ESSAYS ON REGRESSION SPLINE STRUCTURAL NONPARAMETRIC STOCHASTIC PRODUCTION FRONTIER ESTIMATION AND INEFFICIENCY ANALYSIS MODELS

A Dissertation

by

KE LI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2010

Major Subject: Agricultural Economics

Essays on Regression Spline Structural Nonparametric Stochastic Production Frontier Estimation and Inefficiency Analysis Models

ESSAYS ON REGRESSION SPLINE STRUCTURAL NONPARAMETRIC STOCHASTIC PRODUCTION FRONTIER ESTIMATION AND INEFFICIENCY ANALYSIS MODELS

A Dissertation

by

KE LI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Ximing Wu
Committee Members,	David Bessler
	H. Alan Love
	Qi Li
Head of Department,	John P. Nichols

December 2010

Major Subject: Agricultural Economics

ABSTRACT

Essays on Regression Spline Structural Nonparametric Stochastic Production Frontier Estimation and Inefficiency Analysis Models. (December 2010)

Ke Li, B.S., South China University of Technology;

M.S., West Texas A&M University

Chair of Advisory Committee: Dr. Ximing Wu

Conventional Cobb-Douglas and Transcendental Logarithmic production functions widely used in Stochastic Production Frontier Estimation and Inefficiency Analysis have merits and deficiencies. The Cobb-Douglas function imposes monotonicity and concavity constraints required by microeconomic theory. However it is inflexible and implies undesired assumptions as well. The Trans-log function is very flexible and does not imply undesired assumptions, yet it is very hard to impose both monotonicity and concavity constraints. The first essay introduced a class of stochastic production frontier estimation models that impose monotonicity and concavity constraints and suggested models that are very flexible. Researchers can use arbitrary order of polynomial functions or any function of independent variables within the suggested frameworks. Also shown was that adopting suggested models could greatly increase predictive accuracy through simulations. In the second essay we generalized the suggested models with the Inefficiency Analysis technique. In the last essay we extended the models developed in the previous two essays with regression spline and let the data decide how flexible or complicated the model should be. We showed the improvement of deterministic frontier estimation this extension could bring through simulations, as well. Works in this dissertation reduced the gap between conventional structural models and nonparametric models in stochastic frontier estimation field. This dissertation offered applied researchers Stochastic Production Frontier models that are more accurate and flexible than previous ones. It also preserves constraints of economic theory.

DEDICATION

To my parents, without their patience and support this dissertation would never have seen the light of day.

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Ximing Wu, for the opportunities he has provided me and for his guidance since fall, 2006. I deeply appreciate his patience, encouragement, and help when I felt frustrated and depressed. It has really been my pleasure to be his student.

I would also like to thank my committee members, Dr. David Bessler, Dr. H. Alan Love, and Dr. Qi Li, for their help and guidance. I also thank all of the professors and support staff in our Agricultural Economics Department who have made my entire graduate experience excellent.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

TABLE Page

1. INTRODUCTION

In microeconomics, a production function is the maximum output of a technology can produce from a certain combination of inputs. It is usually used for technologies have single or a group of closely related outputs. To obtain valid results from standard optimization problems, i.e. profit maximization and cost minimization problems, microeconomics usually requires production function to be monotone increasing and concave. As one of the most important functions in economics, production function has received a lot of attention since the beginning of econometrics.

Yet before Farrell (1957), econometricians used Least Square(LS) technology to estimate average production function, which is certainly different from production function in microeconomic theory. Since then econometricians have put a lot of effort to close this gap. Data Envelopment Analysis (DEA) was the first practical solution developed. DEA is still used in operation research. The main problem of DEA is that it fails to distinguish random events that affect output from inefficiency. Aigner et al. (1977, hereafter ALS) described a family of composed error term models to solve the problem. This family of models is called Stochastic Efficiency Frontier (SEF) models. SEF models use Maximum Likelihood Estimator (MLE) technology to estimate the model. Therefore, the variety and flexibility of assumed distributions play a key role in estimation accuracy. Followed by the work of Stevenson (1980) and Greene (1990), econometricians now have four distributions at their disposal. Some of them are quite flexible.

This dissertation follows the style of Journal of Econometrics.

Besides inefficiency and production function econometricians are also interested in non-input/input elements that may affect inefficiency. Kumbhakar et al. (1991) made first attempt, then followed by Reifschneider and Stevenson (1991), Battese and Coelli (1992), Caudill and Ford (1993), Huang and Liu (1994), Caudill et al.(1995), and Wang and Schmidt (2002). This set of models was called Inefficiency Analysis (IA) models. Alvarez et al. (2006, hereafter AAOS) gave a good summary in this field.

The functional form of production function, which econometricians are probably most interested in is another story. Right now, econometricians are basically stuck with Cobb-Douglas model and Transcendental Logarithmic model. With coefficients except intercept between 0 and 1, Cobb-Douglas model preserves monotonicity and concavity. However Cobb-Douglas model also imposes assumptions econometricians do not want or rejected by data e.g. unitary elasticity of factor substitution and constant partial and total production elasticities. Trans-log model is very flexible, and does not bear undesired assumptions that Cobb-Douglas function has. Yet it is very hard to impose monotonicity and concavity constraint either (Henningsen, and Henning 2009).

In this dissertation we will introduce a class of flexible functions that impose monotonicity and concavity constraints to SEF and IA models. The dissertation will be organized as follow:

- In the second section, we will introduce a class of flexible functions with monotonicity and concavity constraints. After that we will implement them in SEF models, and discuss estimation and inference method. Then we will show improvement of both deterministic frontier estimation and average technical efficiency recovery by adopting suggested models. In the end of the section we will apply the model on an unbalanced panel dataset of airline industry.
- In the third section, we will implement suggested models in IA models. We will

show improvement of deterministic frontier estimation, average technical efficiency recovery and estimation of the marginal effect of given variable on inefficiency by adopting suggested models. In the last part of this section we will apply the model on the same airline dataset.

- In the fourth section, we will extend the class of flexible functions with monotonicity and concavity constraints into regression spline structure nonparametric models. We will also show this extension could greatly improve estimation accuracy when signal noise ratio is high or sample size is large. After that we will demonstrate the model on the airline dataset.
- In the fifth section, we will conclude the dissertation and discuss future research opportunities.

2. A CLASS OF FLEXIBLE STOCHASTIC PRODUCTION FRONTIER MODELS WITH SHAPE CONSTRAINTS

2.1 Introduction

In economic theory production function is the maximum output level a technology can produce from given set of inputs. Firms may have inefficiencies. Their output levels lie on or below the maximum level. The maximum levels for all input combinations are called production-possibility frontier. Since ALS (1977) developed the first model to estimate stochastic production frontier, it has became an important field in microeconometrics. The general Stochastic Efficiency Frontier model can be written as:

$$y_i = f(x_i)\exp(u_i)\exp(v_i) \tag{1}$$

In equation above,$f(x_i)$ is the deterministic part of production frontier, e^{u_i} is the stochastic part of production frontier. u_i can take any value on the real line. Usually researchers assume it follows a symmetric distribution that centers at zero, such as normal distribution. e^{v_i} is technique efficiency of the ith firm. v_i can only be negative real number. Therefore Technique efficiency will take any value from 0 to 1. Intuitively it represents the percentage of production frontier that a firm can reach.

Econometricians have put a lot of efforts to develop new distributions for inefficiency term. Now we have four distributions at our disposal. The gap between theory and realty is also relatively small. Deterministic frontier itself, which econometricians are probably more interested in, is another story. Econometricians basically have two classes of alternative models for the deterministic frontier. Each of them has their advantage and deficiency.

Microeconomic theory requires production function to be monotone increasing and concave. If either of these two properties is missing the profit maximization procedure and most of equilibrium models will breakdown. Without monotone increasing property, even the cost minimization problem might not have a solution.

As long as all exponents are between 0 and 1, good old Cobb-Douglas production function preserves both properties implicitly. Many econometricians incorporated this functional form in their researches. However Cobb-Douglas production function also implies properties researchers do not desire. For instance it assumes unitary elasticity of factor substitution and partial and total production elasticities that do not change with input.

Trans-log can be considered as a generalized form of Cobb-Douglas function. It is a very flexible functional form, and does not bear undesire assumptions that Cobb-Douglas function does. It becomes by far the most popular model in stochastic production frontier estimation. Yet the flexibility does not come without cost. It is also very hard to impose monotonicity and concavity restrictions on Trans-log model as well. Although implying convex production function, linear production function is used by researchers in a few papers.

In this paper we will suggest three flexible functional forms that embed monotonicity and concavity. This class of models allows econometricians to estimate a flexible production function that incorporate prior knowledge from the theory without the burden of Cobb-Douglas model.

The paper will be organized as follow. In Section 2.2 we will do a brief literature review of stochastic production frontier estimation. In Section 2.3, the flexible forms of production function will be presented. In Section 2.4 we will suggest a three-step procedure to estimate these models as well as derive point estimation and inference of value that might be interested by researchers. In Section 2.5 we will use two models to estimate a simulated arbitrary monotone concave production function. The result will be compared to the result of Cobb-Douglas model. In the same section we will also estimate the Normalized Integrated Squared Error and absolute error of average technical efficiency recovery of first models under different error term settings. In Section 2.6 we will apply the first two models on the airline dataset. In Section 2.7 we will summarize our research and discuss future research possibilities.

2.2 Literature Review

Farrell (1957) first explored the possibility to estimate the frontier production function. Aigner and Chu (1968), Afriat (1972) and Richmond (1974) developed a set of techniques estimating deterministic frontier production using linear or quadratic programming technology. Their methods minimize

$$\sum_{i=1}^{n} |y_i - f(x_i)|$$

or

$$\sum_{i=1}^{n} (y_i - f(x_i))^2$$

subject to

$$y_i \leq f(x_i)$$

Deterministic efficiency frontier method fails to incorporate random elements that affect output. These elements such as natural disaster, worker strike or finding new natural recourses commonly exit in real world. The first attempt to incorporate these random elements was done by Aigner et al.(1976). They suggested a discontinuous distribution for the error term. ALS (1977) constructed the first practical Stochastic Efficiency Frontier estimator.

ALS (1977) assume $-v_i$ follows a Half normal distribution with 0 mean and ${\sigma_v}^2$ variance. Econometricians still widely use this distribution. The log likelihood function is:

$$LL_i = \ln \frac{2}{\sigma^2} + \ln(\phi(\frac{\epsilon_i}{\sigma^2})) + \ln(1 - \Phi(\epsilon_i \lambda \sigma^{-1})) \tag{2}$$

where$\lambda = \sigma_v/\sigma_u, \sigma = \sigma_u^2 + \sigma_v^2$. $\phi()$ is the density of standard normal distribution.$\Phi()$ is the cumulated distribution function of standard normal distribution.

ALS (1977) also briefly considered the exponential distribution for$-v_i$. The log likelihood function under this assumption is:

$$LL_i = \ln(\lambda) + \ln(1 - \Phi(\frac{\epsilon_i}{\sigma_u} - \sigma_u\lambda)) + \epsilon_i\lambda + \frac{1}{2}(\sigma_u\lambda)^2 \tag{3}$$

in which $\frac{1}{\lambda} = \mathrm{E}(-v), \frac{1}{\lambda^2} = \mathrm{Var}(v)$

Both half normal and exponential distribution bear implicit assumption that inefficiency is more concentrated near zero than further away. Stevenson (1980) suggested a more general normal-truncated normal distribution. Rather than assuming the normal distribution generating $-v_i$ centers at zero, truncated normal $-v_i$ allow mean to be any value on real line. The log likelihood of normal-truncated normal is:

$$LL_i = -\ln\sigma - \frac{1}{2}\ln 2\pi - \ln\Phi(\frac{\mu\lambda\sigma}{\sqrt{1+\lambda^2}}) + \ln\Phi(\frac{\mu}{\sigma\lambda} - \frac{\epsilon_i\lambda}{\sigma}) - \frac{1}{2}(\frac{\epsilon_i+\mu}{\sigma})^2 \quad (4)$$

Stevenson also instructed a Normal-Gamma distribution. In his specification the shape parameter P in Gamma distribution can only take integer value. Greene (1990) generalized Gamma distribution framework, by allowing P to take any positive real number. The log likelihood function in his framework is:

$$\begin{aligned} LL_i =& P\ln(\Theta) - \ln(\Gamma(P)) + {\sigma_u}^2\Theta^2/2 + \Theta\epsilon_i \\ &+ \ln(\Phi(-(\epsilon_i + \Theta\sigma_u^2)/\sigma_u)) + \ln(h(P-1,\epsilon_i)) \end{aligned} \quad (5)$$

in which

$$h(r,\epsilon_i) = \mathrm{E}[Q^r|Q>0,\epsilon_i] \quad Q \sim \mathrm{Normal}(-(\epsilon_i+\Theta\sigma_u^2),\sigma_u^2)$$

P is the shape parameter in Gamma distribution.Θ is the rate of Gamma distribution. Larger the P, further away concentration of firm-specific inefficiency is located from zero.

Previous researchers suggested maximum likelihood estimator and method of moment estimator. As we already assumed distribution of u_i as v_i, MLE is at least as efficient and MME. I this paper we will focus on MLE.

2.3 A Class of Flexible Stochastic Production Frontier Models with Shape Constraints

Researchers interested in estimating monotone function since development of the isotonic regression by Bartholomew (1959) and Kruskal (1965) and the Box-Cox transformation by Box and Cox (1964). Various monotone functions were developed. Some sacrificed flexibility. For instance Ramsay (1988) imposed monotonicity by restricting all coefficient of spline estimator to be positive. Ramsay (1998) developed a differential equation based monotone function. Their model is:

$$\hat{y} = \beta_0 + \beta_1 \int_0^x \exp(\int_0^a c(b)\mathrm{d}b)\mathrm{d}a \tag{6}$$

In his setup $c(x)$ is a polynomial or spline function. Sickles and Wu (Personal Discussion) improved the model to impose not only monotonicity but also concavity constrains. They also extended the model to multivariate case. From now on we will call it Additive model. The improved model is:

$$f(x) = \beta_0 + \sum_{j=1}^{p} \beta_j m(c_j(x_j)) \tag{7}$$

in which

$$m(c(x)) = \int_0^x \exp(-\int_0^a \exp(c(b))\ \mathrm{d}b)\ \mathrm{d}a$$

x_j is a vector of jth independent variables. p is number of independent variables in the dataset. $c_j()$ is some arbitrary function. With production function above, we have:

$$\frac{\partial f(x)}{\partial x_j} = \beta_j m'(c_j(x_j)) \tag{8}$$

in which

$$m'(c(x)) = \exp(-\int_0^x \exp(c(b))\ \mathrm{d}b)$$

The sign of partial derivative is determined by the sign of β_j. If on average output will increase when input j increases, we can guarantee output monotone increasing respect to input j in the estimated production function.

$$\frac{\partial^2 f(x)}{\partial {x_j}^2} = \beta_j m''(c_j(x_j)) \tag{9}$$

in which

$$m''(c(x)) = -\exp(c(x))\exp(-\int_0^x \exp(c(b))\ \mathrm{d}b)$$

Likewise if $\beta_j > 0$, we can guarantee $f(x)$ is a concave function of x_j. Moreover the relative concavity of production function is:

$$\frac{\partial^2 f(x)/\partial^2 x_j}{\partial f(x)/\partial x_j} = -\exp(c(x_j)) \tag{10}$$

Substitute equation (7) into (1), we have the full Additive model:

$$y_i = (\beta_0 + \sum_{j=1}^{p} \beta_j m(c_j(x_{i,j})))\exp(u_i)\exp(v_i) \tag{11}$$

To separate the error term from deterministic frontier we need to take natural log on both side.

$$ln(y_i) = ln(\beta_0 + \sum_{j=1}^{p} \beta_j m(c_j(x_{i,j})))) + u_i + v_i \tag{12}$$

In Additive model, we assume all inputs are separable. That is $\partial^2 y/\partial x_k \partial x_j = 0$. This assumption may be undesirable to some researchers. We will suggest another multiplicative functional form to avoid this assumption. From now on we will call it Generalized Cobb-Douglas (GCD) model. The deterministic frontier is defined by:

$$y = \beta_0 \prod_{j=1}^{p} m(c_j(x_j))^{\beta_j} \tag{13}$$

Again $c_j(x)$ can be arbitrary function of x_j. In this paper we will use polynomial specification. The marginal productivity of GCD model is:

$$\frac{\partial y}{\partial x_k} = \beta_0 \times \beta_k \times m'(c_k(x_k)) \times m(c_k(x_k))^{\beta_k - 1} \times \prod_{j=1, j\neq k}^{p} m(c_j(x_j))^{\beta_j} \tag{14}$$

If both β_0 and β_k are positive, we can be sure the marginal productivity of input k is always positive. As all output is non-negative, β_0 will always be greater than zero. The change of marginal productivity of kth input respect to itself is defined by:

$$\begin{aligned} \frac{\partial^2 y}{\partial x_k^2} &= \beta_0 \times \beta_k \times [(\beta_k - 1) \times m'(c_k(x_k,))^2 \times m(c_k(x_k))^{\beta_k - 2} + \\ & m''(c_k(x_k)) \times m(c_k(x_k))^{\beta_k - 1}] \times \prod_{j=1, j\neq k}^{p} m(c_j(x_j))^{\beta_j} \end{aligned} \tag{15}$$

If β_k lies between 0 and 1, we can guarantee the second derivative is negative. However different from Cobb-Douglas model, even if β_k is greater than 1 the second derivative can still be negative. The actual sign of second derivative depends on $c_k()$ and x_k. We cannot give a specific range. Yet one can always evaluate the second derivative after the model is estimated. Similar to Additive model, we need to take natural log to separate the deterministic frontier from error term.

$$\ln(y_i) = \ln(\beta_0) + \sum_{j=1}^{p} \beta_j \ln(m(c_j(x_{i,j}))) + u_i + v_i \tag{16}$$

Both Additive and GCD model preserve monotonicity and concavity conditional on right βs. The third model we suggest is a restricted form of GCD model. By construction, $m()$ cannot take value much lager than 1. Intuitively in GCD model β_0 is the global scaling parameter to match $m()$s to the scale of y. β_j is the individual scaling parameter for input j. Theoretically β_js are not necessary, because $c_j()$s can adjust the relative scales between inputs by themselves. The third model fix β_js to 1. We will call the third model Restricted Generalized Cobb-Douglas(RGCD) model. By eliminating β_js, we enforce the model to preserve monotonicity and concavity. Yet it does not come without cost. β_js play an important role in GCD model. Getting rid of them may reduce estimation accuracy. The impact can be compensated by adding more degree of freedom to $c_j()$s. In some circumstance RGCD can even out perform GCD model. We suggest researchers to use RGCD when number of observations per degree of freedom is low and data well behaves. RGCD model can also be used by researchers really want to impose monotonicity and concavity constraints on a dataset inconsistent with monotonicity or concavity, however the result can be strange and unpredictable. The specification of RGCD model is:

$$y = \beta_0 \prod_{j=1}^{p} m(c_j(x_j)) \tag{17}$$

2.4 Estimation and Inferences

For simplicity, we assume $c(x)$ to be polynomial function in the rest of paper.

$$c(x) = \sum_{j=1}^{k} c_j x^j \tag{18}$$

Estimation methods for three models are virtually the same. We will discuss estimation of Additive model as an example. As the objective function is highly nonlinear, we have two difficulties while estimating the model. The parameter estimation may be sensitive to starting value, moreover for some starting value, the log likelihood function cannot be evaluated. Suggested models have high degree of freedom per independent variable. Estimation is computationally intensive and sensitive to local minimum.

To overcome these difficulties we suggest a three-step estimation procedure. First we estimate the stochastic production frontier model with Cobb-Douglas production function. Then we can use its $\hat{\lambda}$ and $\hat{\sigma}$ as starting value. Our simulation shows that despite to be restrictive, Cobb-Douglas specification recovers distributional parameter very well.

Cobb-Douglas specification will also generate a prediction of deterministic production frontier $\hat{PF}$. We use $\hat{PF}$ as pseudo true production frontier and estimate $\hat{PF}$ with Additive model using the least square estimator. $\hat{\beta}$ and $\hat{c}$ generated by least square estimator will be used as starting value of β and c. In our research 0 is good starting value for c, and 1 is good starting value for β in the least square estimation.

After these two auxiliary regressions our starting value should be reasonably close to the true value. It is very likely the log likelihood function can be evaluated with starting value. In more than 10000 simulations we reported with two different models cross wide ranges of error term settings, non fail to find appropriate starting value. If log likelihood function still cannot be evaluated after the auxiliary regressions, we suggest to use Simulated-annealing algorithm optimize the log likelihood function. Simulated-annealing algorithm is very slow, in our experience, at least 10 times slower than the quasi-Newton method, but it does not require log likelihood function able to be evaluated with starting value.

After appropriate starting values are acquired, we can estimate the model using quasi-Newton method(also known as a variable metric algorithm). In our simulations the average convergence time is 7.2 seconds for a model with 2 variables and 200 observations. Both Simulated-annealing algorithm and quasi-Newton method are supported by the optim() function in R and are fairly easy to use.

As β, λ, σ and c themselves do not have economic meanings, we need to derive point estimations and inferences of values that interested by economists. Applied researchers are commonly interested in marginal productivities, input elasticities, return to scale, Elasticities of Substitution and average TE.

The mean productivity of input j in Additive model equals equation (8) evaluated at sample mean. For Generalized Cobb-Douglas model and RGCD model, the marginal productivities are defined in equation (15). The input elasticities are derived from normalizing mean productivities with input-output ratios. The Return to scale equals to summation of all input elasticities. The elasticity of substitution of ith input to jth input equals input elasticity of ith input divide by that of jth input. Average TE equals $E[exp(-v)|\hat{\theta}]$. $\hat{\theta}$ is a vector of estimated parameters of distribution that $-v$ was assumed to follow. The expression to recover average TE for half-normal distribution is:

$$TE = \int_{-\infty}^{\infty} \exp(-|x|)\phi_{0,\sigma}(x)\mathrm{d}x \tag{19}$$

in which, $\phi_{0,\sigma}$ is pdf of

$$\text{Normal}(0, \sqrt{\frac{\sigma^2\lambda^2}{1+\lambda^2}})$$

Because analytically deriving Jacobian and Hessians needed for calculating Fishers Information Matrices is very tedious, we use "numDeriv" package in R to numerically approximate them. In most simulations Fisher's Information Matrices are near singular. Therefore, we suggest more time consuming but more robust Bootstrapping approach to inference values above. In regression Bootstrapping we use $\hat{y}$ as true EF, and resample $\hat{\epsilon}$ with replacement. By the natural of Stochastic Efficiency Frontier model, our Bootstrapping setup is a little bit different from ordinary:

$$\tilde{y} = \hat{y}\exp(\tilde{\epsilon}) \tag{20}$$

2.5 Monte Carlo Simulations

In this section we will conduct three groups of simulations. First two simulations intended to investigate estimation accuracy of suggested models under separability assumption. Third simulations are conducted without separability assumption.

In first two simulation we will use equation (21) to simulate y.

$$y_i = (4.5 + 18\ln(x_{i,1}) - 2.2x_{i,1}^2 + 16\ln(x_{i,2}) - 2x_{i,2}^2)\exp(u_i + v_i) \tag{21}$$

$$u_i \sim \text{Normal}(0, \sigma_u)$$
$$v_i \sim -|\text{Normal}(0, \sigma_v)|$$

$x_{i,1}$ is generated by $Gamma(100, 10)$ distribution. $x_{i,2}$ is generated by $Gamma(120, 20)$ distribution. Both variables were rescaled into $[1, 2]$ range. With setup in equation (21) we can guarantee the production function to be smooth, monotone increase and concave. We choose this form for deterministic frontier, because we want to test suggested models on some arbitrary smooth monotone concave function.

Each simulation contains 200 observations. In each simulation we also generated additional 200 "naked" observations to evaluate out of sample prediction accuracy of the model. The term "naked" means dependent variable of evaluation observations only contains deterministic frontier i.e. $(4.5 + 18\ln(x_{i,1}) - 2.2x_{i,1}^2 + 16\ln(x_{i,2}) - 2x_{i,2}^2)$. Independent variables in evaluation sample are generated from identical procedure to that in simulation. Intuitively out of sample MSE is an empirical approximation of Normalized Integrated Squared Error (NISE)

$$\text{NISE} = \frac{1}{36}\int_1^6\int_1^6 (\hat{PF}(x_1, x_2) - PF(x_1, x_2))^2 \mathrm{d}F_1(x_1)\mathrm{d}F_2(x_2) \tag{22}$$

Table 1: Predictive Accuracy of Additive, Generalized Cobb-Douglas (GCD) and Cobb-Douglas Models on Efficiency Frontier, with Separability Assumption

	GCD Model			Additive Model			Cobb-Douglas Model		
Lambda	0.7	0.8	0.9	0.7	0.8	0.9	0.7	0.8	0.9
3	0.306	0.130	0.048	0.394	0.268	0.237	0.446	0.309	0.279
	(0.262)	(0.093)	(0.039)	(0.289)	(0.126)	(0.079)	(0.431)	(0.137)	(0.121)
	0.356	0.155	0.065	0.460	0.328	0.315	0.540	0.392	0.372
	(0.308)	(0.111)	(0.054)	(0.345)	(0.203)	(0.157)	(0.504)	(0.276)	(0.267)
5	0.211	0.089	0.038	0.293	0.235	0.237	0.323	0.260	0.263
	(0.147)	(0.055)	(0.020)	(0.130)	(0.086)	(0.089)	(0.141)	(0.096)	(0.100)
	0.253	0.114	0.054	0.369	0.298	0.305	0.413	0.351	0.345
	(0.190)	(0.079)	(0.042)	(0.221)	(0.162)	(0.151)	(0.293)	(0.244)	(0.233)
7	0.161	0.075	0.037	0.270	0.230	0.227	0.299	0.253	0.265
	(0.094)	(0.040)	(0.016)	(0.113)	(0.074)	(0.075)	(0.131)	(0.086)	(0.101)
	0.190	0.104	0.054	0.337	0.289	0.299	0.371	0.322	0.357
	(0.118)	(0.083)	(0.043)	(0.203)	(0.132)	(0.147)	(0.251)	(0.186)	(0.223)

0.7, 0.8, 0.9 are different TE settings
Numbers in parenthesis are standard divinations of value above them.
Numbers on the top without parenthesis are average MSE of in sample prediction.
Numbers at the bottom without parenthesis are average NISE.

We did simulations in three different average TE settings by three different λ settings grid. We compared the deterministic production frontier recovery ability and the distributional parameter recovery ability amount Additive model, GCD model and Cobb-Douglas model. For each point in grid we did 500 simulations.

We used cubic polynomial function for $c(x)$ in all three simulations. Because the model already contains a constant term, we exclude the constant term in $c(x)$. The Additive model used to estimate simulated data is:

$$y_{out} = \beta_0 + \beta_1 m(c_1(x_1)) + \beta_2 m(c_2(x_2)) \tag{23}$$

in which

$$m(c(x)) = \int_0^x \exp(-\int_0^a \exp(c(b))\,\mathrm{d}b)\,\mathrm{d}a$$
$$c_i(x_i) = c_{i,1}(x_i) + c_{i,2}(x_i)^2 + c_{i,3}(x_i)^3$$

In Table 1 is the deterministic frontier recovery accuracy comparison between Additive model, GCD and Cobb-Douglas model. NISE tend to increase when average TE drops, since σ_v is a decreasing function of TE. For instance σ_v corresponding to 0.7 average TE is 0.4982, and for 0.9 average TE σ_v is 0.1361. TE is a scaling factor. For given λ it also controls variance of stochastic frontier.

Compare to mean variance of the true deterministic frontier, which is 2.21 all models did a good job recovering it. It is also clear that Additive and GCD models consistently outperform Cobb-Douglas model, cross all settings. Especially when TE gets from 0.8 to 0.9, NISE of suggested models keep reducing. NISE of Cobb-Douglas model become more or less stagnate. Reason is none models is the true model. Suggested and Cobb-Douglas models are trying to approximate an arbitrary smooth monotone concave function. As suggested models are more flexible, it can proximate the true model with less error. This effect becomes more significant when deterministic frontier is covered by less volatile noise.

It surprised us that GCD consistently outperform Additive model. The setup of first simulation specifically favors Additive model over GCD model i.e. separability assumption.

The standard deviations of NISEs of suggested models also are smaller than Cobb-Douglas model cross all setups. Suggested models' performance is also more stable in terms of deterministic frontier recovery. Relative standard deviation gap gets larger when average TE gets closer to 1 for the same reason stated before.

λ by definition, can serve as the measure of significance of stochastic frontier in disturbance term. Smaller the λ more significant stochastic frontier is. Changing λ does not affect NISE much, when TE is close to 1. When TE gets smaller λ's effect becomes more significant. Overall smaller λ leads more inaccurate estimation.

Similar to deterministic frontier, MAE of average TE in Table 2 is a decreasing function of both average TE and λ. All models estimated average TE rather accurately. When average TE is small performance of all models is very close. When average TE and λ get bigger the gap become more and more significant. Similar to what happens in NISE of deterministic frontier measure, Additive and Cobb-Douglas model fails to continue improve average TE estimation when average TE gets 0.9. Yet Additive Model suffer less approximation error effect.

Table 2: Predictive Accuracy of Additive Model and Cobb-Douglas Model on Technical Efficiency, with Separability Assumption

	GCD Model			Additive Model			Cobb-Douglas Model		
Lambda	0.7	0.8	0.9	0.7	0.8	0.9	0.7	0.8	0.9
3	0.0194	0.0142	0.0097	0.0188	0.0151	0.0173	0.0200	0.0172	0.0228
	(0.0172)	(0.0119)	(0.0087)	(0.0176)	(0.0119)	(0.0098)	(0.0204)	(0.0133)	(0.0144)
5	0.0151	0.0110	0.0074	0.0134	0.0121	0.0176	0.0144	0.0126	0.0223
	(0.0115)	(0.0082)	(0.0062)	(0.0101)	(0.0089)	(0.0103)	(0.0109)	(0.0090)	(0.0115)
7	0.0129	0.0094	0.0068	0.0119	0.0121	0.0179	0.0132	0.0129	0.0238
	(0.0095)	(0.0074)	(0.0058)	(0.0091)	(0.0085)	(0.0098)	(0.0099)	(0.0094)	(0.0131)

0.7, 0.8, 0.9 are different TE settings
Numbers on the top are Mean Absolute Error of Technical Efficiency.
Numbers in parenthesis are standard deviation of Error of Technical Efficiency.
GCD: Generalized Cobb-Douglas

Based on our first simulation suggested models dominates Cobb-Douglas model in deterministic frontier recovery accuracy and average TE recovery accuracy cross all settings. The benefit of using suggested models over Cobb-Douglas model becomes more significant when average TE gets larger. Between suggested models, GCD model outperform Additive model significantly even data agree with separability assumption.

Fixed setup simulation gives us idea how Additive model performs at different points. Yet it does not tell us how model performs between these points. Based on our study of existing literature, average TE most likely lies between 0.3 and 0.95, and λ most likely lies between 1 and 10. We did a random setup simulation. In which average TE and λ will uniformly or close to uniformly distributed between range stated before. Then we constructed surfaces of $\hat{f}(MAE_{TE}|\lambda, TE)$ and $\hat{f}(NISE_{DEF}|\lambda, TE)$, using local constant kernel estimator. In second simulation we also intent to push Additive model to the limit, see what setting will rand Additive model useless. In second simulation we will also use cubic $c(x)$ identical to which used first simulation.

Average TE is controlled through changing σ_v. Relationship between two parameters is not linear. Range of σ_v corresponding to average TE range stated above is from 0.0652 to 2.32. Rather than generate uniform σ_v, we generated sample σ_{unif} using uniform between $exp(-0.0652)$ and $exp(-2.32)$ and rescaled back by $\sigma_v = -ln(\sigma_{unif})$. We did some simulations to make sure average TE's distribution is close enough to uniform. We used standard normal quantile to transform average TE rescaled to 0-1 range. Transformed variables cannot reject null hypothesis of Anderson-Darling normality test in all 100 simulations with sample size of 100 at 10% level, and cannot reject same hypothesis in 21 out of 100 simulations with sample size of 1000 at same level. Therefore simulated average TE is not strictly uniform distributed, but it should be close enough for our purpose. We estimated 5000 simulated samples. For consistency each sample contains 200 observations.

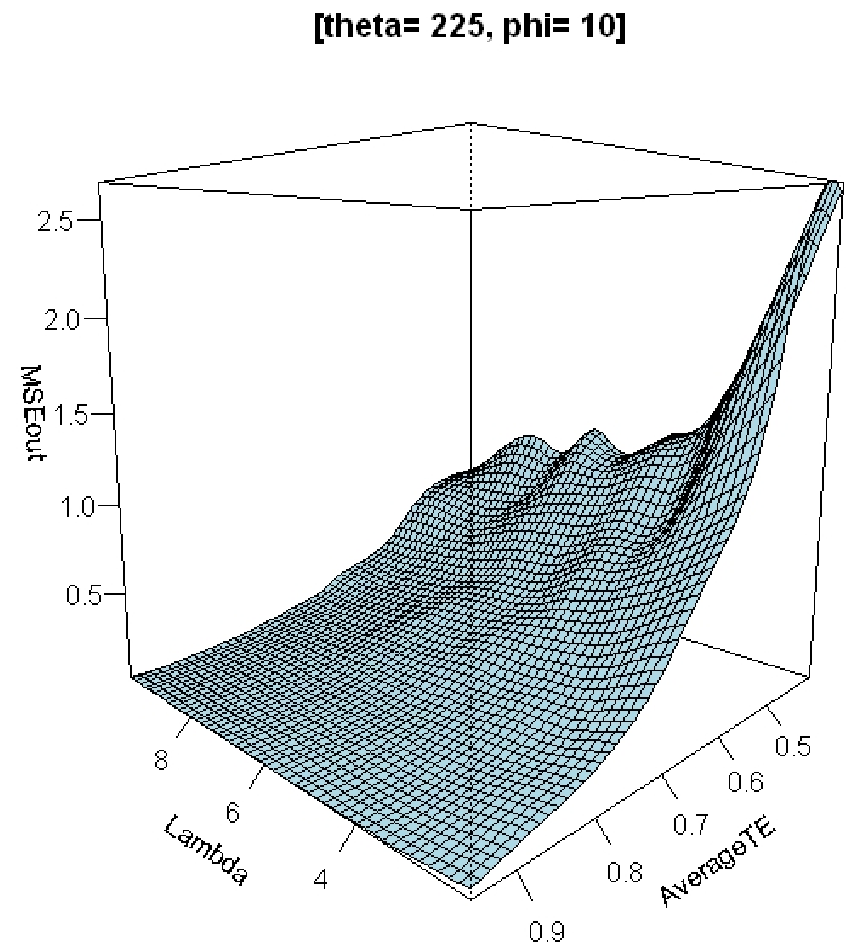

Figure 1: Mean Squared Error (MSE) of out of sample prediction

Because NISE becomes too large when average TE gets below 0.4 and λ gets below 2. We reduced grid to [2-9]$\times$[0.4-0.95], while plotting NISE. In Figure 1 we can see in large part of our setting grid, Additive model performs well. When λ is large NISE is a quite flat and close to linear function of average TE. When λ is small, NISE becomes a much steeper and convex function of average TE. While average TE is close to 1, NISE is almost flat function of λ. While average TE gets smaller, NISE becomes more steep function of λ.

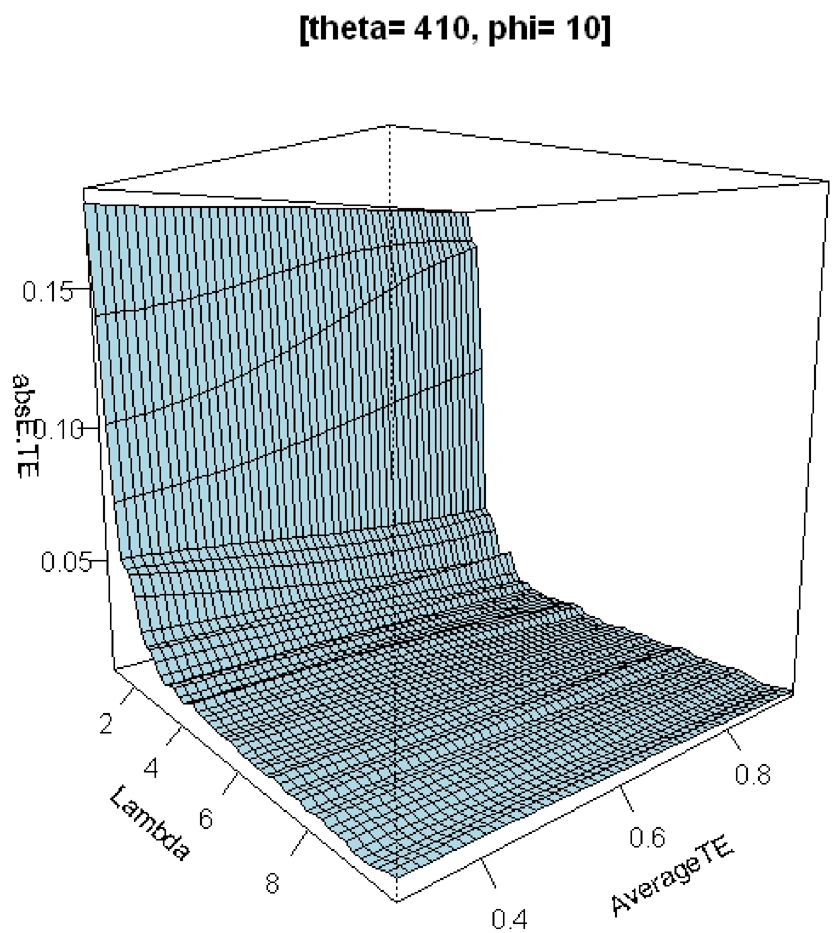

Figure 2: Absolute error of average Technical Efficiency (absE.TE) recovery

Table 3: Predictive Accuracy of Additive, Generalized Cobb-Douglas (GCD) and Cobb-Douglas Models on Efficiency Frontier, without Separability Assumption

	Additive Model			GCD Model			Cobb-Douglas Model		
Lambda	0.7	0.8	0.9	0.7	0.8	0.9	0.7	0.8	0.9
3	0.539	0.362	0.304	0.463	0.217	0.092	0.652	0.465	0.437
	(0.364)	(0.165)	(0.126)	(0.408)	(0.152)	(0.054)	(0.421)	(0.290)	(0.229)
	0.640	0.435	0.395	0.550	0.267	0.129	0.838	0.604	0.574
	(0.497)	(0.262)	(0.220)	(0.511)	(0.195)	(0.087)	(0.912)	(0.448)	(0.404)
5	0.429	0.314	0.297	0.346	0.155	0.071	0.497	0.403	0.511
	(0.258)	(0.128)	(0.126)	(0.258)	(0.085)	(0.036)	(0.259)	(0.277)	(0.270)
	0.530	0.392	0.391	0.415	0.212	0.104	0.626	0.543	0.664
	(0.419)	(0.234)	(0.263)	(0.300)	(0.160)	(0.073)	(0.515)	(0.434)	(0.413)
7	0.360	0.293	0.300	0.276	0.134	0.071	0.457	0.401	0.377
	(0.166)	(0.109)	(0.108)	(0.176)	(0.069)	(0.047)	(0.222)	(0.389)	(0.144)
	0.452	0.366	0.377	0.331	0.181	0.101	0.568	0.524	0.519
	(0.270)	(0.182)	(0.209)	(0.224)	(0.114)	(0.084)	(0.424)	(0.492)	(0.390)

0.7, 0.8, 0.9 are different TE settings
Numbers in parenthesis are standard divinations of value above them.
Numbers on the top without parenthesis are average MSE of in sample prediction.
Numbers at the bottom without parenthesis are average NISE.

In Figure 2 we can see absolute error of average TE recovery is consistently a very flat liner function of average TE. When λ is larger than 4 it is also an almost flat function of λ. When λ gets below 4 absolute error of average TE recovery starts to climb. When λ gets below 2 it rises almost straight up. We also did random setup simulations for GCD model. The result is very similar to that of Additive model. Therefore we will not report the result in this paper.

In third simulation we used a slightly different deterministic frontier. We added a monotone concave cross term in production function, as GCD model do not assume separability. The mean variance of production function with 200 observations is 3.99.

$$\begin{aligned} y_i =& (4.5 + 18\ln(x_{i,1}) - 2.2x_{i,1}^2 + 16\ln(x_{i,2}) - x_{i,2}^2 + \\ & 2.7x_{i,1}x_{i,2} - 0.3(x_{i,1}x_{i,2})^2)\exp(u_i + v_i) \end{aligned} \tag{24}$$

in which

$$\begin{aligned} & u_i \sim \text{Normal}(0, \sigma_u) \\ & v_i \sim -|\text{Normal}(0, \sigma_v)| \end{aligned}$$

In Table 3 are predictive accuracy comparisons on production frontier. Similar to the first simulation, GCD model consistently out performs Additive model, which outperforms Cobb-Douglas model. Gap become larger when average TE and λ get larger. In column 0.9 NISEs of Cobb-Douglas are nearly 5 times of those of GCD.

Table 4: Predictive Accuracy of Additive Model and Cobb-Douglas Model on Technical Efficiency, without Separability Assumption

	Additive Model			GCD Model			Coub-Douglas Model		
Lambda	0.7	0.8	0.9	0.7	0.8	0.9	0.7	0.8	0.9
3	0.0176	0.0145	0.0149	0.0181	0.0152	0.0106	0.0199	0.0170	0.0257
	(0.0162)	(0.0117)	(0.0097	(0.0151)	(0.0132)	(0.0087)	(0.0168)	(0.0147)	(0.0421)
5	0.0144	0.0121	0.0143	0.0141	0.0108	0.0079	0.0144	0.0151	0.0291
	(0.0109)	(0.0093)	(0.0085	(0.0112)	(0.0086)	(0.0075)	(0.0117)	(0.0158)	(0.0158)
7	0.0123	0.0107	0.0150	0.0132	0.0101	0.0079	0.0137	0.0168	0.0219
	(0.0097)	(0.0081)	(0.0089	(0.0097)	(0.0080)	(0.0085)	(0.0103)	(0.0520)	(0.0123)

0.7, 0.8, 0.9 are different TE settings
Numbers on the top are Mean Absolute Error of Technical Efficiency.
Numbers in parenthesis are standard deviation of Error of Technical Efficiency.

One can see in Table 4, average TE recovery accuracy of GCD is consistently higher than Cobb-Douglas. Moreover in column 0.9 GCD did not suffer approximation error effect. Third simulation tells us, GCD model always dominates Cobb-Douglas model in terms of deterministic frontier recovery and average TE recovery.

2.6 An Empirical Application

In this section, we will demonstrate suggested models, by estimating an airline dataset. Data is downloaded from Greene's website . This data set is an extension of Caves et al.(1984). It is an unbalanced panel with 256 observations on 25 firms. Independent variables are materials, fuel, equipment labor and property. The dependent variable is an index of airline out put. Summary statistic of the dataset is in Table 5.

First we will estimate Additive model and GCD model under pooled model. The procedure is identical to what we did in simulation part and what we suggest researcher do to cross sectional dataset. Throughout this section, We will use the cubic $c(x)$ same as what we used in simulation. Our empirical model for additive deterministic frontier is:

Table 5: Summary Statistic of Airline Dataset-A

	Output	Material	Fuel	Equipment	Labor	Property
Mean	0.6288	0.7516	0.5839	0.6517	0.5950	0.6562
StDev	0.5919	0.6430	0.5038	0.5677	0.5082	0.6926
Min	0.0234	0.0645	0.0462	0.0496	0.0632	0.0146
Median	0.4000	0.4733	0.3609	0.3597	0.3588	0.3728
Max	2.4424	2.4479	1.7698	2.1049	1.6919	2.8070
Skewness	0.7945	0.6510	0.6004	0.7171	0.6787	1.1252
Kurtosis	-0.5266	-1.0430	-1.0934	-0.7869	-1.0924	0.3070

StDev:Standard Deviation

$$\begin{aligned} y_{out} = &\beta_0 + \beta_1 m(c_1(Mat.)) + \beta_2 m(c_2(Fue.)) + \beta_3 m(c_3(Eqp.)) \\ &+ \beta_4 m(c_4(Lab.)) + \beta_5 m(c_5(Pro.)) \end{aligned} \tag{25}$$

in which

$$m(c(x)) = \int_0^x \exp(-\int_0^a \exp(c(b))\, \mathrm{d}b)\, \mathrm{d}a$$

$$c_i(x_i) = c_{i,1}(x_i) + c_{i,2}(x_i)^2 + c_{i,3}(x_i)^3$$

In Table 6 is the estimation of interested values less marginal productivity and elasticity of substitutions. Elasticity of substitutions will be reported in different table. Because dataset was already rescaled before we get it, marginal productivities do not bear any economic meaning. The standard deviation and confidence interval is derived from Bootstrapping with 399 replications. The bootstrapping took 3.06 hours, without parallel computing. Because truncated normal model is less restrictive than half normal model, we also estimated the model for comparison. Overall estimation of two models are very close, except for λ. For some reason λ estimation of truncated normal model is very volatile. Fortunately average TE the variable we are interested in is rather stable, and close to half normal model.

In half normal model estimated λ is pretty small, this means stochastic frontier roughly contributed 36.4% of total variation of error term. Average TE is 89.28%. Most of airlines are operating at very high efficiency. Average return to scale is 1.32. A "average" airline is smaller than economic scale.

Input elasticity of labor is negative, and non replication in bootstrapping has positive value. This is inconsistent with the Economic theory. Yet it happens to Cobb-Douglas model as

Table 6: Vital Parameter Estimation of Additive Model

	Half-Normal				Truncated Normal			
	Mean	StDev	95%L	95%U	Mean	StDev	95%L	95%U
Mu	-	-	-	-	0.306	0.049	0.216	0.358
Sig.	0.206	0.027	0.156	0.263	0.174	0.019	0.155	0.194
Lam.	1.211	0.677	0.147	2.562	44.98	77.06	0.460	200.5
TE	0.893	0.038	0.832	0.982	0.891	0.027	0.866	0.949
RTS	1.316	0.063	1.178	1.435	1.593	0.073	1.476	1.673
Mat.*	0.669	0.095	0.484	0.847	0.786	0.121	0.577	0.995
Fue.*	0.403	0.081	0.245	0.564	0.537	0.103	0.376	0.723
Eqp.*	0.337	0.090	0.161	0.501	0.378	0.106	0.191	0.561
Lab.*	-0.263	0.079	-0.410	-0.118	-0.309	0.107	-0.496	-0.127
Pro.*	0.171	0.035	0.104	0.247	0.202	0.042	0.126	0.270

Sig: Sigma
Lam.:Lambda
Mu: Mean of Truncated Normal Distribution
TE: Average Technical Efficiency
RTS: Average Return to Scale
* are average input elasticity for given input
Mat: Material, Fue.: Fuel, Eqp: Equipment, Lab.: Labor, Pro.:Property.
StDev: Standard Deviation, L: Lower Bound, U: Upper Bound

well as other models Greene tried on. We will just take it as how data behave. Elasticities of all other input are positive. Material has largest input elasticity. 1 percent increase in material will lead to 0.67 percent increase in output.

Table 7: Elasticity of Substitution Additive Model Half Normal Distribution

	Mtl.	Fue.	Eqp.	Lab.	Pro.
Mtl.	1	1.7458	2.2070	-2.8848	4.1102
	(0)	(0.5411)	(0.9658)	(1.9022)	(1.1602)
Fue.	0.6164	1	1.3257	-1.7018	2.4785
	(0.1606)	(0)	(0.5864)	(0.8856)	(0.7776)
Eqp.	0.5222	0.8891	1	-1.4406	2.0686
	(0.1829)	(0.3593)	(0)	(0.7964)	(0.7565)
Lab.	-0.3949	-0.6652	-0.8413	1	-1.5935
	(0.1148)	(0.2051)	(0.3681)	(0)	(0.5668)
Pro.	0.2610	0.4463	0.5524	-0.7319	1
	(0.0678)	(0.1536)	(0.2201)	(0.4056)	(0)

All numbers are derived by elasticities of parameters in row divide by elasticities of parameters in column. For example, 1.7458 in row Mtl., column Fue. is derived from $E_{Mtl.}/E_{Fue.}$. Mtl: Material, Fue.: Fuel, Eqp: Equipment, Lab.: Labor, Pro.:Property.

In Table 7, are the elasticities of substitutions cross all inputs estimated by half normal model. Although in theory marginal rate of substitution is a better measure for production function. Inputs in this dataset are indexed from multiple categories and rescaled . Unit of each input is unclear. Thus elasticities of substitution are better measure in this case.

In Table 8 are vital parameter estimation of GCD model. Return to scale and all elasticity estimations are reasonably close to Additive model. Within half normal framework we have large amount of bootstrapping scenario without any technical inefficiency. Within Normal-Truncated normal original estimation and most of bootstrapping scenario do not found technical inefficiency. When we were estimating GCD model with normal-truncated normal distribution , quasi-Newton method cannot find appropriate gradient. We solved this problem by adopting Nelder and Mead method, which does not use gradient or Hessian. We believe the cause of this problem is this model did not find technical inefficiency in the dataset. Nelder and Mead method is also supported by optim() function.

Table 8: Vital Parameter Estimation, Generalized Cobb-Douglas (GCD) Model

	Half-Normal				Truncated Normal			
	Mean	StDev	95%L	95%U	Mean	StDev	95%L	95%U
Mu	-	-	-	-	-65.6	30.7	-123.9	-3.7
Sig.	0.170	0.023	0.135	0.217	2.027	0.753	0.216	3.348
Lam.	0.659	0.524	0.000	1.598	15.42	6.153	1.277	26.18
TE	0.938	0.046	0.871	1.000	-	-	-	-
RTS	1.222	0.068	1.111	1.342	1.208	0.035	1.132	1.277
Mat.*	0.730	0.083	0.565	0.918	0.680	0.073	0.529	0.832
Fue.*	0.405	0.082	0.243	0.570	0.390	0.065	0.258	0.510
Eqp.*	0.322	0.078	0.174	0.467	0.350	0.056	0.240	0.465
Lab.*	-0.422	0.077	-0.593	-0.269	-0.405	0.052	-0.515	-0.299
Pro.*	0.187	0.036	0.121	0.269	0.192	0.030	0.130	0.252

Sig: Sigma
Lam.:Lambda
Mu: Mean of Truncated Normal Distribution
TE: Average Technical Efficiency
RTS: Average Return to Scale
* are average input elasticity for given input
Mat: Material, Fue.: Fuel, Eqp: Equipment, Lab.: Labor, Pro.:Property.
StDev: Standard Deviation, L: Lower Bound, U: Upper Bound

Work of previous researchers allows us to exploit panel feature of the dataset. Pitt and Lee (1981) suggested two random effect models. In the first model they assume inefficiency term is time invariant and follows half-normal distribution. In the second model they assume inefficiency term is serial correlated. Because integration of multivariate distribution is computationally infeasible they suggested a Constrained Seemingly Unrelated Regression (CSUR) approach. With our dataset, equations in CSUR specification can have as few as 12 observations. In each equation we will have at least 21 parameters. Therefore we cannot estimate second random effect model. Schmidt and Sickles (1984) suggested a fixed effect within estimator. A distinct advantage of their estimator is individual inefficiency of each firm can be identified.

Table 9: Vital Parameter Estimation, Generalized Cobb-Douglas (GCD) Model, under Random Effect and Fixed Effect Specifications

	Fixed Effect				Random Effect			
	Mean	StDev	95% L	95% U	Mean	StDev	95% L	95% U
Sig.	0.206	0.028	0.156	0.256	0.270	0.034	0.192	0.335
Lam.	21.90	3.680	15.55	29.473	2.863	0.553	1.857	3.944
TE	0.676	0.046	0.577	0.770	0.983	0.003	0.977	0.989
RTS	1.283	0.039	1.198	1.354	1.508	0.073	1.348	1.642
Mat.*	0.852	0.067	0.748	0.998	1.101	0.087	0.945	1.270
Fue.*	-0.138	0.082	-0.293	0.021	0.257	0.032	0.196	0.324
Eqp.*	0.238	0.094	0.084	0.461	0.346	0.093	0.155	0.541
Lab.*	0.247	0.074	0.099	0.379	-0.271	0.060	-0.385	-0.158
Pro.*	0.084	0.027	0.043	0.146	0.076	0.023	0.034	0.117

Sig: Sigma
Lam.:Lambda
TE: Average Technical Efficiency
RTS: Average Return to Scale
* are average input elasticity for given input
Mat: Material, Fue.: Fuel, Eqp: Equipment, Lab.: Labor, Pro.:Property.
StDev: Standard Deviation, L: Lower Bound, U: Upper Bound

As GCD is very close to Cobb-Douglas function Pitt and Lee (1981) discussed, we use GCD model for fixed and random effect estimation. Additive model can also be estimated in all specifications we mentioned above with small modifications. In Table 9 is the vital parameter of fixed and random effect model. Estimation of random effect model is relatively close to counterpart in pooled model. We find almost no technical inefficiency in airline industrial with random effect model. Estimation of fixed effect model is quite different from all estimations before. It find more than 13 out of 25 firms is below 70% efficiency. Average efficiency is only 67.6%. Input elasticity of Labor becomes positive, which is good. However point estimation of Fuel's input elasticity change to negative, and we cannot reject hypothesis that it equal zero with $\alpha = 5\%$.

The original intension is use RGCD model on the dataset as well. Because dataset especially labor variable does not behave well, RGCD yield a very strange result. We suggest that RGCD model shouldn't be used on a dataset inconsistent with theory. The result can be strange and unpredictable.

2.7 Summary

Our purpose has been present a class of flexible functional forms embedding monotonicity and concavity for stochastic production frontier estimation. Suggested models comprehended Cobb-Douglas model's ability to preserve monotonicity and concavity, and flexibility. Despite highly nonlinear nature of the model, it can be estimated within relatively short time with three-step method suggested. Because Fisher's Information Matrix of Additive model is near singular, we suggested Bootstrapping inference. The time need for 399 replication of model with 5 independent variables and 256 observations is 3.06 hours, without parallel computing. Moreover in our simulation both Additive model and GCD model out performed Cobb-Douglas model in both deterministic frontier recovery and average TE recovery. We also gave researcher a general idea of how our model's accuracy behaves in different error term setting, through a random setup simulation. Then we applied the Additive model and GCD model to an airline dataset with both half-normal and normal-truncated normal distribution. Finally we demonstrated how to estimate GCD model under random and fixed effect specifications.

3. A CLASS OF FLEXIBLE INEFFICIENCY ANALYSIS MODELS WITH MONOTONICITY AND CONCAVITY CONSTRAINTS

3.1 Introduction

In this paper, we are interested in introduce a class of flexible frontier production function with monotonicity and concavity constraint into inefficiency analysis. Since Aigner, Lovell, and Schmidt (1977, hereafter ALS 1977) developed first practical Stochastic Production Frontier model many work has been done in this field. In recent years researchers extend interest to model firm level inefficiency with observable characteristics. Deterministic production frontiers researchers used in previous literatures belong to log linear family. For instance Caudill and Ford (1993) used Cobb-Douglas model, Huang and Liu (1994) used Trans-log model. Although log linear family is easy to estimate, its members does bear undesired theoretical implications.

Microeconomic theory requires production function to be monotone increasing respect to each input and diminishing return to scale. These two assumptions are essential to profit maximization, cost minimization and various equilibrium problems. Conventional Cobb-Douglas production function preserves both properties automatically when all coefficients are between 0 and 1. Yet it also enforces undesired properties such as unitary elasticity of factor substitution and constant partial and total production elasticities.

Trans-log can be considered as a generalized form of Cobb-Douglas function. It is very flexible, and generally does not imply any properties econometricians do not desired. Yet it does not imply monotone increasing respect to each input and diminishing return to scale either.

In this paper we will introduce a flexible model that imposes these two constraints without the burden of Cobb-Douglas model to inefficiency analysis. This paper will be organized as follow. In the next section we will do a brief literature review of inefficiency analysis. In Section 3.3 will introduce two alternative production functions for inefficiency analysis. Then we will derive estimator and likelihood function for inefficiency analysis. After that we will suggest estimation and inference method for interesting parameters. In Section 3.5 we will present result of Monte Carlo simulations and compare the result with Cobb-Douglas model. In Section 3.6 we will apply the model on a airline dataset.

3.2 Literature Review

Half Normal-Normal and Exponential-Normal frameworks developed by ALS (1977) have one common assumption. That is firms are more concentrated in higher efficiency area than in lower efficiency area. This assumption might not be true in reality. Stevenson (1980) generalized their model and developed Truncated Normal-Normal framework. Stevenson (1980) also suggested a Gamma-Normal framework with positive integer shape parameters. Greene (1990) generalized the Gamma-Normal framework to allow positive real shape parameter that is greater than one. Both Truncated Normal-Normal and Gamma-Normal frameworks do not imply undesired assumptions ALS (1977) have. Yet Gamma-Normal framework does not have close form log likelihood function. Estimation is much more computationally intensive than Truncated Normal-Normal framework. Inefficiency analysis is already much more computationally intensive than conventional production frontier problem. Previous researchers adopted Truncated Normal-Normal framework in inefficiency analysis.

Kumbhakar et al. (1991) made first attempt to explain individual inefficiency with observable characteristics. Their model allows mean of truncated normal distribution vary with characteristic variables. Huang and Liu (1994) suggested a similar model with two-step estimation method. Although two step method has the advantage of identifying individual inefficiency, the estimator could be biased. Wang and Schmidt (2002) showed that the bias can be substantial. Therefore in this paper we will focus on one step MLE estimator. Reifschneider and Stevenson (1991) developed a class of models, in which variances of inefficiency term depend on characteristic variables.

Alvarez et al. (2006, hereafter AAOS 2006) summarized results of previous papers. According to them, the nested model of inefficiency analysis is:

$$
\begin{aligned}
y_i &= f(x_i)\exp(u_i - v_i) \qquad (26)\\
u_i &\sim \mathrm{Normal}(0, \sigma_u)\\
v_i &\sim \mathrm{Normal}(\mu_{v,i}, \sigma_{v,i})_+
\end{aligned}
$$

In which

$$
\begin{aligned}
\mu_{v,i} &= \mu_v^* \exp(\gamma z_{\mu,i})\\
\sigma_{v,i} &= \sigma_v^* \exp(\delta z_{\sigma,i})
\end{aligned}
$$

$f(x_i)$ and $exp(u_i)$ are deterministic part and stochastic part of production frontier, which is identical to conventional stochastic efficiency frontier estimation. $exp(v_i)$ is efficiency level of ith firm. Different from conventional stochastic production frontier problem, each firm has their own distribution of efficiency determined by γ_i, δ_i, $z_{\mu,i}$, and $z_{\sigma,i}$. $z_{\mu,i}$, and $z_{\sigma,i}$ are vectors of firm level characteristics that affect efficiency. γ_i and δ_i are vectors of coefficients associated with these characteristics. x_i, $z_{\mu,i}$, and $z_{\sigma,i}$ can overlap each other partially or fully. In previous literatures that we are aware of, researchers assume $z_{\mu,i} = z_{\sigma,i}$. For convenience we adopt this assumption in the rest of the paper.

By putting different constraints on parameters, we can obtain three popular models in inefficiency analysis. When assuming $\gamma = \delta$ we have Scale Stevenson (SS) model used by Wang and Schmidt (2002).

$$v_i \sim \text{Normal}(\mu_v^* \exp(\delta z_i), \sigma_v^* \exp(\delta z_i))_+ \tag{27}$$

It can also be written as:

$$v_i \sim \exp(\delta z_i)\text{Normal}(\mu_v^*, \sigma_v^*)_+$$

Assuming $\delta = 0$, we have KGMHLBC model used by Kumbhakar et al. (1991), Huang and Liu (1994), and Battese and Coelli (1992).

$$v_i \sim \text{Normal}(\mu_v^* \exp(\gamma z_i), \sigma_v^*)_+ \tag{28}$$

When fixing $\mu_v^* = 0$, we got RSCFG model used by Reifschneider and Stevenson (1991), Caudill and Ford (1993), and Caudill et al. (1995).

$$v_i \sim \text{Normal}(0, \sigma_v^* \exp(\delta z_i))_+ \tag{29}$$

It can also be written as:

$$v_i \sim \exp(\delta z_i)\text{Normal}(0, \sigma_v^*)_+$$

3.3 A Class of Flexible Inefficiency Analysis Models with Shape Constraints

Sickles and Wu (Personal Discussion) developed a production function with monotonicity and concavity constraints embedded. Their function is:

$$f(x) = \beta_0 + \sum_{j=1}^{p} \beta_j m(c_j(x_j)) \tag{30}$$

in which

$$m(c(x)) = \int_0^x \exp(-\int_0^a \exp(c(b))\, \mathrm{d}b)\, \mathrm{d}a$$

In equation above $c(\cdot)$ is some arbitrary function. From now on we will call it Additive model. Marginal productivity of input j is:

$$\frac{\partial f(x)}{\partial x_j} = \beta_j m'(c_j(x_j)) \tag{31}$$

in which

$$m'(c(x)) = \exp(-\int_0^x \exp(c(b))\,\mathrm{d}b)$$

As exponential function, $m'(\cdot)$ will always be positive. The second derivative of Additive model is:

$$\frac{\partial^2 f(x)}{\partial {x_j}^2} = \beta_j m''(c_j(x_j)) \tag{32}$$

in which

$$m''(c(x)) = -\exp(c(x))\exp(-\int_0^x \exp(c(b))\,\mathrm{d}b) \tag{33}$$

Likewise $m''(\cdot)$ is guaranteed negative. Thus as long as all β_j is positive $f(x)$ is a monotone increasing function exhibiting diminishing return to scale. Additive model implicitly assume separability between inputs, i.e. $\partial^2 y/\partial x_k \partial x_j = 0$. Some researchers might not want this assumption. To solve this problem, we suggest a multiplicative production function.

$$y = \beta_0 \prod_{j=1}^{p} m(c_j(x_j))^{\beta_j} \tag{34}$$

The functional form of model above is very similar to that of Cobb-Douglas model. From now on we will call it Generalized Cobb-Douglas (GCD) model. First and second derivative of GCD model are:

$$\frac{\partial y}{\partial x_k} = \beta_0 \times \beta_k \times m'(c_k(x_k)) \times m(c_k(x_k))^{\beta_k - 1} \times \prod_{j=1, j\neq k}^{p} m(c_j(x_j))^{\beta_j} \tag{35}$$

$$\frac{\partial^2 y}{\partial x_k^2} = \beta_0 \times \beta_k \times [(\beta_k - 1) \times m'(c_k(x_k,))^2 \times m(c_k(x_k))^{\beta_k - 2} +$$

$$m''(c_k(x_k)) \times m(c_k(x_k))^{\beta_k - 1}] \times \prod_{j=1, j\neq k}^{p} m(c_j(x_j))^{\beta_j} \tag{36}$$

As $m(\cdot)$ and $m'(\cdot)$ are positive and $m''(\cdot)$ is negative, as long as β_js lie between 0 and 1, GCD is guaranteed to exhibit monotone increasing and diminishing return to scale property. More over because $m''(\cdot)$ is negative, $f(x)$ might preserve diminishing return to scale even if β_js is greater than 1. In this situation $\partial^2 y/\partial x_k^2$ depends not only on β_j but also xs and $c(\cdot)$. Researchers need to evaluate over range of all xs to make sure function is concave.

Substituting Additive model into nested model, we have Additive-Nested Inefficiency Analysis (A-NIA)model.

$$y_i = (\beta_0 + \sum_{j=1}^{p} \beta_j m(c_j(x_{i,j}))) \exp(u_i - v_i) \tag{37}$$

in which

$$m(c(x)) = \int_0^x \exp(-\int_0^a \exp(c(b))\, \mathrm{d}b)\, \mathrm{d}a$$

$$u_i \sim \mathrm{Normal}(0, \sigma_u)$$

$$v_i \sim \mathrm{Normal}(\mu_v^* \exp(\gamma z_i), \sigma_v^* \exp(\delta z_i))_+$$

When GCD is combined with nested model we get the estimator for GCD-Nested Inefficiency Analysis (GCD-NIA) model.

$$y_i = \beta_0 \prod_{j=1}^{p} m(c_j(x_{i,j}))^{\beta_j} \exp(u_i - v_i) \tag{38}$$

in which

$$m(c(x)) = \int_0^x \exp(-\int_0^a \exp(c(b))\,\mathrm{d}b)\,\mathrm{d}a$$

$$u_i \sim N(0, \sigma_u)$$

$$v_i \sim N(\mu_v^* \exp(\gamma z_i), \sigma_v^* \exp(\delta z_i))_+$$

In both equations above $c(\cdot)$ is some arbitrary function. To derive estimator for different inefficiency analysis model we only need to impose constraint required. For instance, when we apply $\delta = 0$ to GCD-NIA model, we get:

$$y_i = \beta_0 \prod_{j=1}^{p} m(c_j(x_{i,j}))^{\beta_j} \exp(u_i - v_i) \tag{39}$$

in which

$$u_i \sim \mathrm{Normal}(0, \sigma_u)$$

$$v_i \sim \mathrm{Normal}(\mu_v^* \exp(\gamma z_i), \sigma_v^*)_+$$

which is estimator of GCD-KGMHLBC model. To estimate the model we still need the log likelihood function. Based on log likelihood function derived by Stevenson (1980), we get the likelihood function for nested model.

$$LL_i = -\ln\sigma_i - \frac{1}{2}\ln 2\pi - \ln\Phi(\frac{\mu_i\lambda_i\sigma_i}{\sqrt{1+\lambda_i^2}}) + \ln\Phi(\frac{\mu_i}{\sigma_i\lambda_i} - \frac{\epsilon_i\lambda_i}{\sigma_i}) - \frac{1}{2}(\frac{\epsilon_i+\mu_i}{\sigma_i})^2 \quad (40)$$

in which

$$\epsilon_i = \ln(y_i) - \ln(f(x_i))$$
$$\sigma_i = \sqrt{\sigma_v^{*2}\exp(2\delta z_i) + \sigma_u^2}$$
$$\lambda_i = \frac{\sigma_v^*\exp(\delta z_i)}{\sigma_u}$$
$$\mu_i = \mu_v^*\exp(\gamma z_i)$$

In equation above $f(x_i)$ can be either GCD or Additive model. $\Phi(\cdot)$ is CDF of standard normal distribution.

3.4 Estimation, Parameter Recovery, and Inference

For simplicity, we assume $c(x)$ to be polynomial function in the rest of paper.

$$c(x) = \sum_{p=1}^{k} c_p x^p \quad (41)$$

As suggested models have many parameters to estimate and are highly nonlinear, result may be sensitive to starting value. Log likelihood function cannot be evaluated by many starting values. To overcome these two problems, we suggest a three-step procedure. First we estimate Cobb-Douglas production function using whatever Inefficiency analysis framework we assume, then use predicted $\hat{PF}$ as true PF, and use suggested models to proximate $\hat{PF}$ using least square estimator. 0 is good starting value for c_ps and 1 is good starting value for β_js. Parameter estimations of least square auxiliary regression should serve reasonably well as staring values. Point estimation of γ, δ, σ_v^* and μ_v^* from Cobb-Douglas model can be used as starting value of corresponding parameters in suggested models. After these two auxiliary regressions our starting value should be reasonably close to true value. We can estimate suggested model by maximizing log likelihood function in equation (15) with quasi-Newton algorithm. Quasi-Newton algorithm is supported by "optim()" function in R. It is pretty easy to use.

As βs and cs themselves do not have economic meanings, we need to derive point estimations and inferences of values that are interested by economists. Regarding deterministic production frontier, applied researchers are commonly interested in marginal productivities, input elasticities, return to scale, and Elasticities of Substitution. Mean productivity of suggested models equals first derivative evaluating at sample mean. Input elasticities are derived from normalizing mean productivities with input-output ratios. Return to scale equals to summation of all input elasticities. Elasticity of substitution of ith input to jth input equals input elasticity of ith input divide by that of jth input.

As γs and δs in inefficiency analysis affects σ_vs and/or μ_vs rather than technical efficiency itself, these coefficients does not bear any economic meanings. We need equation below to recover their marginal effect on technical efficiency in nested inefficiency analysis models.

$$\frac{\partial TE}{\partial z_{i,j}} = \int_0^\infty \exp(-x)[\delta_j(x-\mu_{v,i})^2\sigma_{v,i}^{-3} + \gamma_j(x-\mu_{v,i})\sigma_{v,i}^{-2} - \delta_j\sigma_{v,i}^{-1}]\phi(x)\mathrm{d}x \tag{42}$$

in which

$$\phi(\cdot) \text{ is pdf of Normal}(\mu_{v,i}, \sigma_{v,i})$$
$$\mu_{v,i} = \mu_v^* \exp(\gamma z_i)$$
$$\sigma_{v,i} = \sigma_v^* \exp(\delta z_i)$$

Equation above is affected by many factors. A negative coefficient alone does not mean the variable is negative or positive correlated with technical efficiency. it can only be evaluated case by case. Yet in different reduced models we might get some idea about the sign of marginal effect of given variable just from sign of γs or δs. For instance in RSCFG model, μ_v^* is restricted to zero. Thus positive coefficient means that variable is positive correlated with inefficiency.

Because analytically deriving Fishers Information Matrices and Hessians for suggested models is very teddies we use "numDeriv" package in R to numerically approximate them. In most simulations Fisher's Information Matrices are near singular. As in inefficiency analysis models we assume error term is correlated with observationally specified variables, we cannot do regression bootstrapping either. We propose researcher to resample observations with Bootstrapping and reestimate the model on resampled dataset.

3.5 Monte Carlo Simulations

In this section we will implement GCD model in inefficiency analysis and compare results to that of Cobb-Douglas model. In first simulation we use SS model. the equation is:

$$y_i = (4.5 + 18\ln(x_{i,1}) - 2.2x_{i,1}^2 + 16\ln(x_{i,2}) - x_{i,2}^2)\exp(u_i + v_i) \tag{43}$$

$$u_i \sim \text{Normal}(0, \sigma_u)$$

$$v_i \sim -\exp(\delta z_i)\text{Normal}(\mu_v^*, \sigma_v^*)_+$$

in which

$$\delta z_i = 0.6x_{i,1} - 0.6x_{i,2} + 0.6z_{i,1} - 0.6z_{i,2}$$

$x_{i,1}$ is generated by $Gamma(100, 10)$ distribution. $x_{i,2}$ is generated by $Gamma(120, 20)$ distribution. $z_{i,1}$ is generated by $Gamma(80, 25)$ distribution. $z_{i,2}$ is generated by $Gamma(140, 15)$ distribution. All variables were rescaled into $[1, 2]$ range. We use $\sigma_v^* = 0.4982$, $\sigma_u = 0.09964$, and $\mu = 0.5$. This error term setup was inherited from $TE = 0.7$ and $\lambda = 5$ setup in the first essay. The estimator for GCD-SS model is:

$$y_i = \beta_0 \prod_{j=1}^{p} m(c_j(x_{i,j}))^{\beta_j} \exp(u_i - v_i) \tag{44}$$

in which

$$u_i \sim \text{Normal}(0, \sigma_u)$$

$$v_i \sim \exp(\delta z_i)\text{Normal}(\mu_v^*, \sigma_v^*)_+$$

For consistency and simplicity we use third order polynomial for $c(x)$ in the rest of this section.

Table 10: Comparison of Generalized Cobb-Douglas-Scaled Stevenson (GCD-SS) and Cobb-Douglas Models

Observations	MSE	NISE	Sigma.v.e	Mu.e	Delta.1.e	Delta.2.e	Delta.3.e	Delta.4.e
	GCD-SS							
200	1.0101	1.1090	13.28%	24.50%	39.15%	41.25%	32.44%	32.98%
	(1.0097)	(1.2327)	(0.1235)	(0.1746)	(0.3191)	(0.3216)	(0.2603)	(0.2819)
400	0.5959	0.6623	9.24%	19.96%	31.34%	33.25%	22.59%	22.34%
	(0.5470)	(0.6689)	(0.0611)	(0.1442)	(0.2390)	(0.2587)	(0.1706)	(0.1826)
600	0.4176	0.4559	7.84%	17.47%	25.55%	25.24%	19.22%	20.65%
	(0.4259)	(0.4632)	(0.0526)	(0.1299)	(0.1860)	(0.1973)	(0.1466)	(0.1549)
800	0.2779	0.3119	6.22%	13.94%	21.05%	22.46%	17.06%	17.68%
	(0.2903)	(0.3156)	(0.0378)	(0.0947)	(0.1607)	(0.1663)	(0.1292)	(0.1333)
1000	0.2316	0.2568	5.72%	12.51%	19.21%	20.97%	16.17%	16.74%
	(0.2265)	(0.2468)	(0.039)	(0.090)	(0.147)	(0.148)	(0.130)	(0.120)
	Cobb-Douglas							
200	1.5032	1.8104	15.98%	27.83%	49.04%	49.78%	32.24%	34.48%
	(1.6798)	(2.2995)	(0.3147)	(0.2136)	(0.4955)	(0.4421)	(0.2545)	(0.6366)
400	0.9477	1.0927	9.67%	22.72%	32.10%	34.03%	24.29%	24.51%
	(1.0969)	(1.4254)	(0.0683)	(0.1885)	(0.2484)	(0.2686)	(0.1777)	(0.1829)
600	0.5685	0.6520	8.21%	16.54%	27.23%	27.45%	19.68%	21.33%
	(0.5995)	(0.7505)	(0.0612)	(0.1330)	(0.2113)	(0.1961)	(0.1458)	(0.1669)
800	0.4105	0.4735	6.68%	13.82%	22.31%	23.57%	17.39%	17.96%
	(0.3048)	(0.3974)	(0.0396)	(0.1080)	(0.1611)	(0.1784)	(0.1293)	(0.1352)
1000	0.3500	0.4143	5.64%	12.43%	21.31%	19.45%	15.38%	16.48%
	(0.2427)	(0.3323)	(0.0325)	(0.0898)	(0.1620)	(0.1552)	(0.1148)	(0.1224)

MSE: in sample prediction Mean Squared Error of deterministic production frontier. NISE: Normalized Integrated Squared Error. Sigma.v.e: mean percentage absolute error (MPAE)of $\sigma_{v,i}$ recovery. Mu.e: MPAE of $\mu_{v,i}$ recovery. Delta.i.e: MPAE of δ_i Standard deviations are in parenthesis

$$c_i(x_i) = c_{i,1}(x_i) + c_{i,2}(x_i)^2 + c_{i,3}(x_i)^3 \tag{45}$$

In each simulation we also generated additional 200 "naked" observations to evaluate out of sample prediction accuracy of the model. Term "naked" means dependent variable of evaluation observations only contains deterministic frontier i.e. $(4.5+18\ln(x_{i,1})-2.2x_{i,1}^2+16\ln(x_{i,2})-x_{i,2}^2)$. Independent variables in evaluation sample are generated from identical procedure to that in simulation. Intuitively out of sample MSE is an empirical approximation of Normalized Integrated Squared Error (NISE).

$$\text{NISE} = \frac{1}{36}\int_1^6\int_1^6 (\hat{PF}(x_1,x_2) - PF(x_1,x_2))^2 \mathrm{d}F_1(x_1)\mathrm{d}F_2(x_2) \tag{46}$$

As the in SS model we need to estimate second moment of y_i, the result tend to be less accurate than conventional production frontier estimation. Therefore we increased number of observations from 200 to 1000 gradually with 200 observations a step. For each setup we did 500 simulations. Results of first simulation is in Table 10.

GCD-SS model dominate Cobb-Douglas-SS model in terms of production frontier recovery. The gap shrinks when number of observation increases. The gap between MSE and NISE of Cobb-Douglas-SS model is also much larger than that of GCD-SS model. The difference is rather large even when sample size reaches 1000.

In inefficiency analysis, recovering individual inefficiency and δs is also amount primary interests. When sample size is small GCD-SS model dominates Cobb-Douglas-SS model.

When sample size gets larger, results of two models become very close to each other.

Although RSCFG model is less flexible than SS model, it is favored by a number of researchers. In second simulation we will us RSCFG specification. The equation used for simulation is:

$$y_i = (4.5 + 18\ln(x_{i,1}) - 2.2x_{i,1}^2 + 16\ln(x_{i,2}) - x_{i,2}^2)\exp(u_i + v_i) \tag{47}$$
$$u_i \sim \text{Normal}(0, \sigma_u)$$
$$v_i \sim -\exp(\delta z_i)\text{Normal}(0, \sigma_v^*)_+$$

in which

$$\delta z_i = 0.6x_{i,1} - 0.6x_{i,2} + 0.6z_{i,1} - 0.6z_{i,2}$$

The estimator of GCD-RSCFG model is:

$$y_i = \beta_0 \prod_{j=1}^{p} m(c_j(x_{i,j}))^{\beta_j} \exp(u_i - v_i) \tag{48}$$

in which

$$u_i \sim \text{Normal}(0, \sigma_u)$$
$$v_i \sim \exp(\delta z_i)\text{Normal}(0, \sigma_v^*)_+$$

Result of second simulation is in Table 11. With μ fixed to 0, production frontier estimation becomes much more accurate than SS model. GCD-RSCFG model consistently out perform Cobb-Douglas-RSCFG model in this field. Relative gap also gets bigger when sample size increases. For instance, when sample size is 200 in sample MSE of GCD-RSCFG is 63.77% of in sample MSE of Cobb-Douglas-RSCFG model. When sample size gets 1000 the ratio becomes 38.52%.

In terms of distributional parameter recovery GCD-RSCFG also performs better than Cobb-Douglas-RSCFG model. Distributional parameter estimation get worse when sample size increase from 800 to 1000.

3.6 An Empirical Application

In this section we will implement the GCD-SS and GCD-RSCFG models on an airline dataset. Data is downloaded from Greene's website . This data set is a extension of Caves et al.(1984). It is an unbalanced panel with 256 observations on 25 firms. Input variables are materials, fuel, equipment labor and property. The dependent variable is a index of airline out put.

This dataset also contains three characteristic variables that might affect efficiency. They are load factor (Loa.), average stage length (Sta.) and number of points served (Poi.). Load factor is percentage of capacity used for revenue generating activities. For instance percentage of ticket sold. We expected it to be strongly negative correlated with inefficiency. Average stage length is average length of each flight. The unit is mile. Points Served is number of airport each airline serves. Due to our limited understanding of airline industry, we do not have expectation on coefficient of these two variables. Summary statistics of the dataset are in Table 12.

Table 11: Comparison of Generalized Cobb-Douglas (GCD) and Cobb-Douglas Models under RSCFG Specification

Observations	GCD-RSCFG						
	MSE	NISE	Sigma.v.e	Gamma.1.e	Gamma.2.e	Gamma.3.e	Gamma.4.e
200	0.3661	0.4296	16.82%	49.79%	52.16%	40.83%	41.20%
	(0.2647)	(0.3323)	(0.2921)	(0.3758)	(0.3806)	(0.3198)	(0.3214)
400	0.1714	0.2119	8.84%	34.72%	38.80%	29.63%	31.30%
	(0.1003)	(0.1503)	(0.0590)	(0.2733)	(0.2988)	(0.2266)	(0.2356)
600	0.1221	0.1525	7.46%	30.22%	31.63%	25.14%	24.80%
	(0.0626)	(0.0934)	(0.0563)	(0.2459)	(0.2421)	(0.1922)	(0.1973)
800	0.0952	0.1146	6.46%	26.02%	28.76%	22.38%	23.59%
	(0.0470)	(0.0778)	(0.0442)	(0.2011)	(0.2089)	(0.1668)	(0.1724)
1000	0.0774	0.0942	5.41%	23.62%	24.33%	20.89%	19.70%
	(0.0350)	(0.0533)	(0.0278)	(0.1803)	(0.1855)	(0.1557)	(0.1541)
	Cobb-Douglas						
200	0.5741	0.6681	19.67%	60.57%	58.29%	42.88%	41.10%
	(0.3816)	(0.5072)	(0.4658)	(0.4956)	(0.4743)	(0.3184)	(0.3388)
400	0.3309	0.4032	10.54%	37.15%	40.78%	31.37%	31.60%
	(0.1406)	(0.2650)	(0.0788)	(0.2912)	(0.3135)	(0.2458)	(0.2328)
600	0.2608	0.3220	8.00%	32.08%	32.19%	24.02%	27.17%
	(0.0913)	(0.2085)	(0.0537)	(0.2369)	(0.2447)	(0.1839)	(0.2022)
800	0.2257	0.2836	7.41%	28.39%	29.37%	23.32%	24.63%
	(0.0747)	(0.1812)	(0.0563)	(0.2228)	(0.2215)	(0.1785)	(0.1819)
1000	0.2008	0.2532	6.07%	26.21%	26.51%	20.27%	20.55%
	(0.0601)	(0.1680)	(0.0278)	(0.2008)	(0.2047)	(0.1551)	(0.1544)

MSE: in sample prediction Mean Squared Error of deterministic production frontier. NISE: Normalized Integrated Squared Error. Sigma.v.e: mean percentage absolute error (MPAE)of $\sigma_{v,i}$ recovery. Delta.i.e: MPAE of δ_i Standard deviations are in parenthesis RSCFG stands for Reifschneider, Stevenson, Caudill, Ford, and Gropper.

Table 12: Summary Statistic of Airline Dataset-B

	Out	Mat.	Fue.	Eqp.	Lab.	Pro.	Loa.	Sta.	Poi.
Mean	0.629	0.752	0.584	0.652	0.595	0.656	0.548	507.9	72.9
StDev	0.592	0.643	0.504	0.568	0.508	0.693	0.056	290.7	26.0
Min	0.023	0.064	0.046	0.050	0.063	0.015	0.378	120.5	30.0
Med.	0.400	0.473	0.361	0.360	0.359	0.373	0.551	507.9	67.0
Max	2.442	2.448	1.770	2.105	1.692	2.807	0.676	1620	168
Skew.	0.79	0.65	0.60	0.72	0.68	1.13	-0.32	0.84	1.12
Kurt.	-0.53	-1.04	-1.09	-0.79	-1.09	0.31	-0.15	0.93	1.29

StDev:Standard Deviation, Med: Median, Skew.: Skewness, Kurt.: Kurtosis
Mat: Material, Fue.: Fuel, Eqp: Equipment, Lab.: Labor, Pro.:Property.
Out: Output, Loa.:Load Factor, Sta.: Average Stage Length, Poi. Number of Points Served

The estimator we used for GCD-SS model is:

$$\begin{aligned} y_i =&\beta_0 m(c_{Mat}(Mat_i))^{\beta_{Mat}} m(c_{Fue}(Fue_i))^{\beta_{Fue}} \\ &m(c_{Eqp}(Eqp_i))^{\beta_{Eqp}} m(c_{Lab}(Lab_i))^{\beta_{Lab}} \\ &m(c_{Pro.}(Pro._i))^{\beta_{Pro.}} exp(u_i - v_i) \end{aligned} \tag{49}$$

in which

$$\begin{aligned} u_i \sim& N(0, \sigma_u) \\ v_i \sim& exp(\delta_{Loa} Loa_i + \delta_{Sta} Sta_i + \delta_{Poi} Poi_i) N(\mu_v^*, \sigma_v^*)_+ \\ c_i(x_i) =& c_{i,1}(x_i) + c_{i,2}(x_i)^2 + c_{i,3}(x_i)^3 \end{aligned}$$

The estimator we used for GCD-RSCFG model is:

$$\begin{aligned} y_i = &\beta_0 m(c_{Mat}(Mat_i))^{\beta_{Mat}} m(c_{Fue}(Fue_i))^{\beta_{Fue}} \\ & m(c_{Eqp}(Eqp_i))^{\beta_{Eqp}} m(c_{Lab}(Lab_i))^{\beta_{Lab}} \\ & m(c_{Pro.}(Pro._i))^{\beta_{Pro.}} \exp(u_i - v_i) \end{aligned} \tag{50}$$

in which

$$\begin{aligned} u_i \sim &\text{Normal}(0, \sigma_u) \\ v_i \sim &\exp(\delta_{Loa} Loa_i + \delta_{Sta} Sta_i + \delta_{Poi} Poi_i) N(0, \sigma_v^*)_+ \\ c_i(x_i) &= c_{i,1}(x_i) + c_{i,2}(x_i)^2 + c_{i,3}(x_i)^3 \end{aligned}$$

The result of both model is Table 13. In SS model all three characteristic variables are negatively correlated with inefficiency. The marginal effect of load factor is 3.07. 1% increase of load factor efficiency will increase by more than 3%. Logically it may seem problematic. Load factor can go up to 100%. Even observation with the highest load factor can have more than 100% efficiency increase when load factor reaches 100%. Our explanation is although logically load factor can reach 100%, with the technology we studied it cannot get that high. If load factor increases that much we consider it a technology shift rather than a improvement of efficiency. Average stage length and number of points served also have positive impact on efficiency. Their marginal effect is much smaller than that of load factor. This outcome is what we expected.

One may noticed a by product of GCD-SS model. All elasticities of inputs are positive and significantly different from zero now. In all models we and Greene estimated there is always a elasticity is negative. We suspect previous models has misspecification problem. Because input elasticity estimation is quite different from previous models, we show Elasticities of

Table 13: Vital Parameter Estimation

	SS				RSCFG			
	Mean	StDev	95%LCI	95%UCI	Mean	StDev	95%LCI	95%UCI
sig.v	0.0034	0.0002	0.0030	0.0039	0.1060	0.0250	0.0531	0.1563
Lam.	0.0284	0.0020	0.0248	0.0324	0.8878	0.2562	0.3914	1.4371
Mu	0.2418	0.0167	0.2108	0.2756	-	-	-	-
ATE	0.7854	0.0131	0.7591	0.8100	0.9210	0.0176	0.8865	0.9590
Loa.ˆ	3.0713	0.0654	2.9372	3.1722	1.7430	0.8698	0.0309	3.6470
Sta.ˆ	0.0027	0.0001	0.0025	0.0031	0.0012	0.0017	0.0003	0.0047
Poi.ˆ	0.0022	0.0003	0.0014	0.0026	0.0002	0.0017	-0.0031	0.0029
RTS	1.3182	0.0288	1.2663	1.3785	1.1990	0.0470	1.1192	1.2945
Mat.*	0.7150	0.0327	0.6137	0.7464	0.6725	0.0931	0.4904	0.8508
Fue.*	0.0662	0.0080	0.0504	0.0818	0.3634	0.0949	0.1960	0.5568
Eqp.*	0.3921	0.0283	0.3517	0.4751	0.3599	0.0714	0.2112	0.4945
Lab.*	0.0388	0.0338	0.0188	0.1284	-0.391	0.0699	-0.522	-0.234
Pro.*	0.1061	0.0135	0.0987	0.1469	0.1949	0.0335	0.1306	0.2681

sig.v: Standard Deviation of Truncated Normal Distribution
Lam.:Lambda
Mu: Mean of Truncated Normal Distribution
ATE: Average Technical Efficiency
RTS: Average Return to Scale
* are average input elasticity for given input
ˆ are average marginal effect of observable characteristics on inefficiency
Mat: Material, Fue.: Fuel, Eqp: Equipment, Lab.: Labor, Pro.:Property.
Loa.:Load Factor, Sta.: Average Stage Length, Poi. Number of Points Served
StDev: Standard Deviation, LCI: Lower Bound of Confidence Interval, UCI: Upper Bound of Confidence Interval

Substitution Matrix in Table 14. we can see material is by far the most dominant input in frontier production function. Impact of 1% change in material can only be offset by 19.76% of labor.

Table 14: Elasticities of Substitution, Generalized Cobb-Douglas-Scaled Stevenson (GCD-SS) Model

	Mtl.	Fue.	Ept.	Lab.	Pro.
Mtl.	1	11.050	1.836	19.760	6.898
	(0)	(3.052)	(0.203)	(17.392)	(1.809)
Fue.	0.093	1	0.170	1.860	0.635
	(0.014)	(0)	(0.028)	(1.148)	(0.140)
Eqp.	0.551	6.098	1	10.579	3.791
	(0.065)	(2.377)	(0)	(11.158)	(1.220)
Lab.	0.056	0.588	0.102	1	0.387
	(0.058)	(0.508)	(0.110)	(0)	(0.421)
Pro.	0.149	1.637	0.272	2.918	1
	(0.025)	(0.498)	(0.044)	(3.225)	(0)

Mtl: Material, Fue.: Fuel, Eqp: Equipment, Lab.: Labor, Pro.:Property.

3.7 Summary

In this section our purpose has been apply suggested models on inefficiency analysis. We use same three step method as what we used in first section to estimate the model. Our simulation shows suggested model does not only greatly improve deterministic frontier estimation accuracy but also significantly improves estimation accuracy of coefficients of variables for inefficiency analysis in both RSCFG and SS models. Then we applied the model on airline dataset.

4. A REGRESSION SPLINE STRUCTURAL NONPARAMETRIC STOCHASTIC PRODUCTION FRONTIER AND INEFFICIENCY ANALYSIS MODELS

4.1 Introduction

ALS (1977) developed first practical stochastic frontier model. Since then Stochastic Frontier Estimation has become a important field in microeconometrics. Four maximum likelihood frameworks have been developed by econometricians. They are Half Normal-Normal, Exponential-Normal, Truncated Normal-Normal and Gamma-Normal.

Within Stevenson's (1980) Half Normal-Normal framework inefficiency term is very flexible. It also nested Half Normal-Normal and Exponential-Normal framework. Gamma-Normal framework is also very flexible and nested Exponential-Normal framework. Yet its log likelihood function does not have close form, and requires numerical integration. Therefore it is computationally expensive and only a few researchers adopted it.

Researchers also developed different fixed effect and random effect model to utilize advantage of panel datasets. Pitt and Lee (1981) developed two random effect models, and Schmidt and Sickles (1984) suggested a fixed effect within estimator.

Functional form of deterministic frontier is another story. Since ALS (1977) researchers are pretty much stuck with Cobb-Douglas and trans log model. Microeconomic theory requires production function to be monotone increasing and concave. If either of these two properties is missing profit maximization procedure and most of equilibrium models will breakdown. Without monotone increase property, even cost minimization problem might not have a solution.

As long as all exponents are between 0 and 1, good old Cobb-Douglas production function preserves both properties implicitly. Many econometricians incorporated this functional form in their researches. However Cobb-Douglas production function also implies properties researchers do not desire. For instance it assumes unitary elasticity of factor substitution and partial and total production elasticities that do not change with input.

Trans-log can be considered as a generalized form of Cobb-Douglas function. It is a very flexible functional form, and do not bear undesire assumptions that Cobb-Douglas function has. It becomes by far the most popular model in stochastic production frontier estimation. Yet the flexibility does not come without cost. it is also very hard to impose monotone and concave restrictions on Trans-log model.

In section I we have introduced a class of flexible stochastic production frontier models with monotonicity and concavity embedded. In previous sections we used third-order polynomial function. In this section we will generalize the model with regression spline and make the model data driven.

The paper will be organized as follow. In Section 4.2 we will do a brief literature review of stochastic production frontier estimation. In Section 4.3, a structure nonparametric model and Leave One out Cross Validation method to find optimal tuning parameter will be presented. In Section 4.4 we will suggest three method to speed up Cross Validation process. In Section 4.5 we will demonstrate performance improvement of adopting Penalized Likelihood by a grid setup simulation. In Section 4.6 we will apply suggested model on the airline dataset. In Section 4.7 we will summarize our research and discuss future research possibilities.

4.2 Literature Review

Farrell (1957) first explored the possibility to estimate the frontier production function. Aigner and Chu (1968), Afriat (1972) and Richmond (1974) developed a set of techniques estimating deterministic frontier production using linear or quadratic programming technology. Their methods minimize:

$$\sum_{i=1}^{n} |y_i - f(x_i)|$$

or

$$\sum_{i=1}^{n} (y_i - f(x_i))^2$$

subject to

$$y_i \leq f(x_i)$$

Deterministic efficiency frontier method fails to incorporate random elements that affect output. These elements such as natural disaster, worker strike or finding new natural recourses commonly exit in real world. The first attempt to incorporate these random elements was done by Aigner et al. (1976). They suggested a discontinuous distribution for the error term. ALS (1977) constructed the first practical Stochastic Efficiency Frontier estimator.

ALS (1977) assume $-v_i$ follows a Half normal distribution with 0 mean and ${\sigma_v}^2$ variance.

Econometricians still widely use this distribution. The log likelihood function is:

$$LL_i = \ln \frac{2}{\sigma^2} + \ln(\phi(\frac{\epsilon_i}{\sigma^2})) + \ln(1 - \Phi(\epsilon_i \lambda \sigma^{-1})) \tag{51}$$

Where $\lambda = \sigma_v/\sigma_u, \sigma = \sigma_u^2 + \sigma_v^2$. $\phi()$ is the density of standard normal distribution.$\Phi()$ is the cumulated distribution function of standard normal distribution.

ALS (1977) also briefly considered exponential distribution for$-v_i$.The log likelihood function under this assumption is:

$$LL_i = \ln(\lambda) + \ln(1 - \Phi(\frac{\epsilon_i}{\sigma_u} - \sigma_u \lambda)) + \epsilon_i \lambda + \frac{1}{2}(\sigma_u \lambda)^2 \tag{52}$$

in which $\frac{1}{\lambda} = \mathrm{E}(-v)$, $\frac{1}{\lambda^2} = \mathrm{Var}(v)$

Both half normal and exponential distribution bear implicit assumption that inefficiency is more concentrated near zero than further away. Stevenson (1980) suggested a more general normal-truncated normal distribution. Rather than assuming normal distribution generating $-v_i$ centers at zero, truncated allow mean to be any value on real line. The log likelihood of normal-truncated normal is:

$$LL_i = -\ln \sigma - \frac{1}{2} \ln 2\pi - \ln \Phi(\frac{\mu \lambda \sigma}{\sqrt{1 + \lambda^2}}) + \ln \Phi(\frac{\mu}{\sigma \lambda} - \frac{\epsilon_i \lambda}{\sigma}) - \frac{1}{2}(\frac{\epsilon_i + \mu}{\sigma})^2 \tag{53}$$

Stevenson also instructed a Normal-Gamma distribution. In his specification shape parameter P in Gamma distribution can only take integer value. Greene (1990) generalized Gamma distribution framework, by allowing P to take any positive real number. Log likelihood function in his framework is:

$$
\begin{aligned}
LL_i = &P \ln(\Theta) - \ln(\Gamma(P)) + {\sigma_u}^2\Theta^2/2 + \Theta\epsilon_i \\
&+ \ln(\Phi(-(\epsilon_i + \Theta\sigma_u^2)/\sigma_u)) + \ln(h(P-1, \epsilon_i))
\end{aligned} \tag{54}
$$

in which

$$
h(r, \epsilon_i) = \mathrm{E}[Q^r | Q > 0, \epsilon_i] \quad Q \sim \mathrm{Normal}(-(\epsilon_i + \Theta\sigma_u^2), \sigma_u^2)
$$

P is the shape parameter in Gamma distribution.Θ is the rate of Gamma distribution. Larger the P, further away concentration of firm-specific inefficiency is located from zero.

Pitt and Lee (1981) first explored the possibility of utilizing penal property of dataset using two random effect model. Both their models are based on Half Normal-Normal framework. First model assumes inefficiency is time invariant. Second model assumes inefficiency is correlated within each firm. They derived close form log likelihood function for first model. Unfortunately log likelihood function of second model does not have close form and requires T-dimensional numerical integration, which is computationally impossible at the time. They suggested a Constrained Seemingly Unrelated Regression approach to estimate second model.

Schmidt and Sickles (1984) suggested a group of fixed effect within estimators. A distinct advantage of their model is one can identify technical inefficiency for each firm. Their within estimator was the most popular fixed effect model. Greene (2005) reintroduced Pitt and Lee's (1981) second model as true random effect model. He also demonstrated the possibility of estimate the model with multidimensional numerical integration. He also generalized Schmidt and Sickle's (1984) within estimator to a true fixed effect model.

Spline smoothing is one of the most powerful techniques in nonparametric regression. (Eubank 1999 ; Wahba 1990 ; Green and Silverman 1994 ; Gu 2002) In early studies focuses on direct apply spline smoothing to the data with Ridge-Regression.

$$y_i = f(x_i) + \epsilon_i \tag{55}$$

Wahba 1990 first considered transform the nonparametric model with a bounded linear function.

$$y_i = \mathcal{L}_i(f(x_i)) + \epsilon_i \tag{56}$$

Ke and Wang (2004) generalized the model to allow $\mathcal{L}_i$ to be arbitrary function. They call them general smoothing spline nonlinear nonparametric regression models (SSNNRMs).

Researchers interested in estimating monotone function since development of isotonic regression by Bartholomew (1959) and Kruskal (1965) and Box-Cox transformation by Box and Cox (1964). Various monotone functions were developed. In the vast literature of estimating monotone function, monotone spline received a lot of attentions in recent years. Passow (1974) and Passow and Roulier (1977) discussed monotone interpolation splines. Wright and Wegman (1980) developed an approach to fitting monotone smoothing splines. Ramsay (1988) imposed monotonicity by restricting all coefficient of spline estimator to be positive. Kelly and Rice (1990) described a method to impose monotone constraint on B-Splines. Fritsch (1988) proposes an algorithm for calculating a monotone cubic spline. Li al et. (1996) developed a 6 step smoothing procedure for piecewise convex/concave curves. Turlach (1997) discussed general model that imposes different shape constraint on cubic spline by restricting coefficients.

While previous work mainly focused on restricting coefficients, Ramsay (1998) come about the ingenious idea of using a link function with integration to blockade the Chain Rule. His model is:

$$\hat{y} = \beta_0 + \beta_1 \int_0^x \exp(\int_0^a c(b)\mathrm{d}b)\mathrm{d}a \tag{57}$$

In his setup $c(x)$ is a polynomial or spline function. Sickles and Wu (Personal Discussion) improved the model to impose not only monotonicity but also concavity constrains. They also extended the model to multivariate case. From now on we will call it Additive model. The improved model is:

$$f(x) = \beta_0 + \sum_{j=1}^{p} \beta_j m(c_j(x_j)) \tag{58}$$

in which

$$m(c(x)) = \int_0^x \exp(-\int_0^a \exp(c(b))\ \mathrm{d}b)\ \mathrm{d}a$$

x_j is a vector of jth independent variables. p is number of independent variables in the dataset. $c_j()$ is some arbitrary function.

4.3 A Regression Spline Structural Nonparametric Model

In this section we will introduce a nonparametric model embeds monotonicity and concavity constraint required by microeconomic theory. In the first section we generalized Additive model into Generalized Cobb-Douglas (GCD) model to remove separability assumption.

$$y = \beta_0 \prod_{j=1}^{p} m(c_j(x_j))^{\beta_j} \tag{59}$$

Again $c_j(x)$ can be arbitrary function of x. With polynomial specification of $c(\cdot)$, both models are very flexible. As we showed in Section 2 and 3 they strictly dominate Cobb-Douglas model, which imposes same constraint. Yet because by construction $c(\cdot)$ can be arbitrary function, we can replace polynomial $c(\cdot)$ with regression spline, and make the model data-driven. Specification of regression spline $c(\cdot)$ is:

$$c(x) = \sum_{j=1}^{k} c_j x^j + \sum_{i=1}^{L} c_{n,i}(x - x_{n,i})_+^k \tag{60}$$

in which

$$(x - x_{n,i})_+ = max(x - x_{n,i}, 0)$$

In equation above k is exponent of spline. In this paper we will use $k = 3$. L is number of nodes in the regression spline. $x_{n,i}$ is ith nodes in the spline. $c_{n,i}$ is coefficient of ith nodes. $(x - x_{n,i})_+$ is called base of ith nodes.

L can take value from 0 up to number of observations. When L equals 0, $c(\cdot)$ returns to third order polynomial. When L goes up to number of observations, $c(\cdot)$ become third order interpolation. $x_{n,i}$ can even space or even quantile spread cross sample space. In this paper we will use even quantile spread. Even quantile means number of observations between nodes will be equal.

With $c(\cdot)$ defined we still have one problem to implement the model. In theory log likelihood value will never get smaller when L increases. In reality log likelihood will probably always increase when we increase L. If we only maximize log likelihood function, we will get a interpolation model. When L increases from 0 to 1, $\hat{f}(x)$ will have better in sample prediction and may have better out of sample prediction. However when we keep increasing L, eventually we will have better and better in sample prediction and worse and worse out of sample prediction. This phenomenon is called over fitting in econometrics.

To prevent over fitting, AIC and BIC first falls into our sight naturally. To use AIC and BIC in model selection, we just try different Ls and use L with smallest AIC or BIC. It is very intuitive and simple. Yet we find AIC and BIC have severe over penalizing problem in our model. We ran 300 simulations to try L between 0 and 20. In all 300 replications models that have smallest out of sample MSE have L larger than 0. In only 27% of replications AIC chose L larger than 0, and in none replications AIC chooses optimal L. Therefore we advise strongly against using AIC or BIC for model selection in our case.

The 95% quantile of over penalizing factor of AIC is from 1.4 to 4. The 95% quantile of over penalizing factor of BIC is between 8 and 10.6. We define over penalizing factor as the number that number of parameters needs to divide before being used for given criterion.

Even with adjustment above we still advise against using AIC and BIC.

When AIC and BIC failed us, we have to consider other options. Penalized spline (P-Spline) or more specifically penalized likelihood is another approach. The idea penalized likelihood is we first over fit the model with a large L, usually 20-30 would be enough, then we penalize the roughness of estimated function. The general specification of penalized likelihood estimator is:

$$\hat{\theta} = \text{argmax}[\ln(f(x|\theta)) - \lambda_P P(\theta)] \tag{61}$$

$f(\cdot)$ is the likelihood function. λ_P is the tuning parameter. It can take non-negative real value. $P(\theta)$ is the roughness penalty function. Because our model no matter Additive or GCD models are quite different from linear regression spline we have to choose $P(\theta)$ very carefully. For instance, we first tried conventional cubic spline penalty term.

$$P(\theta) = \int_x \hat{g}''(a)^2 \mathrm{d}a \tag{62}$$

$\hat{g}(\cdot)$ is the estimated production frontier. The intuition of cubic spline penalty term is penalizing integrated squared second derivative i.e. visual roughness of estimated function. By construction our model should have negative second derivative, and it does not only depends on spline coefficients but also other parameters. Therefore when we implement the penalty term result is very bad.

Ruppert (2002) suggested a penalty function directly penalize the magnitude of spline coefficients.

$$P(\theta) = \sum_{j=1}^{v} \sum_{i=1}^{L} {c_{n,i,j}}^2 \tag{63}$$

In equation above v is number of independent variables. This penalty function is very intuitive. When λ_P goes to infinite $c(\cdot)$ becomes near polynomial and model converge to model we discussed in Section 2. When λ_P goes to zero $c(\cdot)$ becomes near interpolation. After we chose an appropriate roughness penalty function, we only need find optimal tuning parameter.

By purpose of penalized likelihood, we first considered Leave One out Cross Validation(LOOCV). The idea of LOOCV is very strait forward. In first iteration we leave first observations out as evaluation sample, and use other observations as training sample. We first train the model with training sample and then evaluate trained model with evaluation sample. We can get a log likelihood value from evaluation. process above is repeated such that each observation in the sample is used once as the validation data. Lastly we try different λ_P to maximize summed log likelihood from all evaluations.

When we implement LOOCV in simulation it yielded satisfactory result. Yet it has one problem. By construction LOOCV is very expensive from a computational point of view. In our case this problem gets even worse. In each evaluation of λ_P we need to numerically optimize a 50+ dimension maximization problem n times for a small 2-variable model. In our initial try it took 17 hours to reach a convergence.

Besides LOOCV we also considered Generalized Cross Validation (GCV), Generalized Maximum Likelihood (GML), Generalized Information Criterion (GIC),and parametric direct plug-in tuning parameter (PDPI) to choose optimal tuning parameters. To the best of our knowledge existing GCV and GML method is for least square estimator only. They require error terms to have zero mean (Wahba 1977), which is clear not the case in efficiency frontier analysis.

Ueki and Fueda (2010) developed PDPI method to choose optimal tuning parameter. In their paper they proved under certain regularity conditions PDPI, GIC and LOOCV will generate consistent estimation of optimal λ_P.

PDPI method relies on complicated analytic calculations. It involves third order derivative of parameters, which is requires too much analytic calculation and programming for our purpose. We also tried to approximate derivative with recursive Richardson algorithm. It terms out to be computationally more expensive than LOOCV. Initial try took more than 3 days. Therefore we decide to abandon PDPI method.

GIC is developed by Konishi and Kitagawa (1996). It removed two assumptions required by AIC i.e. "estimation is by maximum likelihood.", and "this is carried out in a parametric family of distributions including the true model." Violation of second assumption is probably the cause of over penalizing problem we met. Estimation of λ_P by GIC is:

$$\lambda_{P,GIC} = \mathrm{argmin}[-2\sum_{i=1}^{n}\ln(f(x_i|\hat{\theta}_\lambda)) + 2\mathrm{tr}[\hat{R}(\hat{\theta}_\lambda)^{-1}\hat{Q}_\lambda(\hat{\theta_\lambda})] \qquad (64)$$

in which

$$\hat{R}(\hat{\theta}_\lambda) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\{\ln(f(x_i|\theta)) - \frac{\lambda_P}{n}P(\theta)\}}{\partial\theta\partial\theta^T}$$
$$\hat{Q}_\lambda(\hat{\theta_\lambda}) = \frac{1}{n}\sum_{i=1}\frac{\partial\{\ln(f(x_i|\theta)) - \frac{\lambda_P}{n}P(\theta)\}}{\partial\theta}\frac{\partial\ln(f(x_i|\theta))}{\partial\theta^T} \tag{65}$$

$\hat{R}(\hat{\theta}_\lambda)$ is the Hessian of PLE estimator divided by $-n$. Quasi-Newton algorithm in optim() will automatically generate numerical approximation of Hessian. The first term in $\hat{Q}_\lambda(\hat{\theta_\lambda})$ is effectively the ith row of Jacobian matrix of $ln(f(x|\theta)) - \frac{\lambda_P}{n}P(\theta)$. The second term in $\hat{Q}_\lambda(\hat{\theta_\lambda})$ is the ith row of Jacobian matrix of $ln(f(x|\theta))$. Jacobian can be approximated using Evaluation of GIC is much faster than LOOCV. Initial try took 2.2 hours to reach convergence.

4.4 Estimation, Parameter Recovery, and Inference

We suggest same three step method used by polynomial model to estimate PLE model. First we estimate stochastic production frontier model with Cobb-Douglas production function. Then we can use its $\hat{\lambda}$ and $\hat{\sigma}$ as starting value. Cobb-Douglas specification will also generate a prediction of deterministic production frontier $\hat{PF}$. We use $\hat{PF}$ as pseudo true production frontier and estimate $\hat{PF}$ with suggested model using least square estimator. $\hat{\beta}$ and $\hat{c}$ generated by least square estimator will be used as starting value of β and c. In our research 0 is good starting value for c, and 1 is good starting value for β in the least square estimation. After appropriate starting values are acquired, we can estimate the model using quasi-Newton method(also known as a variable metric algorithm).

To make the model practical we also need reduce computational expenses. In simulation we implement three method to reduce computational expenses of LOOCV method. First method we use K-Fold Cross Validation (KFCV) to replace LOOCV. K-fold Cross Validation randomly assigns observations into K groups. Of the K groups, a single group is retained as the validation data to evaluate the model, and the remaining K-1 groups are used as training data. The Cross Validation process is then repeated K times, with each of the K groups used exactly once as the evaluation data. As a matter of fact LOOCV can be considered as a special case of KFCV, where K equals number of observations. Generally speaking this method may reduce computational expenses to n/K of original. According to McLachlan et al. (2004) we suggest 10-fold cross-validation for cross sectional dataset. For N×T Panel dataset we suggest N/k-fold CV. N/k should be positive integer. That is in each iteration leave k individuals out as validation sample. We suggest this, because observations within each individual are usually not independent.

In simulation, we found out the objective function of CV is very close to concave function with one maximum. Instead of quasi-Newton algorithm, we suggest a Hill Climbing algorithm to maximize summed log likelihood. With Hill Climbing algorithm, model usually converges within several evaluations. In simulation we also find out increase relative tolerance of quasi-Newton algorithm to 1×10^{-4} will greatly increase speed with negligible impact on final result. With these adjustment we reduce average time need for CV to around 2 minutes in 2-variable 400 observation case. With implementation of second and third method, average time need for GIC method reduced to around 9 minutes.

As β, λ, σ and c themselves do not have economic meanings, we need to derive point estimations and inferences of values that interested by economists. Applied researchers commonly interested in marginal productivities, input elasticities, return to scale, Elasticities of Substitution and average TE. The marginal productivity of GCD-PLE model is:

$$\frac{\partial y}{\partial x_k} = \beta_0 \times \beta_k \times m'(c_k(x_k)) \times m(c_k(x_k))^{\beta_k - 1} \times \prod_{j=1, j\neq k}^{p} m(c_j(x_j))^{\beta_j} \tag{66}$$

in which

$$m'(c(x)) = \exp(-\int_0^x \exp(c(b))\ \mathrm{d}b)$$

Input elasticities are derived from normalizing mean productivities with input-output ratios. Return to scale equals to summation of all input elasticities. Elasticity of substitution of ith input to jth input equals input elasticity of ith input divide by that of jth input. Average TE equals $E[exp(-v)|\hat{\theta}]$. $\hat{\theta}$ is a vector of estimated parameters of distribution that $-v$ was assumed to follow. Expression to recover average TE for half-normal distribution is:

$$TE = \int_{-\infty}^{\infty} \exp(-|x|)\phi_{0,\sigma}(x)\mathrm{d}x \tag{67}$$

In which, $\phi_{0,\sigma}$ is pdf of

$$\text{Normal}(0, \sqrt{\frac{\sigma^2\lambda^2}{1+\lambda^2}})$$

We suggest Bootstrapping approach to inference values above. In regression Bootstrapping we use $\hat{y}$ as true EF, and resample $\hat{\epsilon}$ with replacement. By the natural of Stochastic Efficiency Frontier model, our Bootstrapping setup is a little bit different:

4.5 Monte Carlo Simulations

In this section we will show the performance improvement of switching from Polynomial-GCD estimator to PLE-GCD estimator. We use equation below to simulate ys.

$$y_i = 20 \times \Phi(x_{i,1}, x_{i,2}) \exp(u_i + v_i) \tag{68}$$

in which

$$\begin{aligned}
&\Phi(\cdot) \text{ is CDF of Normal}(0, \Sigma) \\
&\Sigma = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 3 \end{bmatrix} \\
&u_i \sim \text{Normal}(0, \sigma_u) \\
&v_i \sim -|\text{Normal}(0, \sigma_v)|
\end{aligned}$$

$x_{i,1}$ is generated by $\Gamma(100, 10)$ distribution. $x_{i,2}$ is generated by $\Gamma(120, 20)$ distribution. Both variables were rescaled into $[1, 6]$ range. Each simulation contains 400 observations. In each simulation we also generated additional 400 "naked" observations to evaluate out of sample prediction accuracy of the model. Term "naked" means dependent variable of evaluation observations only contains deterministic frontier i.e. $20 \times \Phi(x_{i,1}, x_{i,2})$. Independent variables in evaluation sample are generated from identical procedure to that in simulation. Intuitively out of sample MSE is an empirical approximation of Normalized Integrated Squared Error (NISE).

$$NISE = \frac{1}{36}\int_1^6\int_1^6 (\hat{PF}(x_1,x_2) - PF(x_1,x_2))^2 \mathrm{d}F_1(x_1)\mathrm{d}F_2(x_2) \tag{69}$$

In equation above, $PF(\cdot)$ and $\hat{PF}(\cdot)$ are true and estimated deterministic frontier. $F_i(x_i)$ is CDF of x_i. We did simulations in three different average TE settings by three different λ settings grid. We compared deterministic production frontier recovery ability and distributional parameter recovery ability between PLE specification and polynomial specification. For each point in grid we did 500 simulations. We use third order polynomial for polynomial specification, and use $L = 20$ for PLE specification. The objective function of PLE specification is:

$$PLL = LL(x, c, c_n, \beta) - \lambda_P P(c_n) \tag{70}$$

in which

$$\begin{aligned}
&LL(x, c, c_n, \beta) = \sum_{i=1}^{n} \ln\frac{2}{\sigma^2} + \ln(\phi(\frac{\epsilon_i}{\sigma^2})) + \ln(1 - \Phi(\epsilon_i\lambda\sigma^{-1}))\\
&\epsilon_i = \ln(y_i) - \ln(\beta_0) - \sum_{j=1}^{2}\beta_j \ln[m(c_j(x_{i,j}))]\\
&m(c(x)) = \int_0^x \exp(-\int_0^a \exp(c(b))\ \mathrm{d}b)\ \mathrm{d}a\\
&c_j(x) = \sum_{i=1}^{3} c_{j,i}x^i + \sum_{i=1}^{20} c_{n,j,i}(x - x_{n,j,i})_+^k\\
&P(c_n) = \sum_{j=1}^{2}\sum_{i=1}^{20} {c_{n,i,j}}^2
\end{aligned}$$

The objective function of Polynomial specification is:

$$PLL = LL(x, c, c_n, \beta) \tag{71}$$

in which

$$LL(x, c, c_n, \beta) = \sum_{i=1}^{n} \ln \frac{2}{\sigma^2} + \ln(\phi(\frac{\epsilon_i}{\sigma^2})) + \ln(1 - \Phi(\epsilon_i \lambda \sigma^{-1}))$$

$$\epsilon_i = \ln(y_i) - \ln(\beta_0) - \sum_{j=1}^{2} \beta_j \ln[m(c_j(x_{i,j}))]$$

$$m(c(x)) = \int_0^x \exp(-\int_0^a \exp(c(b))\, \mathrm{d}b)\, \mathrm{d}a$$

$$c_j(x) = \sum_{i=1}^{3} c_{j,i} x^i$$

Table 15: Predictive Accuracy of Penalized Likelihood Estimator (PLE) and Polynomial Model on Technical Efficiency

	PLE			Polynomial		
Lambda	0.7	0.8	0.9	0.7	0.8	0.9
3	0.0126	0.0095	0.0052	0.0121	0.0095	0.0057
	(0.0103)	(0.0074)	(0.0040)	(0.0099)	(0.0075)	(0.0042)
5	0.0086	0.0070	0.0038	0.0093	0.0071	0.0042
	(0.0063)	(0.0057)	(0.0030)	(0.0070)	(0.0052)	(0.0030)
7	0.0083	0.0065	0.0034	0.0082	0.0066	0.0041
	(0.0063)	(0.0047)	(0.0024)	(0.0062)	(0.0049)	(0.0030)

0.7, 0.8, 0.9 are different TE settings
Numbers on the top are Mean Absolute Error of Technical Efficiency.
Numbers in parenthesis are standard deviation of Error of Technical Efficiency.

In Table 15 is the comparison of Average TE recovery accuracy. Both models recover Average TE precisely. Differences between two models are relatively small. When Average TE equals 0.7 results of two model are indistinguishable. When Average TE equals 0.8 accuracy improvement of adopting PLE is negligible. When Average TE gets 0.9 PLE specification has significant advantage over polynomial specification. This result is some what expected. Our primary goal has been improve deterministic frontier. When signal-noise ratio is low, most of predictive error was caused by variance of error terms. Improving deterministic frontier estimation will not help much. When error terms are less volatile gain of deterministic frontier estimator improvement become more significant.

Table 16: Predictive Accuracy of Penalized Likelihood Estimator (PLE) and Polynomial Model on Deterministic Production Frontier

	MPLE			Polynomial		
Lambda	0.7	0.8	0.9	0.7	0.8	0.9
3	0.6095	0.2444	0.0543	0.6263	0.3275	0.1894
	(0.4593)	(0.1842)	(0.0357)	(0.4599)	(0.1917)	(0.0461)
	0.6712	0.2857	0.0639	0.6837	0.3834	0.2340
	(0.5202)	(0.2552)	(0.0511)	(0.5006)	(0.2219)	(0.0961)
5	0.3660	0.1532	0.0355	0.3821	0.2341	0.1734
	(0.2324)	(0.1083)	(0.0228)	(0.2047)	(0.0907)	(0.0326)
	0.4208	0.1811	0.0436	0.4442	0.2780	0.2214
	(0.2961)	(0.1619)	(0.0358)	(0.2810)	(0.1387)	(0.0862)
7	0.2787	0.1164	0.0254	0.3205	0.2164	0.1709
	(0.1942)	(0.0827)	(0.0148)	(0.1422)	(0.0654)	(0.0332)
	0.3420	0.1347	0.0317	0.3783	0.2664	0.2165
	(0.2917)	(0.1080)	(0.0237)	(0.2120)	(0.1340)	(0.0715)

0.7, 0.8, 0.9 are different TE settings
Numbers without parenthesis on the top are average MSE of in sample prediction.
Numbers without parenthesis at the bottom are Integrated Squared Error.
Numbers in parenthesis are standard divinations of value above them.

In Table 16, is comparison of predictive accuracy of PLE and polynomial specification. Similar to Average TE recovery accuracy, improvement of adopting PLE increases when variance of error terms decreases. When Average TE equals 0.7 PLE specification is only marginally superior to polynomial specification. When Average TE equals to 0.8 adopting PLE can reduce Integrated Squared Error close to 50%. When Average TE equals 0.9 adopting PLE can reduce Integrated Squared Error up to 85%.

Table 17: Predictive Accuracy of Penalized Likelihood Estimator (PLE) Model and Polynomial Model on Technical Efficiency and Deterministic Frontier, under Different Sample Size

Observations	PLE		
	MSE	NISE	TE.e
400	0.6095 (0.4593)	0.6712 (0.5202)	1.26% (0.0103)
800	0.3206 (0.2424)	0.3606 (0.2963)	0.89% (0.0069)
1200	0.1994 (0.1499)	0.2275 (0.1734)	0.66% (0.0053)
	Polynomial		
400	0.6263 (0.4599)	0.6837 (0.5006)	1.21% (0.0099)
800	0.3510 (0.2203)	0.3923 (0.2638)	0.88% (0.0068)
1200	0.2617 (0.1419)	0.3041 (0.1426)	0.78% (0.0058)

*Standard Deviations are in Parenthesis.
MSE: In Sample Prediction MSE
NISE: Normalized Integrated Squared Error
TE.e: Mean Absolute Error of Average Technical Efficiency Recovery

In second simulation we fixed error term setting at $\lambda = 3$ and $TE = 0.7$, and compared result with different sample size. In Table 17 we can see PLE model converges to true deterministic frontier much faster than Polynomial model. When sample size equals 400 PLE is only marginally superior to Polynomial model. When sample size reaches 1200 MSE reduction from adopting PLE model goes up to almost 30%.

Average Technical Efficiency estimation of PLE model also converges to true Average Technical Efficiency faster than Polynomial model. When sample size equals 400 MAE of TE of PLE model is 0.05% larger than Polynomial model. When sample size grows to 800 the difference reduces to 0.01%. When sample size reaches 1200 MAE of TE of PLE model is 0.12% smaller than Polynomial model.

Table 18: Comparisons of Tuning Parameter Choosing Method

	GIC			KFCV		
Lambda/TE	NISE	MSE	TE.e	NISE	MSE	TE.e
5/0.8	0.1777 (0.1487)	0.1519 (0.1052)	0.0067 (0.0052)	0.1811 (0.1619)	0.1532 (0.1083)	0.0070 (0.0057)
7/0.9	0.0378 (0.0374)	0.0258 (0.0167)	0.0034 (0.0026)	0.0317 (0.0237)	0.0254 (0.0148)	0.0034 (0.0024)

*Standard Deviations are in Parenthesis.
GIC: Generalized Information Criterion.
KFCV: K-Fold Cross Validation
MSE: In Sample Prediction MSE
NISE: Integrated Squared Error
TE.e: Mean Absolute Error of Average Technical Efficiency Recovery
TE: Technical Efficiency

In the last simulation we compared the performance of two method for tuning parameter selection under two error term settings. In Table 18 we can see the results of two method are quite similar. Therefore researchers can choose either or both of them in applied work.

4.6 An Empirical Application

In this section we will implement the GCD-SS and GCD-RSCFG models on an airline dataset. Data is downloaded from Greene's website . This data set is a extension of Caves et al.(1984). It is an unbalanced panel with 256 observations on 25 firms. Input variables are materials, fuel, equipment labor and property. The dependent variable is a index of airline out put.

Table 19: Summary Statistic of Airline Dataset-C

	Out	Mat.	Fue.	Eqp.	Lab.	Pro.	Loa.	Sta.	Poi.
Mean	0.629	0.752	0.584	0.652	0.595	0.656	0.548	507.9	72.9
StDev	0.592	0.643	0.504	0.568	0.508	0.693	0.056	290.7	26.0
Min	0.023	0.064	0.046	0.050	0.063	0.015	0.378	120.5	30.0
Med.	0.400	0.473	0.361	0.360	0.359	0.373	0.551	507.9	67.0
Max	2.442	2.448	1.770	2.105	1.692	2.807	0.676	1620	168
Skew.	0.79	0.65	0.60	0.72	0.68	1.13	-0.32	0.84	1.12
Kurt.	-0.53	-1.04	-1.09	-0.79	-1.09	0.31	-0.15	0.93	1.29

StDev:Standard Deviation, Med: Median, Skew.: Skewness, Kurt.: Kurtosis
Out.: Output, Mat: Material, Fue.: Fuel, Eqp: Equipment, Lab.: Labor, Pro.:Property.
Loa.:Load Factor, Sta.: Average Stage Length, Poi. Number of Points Served

This dataset also contains three characteristic variables that might affect efficiency. They are load factor (Loa.), average stage length (Sta.) and number of points served (Poi.). Load factor is percentage of capacity used for revenue generating activities. For instance percentage of ticket sold. We expected it to be strongly negative correlated with inefficiency. Average stage length is average length of each flight. The unit is mile. Points Served is number of airport each airline serves. Due to our limited understanding of airline industry, we do not have expectation on coefficient of these two variables. Summary statistics of the dataset are in Table 19.

First we will apply pooled GCD model with Half Normal-Normal and Truncated Normal-Normal specification. The objective function for PLE is:

$$PLL = LL(x, c, c_n, \beta) - \lambda_P P(c_n) \tag{72}$$

in which, For Half Normal-Normal Specification

$$LL(x, c, c_n, \beta) = \sum_{i=1}^{n} \ln \frac{2}{\sigma^2} + \ln(\phi(\frac{\epsilon_i}{\sigma^2})) + \ln(1 - \Phi(\epsilon_i \lambda \sigma^{-1}))$$

For Truncated Normal-Normal Specification

$$\begin{aligned} LL(x, c, c_n, \beta) = & \sum_{i=1}^{n} -\ln\sigma - \frac{1}{2}\ln 2\pi - \ln\Phi(\frac{\mu\lambda\sigma}{\sqrt{1+\lambda^2}}) + \\ & \ln\Phi(\frac{\mu}{\sigma\lambda} - \frac{\epsilon_i\lambda}{\sigma}) - \frac{1}{2}(\frac{\epsilon_i + \mu}{\sigma})^2 \end{aligned} \tag{73}$$

For Both Specification

$$\begin{aligned} \epsilon_i &= \ln(y_i) - \ln(\beta_0) - \sum_{j=1}^{5} \beta_j \ln[m(c_j(x_{i,j}))] \\ m(c(x)) &= \int_0^x \exp(-\int_0^a \exp(c(b))\,\mathrm{d}b)\,\mathrm{d}a \\ c_j(x) &= \sum_{i=1}^{3} c_{j,i} x^i + \sum_{i=1}^{20} c_{n,j,i}(x - x_{n,j,i})_+^k \\ P(c_n) &= \sum_{j=1}^{5} \sum_{i=1}^{20} {c_{n,i,j}}^2 \end{aligned}$$

Because the dataset is an unbalanced panel, 10-Fold CV does not work well for us. We use leave one individual out CV in this application. That is we leave observations from one firm out as validation sample in each iteration. For each specification we did 399 bootstrapping for inference.

Table 20: Vital Parameter Estimation of Generalized Cobb-Douglas Penalized Likelihood Estimator (GCD-PLE) Model under Pooled Specification

	Half-Normal				Truncated Normal			
	Mean	StDev	95%L	95%U	Mean	StDev	95%L	95%U
TE	0.907	0.044	0.851	1.000	0.999	0.006	1.000	1.000
RTS	1.289	0.073	1.123	1.409	1.254	0.041	1.170	1.332
Mat.*	0.771	0.086	0.591	0.936	0.754	0.072	0.607	0.891
Fue.*	0.272	0.084	0.122	0.442	0.292	0.062	0.166	0.425
Equ.*	0.391	0.077	0.248	0.539	0.342	0.064	0.201	0.475
Lab.*	-0.279	0.086	-0.448	-0.092	-0.278	0.069	-0.445	-0.136
Pro.*	0.133	0.024	0.085	0.184	0.144	0.023	0.104	0.202

TE: Average Technical Efficiency
RTS: Average Return to Scale
* are average input elasticity for given input
Mat: Material, Fue.: Fuel, Eqp: Equipment, Lab.: Labor, Pro.:Property.
StDev: Standard Deviation, L: Lower Bound, U: Upper Bound

In Table 20 are the vital parameter estimation of GCD-PLE model under Half Normal-Normal and Truncated Normal-Normal specification. Estimation differences of return to scale and input elasticities between two models are minor. Average TE estimations between two models are very different. Under Half Normal-Normal specification suggested model only have 5.25% bootstrapping replications with Average TE equals one. Under Truncated Normal-Normal specification suggested model found 99% bootstrapping replications have Average TE equals one. The parameter estimations are also close to that of GCD-Polynomial model in the first section.

As the dataset is unbalanced panel we also applied random effect model by Pitt and Lee (1981) and fixed effect model by Schmidt and Sickles (1984). The objective function of random effect model is quite similar to that of pooled model.

$$PLL = LL(x, c, c_n, \beta) - \lambda_P P(c_n) \tag{74}$$

in which

$$\begin{aligned}
LL(x, c, c_n, \beta) &= \sum_{i=1}^{N} \ln 2 - \frac{T_i}{2} \ln(2\pi) - \frac{T_i - 1}{2} \ln {\sigma_v}^2 - \frac{1}{2} \ln({\sigma_v}^2 + T_i {\sigma_u}^2) \\
&- \frac{T_i}{2\sigma_v^2} E(\epsilon_i^2) + \frac{T_i^2 \sigma_u^2}{2\sigma_v^2(\sigma_v^2 + T_i\sigma_u^2)} E(\epsilon_i)^2 + \ln[1 - \Phi(\frac{T_i \sigma_u}{\sigma_v \sqrt{\sigma_v^2 + T_i \sigma_u^2}} E(\epsilon_i))] \\
E(\epsilon_i) &= \frac{1}{T_i} \sum_{t=1}^{T_i} [\ln(y_{i,t}) - \ln(\beta_0) - \sum_{j=1}^{5} \beta_j \ln(m(c_j(x_{i,j,t})))]
\end{aligned} \tag{75}$$

$$\begin{aligned}
E(\epsilon_i^2) &= \frac{1}{T_i} \sum_{t=1}^{T_i} [\ln(y_{i,t}) - \ln(\beta_0) - \sum_{j=1}^{5} \beta_j \ln(m(c_j(x_{i,j,t})))]^2 \\
m(c(x)) &= \int_0^x \exp(-\int_0^a \exp(c(b))\, \mathrm{d}b)\, \mathrm{d}a \\
c_j(x) &= \sum_{i=1}^{3} c_{j,i} x^i + \sum_{i=1}^{20} c_{n,j,i} (x - x_{n,j,i})_+^k \\
P(c_n) &= \sum_{j=1}^{5} \sum_{i=1}^{20} {c_{n,i,j}}^2
\end{aligned}$$

Different from all other models fixed effect model uses Penalized Least Square Estimator.

$$PLS = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \epsilon_{i,t}^2 - \lambda_P P(c_n) \tag{76}$$

in which

$$\epsilon_{i,t} = \ln(y_{i,t}) - \ln(\beta_{0,i}) - \sum_{j=1}^{5} \beta_j \ln(m(c_j(x_{i,j,t})))$$

$$m(c(x)) = \int_0^x \exp(-\int_0^a \exp(c(b))\,\mathrm{d}b)\,\mathrm{d}a$$

$$c_j(x) = \sum_{i=1}^{3} c_{j,i}x^i + \sum_{i=1}^{20} c_{n,j,i}(x - x_{n,j,i})_+^k$$

$$P(c_n) = \sum_{j=1}^{5} \sum_{i=1}^{20} {c_{n,i,j}}^2$$

Table 21: Vital Parameter Estimation of Cobb-Douglas Penalized Likelihood Estimator (GCD-PLE) Model under Fixed Effect and Random Effect Specification

	Random Effect				Fixed Effect			
	Mean	StDev	95%L	95%U	Mean	StDev	95%L	95%U
TE	0.932	0.007	0.917	0.945	0.730	0.029	0.661	0.778
RTS	1.492	0.087	1.331	1.697	1.269	0.041	1.186	1.343
Mat.*	1.051	0.096	0.863	1.237	0.939	0.069	0.798	1.068
Fue.*	0.307	0.084	0.151	0.477	0.040	0.086	-0.115	0.206
Equ.*	0.334	0.100	0.145	0.536	0.319	0.082	0.174	0.492
Lab.*	-0.270	0.105	-0.471	-0.005	-0.097	0.101	-0.279	0.106
Pro.*	0.071	0.028	0.009	0.133	0.068	0.009	0.050	0.085

TE: Average Technical Efficiency
RTS: Average Return to Scale
* are average input elasticity for given input
Mat: Material, Fue.: Fuel, Eqp: Equipment, Lab.: Labor, Pro.:Property.
StDev: Standard Deviation, L: Lower Bound, U: Upper Bound

Fixed effect model assumes firm with highest Technical Efficiency (i.e. $\beta_{0,i}$) is right at the frontier. Efficiency of all other firm is $\beta_{0,i}/\beta^*_{0,i}$.

Table 22: Vital Parameter Estimation of Scaled Stevenson (SS) and RSCFG Inefficiency Analysis Models

	SS				RSCFG			
	Mean	StDev	95% LCI	95% UCI	Mean	StDev	95% LCI	95% UCI
sig.v	0.0054	0.0005	0.0037	0.0061	0.1280	0.0199	0.0870	0.1678
Lam.	0.0457	0.0034	0.0369	0.0513	1.1113	0.2256	0.7284	1.6258
Mu	0.3873	0.0309	0.2954	0.4354	-	-	-	-
TE	0.6774	0.0401	0.6469	0.7411	0.9057	0.0136	0.8791	0.9342
Loa.ˆ	2.5295	0.0829	2.3994	2.6993	1.9175	0.3299	1.3153	2.7108
Sta.ˆ	0.0014	0.0002	0.0013	0.0020	0.0008	0.0003	0.0004	0.0014
Poi.ˆ	0.0023	0.0004	0.0020	0.0035	-0.0007	0.0009	-0.0027	0.0014
RTSˆ	1.5380	0.0610	1.3597	1.6289	1.1929	0.0386	1.1081	1.2591
Mat.*	0.7467	0.0680	0.5100	0.8014	0.7698	0.0733	0.5965	0.9206
Fue.*	0.2375	0.0545	0.2102	0.4391	0.1240	0.0744	-0.0175	0.2857
Eqp.*	0.3643	0.0359	0.3345	0.4687	0.3371	0.0626	0.1999	0.4703
Lab.*	0.0930	0.0601	-0.1322	0.1148	-0.1712	0.0556	-0.2983	-0.0561
Pro.	0.0966	0.0125	0.0795	0.1327	0.1333	0.0193	0.0988	0.1772

sig.v: Standard Deviation of Truncated Normal Distribution
Lam.:Lambda
Mu: Mean of Truncated Normal Distribution
TE: Average Technical Efficiency
RTS: Average Return to Scale
* are average input elasticity for given input
ˆ are average marginal effect of observable characteristics on inefficiency
RSCFG stands for Reifschneider, Stevenson, Caudill, Ford, and Gropper.
Mat: Material, Fue.: Fuel, Eqp: Equipment, Lab.: Labor, Pro.:Property.
Loa.:Load Factor, Sta.: Average Stage Length, Poi. Number of Points Served
StDev: Standard Deviation, LCI: Lower Bound of Confidence Interval, UCI: Upper Bound of Confidence Interval

In Table 21 are parameter estimations of random effect and fixed effect model. Similar to GCD-Polynomial case, results of random effect model and fixed effect model are quite different. Parameter estimations of random effect model are close to that of GCD-Polynomial case. Parameter estimations of fixed effect model are quite different from that of GCD-Polynomial case. It found input elasticities of both labor and fuel not significant when $\alpha = 2.5\%$.

To take full advantage of the dataset, we also applied SS and RSCFG Inefficiency analysis model on the dataset. Result of inefficiency analysis is in Table 22.

4.7 Summary

Our purpose has been extending the suggested flexible functional form to constraint non-parametric model using regression spline. We suggested a penalized likelihood approach to extend the model. We considered three methods to obtain optimal tuning parameter. Amount these three methods KFCV and GIC have reasonable computational cost. Three-step can also be used on estimation of nonparametric version of suggested model. We demonstrated in some circumstances adopting nonparametric version of suggested model could greatly increase estimation accuracy of deterministic frontier. Then we applied the model on airline dataset again.

5. CONCLUSION AND FUTURE RESEARCH

The goal of this essay has been introduce a class of flexible functional form that narrow the gap between micro economic theory and stochastic frontier estimation as well as improve estimation accuracy of deterministic production frontier and technical efficiency. We discussed Additive Model and Generalized Cobb-Douglas (GCD) model. Without assumption of separability GCD is closer to reality. Both functional forms we discussed impose monotonicity and concavity constraint under the same condition as Cobb-Douglas production function. Yet is does not imply undesired assumptions i.e. unitary elasticity of factor substitution and constant partial and total production elasticities, that Cobb-Douglas function does. Suggested function can also be very flexible, as a matter of fact we extend it into nonparametric case use regression spline later in the essay.

In the essay we also discussed a three step method to estimate suggested models. This three step method is relatively easy to implement and can reach convergence in a matter of seconds. Because of the flexibility, suggested model can also greatly improve prediction accuracy of deterministic frontier and technical efficiency.

This class of flexible functional form can also be used on Inefficiency Analysis (IA). Our simulation show adopting suggested model can significantly improve prediction accuracy of estimated deterministic frontier and estimation accuracy of coefficients of variables for inefficiency analysis in both RSCFG and SS models.

In first two essays we use third order polynomial as transformation function ($c(\cdot)$). In reality, third order polynomial may not be flexible enough for some data set. To encounter this challenge, we extended $c(\cdot)$ to nonparametric case using regression spline. In third essay we adopted Penalized Likelihood Estimator (PLE) method to deal with over fitting problem i.e. let data decide how flexible or complicated $c(\cdot)$ should be. To choose optimal tuning parameter for PLE method we considered Leave One Out Cross Validation (LOOCV), K-Fold Cross Validation (KFCV), Parametric Direct Plug-In (PDPI) and Generalized Cross Validation (GIC) method. After our adjustment KFCV and GIC takes reasonable amount of time (minutes), and yield very close result. Our simulation shows this extension could significantly increase estimation accuracy in some circumstances.

This research can be extended in several ways.

- We still need to derive the converge rate of nonparametric extension. Mammen and Thomas-Agnan (1999) derived the converge rate of general case of suggested model with LS estimator. With some additional work we may derive convergence rate of PL estimator.

- Dualities of this class of production function still need further work i.e. the profit function, cost function, conditional demand and unconditional demand.

- Suggested model can be easily extended to cost frontier case. However as the method of estimating cost frontier is almost identical to that of production frontier, contribution on methodology will be limited. We will extend to this direction when practical problem and data present themselves.

- The most promising research opportunity is link other theoretical microeconomic model and nonparametric econometrics. For instance, microeconomics requires Marshallian demand functions to exhibit Slutsky symmetry, negativity, Homogeneity of degree zero, and budget constraint properties. With kernel estimator it is very hard to impose Slutsky symmetry Haag et al. (2009) constraint let along with other constraints above. With smoothing spline we may able to impose all constraint above.

REFERENCES

Afriat, S. N., 1972. Efficiency estimation of production functions. International Economic Review 13 (3), 568–598.

Aigner, D., Amemiya, T., Poirier, D. J., 1976. On the estimation of production frontiers: Maximum likelihood estimation of the parameters of a discontinuous density function. International Economic Review 17 (2), 377–396.

Aigner, D., Chu, S., 1968. On estimating the industry production function. The American Economic Review 58 (4), 826–839.

Aigner, D., Lovell, C., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6 (1), 21–37.

Alvarez, A., Amsler, C., Orea, L., Schmidt, P., 2006. Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. Journal of Productivity Analysis 25 (3), 201–212.

Bartholomew, D., 1959. A test of homogeneity for ordered alternatives. Biometrika 46 (1-2), 36.

Battese, G., Coelli, T., 1992. Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. Journal of Productivity Analysis 3 (1), 153–169.

Box, G., Cox, D., 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological) 26 (2), 211–252.

Caudill, S., Ford, J., 1993. Biases in frontier estimation due to heteroscedasticity. Economics Letters 41 (1), 17–20.

Caudill, S., Ford, J., Gropper, D., 1995. Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. Journal of Business and Economic Statistics 13 (1), 105.

Caves, D., Christensen, L., Tretheway, M., 1984. Economies of density versus economies of scale: Why trunk and local service airline costs differ. The RAND Journal of Economics, 471–489.

Eubank, R., 1999. Nonparametric regression and spline smoothing. CRC Press: Boca Raton.

Farrell, M., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society Series A(General) 120 (3), 253–281.

Fritsch, F., 1988. Monotone piecewise cubic data fitting. In: Conference: 2. Shrivenham International Conference on Algorithms for Approximation, Shrivenham, UK, 12 Jul 1989.

Green, P. J., Silverman, B., 1994. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman& Hall: London.

Greene, W., 1990. A gamma-distributed stochastic frontier model. Journal of Econometrics 46 (1-2), 141–163.

Greene, W., 2005. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of Econometrics 126 (2), 269–303.

Greene, W., July 2010. 2010 a course in stochastic frontier modeling.
URL `http://pages.stern.nyu.edu/ wgreene/`

Gu, C., 2002. Smoothing Spline ANOVA Models. Springer-Verlag: Berlin.

Haag, B., Hoderlein, S., Pendakur, K., 2009. Testing and imposing Slutsky symmetry in nonparametric demand systems. Journal of Econometrics 153 (1), 33–50.

Henningsen, A., Henning, C., 2009. Imposing regional monotonicity on translog stochastic production frontiers with a simple three-step procedure. Journal of Productivity Analysis 32 (3), 217–229.

Huang, C., Liu, J., 1994. Estimation of a non-neutral stochastic frontier production function. Journal of Productivity Analysis 5 (2), 171–180.

Ke, C., Wang, Y., 2004. Smoothing spline nonlinear nonparametric regression models. Journal of the American Statistical Association 99 (468), 1166–1175.

Kelly, C., Rice, J., 1990. Monotone smoothing with application to dose-response curves and the assessment of synergism. Biometrics 46 (4), 1071–1085.

Konishi, S., Kitagawa, G., 1996. Generalised information criteria in model selection. Biometrika 83 (4), 875.

Kruskal, J., 1965. Analysis of factorial experiments by estimating monotone transformations of the data. Journal of the Royal Statistical Society. Series B (Methodological) 27 (2), 251–263.

Kumbhakar, S., Ghosh, S., McGuckin, J., 1991. A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms. Journal of Business and Economic Statistics 9 (3), 279.

Li, W., Naik, D., Swetits, J., 1996. A data smoothing technique for piecewise convex/concave curves. SIAM Journal on Scientific Computing 17, 517.

Mammen, E., Thomas-Agnan, C., 1999. Smoothing splines and shape restrictions. Scandinavian Journal of Statistics 26 (2), 239–252.

McLachlan, G., Do, K., Ambroise, C., 2004. Analyzing microarray gene expression data. Wiley-IEEE: Hoboken, NJ.

Passow, E., 1974. Piecewise monotone spline interpolation. Journal of Approximation Theory 12 (3), 240–241.

Passow, E., Roulier, J., 1977. Monotone and convex spline interpolation. SIAM Journal on Numerical Analysis 14 (5), 904–909.

Pitt, M. M., Lee, L.-F., 1981. The measurement and sources of technical inefficiency in the indonesian weaving industry. Journal of Development Economics 9 (1), 43–64.

Ramsay, J., 1988. Monotone regression splines in action. Statistical Science 3 (4), 425–441.

Ramsay, J., 1998. Estimating smooth monotone functions. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60 (2), 365–375.

Reifschneider, D., Stevenson, R., August 1991. Systematic departures from the frontier: A framework for the analysis of firm inefficiency. International Economic Review 32 (3), 715–23.

Richmond, J., 1974. Estimating the efficiency of production. International Economic Review 15 (2), 515–521.

Ruppert, D., 2002. Selecting the Number of Knots for Penalized Splines. Journal of Computational and Graphical Statistics 11 (4), 735–757.

Schmidt, P., Sickles, R., 1984. Production Frontiers and Panel Data. Journal of Business and Economic Statistics 2 (4), 367.

Stevenson, R., 1980. Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics 13 (1), 57–66.

Ueki, M., Fueda, K., 2010. Optimal tuning parameter estimation in maximum penalized likelihood method. Annals of the Institute of Statistical Mathematics 62 (3), 413–438.

Wahba, G., 1977. Practical approximate solutions to linear operator equations when the data are noisy. SIAM Journal on Numerical Analysis 14 (4), 651–667.

Wahba, G., 1990. Spline models for observational data. Society for Industrial Mathematics: Philadelphia, PA.

Wang, H., Schmidt, P., 2002. One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. Journal of Productivity Analysis 18 (2), 129–144.

Wright, I., Wegman, E., 1980. Isotonic, convex and related splines. The Annals of Statistics 8 (5), 1023–1035.

VITA

Name: Ke Li

Address: 332 Blocker Bldg. 2124 TAMU, College Station, TX 77843-2124

Email Address: Knj5120@gmail.com

Education: B.S., Photoelectronic Information, South China University of Technology 2001

M.S., Finance and Economics, West Texas A&M University 2004

Ph.D., Agricultural Economics, Texas A&M University 2010