherosclerosis (MESA) COPD Study and the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). *Thorax* **69** 987–996.

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

TSCHIRREN, J., HOFFMAN, E. A., MCLENNAN, G. and SONKA, M. (2005a). Intrathoracic airway trees: Segmentation and airway morphology analysis from low-dose CT scans. *IEEE Trans. Med. Imag.* **24** 1529–1539.

TSCHIRREN, J., HOFFMAN, E. A., MCLENNAN, G. and SONKA, M. (2005b). Segmentation and quantitative analysis of intrathoracic airway trees from computed tomography images. *Proc. Am. Thorac. Soc.* **2** 484–7, 503–4.

TSCHIRREN, J., MCLENNAN, G., PALAGYI, K., HOFFMAN, E. A. and SONKA, M. (2005c). Matching and anatomical labeling of human airway tree. *Comput. Biol. Med.* **24** 1540–1547.

WEIBEL, E. R. (2015). How Benoit Mandelbrot changed my thinking about biological form. *Benoit Mandelbrot: A Life in Many Dimensions* **1** 471–487.

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 49–67. MR2212574

ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701

ZHANG, C., JIANG, Y. and CHAI, Y. (2010). Penalized Bregman divergence for large-dimensional regression and classification. *Biometrika* **97** 551–566. MR2672483

ZHAO, P., ROCHA, G. and YU, B. (2009). The composite absolute penalties family for grouped and hierarchical variable selection. *Ann. Statist.* **37** 3468–3497. MR2549566

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469

ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443

K. CHEN
DEPARTMENT OF STATISTICS
UNIVERSITY OF CONNECTICUT
STORRS, CONNECTICUT 06029
USA
E-MAIL: kun.chen@uconn.edu

I. SEETHARAMAN
DEPARTMENT OF STATISTICS
KANSAS STATE UNIVERSITY
MANHATTAN, KANSAS 66506
USA
E-MAIL: inds@ksu.edu

E. A. HOFFMAN
DEPARTMENT OF RADIOLOGY
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242
USA
E-MAIL: eric-hoffman@uiowa.edu

F. JIAO
K.-S. CHAN
DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242
USA
E-MAIL: feiran-jiao@uiowa.edu
kung-sik-chan@uiowa.edu

C.-L. LIN
DEPARTMENT OF MECHANICAL
AND INDUSTRIAL ENGINEERING
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242
USA
E-MAIL: ching-long-lin@uiowa.edu