Taylor & Francis
Taylor & Francis Group

Check for updates

A Conjugate Model for Dimensional Analysis

Weijie Shen and Dennis K. J. Lin

Department of Statistics, The Pennsylvania State University, State College, PA

ABSTRACT

Dimensional analysis (DA) is a methodology widely used in physics and engineering. The main idea is to extract key variables based on physical dimensions. Its overlooked importance in statistics has been recognized recently. However, most literature treats DA as merely a preprocessing tool, leading to multiple statistical issues. In particular, there are three critical aspects: (a) the nonunique choice of basis quantities and dimensionless variables; (b) the statistical representation and testing of DA constraints; (c) the spurious correlations between post-DA variables. There is an immediate need for an appropriate statistical methodology that integrates DA and the quantitative modeling. In this article, we propose a power-law type of "DA conjugate" model that is useful for incorporating dimensional information and analyzing post-DA variables. Adapting the similar idea of "conjugacy" in Bayesian analysis, we show that the proposed modeling technique not only produces flexible and effective results, but also provides good solutions to the above three issues. A modified projection pursuit regression analysis is implemented to fit the additive power-law model. A numerical study on ocean wave speed is discussed in detail to illustrate and evaluate the advantages of the proposed procedure. Supplementary materials for this article are available online.

1. Introduction

Dimensional analysis (DA) is a variable extraction and reduction method that is widely used in physics and engineering due to its generality (see Buckingham 1914; Sonin 2001; Szirtes 2007). It transforms the raw variables with certain physical dimensions into dimensionless variables through a power-law form, and states that the dimensionless variables are better in characterizing the underlying physical mechanics that are invariant to measurement scales. It has been applied to diverse fields, such as in Asmussen and Heebooll-Nielsen (1955), Islam and Lye (2009), and Stahl (1962). However, its importance has not come into the view of statisticians until recently, as in Albrecht et al. (2013), Shen et al. (2014), and Shen, Lin, and Chang (2017), among others. The main benefits of applying DA before statistical designs and analyses can be summarized into the following three points: (i) it is an analytical method based on physics that can be carried out before collecting the data, which is useful even in the design and planning stage; (ii) the reduction of the number of variables helps produce more efficient and parsimonious models; (iii) the DA models satisfying the dimensional constraints are easier to interpret for practitioners and generate scalable results. A comprehensive discussion on the combined use of DA in statistics is presented in Albrecht et al. (2013), Davis (2013), Lin and Shen (2013), Frey (2013), Jones (2013), Piepel (2013), Plumlee, Joseph, and Wu (2013) and the rejoinder by Albrecht et al. (2013). In addition to the practical applications, Shen and Lin (in press) showed that post-DA variables are sufficient and complete to some distribution family and maximal invariant to the scale transformation

of physical dimensions, which supports its utility in statistical applications from a theoretical point of view.

Although the capability and applicability of DA in statistical problems have been well justified, the modeling techniques specifically customized to post-DA variables are still underdeveloped. Most literature treats post-DA variables as the new input variables, then feeds them into statistical procedures, without investigating appropriate models for post-DA variables and the corresponding theoretical properties of the resulting estimates or predictions due to the DA transformation. In particular, there are three critical issues about constructing appropriate models for post-DA variables.

(i) The choice of post-DA variables is not unique. The post-DA variables are actually "independent" dimensionless variables that cannot express each other. However, any function of two dimensionless variables is still dimensionless. Different investigators may choose different basis quantities and deduce different post-DA variables. Even if the same analysis strategies are implemented, the resulting empirical models are very likely to be different because the values of working variables are distinct. The inconsistency lying between DA and statistical procedures possibly increases ambiguity and confusion. Intuitively, we would like to build models on statistically significant dimensionless variables so that the characterization of the dependence relationship relies on important variables and does not involve those that have little or no effect. Currently, there are no existing procedures to systematically select "statistically optimal" post-DA variables for a given model.

(ii) DA is a framework to extract the dimensional constraints out of the physical variables and produce post-DA variables that are free from the constraints that are necessary for physically valid models. Such constraints hold in theories but it is not uncommon to observe the empirical data revealing evidence against the constraints. In those cases, it does not necessarily imply DA is wrong. The violation of theories possibly indicates the missing of other potentially important physical quantities including physical constants as discussed by Albrecht et al. (2013). Therefore, a good statistical procedure should be able to test whether the constraints hold from an empirical perspective. However, if DA is treated as a pure preprocessing tool, it will be difficult to represent the constraints in the model and set up a hypothesis test on them.

(iii) Transforming the response variables by explanatory variables will possibly generate the spurious correlation. The concept of "spurious correlation" was initiated by Pearson (1897) to describe the correlation between ratios of absolute measurements, which arises because of using ratios, rather than any actual correlations between measurements. This phenomenon is worth attention in DA transformation. In fact, the post-DA variables are ratios and have different dependence structure from the original variables. Even if one explanatory variable has no effect on the response variable, its correlation with other explanatory variables by which the response variable is transformed will generate an artificial explanatory power to the transformed response. In other words, the significant presence of an explanatory variable may be due to its ability to explain other explanatory variables but not the response. Further discussions on spurious correlation related to DA can be found in the Shen et al. (2014) and Piepel (2013).

In this article, we propose the concept of "conjugate model" for post-DA analysis. The idea of "conjugacy" between probability models and parameter priors in Bayesian analysis is adapted and modified to represent a model family that satisfies a closure property under the operation of applying DA. A specific type of conjugate model is of an additive power-law form. The model is of good generality and resolves the above three issues in the following ways: (i) different choices of basis quantities lead to exactly the same probability model, just with different parameterizations. The same data will produce the same fitted model without worries on which post-DA variables to be included. Important dimensionless variables are identified and used automatically. (ii) The dimensional constraints from DA can be expressed through a set of linear constraints on the parameters. The hypothesis tests on the validity of DA are easily formulated and conducted. (iii) The spurious correlation is avoided by directly modeling on the original response variable rather than the transformed one. A *modified projection pursuit regression* is implemented as the estimation algorithm. Our contribution is to be the first to establish an appropriate statistical analysis methodology that incorporates DA as an integrated part of modeling, instead of treating it as a preprocessing tool. The proposed model resolves the three existing issues on DA modeling in the literature (namely, nonunique choices of dimensionless variables, tests on DA validity and spurious correlations), and is shown to be flexible, consistent, and efficient.

The rest of the article is organized as follows. In Section 2, we introduce the definitions and procedures of dimensional analysis with an illustrative example. In Section 3, we define the "DA conjugate model," and propose the single and additive power-law families that satisfy the DA conjugacy. The above three main concerns are shown to be resolved. A modified projection pursuit regression analysis is then implemented to fit the proposed model. In Section 4, a numerical experiment of ocean wave speed is conducted to illustrate the additive power-law model, its generality, and its favorable performance. (Another study on the particle physics data is included in the supplementary material, which demonstrates the single power-law model, the three modeling issues, and the test of DA validity.) Section 5 summarizes concluding remarks and prospective researches.

2. Introduction of Dimensional Analysis

2.1 Physical Background of DA

The dimension in physics refers to the physical classification of a quantity's type. SI system (International System of Units) defines seven fundamental physical dimensions, namely, length [**L**], mass [**M**], time [**T**], electrical current [**I**], absolute temperature [**Θ**], amount of substance [**N**], and luminous intensity [**J**]. Other physical dimensions, which are expressed through combinations of fundamental physical dimensions, are called derived dimensions. For instance, speed has the dimension length per time [**L/T**]. A measurement system consists of the definitions of units for each dimension. In SI system of units, we measure length by meters, time by seconds and so on. Other derived physical dimensions are measured correspondingly, such as speed by meters per second. An extension to the dimensions above was proposed by Siano (1985a, 1985b) in his orientational analysis, which basically treats components of a physical vector as separate dimensions. Therefore, practically groups of independent dimensions (unable to derive each other) may consist of more than seven dimensions.

Dimensional analysis (DA) stems from the principle that a physical phenomenon must be independent of units used in the measurement. Thus, any meaningful equation or inequality should have exactly the same dimensions on both sides. A foundational theorem of DA is the Buckingham's Π-theorem (Buckingham 1914). It states that a physically valid equation involving n variables can be reduced to an equation with $p = n - k$ variables, where k is the number of variables whose dimensions are independent and representative. These k variables are called *basis quantities*, as they form a basis in terms of dimensions. DA provides a scheme to select the basis quantities and transform the other variables into dimensionless. Equivalently, it is also a scheme that generates dimensionless variables, where each of them is not a function of the rest.

2.2 Procedures of DA and Three Modeling Issues

Suppose $X_1, \ldots, X_n$ are positive variables with dimensions $\mathbf{D_1}, \ldots, \mathbf{D_n}$ and $e_1, \ldots, e_m$ are fundamental dimensions.

Thus, $\mathbf{D_i} = \prod_{j=1}^{m} e_j^{d_{ij}}$. Let $D = (d_{ij})$ be the dimensional matrix whose (i, j) element is d_{ij}, and $k = \text{rank}(D)$. Without loss of generality, suppose the first k rows are linearly independent and the other rows can be linear represented as $d_{tj} = \sum_{i=1}^{k} b_{ti} d_{ij}$ for $t = k+1, \ldots, n$; $j = 1, \ldots, m$. When $X_1, \ldots, X_k$ are taken as basis quantities, $X_{k+1}, \ldots, X_n$ can be transformed into dimensionless as $\pi_t = X_t X_1^{-b_{t1}} \ldots X_k^{-b_{tk}}$ for $t = k+1, \ldots, n$. The general function $f(X_1, \ldots, X_n) = 0$ can be written as $g(X_1, \ldots, X_k, \pi_{k+1}, \ldots, \pi_n) = f(X_1, \ldots, X_k, \pi_{k+1} \prod_{i=1}^{k} X_i^{b_{k+1,i}}, \ldots, \pi_n \prod_{i=1}^{k} X_i^{b_{n,i}}) = 0$. By the Buckingham's Π-theorem, $g(X_1, \ldots, X_k, \pi_{k+1}, \ldots, \pi_n) = g(\pi_{k+1}, \ldots, \pi_n) = 0$.

There are several important issues with DA. First, the choices of valid basis quantities $X_1, \ldots, X_k$ are not unique and the resulting dimensionless variables $\pi_{k+1}, \ldots, \pi_n$ depend on the selected $X_1, \ldots, X_k$. One can always multiply two dimensionless variables and get a third one. Generally, for any function q, $q(\pi_{k+1}, \ldots, \pi_n)$ is a dimensionless variable. In practice, specialists and technicians may have a list of common dimensionless variables in their disciplines with specific physical meanings. Such choice depends on domain knowledge and is often subjective. Another option is to select variables that best explain the system or trend in terms of parsimony and significance. This article is perhaps the first work to propose a statistical modeling procedure that is invariant of the choices of basis quantities and implicitly selects the most appropriate post-DA variables. Second, data may show empirical evidence against the theoretical dimensional constraints: there may exist significant effects from basis quantities, and the reduction is not valid. In fact, DA requires that all relevant variables should be included in the original model. Otherwise, the key variables may be ruled out due to the lack of presentation in dimensions. See Albrecht et al. (2013) for details. If the basis quantities are highly significant, there is high possibility of missing such relevant key variables. This article is perhaps the first one to develop inference methods that are capable of testing whether DA assumptions are violated, and whether basis quantities should be kept within the model in the statistical literature. Finally, models based on transformed $\pi_{k+1}, \ldots, \pi_n$ may result in spurious effects of $X_1, \ldots, X_n$. For example, suppose the true predictive model is $X_1 \equiv c = f(X_2, X_3)$, that is, X_1 is a constant c and does not depend on X_2, X_3, with X_1, X_2, X_3 having the same physical dimension $\mathbf{D}$. After DA, we obtain dimensionless $\pi_1 = X_1/X_2, \pi_3 = X_3/X_2$ and model by $\pi_1 = g(\pi_3)$, that is, $X_1/X_2 = g(X_3/X_2)$. Although, X_1 and X_3 are not correlated (because $X_1 = c$ is a constant), π_1 and π_3 are likely to have significant positive correlation just because they share X_2 in common: small X_2 generates large π_1 and π_3, while large X_2 generates small π_1 and π_3. Estimated from such data, the model may possibly take the form of $X_1/X_2 = \hat{k} X_3/X_2$ for some significant coefficient $\hat{k}$, leading to the final model $X_1 = \hat{k} X_3$. The seemingly significant effect from X_3 is actually spurious and X_3 should not be in the model. Detail discussion can be found in Shen et al. (2014) and Piepel (2013). In fact, the relationship discovered between dimensionless variables is valid. The issue lies within the back transformation into dimensional variables and the inference made outside the dimensionless framework. Our method focuses on modeling the original response (X_1 in the above case) and circumvents the issue of spurious correlation. To demonstrate the above modeling issues in a more realistic setting, an ocean wave speed example is discussed as follows.

2.3 Illustrative Example of DA

Understanding the speed of ocean waves is crucial for the prediction of a natural disaster like tsunami. There are multiple sources of driving forces that generate wave motions on the sea surface, such as wind, gravity, and rotation of the earth. Therefore, it is a complicate system whose analytical behavior may not be tractable. Suppose we consider the wave whose restoring force is the gravity of the earth, that is, gravity waves. The phase speed of the gravity waves v is of the main interest. Our predictors are g the gravity constant of the earth, λ the wavelength, ρ the density of water, H the depth of the sea. The function we want to estimate is f, where $v = f(g, \lambda, \rho, H)$. Note that in practice this collection of variables will be achieved from expert knowledge of the specialized field. It is completely valid to include other potential variables of interest into the model, such as the speed of sound in the water. Here, we use the model $v = f(g, \lambda, \rho, H)$ as an example.

The first step of DA is to identify the dimensions of variables. The fundamental dimensions involved in the system are length, mass, and time. The dimensions of the variables and the dimensional matrix R can be found in Table 1. The second step is to choose the basis quantities. We select g, λ, ρ. The third step is to transform other variables into dimensionless using basis quantities. Thus, we get $\pi_v = v/\sqrt{g\lambda}$, and $\pi_H = H/\lambda$. The final step is to rewrite the function of interest. $\pi_v = h(g, \lambda, \rho, \pi_H) = h(\pi_H)$ because of the Buckingham's Π-theorem, which results in the DA model $v = \sqrt{g\lambda} \cdot h(H/\lambda)$. (Throughout this article, we will use · for a proper multiplication of two quantities, which can be a scaler times a scaler, a scaler times a matrix or a matrix multiplication.)

The actual wave speed can be approximated analytically by $v = \sqrt{g\lambda/2\pi} \cdot \sqrt{\tanh(2\pi H/\lambda)}$ as in Phillips (1977). Thus, the true function of h actually takes the form of $h(x) = \sqrt{\tanh(2\pi x)/2\pi}$. From an empirical point of view, DA helps us remove the nuisance variable ρ and reduce the number of variables to 1 to estimate h. From a physical point of view, the dimensionless sea depth $\pi_H = H/\lambda$ actually characterizes the key feature of the sea waves. When in the deep water (as in the ocean), $H \gg \lambda$, $\pi_H \gg 1$. Thus, $\tanh(2\pi \pi_H) \approx 1$ and the wave speed $v \approx \sqrt{g\lambda/2\pi}$, mainly depends on the wavelength; while in the shallow water (as along the coastline), $H \ll \lambda$, $\pi_H \ll 1$. Thus, $\tanh(2\pi \pi_H) \approx 2\pi \pi_H$ and the wave speed $v \approx \sqrt{gH}$, mainly depends on the depth of the water. As a result, DA also induces better physical interpretability.

Table 1. Dimensions of variables in ocean waves example.

Variables	Description	Dimensions	SI units	$r_{i,\text{Length}}$	$r_{i,\text{Mass}}$	$r_{i,\text{Time}}$
v	Wave speed	$\mathbf{LT}^{-1}$	m/s	1	0	−1
g	Gravity constant	$\mathbf{LT}^{-2}$	m/s^2	1	0	−2
λ	Wavelength	$\mathbf{L}$	m	1	0	0
ρ	Water density	$\mathbf{ML}^{-3}$	kg/m^3	−3	1	0
H	Sea depth	$\mathbf{L}$	m	1	0	0
γ	Surface tension	$\mathbf{MT}^{-2}$	kg/s^2	0	1	−2

If the wavelength λ is very small, the effect of surface tension should also be taken into account. Denote γ as the surface tension of water–air interfaces whose dimension can also be found in Table 1. Then, without DA, the function of interest is $v = f(g, \lambda, \rho, H, \gamma)$. With DA, another dimensionless variable $\pi_\gamma = \gamma/g\lambda^2\rho$ is generated to include the capillary effect. Therefore, $\pi_v = h(\pi_H, \pi_\gamma)$, that is, $v = \sqrt{g\lambda} \cdot h(H/\lambda, \gamma/g\lambda^2\rho)$. According to Phillips (1977), the phase speed of gravity-capillary waves is $v = \sqrt{(g\lambda/2\pi + 2\pi\gamma/\rho\lambda) \tanh(2\pi H/\lambda)}$, with $h(x, y) = \sqrt{(1/2\pi + 2\pi y) \tanh(2\pi x)}$. The capillary wavelength is defined as $\lambda_c = 2\pi\sqrt{\gamma/g\rho} = \lambda \cdot 2\pi\sqrt{\pi_\gamma} \approx 1.7$ cm for water–air interfaces. Roughly speaking, the surface tension is negligible if $\lambda_c << \lambda$. The derived dimensionless $\pi_\gamma = (\lambda_c/2\pi\lambda)^2$, characterizing their ratio, determines the impact of surface tension on the wave speed.

The proposed three issues in modeling can be clearly illustrated in this example. Consider the case of gravity waves with negligible influence from the surface tension. First, instead of (g, λ, ρ), one can select (g, H, ρ) as the basis quantities. This leads to dimensionless quantities $\pi_v^* = v/\sqrt{gH}$ and $\pi_\lambda^* = \lambda/H$. Modeling on $\pi_v^* = h^*(\pi_\lambda^*)$ results in $h^*(x) = \sqrt{x \tanh(2\pi/x)/2\pi}$. While h is close to a piecewise linear function, h^* resembles a logarithmic function. Even estimating from the same dataset, we can expect that the two models ($\hat{h}$ and $\hat{h}^*$) will produce different results. Second, suppose that the water density ρ turns out to have noticeable impact on the wave speed v from the data perspective. Then, the dimensional constraints that rule out the dependence between v and ρ may be violated. A formal hypothesis test needs to be conducted to decide whether such violation is due to random errors or systematic biases. If the effect of ρ turns out to be statistically significant, it should be kept in the model and viewed as an indicator of missing key variables with the dimension mass (**M**). Third, the DA model we assume is $v/\sqrt{g\lambda} = h(H/\lambda)$. Under the situation of deep water with $H >> \lambda$ and $\pi_H >> 1$, H hardly affects the wave speed v. See Figure 1(a). With $g = 9.8$ m/s^2, $\rho = 1.025 \times 10^3$ kg/m^3, and $\gamma = 0.074$ kg/s^2, under different values of λ (in meters), v does not depend on H. However, because π_v and π_H share λ in common in the denominator, it is highly possible that π_v and π_H have strong relationship, inducing a spurious significant effect from H. See Figure 1(b) for the relationship of

π_v and π_H under the same conditions as Figure 1(a). With the same range of H from 1 to 1.5 m, π_v and π_H show a strong positive correlation. From that, models such as $v = \hat{k} \cdot H/\lambda \cdot \sqrt{g\lambda}$ might be concluded, leading to a false interpretation of the effect of the sea depth H in the wave speed v.

3. A Conjugate Model for Dimensional Analysis

Although practical successes of implementing DA in statistical problems prove DA to be a useful reduction methodology, the studies on appropriate statistical models customized to post-DA variables remain absent. As shown in the previous section, the potential issues in modeling would generate false inferences and predictions, if DA is treated as merely a preprocessing tool. In this section, we propose a DA conjugate family and an additive power-law model to resolve the issues. In this framework, it can be shown that the dimensional constraints can be represented by a set of linear constraints on the parameters, leading to a model that is (i) invariant to choices of basis quantities, (ii) easy to test dimensional constraints, and (iii) free from spurious correlation.

3.1 DA Conjugate Family and a Power-Law Model Family

A probability model family that is invariant to changes of measurement system is called *a dimensionless model*. Typically, invariant models are built on invariant statistics. In the case of DA, the (maximal) invariant statistics are exactly the post-DA dimensionless variables (Shen and Lin in press). Among all dimensionless models, a special type of parametric family is of most interest and we call it "*DA Conjugate Family*." It has the property that if post-DA variables $\pi_{k+1}, \ldots, \pi_n$ are modeled under a DA conjugate family, then the model expressed in original variables $(X_1, \ldots, X_n)$ remains within one same family, regardless of the specific form of DA transformation that has been carried out.

Definition 1. Suppose $X_1, \ldots, X_n$ are physical quantities and $\mathcal{M}$ is an DA transformation. Therefore, $(\pi_{k+1}, \ldots, \pi_n) = \mathcal{M}(X_1, \ldots, X_n)$ are the corresponding post-DA variables. Suppose $\mathcal{F} = \{P_\theta, \theta \in \Theta\}$ is a parametric family on $(X_1, \ldots, X_n)$, $\mathcal{G} = \{Q_\lambda, \lambda \in \Lambda\}$ is a parametric family on $(\pi_{k+1}, \ldots, \pi_n)$. The

(a) Relationship of v and H.

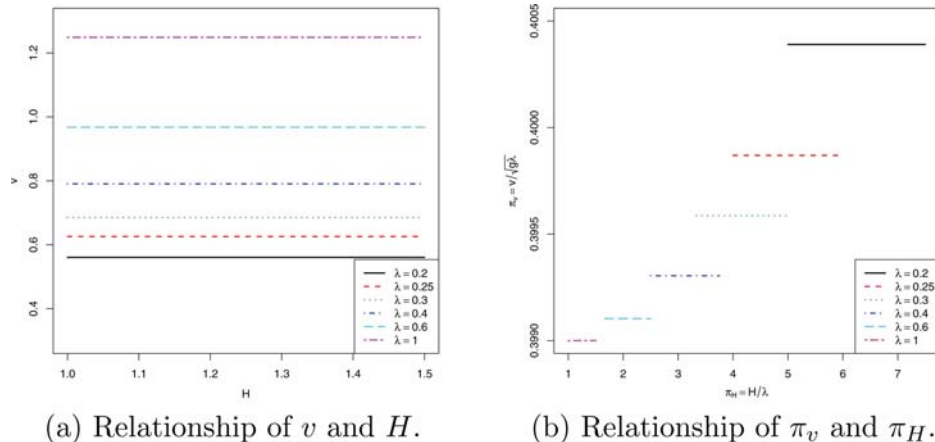(b) Relationship of π_v and π_H.

Figure 1. Spurious correlation among dimensionless variables in the ocean wave speed example. $v = \sqrt{(g\lambda/2\pi + 2\pi\gamma/\rho\lambda) \tanh(2\pi H/\lambda)}$ (Phillips 1977) with $g = 9.8$ m/s^2, $\rho = 1.025 \times 10^3$ kg/m^3, and $\gamma = 0.074$ kg/s^2. H and λ have units of m (meters) and v has a unit of m/s (meters per second).

pair $(\mathcal{F}, \mathcal{G})$ is called a DA conjugate family, if for any DA transformation $\mathcal{M}$, there exists Θ_0, such that $\mathcal{G} = \{Q_\lambda, \lambda \in \Lambda\} = \{P_\theta \circ \mathcal{M}^{-1}, \theta \in \Theta_0\}$, where $\Theta_0 \subseteq \Theta$ is called a dimensionally homogenous parameter subspace.

Specifically, $\mathcal{F}$ is a pre-DA model on n-dimensional vector $(X_1, \ldots, X_n)$, while $\mathcal{G}$ is a post-DA model on $(n-k)$-dimensional vector $(\pi_{k+1}, \ldots, \pi_n)$. Therefore, $\mathcal{G}$ is a dimensionless model. If $\mathcal{G}$ is actually composed of induced measures (by $\mathcal{M}$) from a subfamily $\mathcal{F}_0 = \{P_\theta, \theta \in \Theta_0\}$ of $\mathcal{F}$, then $(\mathcal{F}, \mathcal{G})$ is a DA conjugate family. Notice that the subspace Θ_0 and subfamily $\mathcal{F}_0$ do not depend on the specific form of DA transformation $\mathcal{M}$, as long as it gives a complete set of dimensionless variables. However, they do depend on the physical dimensions of $X_1, \ldots, X_n$.

The idea of defining DA conjugate model is to bring the modeling scope from $(\pi_{k+1}, \ldots, \pi_n)$ back to $(X_1, \ldots, X_n)$. Instead of questioning the best set of basis quantities and dimensionless variables to use, we model on the original variables, with the dimensional constraints represented in the parameters. Although different people may have different choices of $\mathcal{M}$, as long as they model by $\mathcal{G}$, the same results can be achieved as if modeling by $\mathcal{F}_0$ on original $(X_1, \ldots, X_n)$ without DA transformation.

Furthermore, it is analogous to the conjugacy concept in Bayesian inference. In fact, pre-DA model is like the "prior"; DA, which is the physical information, resembles the "likelihood," which is the empirical information in Bayesian inference; post-DA model is like the "posterior." The "prior" model gets updated with the physical information from DA, and turns into a "posterior" model. The distinction here is that the DA conjugate model is on data variables and the conjugate prior/posterior distribution is on parameters.

There are certain advantages of using DA conjugate families in DA modeling. First, the consistency in representation in terms of the original variables benefits the comparisons of models with different sets of basis quantities, leading to a more comprehensive view of significant factors. Second, the resulting model can be compared to non-DA models directly, to assess the impact and validity of DA. Dimensional constraints are represented by constraints on the parameters (instead of variables), which makes hypothesis testing of DA a straightforward task as $H_0 : \theta \in \Theta_0$ versus $H_a : \theta \in \Theta$. Third, modeling in original variables with constrained parameters (instead of transformed variables with free parameters) avoids spurious correlations. Finally, if DA holds, we focus on a lower dimensional subspace of parameters within the same parametric family, resulting in more power.

Sometimes, it is convenient to assume that $\mathcal{F}$ and $\mathcal{G}$ belong to the same type of model families, but on different number of variables (such as linear regressions). In this case, the collection of such model families is defined as a *strict DA conjugate family*.

Definition 2. Suppose $\mathcal{F}^d$ is a parametric family on d-dimensional vector. The collection $\mathfrak{F} = \{\mathcal{F}^d : d = 1, 2, \ldots\}$ is a strict DA conjugate family if the pair $(\mathcal{F}^{d_1}, \mathcal{F}^{d_2})$ is a DA conjugate family for any $d_1 \geq d_2$.

An example of a strict DA conjugate family that models univariate response is the *single power-law family with multiplicative errors* $\mathcal{P}$: $Y = \prod_{i=1}^{n} X_i^{\beta_i} \cdot \epsilon$, where Y is response, X_i's are n predictors, and ϵ is an independent random error, following lognormal distribution, for example. Suppose $X_1, \ldots, X_k$ are basis quantities, and $Y, X_{k+1}, \ldots, X_n$ have corresponding post-DA variables $\pi_0, \pi_{k+1}, \ldots, \pi_n$. Applying $\mathcal{P}$ to $\pi_0, \pi_{k+1}, \ldots, \pi_n$, we have $\pi_0 = \prod_{t=k+1}^{n} \pi_t^{\beta_t} \cdot \epsilon$. Since $\pi_0 = Y X_1^{-b_{01}} \ldots X_k^{-b_{0k}}$, $\pi_t = X_t X_1^{-b_{t1}} \ldots X_k^{-b_{tk}}$ for $t = k+1, \ldots, n$, expressing the model in the original $Y, X_1, \ldots, X_n$ yields $Y = \prod_{i=1}^{n} X_i^{g_i(\beta_{k+1}, \ldots, \beta_n)} \cdot \epsilon$, where $g_i(\beta_{k+1}, \ldots, \beta_n) = b_{0,i} - \sum_{t=k+1}^{n} b_{t,i} \beta_t$ for $i = 1, \ldots, k$ and $g_i(\beta_{k+1}, \ldots, \beta_n) = \beta_i$ for $i = k+1, \ldots, n$. Any values of parameters β_i's modeling $\pi_0, \pi_{k+1}, \ldots, \pi_n$ produce an equivalent model based on $Y, X_1, \ldots, X_n$ within the same family $\mathcal{P}$ but with a different set of parameters $g_1, \ldots, g_n$. In general, an appropriate scale family is capable to serve as a strict DA conjugate family.

The proposed single power-law family $\mathcal{P}$ solves the three modeling issues. First, different choices of basis quantities yield exactly the same model with particular sets of parameters β_i. In fact, we have the following theorems.

Theorem 1. $\prod_{i=k+1}^{n} \pi_i^{\alpha_i}$ spans all possible dimensionless variables of the form $X_1^{\alpha_1} \ldots X_n^{\alpha_n}$.

Theorem 2. Suppose based on two sets of basis quantities, DA leads to two sets of dimensionless variables $\{\pi_0, \pi_{k+1}, \ldots, \pi_n\}$ and $\{\pi_0', \pi_{k+1}', \ldots, \pi_n'\}$. The following two models are constructed: $\pi_0 = \prod_{t=k+1}^{n} \pi_t^{\beta_t} \cdot \epsilon$ and $\pi_0' = \prod_{t=k+1}^{n} \pi_t'^{\beta_t'} \cdot \epsilon$. Then, expressed in terms of $Y, X_1, \ldots, X_n$, the above two models are equivalent with different parameterizations.

Practically, although different people may conduct DA in a distinct way, as long as the power-law model is used, they will achieve the same results, including estimates, predictions, and inferences.

Second, by using the single power-law model, the DA constraints can be expressed as a set of linear constraints. In fact, without DA, applying $\mathcal{P}$ to $Y, X_1, \ldots, X_n$ leads to the model $Y = \prod_{i=1}^{n} X_i^{\beta_i} \cdot \epsilon$. Through applying $\mathcal{P}$ to π_0, π_{k+1}, π_n, the resulting model is of the form $Y = \prod_{i=1}^{n} X_i^{g_i(\beta_{k+1}, \ldots, \beta_n)} \cdot \epsilon$, where $g_i(\beta_{k+1}, \ldots, \beta_n) = b_{0,i} - \sum_{t=k+1}^{n} b_{t,i} \beta_t$ for $i = 1, \ldots, k$ and $g_i(\beta_{k+1}, \ldots, \beta_n) = \beta_i$ for $i = k+1, \ldots, n$, as shown before. Thus, implementing DA literally puts a set of linear constraints $(\beta_i = b_{0,i} - \sum_{t=k+1}^{n} b_{t,i} \beta_t$, for $i = 1, \ldots, k)$ on the parameters β_i from the original model $Y = \prod_{i=1}^{n} X_i^{\beta_i} \cdot \epsilon$. Testing validity of DA is equivalent to testing whether these linear constraints hold empirically for the current data.

Third, when modeling post-DA variables $\pi_0, \pi_{k+1}, \ldots, \pi_n$, the final equation to be fitted is $Y = \prod_{i=1}^{n} X_i^{g_i(\beta_{k+1}, \ldots, \beta_n)} \cdot \epsilon$, where $g_i(\beta_{k+1}, \ldots, \beta_n) = b_{0,i} - \sum_{t=k+1}^{n} b_{t,i} \beta_t$ for $i = 1, \ldots, k$ and $g_i(\beta_{k+1}, \ldots, \beta_n) = \beta_i$ for $i = k+1, \ldots, n$. Since the response is the original Y without any transformations by X_i's, the spurious correlation is circumvented.

3.2 Reparameterization

The proposed single power-law family consists of models that are invariant to the selection of basis quantities and express dimensional constraints as a set of linear constraints on the

parameters. These properties can be more clearly illustrated through the following reparameterization.

Assume a conjugate power-law model $Y = \prod_{i=1}^{n} X_i^{\beta_i} \epsilon$ for a physical system with k fundamental dimensions, $e_1, \ldots, e_k$. Y has dimension $\mathbf{D_0} = \prod_{j=1}^{k} e_j^{d_{0j}}$ and X_i have dimensions $\mathbf{D_i} = \prod_{j=1}^{k} e_j^{d_{ij}}$, respectively. We call d_{ij} as dimensional coefficients. The dimensionally homogenous assumption leads us to the following equations:

$$d_{0,j} = \sum_{i=1}^{n} \beta_i d_{ij}, \text{ for } j = 1, \ldots, k.$$

The above linear system contains n variables β_i and k equations. Without loss of generality, suppose they are linearly independent, that is, $k = \text{rank}(D)$, (otherwise, find $e_1, \ldots, e_k$ such that $k = \text{rank}(D)$). Then we have $n - k$ dimensions left to be estimated from the data.

Denote $\mathbf{d_j} = (d_{1j}, \ldots, d_{nj})^T$ as the dimensional coefficient vector of predictors for dimension j, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^T$ as the model parameter vector of predictors. $\mathcal{D} = \text{Span}\{\mathbf{d_t} : t = 1, \ldots, k\}$. $k = \text{rank}(D)$ implies that $\mathbf{d_j}$ are linearly independent and $k = \dim(\mathcal{D})$. The above equations can be written as

$$d_{0,j} = \boldsymbol{\beta}^T \cdot \mathbf{d_j}, \text{ for } j = 1, \ldots, k. \qquad (1)$$

Let $\mathbf{P_{-j}}$ be the projection matrix onto the linear space $\mathcal{D}_{-j}$ spanned by $\{\mathbf{d_t} : t \neq j\}$. Then $\mathbf{d_j} = \mathbf{P_{-j}d_j} + \mathbf{P_{-j}^{\perp}d_j}$ can be decomposed into two parts: the part spanned by $\mathcal{D}_{-j}$ and the part orthogonal to that. Denote $\mathbf{b^j} = \frac{\mathbf{P_{-j}^{\perp}d_j}}{\|\mathbf{P_{-j}^{\perp}d_j}\|_2^2}$ as the standardized normal vector for $j = 1, \ldots, k$. Note that $\mathbf{P_{-j}^{\perp}d_j}$ can be achieved as the residual vector of the linear regression $\mathbf{d_j} \sim \sum_{t \neq j} \mathbf{d_t}$ and $\mathbf{b^j}$ is a scaled version of it. $\mathbf{b^j}$ can be viewed as the dual basis of $\mathcal{D}$ with basis $\{\mathbf{d_t} : t = 1, \ldots, k\}$ in the Euclidean (Hilbert) space. In fact, it is straightforward to show the duality

Property: $\mathbf{b^j}^T \cdot \mathbf{d_j} = 1$ and $\mathbf{b^j}^T \cdot \mathbf{d_t} = 0, \forall t \neq j$.

In addition, Theorem 3 shows that $\{\mathbf{b^j}\}$ is a basis of $\mathcal{D}$.

Theorem 3. Denote $\mathbf{d_j}$ and $\mathbf{b^j}$ as above, if $\mathbf{d_j}$ are linearly independent, then $\mathbf{b^j}$ are linearly independent and $\mathcal{D} = \text{Span}\{\mathbf{d_j}, j = 1, \ldots, k\} = \text{Span}\{\mathbf{b^j}, j = 1, \ldots, k\}$.

Take an orthonormal basis of the orthogonal linear space $\mathcal{D}^{\perp}$, and denote as $\{\mathbf{b^{k+1}}, \ldots, \mathbf{b^n}\}$. Then $\{\mathbf{b^1}, \ldots, \mathbf{b^n}\}$ constitutes a basis for $\mathbb{R}^n$. We reparameterize the n-dimensional parameter $\boldsymbol{\beta}$ by Λ as $\boldsymbol{\beta} = \lambda^1 \mathbf{b^1} + \cdots + \lambda^n \mathbf{b^n} = \mathbb{B}\Lambda$, where $(\lambda^1, \ldots \lambda^n)^T = \Lambda$ is an n-dimensional parameter and $\mathbb{B} = (\mathbf{b^1}, \ldots, \mathbf{b^n})$ is a fixed matrix. Equation (1) turns out to be

$$d_{0,j} = \lambda_j, \quad \text{for } j = 1, \ldots, k. \qquad (2)$$

Equation (2) implies that the DA procedure can be interpreted as a prior knowledge on the first k elements of Λ after reparameterization. Furthermore, identical projection and reparameterization are achieved no matter which basis quantities are chosen, as long as the single power-law model is used. From a Bayesian points of view, we may further suggest that informative priors should be assigned to $\lambda_1, \ldots, \lambda_k$ (concentrating on $d_{0,1}, \ldots, d_{0,k}$, respectively), while uninformative priors be assigned to the rest $\lambda_{k+1}, \ldots, \lambda_n$.

Notice that although $\mathcal{D}^{\perp}$ is unique, $\mathbf{b^{k+1}}, \ldots, \mathbf{b^n}$ are not unique. Choice of them may depend on other knowledge such as PCA. Under a general model setting, similar results can be derived, but no general formulas are available.

The previous derivation suggests that the power-law model has a compatible structure with post-DA variables and resolves those three modeling issues. In the following two sections, two specific power-law models, single and additive power-law models, are discussed in detail. The hypothesis testing of DA validity is introduced with explicit formulation in the context of single power-law model. A modified projection pursuit regression procedure is implemented as the algorithm to estimate the additive power-law model.

3.3 Single Power-Law Model

The benefits shown in the previous section lead us to $\mathcal{P}$, the *single power-law model*

$$Y = \prod_{i=0}^{n} X_i^{\beta_i} \epsilon, \qquad (3)$$

where $X_0 = \exp(1)$ to capture a constant multiplier and ϵ is positive (say with mean 1). The response Y therefore ought to follow a distribution in a positive valued scaling family $\{P_a : P_a = P_1 \circ a^{-1}, a > 0, P_1 \text{ positive measure}\}$: for example, the exponential distribution, the half-normal distribution, and the lognormal distribution with fixed unknown variance. In fact, model (3) is a special generalized linear model: we take the $\ln X_i$ as predictors and β_i as corresponding parameters. The link function is $g(u) = e^u$ and $E[Y|X_i] = g(\sum_{i=0}^{n} \beta_i \ln X_i)$. The inference of model (3) can be conducted by the conventional procedure for generalized linear model. By specifying the distribution of Y, the maximum likelihood estimator for the parameters can be derived. The asymptotic variance of the estimator can be achieved by the inverse Fisher information matrix. Statistical testings of individual parameters can be formulated by establishing z-statistic and finding its null distribution or the asymptotic version of that. Variable selections can be implemented by using penalized likelihood. Goodness-of-fit test can be conducted by Kolmogorov–Smirnov type of nonparametrical test or by Shapiro–Wilk type of parametrical test. A simple example is illustrated below.

Assume model $\pi_0 = e^{\beta_0} \cdot \prod_{i=k+1}^{n} \pi_i^{\beta_i} \cdot \epsilon$, where ϵ follows the lognormal distribution with $\ln N(0, \sigma^2)$. Notice that taking logarithms on both sides, we have $\ln(\pi_0) = \beta_0 + \sum_{i=k+1}^{n} \beta_i \ln(\pi_i) + \ln \epsilon, \ln \epsilon \sim N(0, \sigma^2)$, which is a linear regression model. Expressed in original variable, the model is $Y = e^{\tilde{\beta}_0} \cdot \prod_{i=1}^{n} X_i^{\tilde{\beta}_i} \cdot \epsilon$, where $\tilde{\beta}_0 = \beta_0$, $\tilde{\beta}_i = g_i(\beta_{k+1}, \ldots, \beta_n)$ for $i = 1, \ldots, n$ as shown before. Taking logarithms, $\ln(Y) = \tilde{\beta}_0 + \sum_{i=1}^{n} \tilde{\beta}_i \ln(X_i) + \ln \epsilon, \quad \ln \epsilon \sim N(0, \sigma^2)$. The only difference from the multiple linear regression is that the coefficients after reparameterization are subject to the dimensional constrains. One can either model through β_i for unique unconstrained presentation, or model through $\tilde{\beta}_i$ under the linear constrains (1). For a general distribution family, after constructing the likelihood, the estimation of coefficients can be achieved through Lagrange multiplier method. However, for this simple case, analytic solutions can

be derived. Denote $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_0, \ldots, \tilde{\beta}_n)^T$, $D_0 = (d_{0,1}, \ldots, d_{0,k})^T$, $\tilde{\mathbf{X}} = (\mathbf{1}, \ln(X_1), \ldots, \ln(X_n))$, $\tilde{Y} = \ln(Y)$, and

$$\mathbf{D} = \begin{pmatrix} 0 & \cdots & 0 \\ \mathbf{d_1} & \cdots & \mathbf{d_k} \end{pmatrix}.$$

The least-square estimate subject to the dimensional constraints (1), that is, $\mathbf{D}^T \tilde{\boldsymbol{\beta}} = D_0$, is

$$\hat{\tilde{\boldsymbol{\beta}}} = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \mathbf{D} (\mathbf{D}^T (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \mathbf{D})^{-1} D_0$$
$$+ [\mathbf{I} - (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \mathbf{D} (\mathbf{D}^T (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \mathbf{D})^{-1} \mathbf{D}^T] (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \tilde{Y}.$$

One special hypothesis testing problem on the parameters is of great interest because it can be interpreted as a testing on the DA validity. As reasoned before, the violation of DA validity from the data point of view could be an indicator of missing important physical quantities, and thus worth exploring.

In the setting of the single power-law model (3) with specified distributional assumptions on the response, the hypothesis test of DA validity can be done through likelihood ratio test of linear constraints on coefficients in a straightforward way. Suppose the likelihood function is $L(\boldsymbol{\beta})$ and the log-likelihood function is $l(\boldsymbol{\beta})$. The maximum likelihood estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}$ without dimensional constraints and $\hat{\tilde{\boldsymbol{\beta}}}$ with dimensional constraints. The hypothesis testing of DA validity is formulated as

$$H_0 : \mathbf{D}^T \boldsymbol{\beta} = D_0; \quad \text{vs.} \quad H_a : \mathbf{D}^T \boldsymbol{\beta} \neq D_0.$$

Under the null distribution that DA is valid, the test statistics

$$2(l(\hat{\boldsymbol{\beta}}) - l(\hat{\tilde{\boldsymbol{\beta}}})) \sim \chi_k^2$$

follows a χ^2 distribution with degrees of freedom k, which is the rank of the dimension matrix.

In the previous example with ϵ following a lognormal distribution, the testing procedure can be explicitly written out. $l(\boldsymbol{\beta}) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} - \frac{N}{2} \ln(||\tilde{Y} - \tilde{\mathbf{X}}\boldsymbol{\beta}||^2) + \frac{N}{2} \ln(N)$, $\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \tilde{Y}$, while $\hat{\tilde{\boldsymbol{\beta}}}$ is shown in the previous section:

$$2(l(\hat{\boldsymbol{\beta}}) - l(\hat{\tilde{\boldsymbol{\beta}}})) = N \ln \left(\frac{||\tilde{Y} - \tilde{\mathbf{X}}\hat{\tilde{\boldsymbol{\beta}}}||^2}{||\tilde{Y} - \tilde{\mathbf{X}}\hat{\boldsymbol{\beta}}||^2} \right) \quad (4)$$

$$= N \ln \left(1 + \frac{||\tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \mathbf{D} (\mathbf{D}^T (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \mathbf{D})^{-1} (\mathbf{D}^T \hat{\boldsymbol{\beta}} - D_0)||^2}{||\tilde{Y} - \tilde{\mathbf{X}}\hat{\boldsymbol{\beta}}||^2} \right),$$

which should follow χ^2 distribution with degrees of freedom $k = \text{rank}(\mathbf{D})$ under the null hypothesis that DA is valid. Small p-value indicates violation of the dimensional constraints and suggests further investigations on issues such as missing key variables.

3.4 Additive Power-Law Model

The single power-law model (3) is not so generic to capture the arbitrary functional dependence f between Y and X_i's. We propose a more general DA conjugate model in the following:

$$E(Y|X) = \beta_0^1 \prod_{i=1}^n X_i^{\beta_i^1} + \beta_0^2 \prod_{i=1}^n X_i^{\beta_i^2} + \cdots + \beta_0^p \prod_{i=1}^n X_i^{\beta_i^p}. \quad (5)$$

Or equivalently,

$$E(Y|X) = \sum_{j=1}^p \beta_0^j \exp(\beta^j \cdot \tilde{X}), \quad (6)$$

where $\beta^j = (\beta_1^j, \ldots, \beta_n^j)$ and $\tilde{X} = (\ln(X_1) \ldots, \ln(X_n))^T$. It belongs to the additive index model with fixed ridge functions, and we call it *additive power-law model*. Here, we do not have further constraints on the distribution of Y except the first conditional moment.

The additive power-law model (5) offers a good approximation to the general relationship between Y and X. Suppose that there is a smooth function f such that $E(Y|X) = f(X_1, \ldots, X_n)$. We have the kth-order Taylor's expansion on $(0, 0, \ldots, 0)$ to approximate f:

$$f^{(k)}(X_1, \ldots, X_n) = f(0, 0, \ldots, 0) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(0, 0, \ldots, 0)X_i$$

$$+ \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(0, 0, \ldots, 0)X_i X_j + \cdots$$

$$+ \sum_{i_1, \ldots, i_k=1}^n \frac{\partial^k f}{\partial x_{i_1} \ldots \partial x_{i_k}}(0, 0, \ldots, 0)X_{i_1} \ldots X_{i_k}. \quad (7)$$

The Taylor's expansion with finite terms in (7) turns out to be a special case of the additive power-law model (5). However, model (5) allows powers β_i^k to be real numbers (rather than just nonnegative integers). As a result, any smooth function can be approximated by the proposed model to a given degree with sufficiently large p. Usually, a small p is enough to offer an efficient and parsimonious model with real numbered β_i^k. Note that it is easy to verify that (5) is indeed a strict DA conjugate model and solves the three modeling issues aforementioned. (See Theorem 4.) In real-world applications, we recommend this additive power-law model (5) over the single power-law model (3), since it is more flexible and useful.

Theorem 4. (Additive power-law model is a strict DA conjugate model.)

Suppose $X_1, \ldots, X_k$ are basis quantities and $Y, X_{k+1}, \ldots, X_n$ are transformed into post-DA variables by the function $\mathcal{M}$: $\pi_0 = Y X_1^{-b_{01}} \ldots X_k^{-b_{0k}}$, $\pi_t = X_t X_1^{-b_{t1}} \ldots X_k^{-b_{tk}}$ for $t = k+1, \ldots n$. Suppose the pth-order n-variable additive power-law model $\mathcal{P}_p^n$ is defined as

$$E(Y|X) = \beta_0^1 \prod_{i=1}^n X_i^{\beta_i^1} + \beta_0^2 \prod_{i=1}^n X_i^{\beta_i^2} + \cdots + \beta_0^p \prod_{i=1}^n X_i^{\beta_i^p},$$

$$\beta_i^q \in \mathbb{R} \text{ for } i = 0, \ldots, n; q = 1, \ldots, p.$$

Then the collection $\mathfrak{P}_p = \{\mathcal{P}_p^1, \mathcal{P}_p^2, \ldots, \mathcal{P}_p^n, \ldots\}$ is a strict DA conjugate family. That is, as probability models on $(Y, X_1, \ldots, X_n)$, $\mathcal{P}_p^{k+1}(\pi_0, \pi_{k+1}, \ldots, \pi_n) = \mathcal{P}_p^{k+1} \circ \mathcal{M}(Y, X_1, \ldots, X_n) \subseteq \mathcal{P}_p^n(Y, X_1, \ldots, X_n)$.

There are some identifiability issues for this model: the labeling for the indices is arbitrary. Thus, there will be $p!$ permuted versions that result in the same fitting. Such issue may not be essential for parameter estimations, but will produce difficulties

in estimating parameters' variances when conducting statistical inference. Investigation on the labeling problem is still an ongoing research topic. Here, the order is determined by the explanatory capability of the index. The first index explains the most variance/deviation of Y; the second index explains the most variance/deviation of the residual after fitting the first index, and so forth. Practically, we fit term by term until the added index has no significant improvement. On the other hand, if one thinks a single power-law model may be enough, we can stop at our first step and turn to the model $E(Y|X) = \prod_{i=0}^{n} X_i^{\beta_i}$, where no identifiability issues are present. Below, we introduce a *modified projection pursuit regression* as a detail procedure for estimating β_i^j in (5).

The projection pursuit regression proposed by Friedman and Stuetzle (1981) is a nonparametric multiple regression technique. It aims for solving $E(Y|X) = \sum_{j=1}^{p} S_{\beta^j}(\beta^j \cdot X)$, where β^j's are unknown parameters and S_{β^j}'s are unknown functions. If S_{β^j}'s are fixed to be exponential functions; X replaced with $\tilde{X}$, then the regression surface is exactly model (6). Therefore, the additive power-law model can be solved by the modified projection pursuit regression algorithm with known S_{β^j}.

Without distributional assumptions on Y, we resort to a regression approach instead of a likelihood approach. The main benefits include more freedom in the model and wider applications. These are especially important because the variation and the distribution of the response Y are often unknown. Here, we use least absolute deviations as the criterion in the projection pursuit regression. The least absolute deviations criterion is believed to be more robust to outliers, than the least-square criterion. Due to the multiplicative nature of the relationship between Y and X_i's, the propagation of errors from X_i's to Y is likely to generate residuals that are inhomogenous and associated with the conditional means. For example, if the observational errors are multiplicative, $Y = \sum_{j=1}^{p} \beta_0^j \prod_{i=1}^{n} (X_i \epsilon_i)^{\beta_i^j} = \sum_{j=1}^{p} \beta_0^j \prod_{i=1}^{n} X_i^{\beta_i^j} \prod_{i=1}^{n} \epsilon_i^{\beta_i^j}$ as $\sum_{j=1}^{p} \beta_0^j (\prod_{i=1}^{n} X_i^{\beta_i^j}) \varepsilon_j$. Even when $p = 1$ and $E(\varepsilon_j) = 1$, we have $\text{var}(Y|X) = [E(Y|X)]^2 \cdot \text{var}(\epsilon)$. The conditional variance is proportional to the squared conditional mean. If the errors ϵ_i are additive and small compared to values of X_i, then $Y = \sum_{j=1}^{p} \beta_0^j \prod_{i=1}^{n} (X_i + \epsilon_i)^{\beta_i^j} \approx \sum_{j=1}^{p} \beta_0^j \prod_{i=1}^{n} [X_i^{\beta_i^j} (1 + \epsilon_i \beta_i^j / X_i)] \approx \sum_{j=1}^{p} \beta_0^j \prod_{i=1}^{n} X_i^{\beta_i^j} (1 + \sum_{i=1}^{n} \epsilon_i \beta_i^j / X_i)$ as $\sum_{j=1}^{p} \beta_0^j (\prod_{i=1}^{n} X_i^{\beta_i^j})(1 + \varepsilon_j')$, just like the multiplicative case. When $p = 1$ and $E(\varepsilon_j') = 0$, $\text{var}(Y|X) = [E(Y|X)]^2 \cdot \text{var}(\epsilon)$. The least absolute deviations protect against possible outliers (resulting from the inhomogenous errors) and are more robust in the projection pursuit regression. Note that the least absolute deviations estimates correspond to the maximum likelihood estimates under additive Laplace distributed errors. The complete procedure is as follows. Suppose the sample size is N, $Y = (y_1, \ldots, y_N)^T$, and $X_i = (x_{i,1}, \ldots, x_{i,N})^T$.

1. Initialize current residuals and term counter

$$r_t \leftarrow y_t, \quad \text{for } t = 1, \ldots, N; \quad j \leftarrow 1.$$

2. Search for the next term in the model and the estimate $\hat{\beta}^j = (\hat{\beta}_0^j, \ldots, \hat{\beta}_n^j)$, by minimizing the least absolute deviations criterion

$$\sum_{t=1}^{N} \left| r_t - \beta_0^j \prod_{i=1}^{n} x_{i,t}^{\beta_i^j} \right|. \quad (8)$$

If the model is applied with DA, the estimate $\hat{\tilde{\beta}}^j = (\hat{\tilde{\beta}}_0^j, \ldots, \hat{\tilde{\beta}}_n^j)$ are achieved under dimensional constrains (1), that is, $d_{0,k} = \sum_{i=1}^{n} \hat{\tilde{\beta}}_i^j d_{ik}$.

3. Termination. If the criterion (8) is smaller than a threshold value, stop. Otherwise, update the current residuals and the term counter

$$r_t \leftarrow r_t - \hat{\beta}_0^j \prod_{i=1}^{n} x_{i,t}^{\hat{\beta}_i^j}, \quad \text{for } t = 1, \ldots, N; \quad j \leftarrow j + 1$$

and go to Step 2.

The optimization problem related to the least absolute deviations (8) among other M-estimators has been well investigated in the literature. Among other alternatives, the iteratively reweighted least-square method first proposed by Schlossmacher (1973) has been widely used. Green (1984) discussed its applications to general distributions or criteria, nonlinear parameterizations, and dependent observations. Considering the nonlinear relationship, we suggest its modification proposed by Phillips (2002), which is formulated through the generalized EM algorithm. Notice that the problem can also be formulated as a linear programming with nonlinear constraints. General robust optimization techniques can also be applied such as the augmented Lagrangian methods (Hestenes 1969; Powell 1969) and the Nelder and Mead (1965) method.

The algorithm termination rule can be specified in several different ways. Most are analogous to those for the step-wise regression. The threshold value we use here is calculated in terms of the percentage of explained deviations: if $\sum_{t=1}^{N} |\hat{\beta}_0^j \prod_{i=1}^{n} x_{i,t}^{\hat{\beta}_i^j}| / \sum_{t=1}^{N} |r_t|$ is sufficiently small, the algorithm is terminated. Other alternatives using likelihood information are also available. These include Akaike information criterion (AIC), Bayesian information criterion (BIC), and likelihood ratio test of significance of additional terms.

As for the hypothesis testing of DA validity, the formulation is similar as before:

$$H_0 : \mathbf{D}^T \boldsymbol{\beta}^j = D_0; \quad \text{vs.} \quad H_a : \mathbf{D}^T \boldsymbol{\beta}^j \neq D_0, \quad \text{for } j = 1, \ldots, p.$$

Given some distributional assumption, under the null hypothesis that DA is valid, the test statistics

$$2(l(\hat{\boldsymbol{\beta}}^1, \ldots, \hat{\boldsymbol{\beta}}^p) - l(\hat{\tilde{\boldsymbol{\beta}}}^1, \ldots, \hat{\tilde{\boldsymbol{\beta}}}^p)) \sim \chi_{k \cdot p}^2$$

follows a χ^2 distribution with degrees of freedom $k \cdot p$, where k is the rank of the dimension matrix, p is the number of terms.

4. Numerical Study

In this article, the proposed models are implemented in two applications, where the benefits and limitations of respective models are investigated in detail. The first example is in the supplementary material due to the length of the article. It uses "prim7" dataset in particle physics and illustrates the three modeling issues if DA is only perceived as a preprocessing tool. The
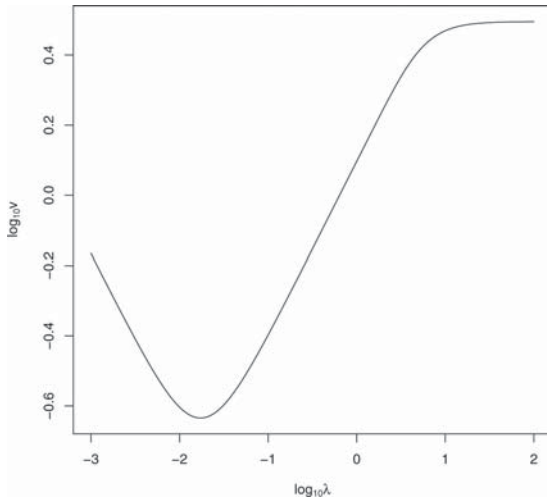
Figure 2. The relationship between the wave speed v and the wave length λ in the log-scale of respective units, meters per second and meters.

single power-law model is implemented to show the ease of use of the proposed framework, and the need of the more general additive power-law model. The second simulation example shown below uses the DA additive power-law model on the "Ocean Wave Speed" problem. It focuses on the algorithm and the performance. The purpose is to obtain a good DA model that at the same time circumvents the modeling issues. The results demonstrate that the additive power-law model is flexible and efficient, and thus it is recommended.

In this study, we revisit the example of ocean wave speed in Section 2.3, to illustrate the implementation and performance of additive power-law model. As stated in Phillips (1977), the phase speed of gravity-capillary waves is $v = \sqrt{(g\lambda/2\pi + 2\pi\gamma/\rho\lambda)\tanh(2\pi H/\lambda)}$. Figure 2 shows the functional dependence between the speed v and wave length λ in the log-scale with $H = 1$ m, as well as the constants $g = 9.8$ m/s^2, $\rho = 1.025 \times 10^3$ kg/m^3, and $\gamma = 0.074$ kg/s^2. When λ is smaller than $\lambda_c = 1.73 \times 10^{-2}$ m (-1.76 in the log-scale), the surface tension is the dominant driving force. Longer waves have smaller phase speed. When $\lambda_c << \lambda << H$, the gravity turns out to dominate the kinetics in the waves. Longer waves propagate faster. λ_c is about the wave length with the minimal phase speed. When $\lambda >> H$, the horizontal movement of water is restrained by the sea bed, leading to a constant speed regardless of the wave length. Here, we consider the case where the depth of the ocean H is very large ($H >> \lambda$, usually H around 4×10^3 m), and the phase speed follows the functional form:

$$v = \sqrt{(g\lambda/2\pi + 2\pi\gamma/\rho\lambda)}\epsilon, \qquad (9)$$

where ϵ denotes random errors. It is reasonable to assume a multiplicative error here because v is always positive and larger values of v may be observed with larger errors. Through the simulation, we want to study whether the proposed additive power-law model is able to fit the data well and whether it discovers both the gravity and capillary components.

To capture the capillary waves where the corresponding wave length is in a much smaller scale, the design of λ consists of 100 equally spaced grid points in the log-scale instead of the original scale (from -3 to 2, that is, 10^{-3} m $\sim 10^2$ m). The responses v's

are generated by (9) with constants $g = 9.8$ m/s^2, $\rho = 1.025 \times 10^3$ kg/m^3, $\gamma = 0.074$ kg/s^2 and independent errors following log-normal distribution with $\mu = 0$, $\sigma = 0.3$. The problem here is to discover the functional relationship of $v = f(\lambda, g, \rho, \gamma)$.

There are four explanatory variables and typically we need a design that varies the values of all four of them to distinguish respective effects. However, g, ρ, γ are all physical constants that are related to the experimental environment and difficult to change. With the help of DA, the effect of these physical constants can be automatically discovered and incorporated into the results without actually varying them. With some physical assumptions, the final empirical model may cover extreme physical conditions with unusual physical constants g, ρ, γ under which experiments are impractical, such as calculating the speed of waves through mercury on the moon.

Suppose the following additive power-law model is used:

$$E(v|\lambda, g, \rho, \gamma) = \sum_{j=1}^{p} \beta_0^j \lambda^{\beta_1^j} g^{\beta_2^j} \rho^{\beta_3^j} \gamma^{\beta_4^j}. \qquad (10)$$

The dimensional constraints induce the parameterization as

$$\begin{cases} \beta_1^j + \beta_2^j - 3\beta_3^j = 1 \cdots [\mathbf{L}] \\ -2\beta_2^j - 2\beta_4^j = -1 \cdots [\mathbf{T}] \\ \beta_3^j + \beta_4^j = 0 \cdots [\mathbf{M}] \end{cases} \rightarrow \begin{cases} \beta_2^j = 0.5\beta_1^j + 0.25 \\ \beta_3^j = 0.5\beta_1^j - 0.25 \\ \beta_4^j = 0.25 - 0.5\beta_1^j \end{cases}$$
$$(11)$$

for all $j = 1, \ldots, p$. The model (10) reduces to $E(v|\lambda) = \sum_{j=1}^{p} \alpha_0^j \lambda^{\beta_1^j}$ and $\beta_0^j = \alpha_0^j / g^{\beta_2^j} \rho^{\beta_3^j} \gamma^{\beta_4^j}$. After fitting $E(v|\lambda) = \sum_{j=1}^{p} \alpha_0^j \lambda^{\beta_1^j}$ and achieving estimated $\hat{\alpha}_0^j$ and $\hat{\beta}_1^j$, all other parameter estimates $\hat{\beta}_0^j, \hat{\beta}_2^j, \hat{\beta}_3^j, \hat{\beta}_4^j$ in (10) are calculated accordingly.

We implement the proposed modified projection pursuit regression to fit the parameters. By the first iteration, the estimated model is $\hat{E}(v|\lambda) = 1.4434\lambda^{0.4035}$. Figure 3(a) shows the scatterplot and the model fitness after the first iteration. On the log scale, the fitted values lie on a line. Apparently, it captures the major trend of data in the gravity wave region. The second iteration attaches an additional term of $6.020 \times 10^{-4}\lambda^{-0.9766}$; while the third iteration adds a term of $-0.05500\lambda^{0.06590}$. The estimated model then becomes $\hat{E}(v|\lambda) = 1.4434\lambda^{0.4035} + 6.020 \times 10^{-4}\lambda^{-0.9766} - 0.05500\lambda^{0.06590}$. Figure 3(b) displays the model fitness after the second and third iteration. It can be seen that the second added term describes the "abnormal dispersion" behavior in the capillary wave region, capturing the trend to have faster waves with smaller wave lengths. The third iteration adjusts the curves in the dipping region around λ_c, by canceling out the attempt in the first iteration to fit the "abnormal dispersion" that has been captured in the second iteration, and moving fitness curve downward to better represent the data. In the fourth iteration, the additional term with estimated coefficients $\hat{\alpha}_0^4 = 6.104 \times 10^{-6}$ and $\hat{\beta}_1^4 = 5.000 \times 10^{-2}$ makes almost no improvement on the explained deviation ($< 0.002\%$). Therefore, the iteration is stopped at the $p = 3$. Expressing by physical constants g, ρ, and γ, the final estimated model is

$$\hat{E}(v|\lambda, g, \rho, \gamma) = 0.8148\lambda^{0.4035} g^{0.4517} \rho^{-0.0483} \gamma^{0.0483}$$
$$+ 1.1626\lambda^{-0.9766} g^{-0.2383} \rho^{-0.7383} \gamma^{0.7383}$$
$$- 0.2272\lambda^{0.0659} g^{0.2830} \rho^{-0.2170} \gamma^{0.2170}. \quad (12)$$

(a) Model fitness after the first iteration.   (b) Model fitness after the second and third iteration.   (c) Final model fitness in original scale.
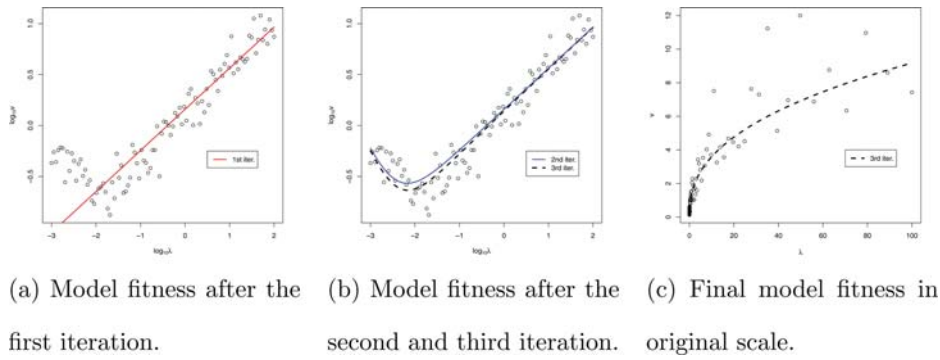
Figure 3. Scatterplot of the data and additive power-law model fitness through iterations of modified projection pursuit regression.

From Figure 3(b), it is safe to conclude that model (12) fits the data reasonably well. It successfully discovers both the gravity and capillary wave behaviors in different wave length regions. This result is significant because the model is actually fitted in the original scale of v and λ. Figure 3(c) displays the data and estimated curve in the original scale, where the actual calculations take place. Most deviations are described after the first iteration. It is noticeable how the algorithm accurately captures the dipping behavior in such a small scale within $\lambda < 0.1$ and $v < 1$.

5. Conclusion

DA has been used in various quantitative studies for almost 100 years. However, most literature treats DA as a preprocessing tool, which results in modeling issues, such as the nonunique choices of dimensionless variables, the representation and testing on DA validity, and the spurious correlations. A statistical modeling method tailored for DA is an immediate need. In this article, we propose a power-law type of "conjugate model" to solve the three critical issues in statistical modeling with DA. By implementing the proposed modeling procedure, (a) different choices of basis quantities induce the same model with different parameterizations. The preference of basis quantities is implicit and does not affect statistical modeling. (b) The dimensional constraints and Buckingham's Π-theorem can be represented by the constraints on the parameters in a straightforward way. The testing of the physical principles based on empirical data becomes available and intuitive. (c) By modeling the original response, the spurious correlations of post-DA variables are avoided. The projection pursuit regression is modified to fit the additive power-law model. The numerical examples show that the proposed models and corresponding algorithms perform quite well in practice. Our contribution here is to propose a genuine framework to integrate DA as a part of probabilistic modeling, and to solve the three existing modeling issues found in the literature. Hopefully, this work could provide some guidance in constructing appropriate analysis techniques to incorporate DA and other physical information.

There may be several directions that are worth exploring to improve the procedure. From the algorithmic aspect, backward adjustment of previous coefficients in the subsequent iterations may improve the efficiency and reduce the number of terms needed. On the other hand, the coefficients β^j may be preferred to have sparse components so that the effects of individual X_i's can be separated. Following the concept of group LASSO,

additional penalized terms on the L_1-norm of β_i^j's can be used in the least absolute deviation criterion. This is currently under investigation.

From the modeling aspect, one may prefer a Bayesian framework to include DA and its physical principles merely as prior information, instead of a mandatory constraint. It is particularly applicable when hypothesis of DA validity is rejected based on the empirical data. Under the notations in Section 3.2, suppose $Y = \prod_{i=1}^{n} X_i^{\beta_i} \epsilon$. Let $\tilde{X} = (\ln X_1, \ldots, \ln X_n)$, then $Y = \exp\{\tilde{X}\boldsymbol{\beta}\}\epsilon = \exp\{\tilde{X}\mathbb{B}\Lambda\}\epsilon = \exp\{X_T\Lambda\}\epsilon$, where $\tilde{X}_T = \tilde{X}\mathbb{B}$ is the transformed data matrix. Section 3.2 suggests that DA is a prior knowledge of $\lambda_1 = d_{0,1}, \ldots, \lambda_k = d_{0,k}$ after reparameterization. An atomic prior of $(\lambda_1 = d_{0,1}, \ldots, \lambda_k = d_{0,k})$ yields standard Bayesian modeling approach based on $\pi_0, \pi_{k+1}, \ldots, \pi_n$ that satisfies dimensional constraints. If DA only serves as an informative prior instead of a necessary constraint, one may assign priors concentrating on $d_{0,1}, \ldots, d_{0,k}$. For instance, $\lambda_j \sim N(d_{0,j}, \sigma_0^2)$ for $j = 1, \ldots, k$, while $\lambda_j \sim N(d_{0,j}, \sigma^2)$ for $j = k+1, \ldots, n$, where the relative magnitude of σ_0^2/σ^2 reflects the strength of belief in DA. Such a conservative choice is often more robust and consistent in practice when DA is available but in question.

Supplementary Materials

Acknowledgment

References

Albrecht, M. C., Nachtsheim, C. J., Albrecht, T. A., and Cook, R. D. (2013), "Experimental Design for Engineering Dimensional Analysis," *Technometrics*, 55, 257–270; with Rejoinder 292–295. [79,81]

Asmussen, E., and Heebooll-Nielsen, K. (1955), "A Dimensional Analysis of Physical Performance and Growth in Boys," *Journal of Applied Physiology*, 7, 593–603. [79]

Buckingham, E. (1914), "On Physically Similar Systems; Illustrations of the Use of Dimensional Equations," *Physical Review*, 4, 345–376. [79,80]

Davis, T. P. (2013), "Comment: Dimensional Analysis in Statistical Engineering," *Technometrics*, 55, 271–274. [79]

Frey, D. D. (2013), "Comments: Dimensional Analysis and Experimentation as a Catalyst to Learning From Data," *Technometrics*, 55, 271–274. [79]

Friedman, J. H., and Stuetzle, W. (1981), "Projection Pursuit Regression," *Journal of the American Statistical Association*, 76, 817–823. [86]

Green, P. J. (1984), "Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and Some Robust and Resistant Alternatives," *Journal of the Royal Statistical Society*, Series B, 46, 149–192. [86]

Hestenes, M. R. (1969), "Multiplier and Gradient Methods," *Journal of Optimization Theory and Applications*, 4, 303–320. [86]

Islam, M., and Lye, L. M. (2009), "Combined Use of Dimensional Analysis and Modern Experimental Design Methodologies in Hydrodynamics Experiments," *Ocean Engineering*, 36, 237–247. [79]

Jones, B. (2013), "Comments: Enhancing the Search for Compromise Designs," *Technometrics*, 55, 278–280. [79]

Lin, D. K. J., and Shen, W. (2013), "Comments: Experimental Design for Engineering Dimensional Analysis," *Technometrics*, 55, 281–285. [79]

Nelder, J. A., and Mead, R. (1965), "A Simplex Method for Function Minimization," *Computer Journal*, 7, 308–313. [86]

Pearson, K. (1897), "Mathematical Contributions to the Theory of Evolution. On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs," *Proceedings of the Royal Society of London*, 60, 489–498. [80]

Phillips, O. M. (1977), *The Dynamics of the Upper Ocean* (2nd ed.), Cambridge: Cambridge University Press. [87]

Phillips, R. F. (2002), "Least Absolute Deviations Estimation via the EM Algorithm," *Statistics and Computing*, 12, 281–285. [86]

Piepel, G. F. (2013), "Comments: Spurious Correlation and Other Observations on Experimental Design for Engineering Dimensional Analysis," *Technometrics*, 55, 286–289. [79,81]

Plumlee, M., Joseph, V. R., and Wu, C. F. J. (2013), "Comments: Alternative Strategies for Experimental Design," *Technometrics*, 55, 289–292. [79]

Powell, M. J. D. (1969), "A Method for Nonlinear Constraints in Minimization Problems," in *Optimization*, ed. R. Fletcher, New York: Academic Press, pp. 283–298. [86]

Schlossmacher, E. J. (1973), "An Iterative Technique for Absolute Deviations Curve Fitting," *Journal of the American Statistical Association*, 68, 857–859. [86]

Shen, W., Davis, T., Lin, D. K. J., and Nachtsheim, C. J. (2014), "Dimensional Analysis and Its Applications in Statistics," *Journal of Quality Technology*, 46, 185–198. [79,81]

Shen, W., and Lin, D. K. J. (in press), "Statistical Theories for Dimensional Analysis," *Statistica Sinica*. [79,82]

Shen, W., Lin, D. K. J., and Chang, C.-J. (2017), "Dimensional Analysis and Its Applications in Computer Experiment," *Quality Engineering*. doi:10.1080/08982112.2017.1320726 [79]

Siano, D. (1985a), "Orientational Analysis — A Supplement to Dimensional Analysis - I," *Journal of the Franklin Institute*, 320, 267–283. [80]

—— (1985b), "Orientational Analysis, Tensor Analysis and the Group Properties of SI Supplementary Units - II," *Journal of the Franklin Institute*, 320, 285–302. [80]

Sonin, A. A. (2001), *The Physical Basis of Dimensional Analysis* (2nd ed.). Cambridge, MA: MIT Press. [79]

Stahl, W. R. (1962), "Similarity and Dimensional Methods in Biology," *Science*, 137, 205–212. [79]

Szirtes, T. (2007), *Applied Dimensional Analysis and Modeling* (2nd ed.), Oxford: Elsevier Butterworth-Heinemann. [79]