Chapter 2
Energy Performance Analysis: Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DES) for Energy Performance Analysis

Abstract Energy performance analysis in the car manufacturing industry is intriguing. The car manufacturing industry, one of the largest energy consuming industries, has been making a considerable effort to improve its energy intensity by implementing energy efficiency programs, in many cases supported by government research or financial programs. While many car manufacturers claim that they have made substantial progress in energy efficiency improvement over the past years through their energy efficiency programs, the objective measurement of energy efficiency improvement has not been studied due to the lack of suitable quantitative methods. This chapter proposes stochastic and deterministic frontier benchmarking models such as the stochastic frontier analysis (SFA) model and the data envelopment analysis (DEA) model to measure the effectiveness of energy saving initiatives in terms of the technical improvement of energy efficiency for the automotive industry, particularly vehicle assembly plants. Illustrative examples of the application of the proposed models are presented and demonstrate the overall benchmarking process to determine best practice frontier lines and to measure technical improvement based on the magnitude of frontier line shifts over time. Log likelihood ratio and Spearman rank-order correlation coefficient tests are conducted to determine the significance of the SFA model and its consistency with the DEA model. ENERGY STAR® EPI (Energy Performance Index) are also calculated. This chapter also provides a short instruction to Excel Solver by illustrating three examples: (1) SFA parameters estimation (2) DEA LP problem and (3) traveling compressed air expert problem, with an attempt to help readers learn and use GRG method, Simplex LP method and evolutionary method, respectively.

2.1 Background of Energy Performance Analysis

The growing awareness of global energy demand issues has become one of major contributors to create the concept of sustainability. The concept of sustainability was first used to describe an economic vision in equilibrium with basic ecological support systems in the 1970s. The concept has since been applied to a wide range of

S.-C. Oh and A.J. Hildreth, *Analytics for Smart Energy Management*,
Springer Series in Advanced Manufacturing, DOI 10.1007/978-3-319-32729-7_2

areas, including the car manufacturing industry, thus, motivating the change in energy consumption trends.

The typical vehicle manufacturing plants of car companies consume energy at different rates, depending on many external or internal factors, such as plant utilization, heating degree days (HDD) and cooling degree days (CDD), which are positively correlated to such factors as heating and cooling energy requirements, product type and size. Although car companies recognize that energy consumption is a large but mandatory expense, most of them have recently invested in energy saving initiatives for their plants every year to reduce energy consumption inspired by the concept of sustainability and its implication for firm values such as enhanced brand value or cost savings in energy. A notable fact is that those energy saving initiatives have been, in many cases, supported by government research or financial programs (e.g., R&D and funding programs offered by US Department of Energy Office of Energy Efficiency and Renewable Energy) because those initiatives are also aligned with the government's energy saving polices. The benefits from energy demand reduction could be significant, ranging from energy conservation and reduced environmental impact to an enhanced competitive position.

Nonetheless, as the benefits from reducing energy demand are significant, many car companies have invested considerably in strategic energy saving initiatives with the support of government R&D or financial subsidies. Now, as a logical following step, car companies and the government endeavor to investigate whether the implemented energy saving initiatives have been effective and further, institutionalized as a managed process or as a part of organizational capability because they seek to determine whether their investment or subsidies were justified and whether they have been recovered. An industry or a company, if the energy saving initiative are implemented and fully institutionalized, starts to have the potential to deliver sustained energy savings, thereby demonstrating best practices in decreasing energy intensity (kWh/vehicle in the context of car manufacturing industry). When the industry or company obtains the potential to deliver sustained energy savings and the potential is expressed as best practices, a structural technical improvement in the industry (or company) is considered to have been made. Therefore, it is possible to use the term technical improvement as a performance indicator to identify the effectiveness of energy saving initiatives, in other words, the extent to which strategic energy saving initiatives become institutionalized or part of organizational standard processes. The challenge is the lack of suitable quantitative methods to measure a structural technical improvement objectively. This chapter applies a benchmarking approach to measure technical improvement. A benchmark is a process for identifying best practices in an industry (or a large company controlling many individual producers insides) and estimating each industry's or company's efficiency by measuring the difference between actual performance and best practices. In the context of the car manufacturing industry, the difference between the actual energy use at a plant and its best practice, i.e., the lowest achievable energy use, is considered. The problem is that what is the best achievable is influenced by different operating conditions of plants (e.g., heating or cooling energy requirements, product size, or plant utilization), thus, the measuring of best practices must

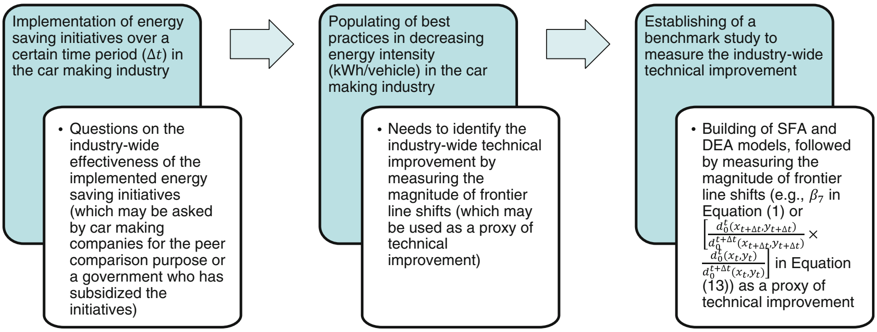


Fig. 2.1 The idea process for measuring the effectiveness of energy saving initiatives

account for these different operation conditions. A suitable benchmark model should normalize these conditions and identify a frontier line that connects the best practices in the industry. This work utilized a benchmarking approach, with the shifts of a frontier line between the time period from t and t + Δt used as a proxy to measure a structural technical improvement. Figure 2.1 depicts the idea process for measuring the effectiveness of energy saving initiatives with a benchmarking process.

This chapter aims to determine and to measure the effectiveness of energy reduction initiatives in terms of a technical improvement that corresponds to a certain structural change in industry-wide energy efficiency between two distinct time periods, namely, by proposing benchmarking models: SFA (stochastic frontier analysis) models based on Hicksian neutral technological change concept and DEA (data envelopment analysis) models incorporating the Malmquist Productivity Change Index (Sect. 2.2 discusses Hicksian concept and Malmquist index in detail). Through the SFA and DEA benchmarking processes, it is possible to identify best practice frontier lines and to analyze the technical improvement based on the magnitude of the frontier line shifts over time.

2.1.1 Background of the Auto Manufacturing Process and the Energy Consumption

A typical automobile manufacturing process generally consists of three main processes: body shop, paint shop, and general assembly. The body shop transforms raw materials into the structure of the vehicle. Then, the paint shop applies a protective and visual coating to the product. Finally, the general assembly assembles all sub-components, such as the engine and seats, into the vehicle.

Two main types of energy utility used in a typical vehicle assembly plant are electricity and fuel (including natural gas). In general, fuel is used for direct heating or to generate steam that is considered as a secondary utility similar to compressed air in vehicle assembly plants. Steam is then used mainly in painting but is also utilized for space heating, car wash and other non-manufacturing activities. Electricity is the main energy source in vehicle assembly plants, and its main uses are painting, HVAC (heating, ventilation, and air conditioning), lighting, compressed air systems, and welding and materials handling/tools.

It is possible to more holistically understand the factors affecting energy consumption by checking the consistency of the analytical results from two different models, SFA and DEA (Lin and Tseng 2005). Previous findings from performance benchmarking literature indicate that DEA and SFA have comparative advantages against each other, thereby offering the possibility of complementary use. In general, DEA is preferable in applications in which the frontier model cannot be expressed in algebraic form or does not have a known inefficiency distribution. The SFA method is preferable when certain classical assumptions are satisfied regarding the composite error terms, including the contributions from the inefficiency distribution and measurement errors. Often, SFA and DEA estimates are highly correlated in terms of rank order, regardless of inefficiency and random error variation, meaning that the feasibility and robustness of the model estimation can be demonstrated by showing a high correlation between two models. Hence, in this chapter, the Spearman rank correlation is used to check the consistency of two different models. This chapter also calculates ENERGY STAR® plant energy performance indicator values based on the SFA models.

The chapter is organized as follows: Introduction section surveys some efforts and studies related to energy use in the automotive industry and overviews benchmarking models including parametric and non-parametric approaches. Sections 2.2 and 2.3 describe SFA and DEA and the concept of technical improvement in additional detail with graphics and proposes benchmarking models to assess the significance of technical improvements in energy use alongside background data about energy consumption in vehicle manufacturing processes. Section 2.4 provides illustrative studies by using hypothetical but representative panel data sets (note: panel data refer to a group of cross-sectional data sets separated into periods of time, thus, appearing as a combination of cross-sectional and time series data sets). For confidentiality reasons, hypothetical data sets are used for the studies. In addition to implementing models, the final proposed models are analyzed and validated. Section 2.5 concludes this chapter. Appendix A shows the derivation of the log likelihood function and first-order partial derivatives for cost frontier model and the resulting parameters obtained from SFA and DEA models. Appendix B provides a short instruction to Excel Solver by illustrating three examples: (1) SFA parameters estimation (2) DEA LP problem and (3) traveling compressed air expert problem, with an attempt to help readers learn and use GRG method, Simplex LP method and evolutionary method, respectively. Note that this chapter expands on previous researches (Oh and Al 2014) by adding detailed procedures of deriving the log likelihood function and first-order partial derivatives for cost frontier model and a short instruction to Excel Solver.

2.1.2 Literature Review

Several studies related to demand, supply and management for energy use in the car manufacturing industry have been conducted.

Galitsky and Worrell (2008) collected energy efficiency improvement opportunities available to car manufacturers. They identified many energy efficiency improvement opportunities for each automotive manufacturing operation. Boyd (2005) developed plant-level energy performance indicators (EPIs) in support of the Environmental Protection Agency's ENERGY STAR program in which 35 automotive manufacturing plants of five auto companies had participated. The participating plants were plants having only body welding, assembly and painting operations. Sullivan et al. (2010) discussed calculating the environmental burdens of the part manufacturing and vehicle assembly stage of the vehicle life cycle. Their approach is bottom-up, with a particular focus on energy consumption and CO_2 emissions. They applied their models to both conventional and advanced vehicles, the latter of which include aluminum-intensive, hybrid electric, plug-in hybrid electric and all-electric vehicles. Oh and Hildreth (2014) proposed a novel decision model based on activity based costing (ABC) and stochastic programming that was developed to accurately evaluate the impact of load curtailments and to determine whether to accept an energy load curtailment offer in the smart grid.

Many previous studies on SFA and DEA, as well as the comparison of their differences are available. In research on SFA, Aigner et al. (1977) and Meeusen and Broeck (1977) proposed the stochastic frontier production function independently. The original model specification considered a production function specified for cross-sectional data in which an error term is divided into two components, one to account for random effects and another to account for technical inefficiency. Subsequently, the original model specification has been used in a large number of empirical applications over the past decades and has also been altered or extended in several ways. One extension is the two-stage estimation procedure to measure the technical change over two time periods in which firm-level efficiencies are predicted using the estimated stochastic frontiers, after which the predicted firm-level efficiencies are regressed upon firm-specific variables (such as managerial skill level change and first decision maker's characteristics) to distinguish reasons for technical changes over time. However, the two-stage estimation procedure has been criticized because it is inconsistent with its assumptions regarding the independence of the inefficiency effects over two time periods. This work follows the model specifications proposed by Battese and Coelli (1995) that addressed the issues inherent to the two-stage procedure.

Regarding research on DEA, Charnes et al. (1978) proposed the constant returns of scale (CRS) restricted DEA model by combining the Farrell efficiency rating concept and a non-parametric mathematical programming better known as CCR (Charnes-Cooper-Rhodes) model, named after its inventors. The CCR model was updated by Banker et al. (1984), who relaxed the constant returns of scale restriction to be variable returns to scale (VRS), thereby able to evaluate both the technical efficiency and the scale efficiency of decision making units (DMUs). The DEA model with the VRS concept is also called a BCC (Banker-Charnes-Cooper) model, likewise named after its inventors. To implement the VRS concept, the BCC model added an additional constraint to the CCR model, that is, the convexity restriction. When a panel data set is available and one would like to measure the technical improvement using DEA models, the Malmquist total factor productivity (TFP) index can be used to reveal a positive or negative technical change across consecutive years. The Malmquist TFP index (Färe et al. 2011) requires four distance function values, and each distance function has an equivalent DEA model. This chapter discusses those four distance functions in detail in Sect. 2.3.

Despite the fact that both SFA and DEA methods are benchmarking methods based on efficiency frontier analysis, they differ markedly. SFA is a parametric model that requires a modeler's assumption in building models. SFA is well suited to separate firms' inefficiency from statistical noise. By contrast, DEA is a non-parametric model not subject to a modeler's assumption and useful when multiple inputs and outputs should be incorporated, but susceptible when outliers in the data set exist. Lin and Tseng (2005) compared SFA and DEA extensively and summarized the differences.

Although the literature on the various methods to establish a benchmark including SFA and DEA is vast, those methods can be categorized into four approaches for benchmarking, as specified in Table 2.1. Regarding examples in the table, OLS (Ordinary Least Squares) means a linear regression model that aims to find a line such that the sum of squares of the errors of a line passing through the data is minimized. OLS reveals overall sample-based information, representing average practices. Corrected OLS aims to find a frontier line by shifting an OLS line up (production model) or down (cost model) until a single observation with a measured

Table 2.1 Four benchmarking approaches—modified from (Productivity Commission 2013)

Approach	Brief description	Examples
Statistical methods	Parametric modeling that requires parameter estimation, with data allowing for imprecision; the frontier line could be a production or a cost function	Ordinary least squared error (OLS), corrected OLS, SFA, structural time series
Non-parametric methods	Non-parametric modeling without any assumptions regarding population distributions (inefficiency distribution, measurement error distribution)	Total factor productivity indexes, DEA
Hybrid methods	A method combining non-parametric and parametric methods using a reinforced learning algorithm	Stochastic DEA (Daraio 2012)
Engineering model methods	Creating an artificial reference model as "bottom-up" based on expert knowledge and information to use as a benchmark	Swedish NPAM (network performance assessment model), bottom-up energy model

efficiency index of one remains. Structural time series models are upgraded time series models incorporating distinct parameters that may shift over time because of structural shifts, such as slowly declining or increasing productivity growth. A stochastic DEA model follows a linear programming model, such as DEA, but is extended to account for the influence of statistical noise.

2.1.3 Energy Performance Assessment

The following sections outline two primary methods to measure technical or efficiency change: SFA and DEA. SFA and DEA models are commonly represented by a form of frontier line that can be considered an optimal combination of outputs producible from a set of inputs (or an optimal combination of outputs with the lowest inefficiency). Observed shifts of the frontier line from one point in time to another suggest technical improvement, thereby implying, moreover, an institutionalized structural technological change in a given industry or company.

The rationale for developing two models concurrently is the fact that SFA and DEA have competitive advantages against each other and could be used complementarily. In detail, when the DEA frontier estimate is biased high because of outlier data beyond the true frontier, the DEA method erroneously extends the estimated frontier outward. If the SFA method can distinguish between inefficiency and noise with sufficient accuracy, then this method can be used to detect the DEA outlier problem. Similarly, DEA can be used to detect the type-II error in SFA when the SFA frontier line reduces to a standard linear regression line. Figure 2.2 illustrates various relationships between energy intensity and non-energy factors (where the best practice indicates the lowest energy use achievable at the given operation conditions), with Fig. 2.2a, b depicting a concave-up increasing energy intensity and a concave-up decreasing energy intensity, respectively. The

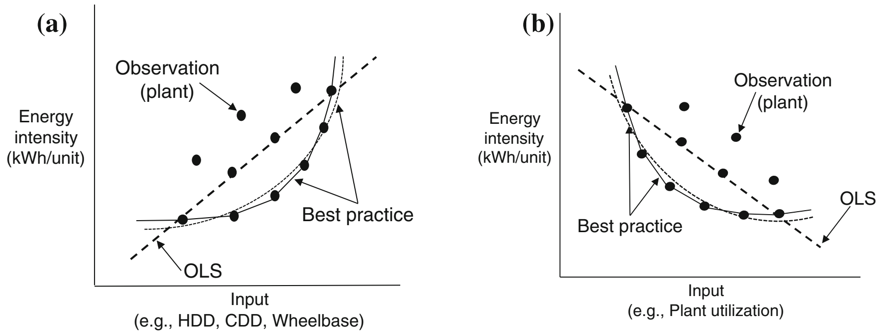


Fig. 2.2 Various relationships between energy intensity and non-energy factors (based on a cross-sectional data set). **a** Concave-up increasing energy intensity; **b** Concave-up decreasing energy intensity

concave-up increasing patterns may be observed when the energy intensity increases as the input variables (e.g., HDD, CDD, or wheelbase) increase, while the concave-up decreasing patterns may be observed when the input variables (e.g., plant utilization) have a negative relationship with the energy intensity.

It is pertinent to observe that the plant energy efficiency at one point in time is subject to the impact of a structural technical improvement as follows:

- The frontier line may shift independently of a set of observations where plants appear less efficient in the (t + Δt)-th year than in the t-th year. This occurrence happens when a technical improvement is made in the industry (or company) during the time period between the t-th and the (t + Δt)-th years, but the energy performances of target assessing plants remains unchanged and thus, the latter's energy efficiency appears less efficient because the difference between the actual efficiency score and the best practice score increases. In the Malmquist literature, this occurrence is called technical change. Figure 2.3a depicts this case.
- A set of observations may move independently closer to a frontier line while the frontier line remains unchanged during the period between the t-th and (t + Δt)-th years. This occurrence happens when a technical improvement has not been made during the time period, but the target assessing plants have improved their energy performance during the same time period and, thus, their energy efficiency appears more efficient in the (t + Δt)-th year than in the t-th year because the difference between the actual energy use and the best practice decreases. In the Malmquist literature, this occurrence is called efficiency change. Figure 2.3b depicts this case.

Aside from the two cases above, both a frontier line shift and a positive movement of a set of observations can happen simultaneously, in which case it may not be easy to differentiate the energy performance improvement of individual plants because the efficiency improvement of individual plants can be offset by the technical improvement of the industry. While SFA is likely to have trouble in

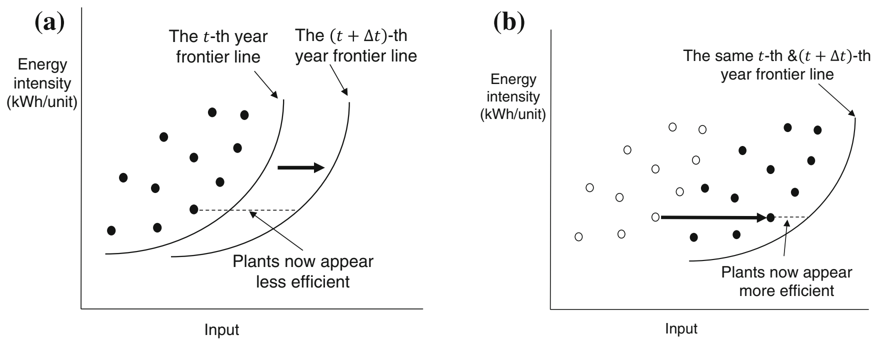


Fig. 2.3 Two main sources affecting changes in the plant energy efficiency over time. **a** Shifts in frontier line independent of a set of observations; **b** Movement of a set of observations closer to the frontier line

distinguishing technical improvements from efficiency improvements, DEA can do so by implementing the Malmquist total factor productivity (TFP) index, which will be discussed in detail in Sect. 2.3.

This study uses Spearman's rank-order correlation coefficient test to determine the consistency in ranks between SFA and DEA models in the illustrative study. The rationale for using this test is that though efficiency levels (or scores) differ between models, these methods may nonetheless generate similar rankings. If the two models' rankings are completely different, then any action taken based on the assessment may be temporary and depend on which frontier model is employed.

2.2 SFA for Energy Performance Analysis

The SFA models in this study follow the model specification proposed by Battese and Coelli (1995) and expands the cross-sectional data model specified by Boyd (2005) to a panel data model by incorporating two cross-sectional data sets. The YEAR variable involved in the resultant SFA models accounts for a Hicksian neutral technological change model. Note that Hicksian models assume special parameters that may shift the frontier line due to structural change, such as the year of the observation. Although this concept does not sufficiently account for the balance between parameters, this work uses the Hicksian neutral technical change concept because the balance between parameters is likely to remain unchanged for the time period until a technical improvement occurs. This stability occurs because the parameters used in this chapter (e.g., HDD, CDD, wheelbase, utilization) are exogenous variables, thus effecting different operation conditions in different plants. This study developed two stochastic frontier models for electricity and fuel because they are the main energy utilities consumed in vehicle manufacturing plants (note: the background on the inclusion of each term in each model is discussed in Boyd (2005). For example, why are quadratic terms for HDD and CDD included in the electricity model and the quadratic term of plant utilization included in fuel model? Why is the wheelbase of a vehicle used as a control variable rather than some other variable(s) that may also reflect the vehicle size?). The proposed SFA model for electricity is:

$$\begin{aligned} E_i/Y_i = A + \beta_1 WBASE_i + \beta_2 HDD_i + \beta_3 HDD_i^2 \\ + \beta_4 CDD_i + \beta_5 CDD_i^2 + \beta_6 Util_i + \beta_7 Year_i + u_i - v_i \end{aligned} \tag{2.1}$$

where:

E_i	Total site electricity use at plant i in kWh;
Y_i	Number of vehicles produced;
$WBASE_i$	Wheelbase (the distance between its front and rear wheels) of the largest vehicle produced in the plant in inch;
HDD_i	Thousand heating degree days for the plant location and year;

HDD_i^2	HDD_i squared;
CDD_i	Thousand cooling degree days for the plant location and year;
CDD_i^2	CDD_i squared;
$Util_i$	Plant utilization rate, defined as output/capacity, where the denominator, capacity is a normalized capacity defined as equal to capacity line rate (or job per hour) × 235 working days × 16 working hours per day;
$Year_i$	t and $t + \Delta t$ where Δt is the time period at which a significant technical improvement in energy efficiency is observed; and
β	Vector of parameters to be estimated

Note that HDD is a metric for quantifying the amount of heating that buildings in a particular location require for a certain period (e.g., a specific month or year) such that $\mathrm{HDD} = \sum_{no.days} \max(0.65\,°\mathrm{F}(\text{or } 60\,°\mathrm{F}) - \text{average day temperature})$. Similar to HDD, CDD is a metric for quantifying the amount of cooling that buildings in a particular location require for a certain period (e.g., a specific month or year) such that $\mathrm{CDD} = \sum_{no.days} \max(0, \text{average day temperature} - 65\,°\mathrm{F}(\text{or } 60\,°\mathrm{F}))$. Note that this study scales HDD and CDD by 1000. The variable v represents a measurement error to be distributed as a symmetric normal distribution, and $N(0, \sigma_v^2)$ and the variable u account for a technical inefficiency to be distributed as a half normal distribution, $N^+(0, \sigma_u^2)$. Meanwhile, the proposed SFA model for fuel is:

$$F_i/Y_i = A + \beta_1 WBASE_i + \beta_2 HDD_i + \beta_3 HDD_i^2 + \beta_4 Util_i + \beta_5 Util_i^2 + \beta_6 Year_i + u_i - \mathrm{v}_i \tag{2.2}$$

where, all the notations are specified identically to Eq. (2.1) except that F_i is the total site fuel use at plant i in 10^6 BTU. Note that this fuel model may not account for the real operation if the given plant uses steam-powered absorption chillers for air conditioning. Such chillers contribute more to the "fuel" load than the "electricity" load. If it is the case, CDD should be included in this model.

Equations (2.1) and (2.2) require several parameters to be solved, such as β, σ_v^2 and σ_u^2. This work uses the maximum likelihood method for parameter estimation and utilizes the parameterization of Battese and Corra (1977), who replaced σ_v^2 and σ_u^2 with $\varepsilon = u - \mathrm{v}, \sigma = \sigma_u^2 + \sigma_v^2, \lambda = \sqrt{\frac{\sigma_u^2}{\sigma_v^2}}$ and $\gamma = \frac{\sigma_u^2}{(\sigma_v^2 + \sigma_u^2)}$. This parameterization is useful for calculating the maximum likelihood estimates because the parameter γ is now confined to exist between 0 and 1, a range that can be more easily searched to provide a good estimate in an iterative maximization process. The first step of the maximum likelihood method is defining the log-likelihood function of the model and the log of the density function for ε:

$$log\,\varphi_\varepsilon(\varepsilon) = -\frac{1}{2} log\left(\frac{\pi}{2}\right) - \frac{1}{2} log\,\sigma^2 + log\Phi\left(\frac{\varepsilon\lambda}{\sqrt{\sigma^2}}\right) - \frac{1}{2}\frac{\varepsilon^2}{\sigma^2}$$

with N independent observations, the log of the joint density function $\varepsilon_1, \ldots, \varepsilon_N$ is:

$$log\varphi(\varepsilon_1, \ldots, \varepsilon_N) = \sum_{i=1}^{N} log\, \varphi_\varepsilon(\varepsilon_i)$$

$$= -\frac{1}{2}N\, log\left(\frac{\pi}{2}\right) - \frac{1}{2}N\, log\, \sigma^2 + \sum_{i=1}^{N} log\Phi\left(\frac{\lambda\varepsilon_i}{\sqrt{\sigma^2}}\right) - \frac{1}{2\sigma^2}\sum\nolimits_{i=1}^{N} \varepsilon_i^2$$

To emphasize that the error term ε depends on the parameter (vector) β, the log likelihood function can be expressed alternatively as:

$$\begin{aligned} l(\beta, \sigma^2, \lambda) = & -\frac{1}{2}N\log\left(\frac{\pi}{2}\right) \\ & -\frac{1}{2}N\log\sigma^2 + \sum_{i=1}^{N}\log\Phi\left(\frac{\lambda(y_i - f(x_i;\beta))}{\sqrt{\sigma^2}}\right) \\ & -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - f(x_i;\beta))^2. \end{aligned} \tag{2.3}$$

The function $l(\beta, \sigma^2, \lambda)$ is the log-likelihood function, which depends on parameters to be estimated (in this case β, σ^2 and λ) and on the data $(x_1, y_1), \ldots, (x_N, y_N)$. The derivation of the log likelihood function is available in Appendix A following Bogetoft and Otto (2011). With σ^2 replaced with $\frac{1}{N}\sum_{i=1}^{N}(y_i - f(x_i;\beta))^2$, first-order partial derivatives for the function can be obtained.

First, the partial derivative of $l(\beta, \lambda)$ with respect to β is:

$$\begin{aligned} \frac{\partial}{\partial\beta_j} l(\beta, \lambda) = & -\frac{\lambda}{\sigma}\sum_{i=1}^{N}\frac{\phi\left(\frac{\lambda\varepsilon_i}{\sigma}\right)}{\Phi\left(\frac{\lambda\varepsilon_i}{\sigma}\right)}X_{ji} \\ & + \frac{\sum_{i=1}^{N}\varepsilon_i X_{ji}}{\sigma^2}\left(1 + \frac{\lambda}{N\sigma}\sum_{i=1}^{N}\frac{\phi\left(\frac{\lambda\varepsilon_i}{\sigma}\right)}{\Phi\left(\frac{\lambda\varepsilon_i}{\sigma}\right)}\varepsilon_i\right). \end{aligned} \tag{2.4}$$

Second, the partial derivative of $l(\beta, \lambda)$ with respect to λ is:

$$\frac{\partial}{\partial\lambda} l(\beta, \lambda) = \sum_{i=1}^{N}\frac{\phi\left(\frac{\lambda\epsilon_i}{\sigma}\right)}{\Phi\left(\frac{\lambda\epsilon_i}{\sigma}\right)}\frac{\epsilon_{\mathrm{i}}}{\sigma}. \tag{2.5}$$

Coelli et al. (2005) suggested a one-sided likelihood-ratio test to determine whether the variation in inefficiency (u_i) is significant. The purpose of the test is to compare the parameter estimates in an ordinary least square regression model (OLS) with respect to the null-hypothesis, $H_0{:}\gamma = \frac{\sigma_u^2}{(\sigma_v^2 + \sigma_u^2)} = 0$, and the parameter estimates in SFA under the alternative hypothesis, $H_1{:}\gamma > 0$. The test value is calculated using Eq. (2.6).

$$LR = -2\left\{ln\left[\frac{L(OLS)}{L(SFA)}\right]\right\} = -2\{ln[L(OLS)] - ln[L(SFA)]\} \tag{2.6}$$

where, $L(OLS)$ and $L(OLS)$ are the values of the likelihood function under OLS and SFA, respectively. In the illustrative study, this study will calculate and compare the *LR* statistic with $\chi^2_{1-2\alpha}(1)$, then determine to accept or reject the null hypothesis. In other words, if the *LR* statistic exceeds $\alpha\%$ critical value, we reject the null hypothesis of no inefficiency effects. If the null hypothesis $H_0{:}\gamma = 0$ is accepted, it would indicate that σ_u^2 is zero and hence that the inefficiency term u_i should be removed from the model, thus, specifying parameters that can be consistently estimated using OLS.

This study developed an Excel spreadsheet tool to obtain the maximum likelihood estimation of subset parameters in the aforementioned SFA models rapidly and intuitively. The tool can accommodate panel data, a half-normal inefficiency distribution and a normal measurement error distribution. Section 2.4 will show what the tool looks like. Regarding an energy performance indicator developed by a credential governmental organization, the U.S. Environmental Protection Agency (EPA) introduced energy performance indicators (EPIs) through its ENERGY STAR program to encourage a variety of U.S. industries to use energy more efficiently. One of the EPIs was developed for a plant-level energy performance indicator to benchmark manufacturing energy use in the automobile industry based on the SFA model (Boyd 2005). Because a typical SFA model has a composite error term including symmetric (normal) measurement errors denoted by v_i and one-sided (half-normal) inefficiencies denoted by u_i, the frontier model takes the form of the following equation, as in Eqs. (2.1) and (2.2):

$$E_i/Y_i = f(X;\beta) + \varepsilon_i \tag{2.7}$$

where, $\varepsilon_i = u_i - \mathrm{v}_i, v_i \sim N(0,\sigma_v^2)$ and $u_i \sim N^+(0,\sigma_u^2)$. In addition, E_i is the energy use of company i; Y_i is the measured production or service measured of company i; X_i is the economic decision variables (i.e., labor-hours worked, materials processed, plant capacity, or utilization rates) or external factors (i.e., heating and cooling energy loads); and β is the vector of parameters to be estimated statistically.

Given company data, Eq. (2.7) can be expressed as Eq. (2.8), thereby providing a way to compute the difference between the actual energy use and the predicted frontier energy use:

$$E_i/Y_i - f(X;\beta) + \mathrm{v}_i = u_i \tag{2.8}$$

Then, the EPI of company i is calculated from the probability distribution of u_i as follows:

$$\begin{aligned} EPI &= probability(energy\ inefficiency \geq E_i/Y_i - f(X;\beta) + \mathrm{v}_i) \\ &= 1 - F(E_i/Y_i - f(X;\beta) + \mathrm{v}_i) \end{aligned} \tag{2.9}$$

$F()$ is the cumulative probability density function of the appropriate one-sided density function for u_i (e.g., gamma, exponential, truncated normal, and other functions). The value $1 - F()$ in Eq. (2.9) defines the EPI score and may be interpreted as a percentile ranking of the company's energy efficiency. However, in practice, the only measureable value is $u_i - \mathrm{v}_i = E_i/Y_i - f(X;B)$. By implication, the EPI score $1 - F(u_i - \mathrm{v}_i)$ is affected by the random component of v_i, that is, the score will reflect the random influences that are not accounted for by the function $F()$. Because this ranking is based on the distribution of inefficiency for the entire industry, but normalized to the specific regression factors of the given company, this statistical model enables the user to answer the hypothetical but practical question, "How does my company compare to everyone else's in my industry, if all other companies were similar to mine?". This study will calculate the EPI scores of each plant based on the proposed SFA models in Sect. 2.4. Yee and Oh (2012) used the EPI score as described in this section for selecting the optimal supply partner for composing semantic web services, when performance metrics for sustainable supply chain are important for automatic business composition, particularly at the service matchmaking phase.

2.3 DEA for Energy Performance Analysis

When a panel data set is available and one is interested in measuring the technical improvement in energy efficiency, the Malmquist total factor productivity (TFP) index can be used to reveal a positive or negative technical change across two distinct years such as t and $t + \Delta t$. One advantage of using the Malmquist TFP index is that it can be decomposed into a structural technical change (improvement or deterioration) and a technical efficiency change, where the structural technical change may account for the technical improvement (e.g., frontier line shifts between two distinct years), while the efficiency change indicates how well companies are improving to the frontier line. For example, when a frontier line shifts independently of the DMU set, DMUs appear less efficient, reflecting a positive technical change. By contrast, when a set of DMUs moves independently closer to the frontier line, DMUs appear more efficient, resulting in a positive technical efficiency change. If the frontier line shifts to a higher efficiency and simultaneously, a set of DMUs shifts to a higher efficiency, a positive TFP has occurred. Depending on the orientation used to measure the efficiency, (i.e., either output oriented or input oriented) the TFP indices differ. Recently, a new approach adopting a directional distance function was introduced to provide a flexibility in measurement by allowing negative input and output quantities. For more details on the underlying theory and application of directional distance function, see Nin et al. (2003).

For the consistency between SFA and DEA models, a new vector variable $Z_i = \left(\mathrm{HDD}_i, \mathrm{CDD}_i, \mathrm{Wheelbase}_i, \frac{1}{\mathrm{Utilization}_i}\right)$ is introduced to represent the systematic external factors given for i-th company or plant. Note that Z_i takes the inverse of utilization because this study is based on the assumption of strong disposability where all the variables must have a non-decreasing relationship with the energy intensity. Then, our interest in defining the minimum energy intensity requirement to produce one unit of vehicle under the given external condition to i-th plant is expressed in the following function:

$$(E_i/Y_i)^* = inf\{can\ process\ Z_i\ to\ produce\ one\ unit\ of\ vehicle\} \tag{2.10}$$

Equation (2.10) motivates the minimal energy density requirement in terms of micro-economic concept. It is possible to connect this motivation expressed in Eq. (2.10) with the interpretation of input distance function that we need to calculate TFP indices. For more specific details of the theoretical development on this connection, see Boyd (2008). An input oriented distance function corresponding to Eq. (2.10) is as follows:

$$D_I(Z_i, E_i/Y_i) = sup\left\{\emptyset : \left(\frac{E_i/Y_i}{\emptyset}\right) can\ process\ Z_i\ to\ produce\ one\ unit\ of\ vehicle\right\} \tag{2.11}$$

Since a distance function is defined, it is possible to calculate the TFP index. In our context, the TFP index requires four distance function values, specifically, $D_I^t(Z_t, E_t/Y_t)$, $D_I^{t+\Delta t}(Z_t, E_t/Y_t)$, $D_I^t(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t})$, and $D_I^{t+\Delta t}(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t})$, where the notation $D_I^t(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t})$ represents the distance from the period $t+\Delta t$ observation to the period t technology. Vector forms, Z_t and E_t/Y_t represent $(Z_{1t}, Z_{2t}, \ldots, Z_{Nt})$ and $(E_{1t}/Y_{1t}, E_{2t}/Y_{2t}, \ldots, E_{Nt}/Y_{Nt})$, respectively. The subscript "I" has been introduced to remind that this is an input-orientated measures. Note that each distance function has an equivalent DEA model. For example, $D_I^t(Z_t, E_t/Y_t)$ is identical to the following DEA model:

$$\begin{aligned} & D_I^t(Z_t, E_t/Y_t) = \min_{\phi,\lambda}\phi \\ s.t.\quad & -Z_{it} + Z_t\lambda \geq 0, \\ & -\phi(E_{it}/Y_{it}) + (E_t/Y_t)\lambda \leq 0 \\ & \lambda \geq 0 \end{aligned} \tag{2.12}$$

The remaining three DEA models are simple variants of this form. Table 2.2 summarizes all the forms.

LP (Linear Program) (2.12) is used to calculate the efficiency of the t-th time period relative to t-th time period technology, while LP (2.13) is used to calculate the efficiency of $(t+\Delta t)$-th time period relative to $(t+\Delta t)$-th time period technology. Similarly, LP (2.14) is used to calculate the efficiency of the $(t+\Delta t)$-th time period relative to t-th time period technology, while LP (2.15) is used to calculate the efficiency of the t-th time period relative to $(t+\Delta t)$-th time period technology. Once $D_I^t(Z_t, E_t/Y_t)$, $D_I^{t+\Delta t}(Z_t, E_t/Y_t)$, $D_I^t(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t})$, and

Table 2.2 DEA models required to calculate Malmquist TFP indices

Input oriented envelopment forms	
$D_I^{t+\Delta t}(E_{t+\Delta t}/Y_{t+\Delta t}, Z_{t+\Delta t}) = \min_{\phi,\lambda}\phi,$ $s.t.\quad -Z_{it+\Delta t} + Z_{t+\Delta t}\lambda \geq 0,$ $-\phi(E_{it+\Delta t}/Y_{it+\Delta t}) + E_{t+\Delta t}/Y_{t+\Delta t}\lambda \leq 0,$ $\lambda \geq 0.$	(2.13)
$D_I^{t}(E_{t+\Delta t}/Y_{t+\Delta t}, Z_{t+\Delta t}) = \min_{\phi,\lambda}\phi$ $s.t.\quad -Z_{it+\Delta t} + Z_{t}\lambda \geq 0,$ $-\phi(E_{it+\Delta t}/Y_{it+\Delta t}) + E_{t}/Y_{t}\lambda \leq 0,$ $\lambda \geq 0.$	(2.14)
$D_I^{t+\Delta t}(Z_t, E_t/Y_t) = \min_{\phi,\lambda}\phi,$ $s.t.\quad -Z_{it} + Z_{t+\Delta t}\lambda \geq 0,$ $-\phi(E_{it}/Y_{it}) + E_{t+\Delta t}/Y_{t+\Delta t}\lambda \leq 0,$ $\lambda \geq 0.$	(2.15)

$D_I^{t+\Delta t}(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t})$ are obtained, the Malmquist TFP index can be calculated and then rearranged such that it is equivalent to the product of a technical efficiency change index and an index of technical change.

$$\begin{aligned} m_I(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t}, Z_t, E_t/Y_t) &= \left[\frac{D_I^t(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t})}{D_I^t(Z_t, E_t/Y_t)} \times \frac{D_I^{t+\Delta t}(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t})}{D_I^{t+\Delta t}(Z_t, E_t/Y_t)}\right]^{1/2} \\ &= \frac{D_I^{t+\Delta t}(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t})}{D_I^t(Z_t, E_t/Y_t)} \left[\frac{D_I^t(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t})}{D_I^{t+\Delta t}(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t})} \times \frac{D_I^t(Z_t, E_t/Y_t)}{D_I^{t+\Delta t}(Z_t, E_t/Y_t)}\right]^{1/2} \end{aligned} \tag{2.16}$$

The first and second term of Eq. (2.16) correspond to an efficiency change and a structural technical change, respectively, as follows:

$$\textit{Efficiency change} = \frac{D_I^{t+\Delta t}(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t})}{D_I^t(Z_t, E_t/Y_t)} \tag{2.17}$$

Meanwhile,

$$\textit{Technical change} = \left[\frac{D_I^t(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t})}{D_I^{t+\Delta t}(Z_{t+\Delta t}, E_{t+\Delta t}/Y_{t+\Delta t})} \times \frac{D_I^t(Z_t, E_t/Y_t)}{D_I^{t+\Delta t}(Z_t, E_t/Y_t)}\right]^{1/2} \tag{2.18}$$

Note that the ϕ and λ are likely to assume different values in the four DEA models in Table 2.2. Furthermore, these four models must be calculated for each plant in the sample. Thus, if there are 10 plants and two time periods, then 40 linear programing problems must be solved. To streamline this multiple calculation procedure, this study developed an Excel spreadsheet tool as does for the SFA models. The developed tool uses VBA in Excel and automates iterations for solving multiple linear programing models. Section 2.4 will show what the tool looks like.

2.4 Illustrative Study

This chapter uses artificial data sets for illustrative studies because of intellectual property issues. The data sets were generated to resemble real-world data as close as possible. Although SFA and DEA are generally conducted with real industry data to suggest new insights or interesting finds, the authors believe that the use of artificial data sets will not be detrimental to the overall purpose of this chapter that is to demonstrate the benchmarking process from building frontier models to identifying any structural technical improvement. The generated artificial data sets are listed in Table 2.3 in which two different years' data (years t and $t + \Delta t$) for 10 vehicle

Table 2.3 Plant data used in the illustrative studies

Plant	*t*-th year					
	Wheel base (inch)	HDD (1000)	CDD (1000)	Util	Electricity intensity (kWh/unit)	Fuel intensity (10^6 BTU/unit)
1	133.50	6.69	1.22	1.19	914.64	2.18
2	105.75	6.20	1.48	1.27	1242.57	2.97
3	155.32	5.17	2.91	1.07	2098.37	5.01
4	112.01	5.40	1.71	1.60	1212.36	2.90
5	130.63	3.22	3.03	1.78	1589.27	3.80
6	133.50	6.47	1.41	1.94	1336.01	3.19
7	105.87	6.12	1.43	1.58	1553.32	3.71
8	155.51	5.33	3.09	0.77	1714.51	4.10
9	112.24	5.85	2.35	0.80	1548.68	3.70
10	130.63	3.03	3.43	1.13	1718.41	4.10
Plant	($t + \Delta t$)-th year					
	Wheel base (inch)	HDD (1000)	CDD (1000)	Util	Electricity intensity (kWh/unit)	Fuel intensity (10^6 BTU/unit)
1	133.50	5.83	1.96	0.73	1272.92	3.04
2	105.75	5.87	1.84	2.09	784.46	1.87
3	155.32	3.86	2.82	1.13	1950	4.66
4	112.01	4.53	2.52	2.00	921.72	2.20
5	130.63	2.51	3.99	0.50	1384.39	3.31
6	133.50	5.87	1.46	1.77	1008.52	2.41
7	105.87	6.17	1.11	0.51	789.69	1.89
8	155.51	4.32	3.23	0.77	1898.7	4.54
9	112.24	4.72	2.03	2.22	1100.04	2.63
10	130.63	2.72	3.75	2.29	995.48	2.38

assembly plants are considered. Regarding the scope of assembly plant, the authors are only considering body shop, paint shop and GA. In fact, these areas vary widely in terms of work volume, labor hours or energy usage depending on their level of in-house *versus* outsourced tasks. The data are generated with an in-house case assumed. In addition, the authors assumed that the major energy-consuming operations are similar among plants. For example, plants are assumed to use electricity-powered chiller, solvent-borne paint system, gas-fired direct heating system, and air conditioning in place.

This chapter uses a commercially available spreadsheet package, Excel, to build the SFA and DEA models. Excel provides an add-on tool called Solver with different solving method options such as Simplex or GRG (Generalized Reduced Gradient). Using the GRG solver method facilitates the maximum likelihood estimation of subset parameters of the proposed SFA models. The example in Fig. 2.4 illustrates a case in which the tool accommodates plant-level input panel data on electricity and builds a model corresponding to Eq. (2.1), thus, estimating parameters for the half-normal inefficiency distribution and the normal measurement error distribution.

The estimated parameters for the electricity and fuel SFA models are shown in Table 2.4 where β_6 and β_7 are the coefficient representing YEAR in the fuel SFA model and in the electricity SFA model, respectively. The one-sided likelihood-ratio test values (*LR*) for both models reveal that the models are adequate at the 99.5 % significance level and that the models have very little error attributable to random noise, with most departures attributable to inefficiency. Therefore, the null-hypothesis, $H_0{:}\gamma = \frac{\sigma_u^2}{(\sigma_v^2 + \sigma_u^2)} = 0$, is rejected, and the alternative

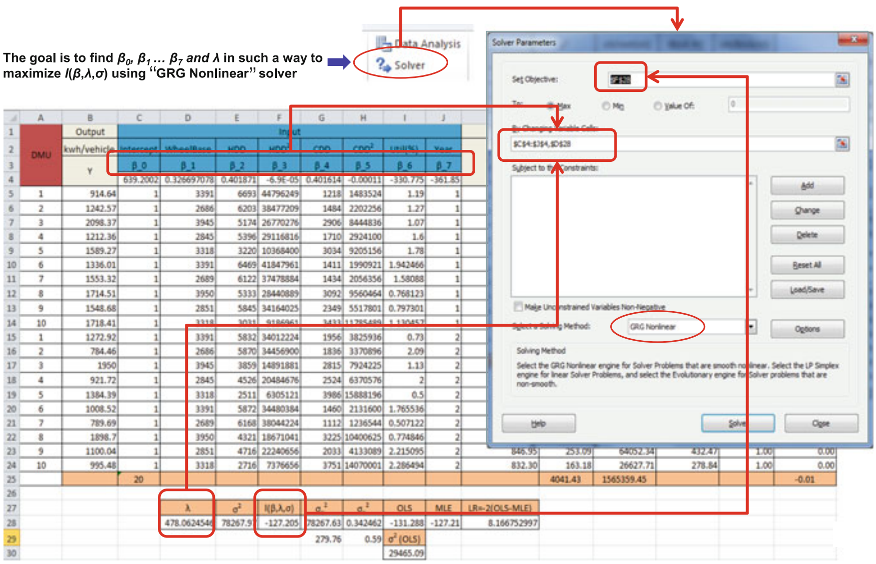


Fig. 2.4 SFA model estimation using MS-Excel Solver with "GRG Nonlinear" selected

Table 2.4 Parameter estimates for the SFA models

Variables	Estimates for the electricity SFA model (standard error; *t*-ratio)	Estimates for the fuel SFA model (standard error; *t*-ratio)
β_0	650.49	0.67
β_1	8.22 (6.05; 1.36)	0.02 (0.00; 10.15)***
β_2	394.87 (1159.16, 0.34)	1.53 (1.38; 1.11)
β_3	−68.34 (123.01; −0.56)	−0.21 (0.15; −1.40)
β_4	410.22 (1297.51; 0.32)	−0.21 (2.83; −0.07)
β_5	−107.80 (268.42; −0.4)	−0.16 (1.03; −0.16)
β_6	−331.85 (173.66; −1.91)**	−0.86 (0.53; −1.60)*
β_7	−361.87 (191.46; −1.89)**	NA
σ_u	279.79	0.68
σ_v	0.55	0.00
$\lambda = \sqrt{\frac{\sigma_u}{\sigma_v}}$	505.96	614.92
$L(OLS)$	−131.28	−10.29
$L(SFA)$	−127.21	−6.81
LR	$8.17 > \chi^2_{1-2\times0.005}(1) = 6.635$	$6.97 > \chi^2_{1-2\times0.005}(1) = 6.635$

Notations for significance level in a two-tailed test: ***(99 %); **(90 %); *(85 %)

hypothesis $H_1{:}\gamma > 0$ with technical inefficiency effect is accepted for both the electricity and fuel SFA models. This statistical results show that a structural technical improvement in electricity (β_7 of the electricity SFA model) and fuel (β_6 of the fuel SFA model) occurred during the period. Furthermore, β_7 and β_6 are statistically significant at the 90 % level $(-1.91 < t_{0.95}(12) = -1.782)$ and the 85 % level $(t_{0.95}(13) = -1.771 < -1.6 < t_{0.9}(13) = -1.350)$ in a two-tailed test, respectively. These results indicate that, all other factors being equal, an average reduction of 330.77 (kWh) and 253.55 (kWh) in the electricity and fuel per vehicle has occurred, leading to efficiency gains of \$41.73/vehicle (note: the calculation assumes \$0.1/kWh for electricity and \$0.03413/kWh (=0.03413 therm/kWh × \$1/therm) for natural gas). This magnitude of efficiency gains may seem small in the unit cost of production but may offer considerable energy cost savings and significantly reduce the environmental impact when the total production is considered. For example, let us assume that a car manufacturing company produces nine million cars per year and must solely purchase CO_2 credits from a market to emit CO_2. Given these condition, if the company achieved the aforementioned magnitude of efficiency gains, then the total cost savings from energy reduction and a reduced environment impact would be \$428 M (note: \$428 M ≈ 9,000,000 × [\$41.73 + (330.77 + 253.55 kWh)/1000 × \$10]; the CO_2 credit price in the market is assumed to be \$10 per CO_2 ton). ENERGY STAR® plant energy performance indicator (EPI) values are also calculated, and the results are summarized in Table 2.5.

Note that DMUs 6 and 7 show the lower efficiency in Table 2.5 in *t*-th Year. These low efficiencies are caused by the large difference between their average practices and best practices. These results, however, also indicate that DMUs 6 and 7 have higher potentials to further improvement in energy savings.

Table 2.5 SFA results in terms of EPI

DMU	*t*-th year		(t + Δt)-th year	
	Electricity (%)	NG (%)	Electricity (%)	NG (%)
1	100	100	66	42
2	26	31	24	28
3	16	16	26	14
4	98	100	100	100
5	99	86	81	100
6	6	6	40	40
7	1	2	99	92
8	99	100	26	24
9	35	30	37	31
10	93	96	56	83
Mean	57	57	56	55

Using the Simplex solver, this study developed a spreadsheet tool for DEA, too. The developed tool uses VBA in Excel and automates iterations for solving multiple linear programing models. Briefly, with respect to automation logic, the tool uses "For" loop to automate iterations of solving multiple linear programing models in which the Solver with the "Simplex" optimization option calculates the efficiency for each DMU and the results are recorded in a table using the copy/paste function (note: the three major functions used in the loop statement of the VBA programing are as follows: (1) "SolverOk"—defines the objective function and the decision variables; (2) "SolverAdd"—defines model constraints; and (3) "SolverSolv"—runs Solver). Figure 2.5 illustrates an example in which the tool accommodates

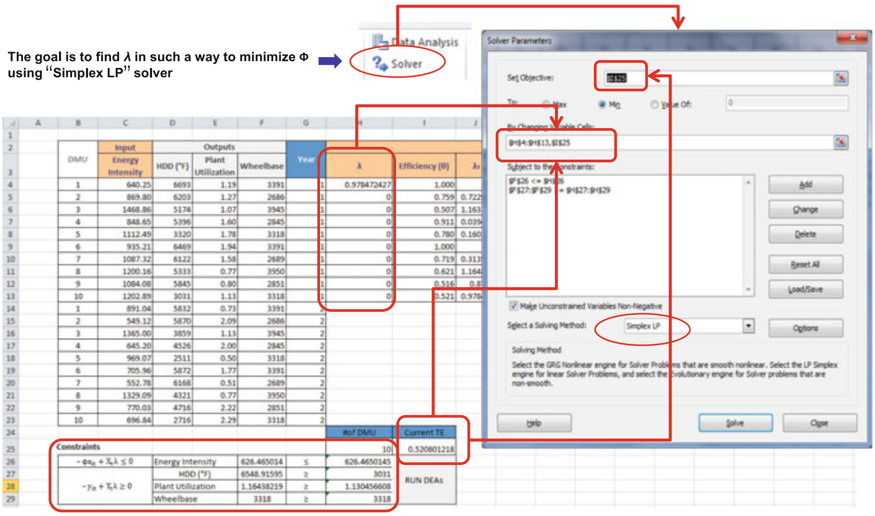


Fig. 2.5 DEA model implementation using MS-Excel Solver with "Simplex LP"

Table 2.6 Malmquist index summary (electricity)

DMU	Efficiency change (%)	Technical change (%)	Total factor productivity change (%)
1	78	106	82
2	123	155	190
3	77	138	106
4	92	166	152
5	76	175	133
6	98	131	128
7	135	124	166
8	62	147	91
9	85	172	147
10	100	187	187
Mean	92	149	139

plant-level input panel data on fuel corresponding to LP (2.7). Tables 2.6 and 2.7 present the Malmquist indices obtained by solving the DEA models for electricity and fuel, respectively. Three indices are presented for each firm, such as efficiency change (relative to a CRS technology), technical change, and total factor productivity change. It should be noted that the technical change of each model from t to $t + \Delta t$ increases (greater than 100 %), indicating that there has been a structural technical improvement in energy performance over the years. The DEA efficiency at each year is also calculated and summarized in Tables 2.8 and 2.9.

In order to measure the consistency between the SFA and the DEA approaches on the efficiency ranking results for firms, a Spearman's rank correlation coefficient test was conducted. Spearman's rank correlation coefficient values are 0.9 and 0.25 for t and $t + \Delta t$, respectively, in electricity, and 0.3 and 0.38 for t and $t + \Delta t$,

Table 2.7 Malmquist index summary (natural gas)

DMU	Efficiency change (%)	Technical change (%)	Total factor productivity change (%)
1	78	92	72
2	132	146	192
3	117	92	108
4	99	148	146
5	90	109	98
6	98	129	127
7	139	122	169
8	98	92	90
9	147	132	194
10	187	122	229
Mean	113	121	143

Table 2.8 DEA results (electricity)

DMU	$d_0^t(x_t, y_t)$ (%)	$d_0^{t+\Delta t}(x_t, y_t)$ (%)	$d_0^t(x_{t+\Delta t}, y_{t+\Delta t})$ (%)	$d_0^{t+\Delta t}(x_{t+\Delta t}, y_{t+\Delta t})$ (%)
1	100	108	94	78
2	82	65	192	100
3	77	55	81	59
4	100	69	173	92
5	100	62	144	76
6	100	74	125	98
7	74	52	107	100
8	99	68	91	61
9	89	58	146	76
10	100	58	202	100
Mean	92	67	135	84

Table 2.9 DEA results (natural gas)

DMU	$d_0^t(x_t, y_t)$ (%)	$d_0^{t+\Delta t}(x_t, y_t)$ (%)	$d_0^t(x_{t+\Delta t}, y_{t+\Delta t})$ (%)	$d_0^{t+\Delta t}(x_{t+\Delta t}, y_{t+\Delta t})$ (%)
1	100	109	72	78
2	76	65	183	100
3	51	55	55	59
4	91	69	149	90
5	78	61	65	70
6	100	75	123	98
7	72	52	107	100
8	62	67	56	61
9	52	54	138	76
10	52	56	158	97
Mean	73	66	111	83

respectively, in natural gas. All of the rank correlation coefficient values are positive, indicating that the ranks of the SFA and DEA results have moderate (in t) and small (in $t + \Delta t$) positive linear relationships.

It makes sense to compare the estimated parameters to those of existing estimated models in terms of value and sign as part of cross-validation if there have been similar estimation works. The 2000 and 2005 models elicited by Boyd (2014) have the identical model configuration with this study. Therefore, a comparison on the estimated parameters was conducted between those models and the results are summarized in Tables 2.10 and 2.11. One challenge against the comparison was that the datasets of two models are significantly different. The 2000 and 2005 models were based on real data composed by collecting some sample plant data from major car making companies in U.S. while this study generated an artificial

Table 2.10 Comparison of electricity SFA model parameters

Parameter	This study (based on simulated data)	2000 model	2005 model	Direction of the relationship
Constant	650.49	369.39	−91.84	N/A
Wbase	8.22	2.77	2.03	↗
HDD	394.87	−48.41	163.06	↗
HDD^2	−68.34	4.79	−15.17	
Util	−331.85	−138.61	−112.54	↘
CDD	410.22	−59.32	−223.89	↗
CDD^2	−107.80	41.91	86.61	

Table 2.11 Comparison of fuel SFA model parameters

Parameter	This study (based on simulated data)	2000 model	2005 model	Direction of the relationship
Constant	0.67	3.827	−0.526	N/A
Wbase	0.02	0.00322	0.019	↗
HDD	1.53	−0.545	0.439	↗
HDD^2	−0.21	0.11		
Util	−0.21	−6.788	−0.072	↘
$Util^2$	−0.16	2.399		

dataset by simulating a population that resembles GM plants located in a specific region. Due to the large difference between datasets, the differences in magnitude between parameter values exist. However, the orders of magnitude between parameter values are in the same range and the directions of relationships between systematic external factors and energy intensity (i.e., signs of estimated parameters) turned out consistent. The authors again want to clarify that the datasets used in this chapter are simulated and should not be taken to be applicable to the industry, but are only illustrative of the proposed models.

It seems that it would be more useful to compare best practices with inefficient practices to identify energy reduction opportunities after computing numerical efficiencies and locating the best and inefficient performance plants. Finding energy reduction opportunities must be preceded by understanding high energy cost drivers for inefficient plants. For this purpose, Oh and Hildreth (2013), Jurek et al. (2012), and Oh et al. (2011) proposed activity-based decision steps including a step of comparing hourly average energy use of each activity between best practice plants and less efficient plants followed by figuring out which activity are problematic cost drivers for less efficient plants.

2.5 Summary

This chapter proposes a benchmarking process using stochastic and deterministic frontier analysis models, specifically, SFA and DEA, to identify industry-wide or company-wide structural technical improvement in energy efficiency with a focus on the car manufacturing industry. The quantitative identification of technical improvement in energy efficiency is important to help car manufacturing companies evaluate the effectiveness of the various energy efficiency programs that they may have implemented, in many cases supported by government R&D or financial programs. This study proposed SFA models that incorporate the Hicksian neutral technological change concept and DEA models implemented to calculate Malmquist Productivity Change indices. Illustrative examples of the proposed models are presented to demonstrate the overall benchmarking process to find frontier lines and to measure the shifts of the frontier line that were used to proxy the structural technical improvement in energy efficiency. A log likelihood ratio test and a Spearman rank-order correlation coefficient test were conducted to test the significance of the SFA model and its consistency with the DEA model, respectively. ENERGY STAR® plant energy performance indicator values were also calculated. The results of the analysis based on the SFA models calculated total efficiency gains of $41.73/vehicle during the tested period. The tools developed for illustrative examples are available upon request at authors.

Regarding future work, one priority is to enhance the proposed SFA and DEA models to enable them to account for structural technological change by including the time-varying behavior of the inefficiency effects, thereby identifying more extensive factors affecting the technical change. Additionally, the implementation of a directional distance function in calculating the Malmquist TFP indices is of interest.

2.6 Exercises

1. Figures 2.4 and 2.5 illustrate how to implement SFA and DEA models in MS-Excel Solvers including "GRG Non-linear" and "Simplex LP" solvers. Refer to Appendix B ("Getting Started with Excel Solver for SFA and DEA Analyses") and replicate the illustrated work.
2. Equation (2.3) is a derivation for the case that a SFA model demonstrates half-normal inefficiency and produces a normal measurement error. The product of the densities of a half-normal inefficiency distribution $(u \sim N^+(0, \sigma_u^2))$ and a normal measurement error distribution $((v \sim N(0, \sigma_v^2)))$ is:

$$\varphi_v(\varepsilon - u)\varphi_u(u) = \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{-\frac{1}{2}\frac{(\varepsilon-u)^2}{\sigma_v^2}} \frac{2}{\sqrt{2\pi\sigma_u^2}} e^{-\frac{1}{2}\frac{u^2}{\sigma_u^2}}$$
$$= \frac{1}{\pi\sqrt{\sigma_u^2\sigma_v^2}} e^{-\frac{1}{2}\frac{(\sigma_u^2+\sigma_v^2)u^2 - 2\sigma_u^2\varepsilon u + \sigma_u^2\varepsilon^2}{\sigma_u^2\sigma_v^2}}.$$

Starting from the equation above, drive Eqs. (2.3)–(2.5). Refer to Appendix A ("Derivation of the log likelihood function and first-order partial derivatives for cost frontier model") for the solution.

3. In reality, many firms are involved in production activities that often generate undesirable outputs such as harmful side products (e.g., pollution, waste, noise, and etc.) to the environment. Those undesirable outputs can be disposable and the disposability assumptions for undesirable products cause difficulties in the measurements of the overall performance of firms. Meanwhile, DEA has been considered as a successful means to address the disposability issues. Read the paper (Toshiyuki and Mika 2012) and gain the insight on the various DEA models to assess the energy and environment performance in the light of disposability assumptions.
4. Section 2.2 introduces EPI (Energy Performance Indicator) which is developed as part of EPA's ENERGY STAR program to measure the energy performance of companies who usually manufacture consumer products. Technically, EPI is calculated as a percentile ranking of the energy efficiency of the given company by adopting a parametric modeling-based benchmarking technique so-called Stochastic Frontier Analysis (SFA). Suppose that the one-sided inefficiency distribution for u_i follows a half-normal distribution such that $u_i \sim N^+(0, \sigma_u^2)$, where we additionally assume that $\sigma_u = 0.5$ MWH/Unit and the best practice company in the industry has 2 MWH/Unit as its energy performance. With these assumptions, answer the following questions by referring to the procedures set forth in Sect. 2.2:

- If your company's energy performance is 5 MWH/Unit, what is your company's EPI in the industry?
- If your company wants to achieve 75 % EPI in the industry, what energy performance (MWH/Unit) should your company meet?

Appendix A: Derivation of the Log Likelihood Function and First-Order Partial Derivatives for Cost Frontier Model

This derivation is for the case in which a SFA model demonstrates half-normal inefficiency and produces a normal measurement error. The derivation of the log likelihood function is modified from Bogetoft and Otto (2011). The product of the densities of a half-normal inefficiency distribution $(u \sim N^+(0, \sigma_u^2))$ and a normal measurement error distribution $(v \sim N(0, \sigma_v^2))$ is:

$$\varphi_v(\varepsilon - u)\varphi_u(u) = \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{-\frac{1}{2}\frac{(\varepsilon-u)^2}{\sigma_v^2}} \frac{2}{\sqrt{2\pi\sigma_u^2}} e^{-\frac{1}{2}\frac{u^2}{\sigma_u^2}}$$
$$= \frac{1}{\pi\sqrt{\sigma_u^2\sigma_v^2}} e^{-\frac{1}{2}\frac{u^2}{\sigma_u^2} - \frac{1}{2}\frac{(\varepsilon-u)^2}{\sigma_v^2}}$$
$$= \frac{1}{\pi\sqrt{\sigma_u^2\sigma_v^2}} e^{-\frac{1}{2}\frac{(\sigma_u^2+\sigma_v^2)u^2 - 2\sigma_u^2\varepsilon u + \sigma_u^2\varepsilon^2}{\sigma_u^2\sigma_v^2}}.$$

The integration of the aforementioned joint density is given by the following:

$$\varphi_\varepsilon(\varepsilon) = \int_0^\infty \varphi_v(\varepsilon - u)\varphi_u(u)du$$
$$= \frac{1}{\pi\sqrt{\sigma_u^2\sigma_v^2}} \int_0^\infty e^{-\frac{1}{2}\frac{(\sigma_u^2+\sigma_v^2)u^2 - 2\sigma_u^2\varepsilon u + \sigma_u^2\varepsilon^2}{\sigma_u^2\sigma_v^2}} du$$
$$= \frac{1}{\pi\sqrt{\sigma_u^2\sigma_v^2}} \int_0^\infty e^{-\frac{1}{2}\left(\left(\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}\right)u^2 - \frac{2}{\sigma_v^2}\varepsilon u + \frac{1}{\sigma_v^2}\varepsilon^2\right)} du$$
$$= \frac{1}{\pi\sqrt{\sigma_u^2\sigma_v^2}} \int_0^\infty e^{-\frac{1}{2}\left(\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}\right)\left(u^2 - \frac{2\varepsilon u}{\left(\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}\right)\sigma_v^2} + \frac{\varepsilon^2}{\left(\left(\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}\right)\sigma_v^2\right)^2} - \frac{\varepsilon^2}{\left(\left(\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}\right)\sigma_v^2\right)^2}\right)} e^{-\frac{\varepsilon^2}{2\sigma_v^2}} du$$
$$= \frac{1}{\pi\sqrt{\sigma_u^2\sigma_v^2}} \int_0^\infty e^{-\left(\frac{1}{\sqrt{2}}\sqrt{\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}}\left(u - \frac{\varepsilon}{\left(\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}\right)\sigma_v^2}\right)\right)^2} e^{\left(\frac{\varepsilon^2}{2\left(\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}\right)(\sigma_v^2)^2} - \frac{\varepsilon^2}{2\sigma_v^2}\right)} du.$$

Let $t = \frac{1}{\sqrt{2}}\sqrt{\frac{1}{\sigma_u^2} + \frac{1}{\sigma_v^2}}\left(u - \frac{\varepsilon}{\left(\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}\right)\sigma_v^2}\right)$, $u = \frac{\sqrt{2}}{\sqrt{\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}}}t + \frac{\varepsilon}{\left(\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}\right)\sigma_v^2}$ and $du = \frac{\sqrt{2}}{\sqrt{\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}}}dt$. Then, if $u \to \infty$, then $t \to \infty$. If $u = 0$, then $t = \frac{-\varepsilon}{\sqrt{2}\sqrt{\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}}\sigma_v^2}$.

$$\varphi_\varepsilon(\varepsilon) = \frac{1}{\pi\sqrt{\sigma_u^2\sigma_v^2}}\left(\int_{\frac{-\varepsilon}{\sqrt{2}\sqrt{\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}}\sigma_v^2}}^{\infty} e^{-t^2}\frac{\sqrt{2}}{\sqrt{\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}}}dt\right)\mathrm{e}^{-\frac{1}{2}\frac{\varepsilon^2}{(\sigma_u^2+\sigma_v^2)}}$$

$$= \frac{1}{\pi\sqrt{\sigma_u^2\sigma_v^2}}\frac{\sqrt{2}}{\sqrt{\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}}}\frac{\sqrt{\pi}}{2}\left(\frac{2}{\sqrt{\pi}}\int_{\frac{-\varepsilon}{\sqrt{2}\sqrt{\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}}\sigma_v^2}}^{\infty}\mathrm{e}^{-t^2}dt\right)\mathrm{e}^{-\frac{1}{2}\frac{\varepsilon^2}{(\sigma_u^2+\sigma_v^2)}}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_u^2+\sigma_v^2}}\left(1+erf\left(\frac{\varepsilon}{\sqrt{2}\sqrt{\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}}\sigma_v^2}\right)\right)\mathrm{e}^{-\frac{1}{2}\frac{\varepsilon^2}{(\sigma_u^2+\sigma_v^2)}}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_u^2+\sigma_v^2}}\left(1+erf\left(\frac{\varepsilon}{\sqrt{2}\sqrt{\sigma_v^2+\sigma_u^2}}\sqrt{\frac{\sigma_u^2}{\sigma_v^2}}\right)\right)\mathrm{e}^{-\frac{1}{2}\frac{\varepsilon^2}{(\sigma_u^2+\sigma_v^2)}}$$

$$\left(\text{Set } \sigma^2 = \sigma_v^2+\sigma_u^2 \quad \text{and} \quad \lambda = \sqrt{\frac{\sigma_u^2}{\sigma_v^2}}\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}}\left(1+erf\left(\frac{\varepsilon}{\sqrt{2}\sqrt{\sigma^2}}\lambda\right)\right)\mathrm{e}^{-\frac{1}{2}\frac{\varepsilon^2}{\sigma^2}} \quad \left(erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x \mathrm{e}^{-x^2}dx\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}}2\Phi\left(+\frac{\lambda\varepsilon}{\sqrt{\sigma^2}}\right)\mathrm{e}^{-\frac{1}{2}\frac{\varepsilon^2}{\sigma^2}} \quad (\Phi \textit{ is the normal cumulative density function})$$

$$= \frac{\sqrt{2}}{\sqrt{\pi\sigma^2}}\Phi\left(+\frac{\lambda\varepsilon}{\sqrt{\sigma^2}}\right)\mathrm{e}^{-\frac{1}{2}\frac{\varepsilon^2}{\sigma^2}}.$$

where $\sigma^2 = \sigma_v^2+\sigma_u^2$, and $= \sqrt{\frac{\sigma_u^2}{\sigma_v^2}}$. The error function $erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x \mathrm{e}^{-t^2}dt$ has the following property: $erf(-x) = -erf(x)$. Its relationship to the normal distribution is given by $\Phi(x) - \frac{1}{2} = \frac{1}{\sqrt{2\pi}}\int_0^x \mathrm{e}^{-\frac{1}{2}t^2}dt = \frac{1}{2}erf\left(\frac{x}{\sqrt{2}}\right)$ such that $\Phi(x) = \frac{1}{2}\left(1+erf\left(\frac{x}{\sqrt{2}}\right)\right)$. Then, the log of this density is:

$$\log\varphi_\varepsilon(\varepsilon) = -\frac{1}{2}\log\left(\frac{\pi}{2}\right) - \frac{1}{2}\log\sigma^2 + \log\Phi\left(\frac{\varepsilon\lambda}{\sqrt{\sigma^2}}\right) - \frac{1}{2}\frac{\varepsilon^2}{\sigma^2}.$$

With K independent observations and K firms, the joint density is $\varphi(\varepsilon_1,\ldots,\varepsilon_K) = \prod_{k=1}^{K}\varphi_\varepsilon(\varepsilon_K)$ and the log of the joint density is:

$$\begin{aligned}
\log \varphi(\varepsilon_1, \ldots, \varepsilon_K) &= \sum_{k=1}^{K} \log \varphi_\varepsilon(\varepsilon_k) \\
&= -\frac{1}{2} K \log\left(\frac{\pi}{2}\right) \\
&\quad -\frac{1}{2} K \log \sigma^2 + \sum_{k=1}^{K} \log \Phi\left(\frac{\lambda \varepsilon_k}{\sqrt{\sigma^2}}\right) - \frac{1}{2\sigma^2} \sum_{k=1}^{K} \varepsilon_k^2.
\end{aligned}$$

We can rewrite this equation to emphasize that the error term ϵ depends on the parameter (vector) β, such that the log likelihood function is given by:

$$\begin{aligned}
l\left(\beta, \sigma^2, \lambda\right) &= \log \varphi_\varepsilon\left(\varepsilon_1(\beta), \ldots, \varepsilon_K(\beta); \sigma^2, \lambda\right) \\
&= \log \varphi_\varepsilon\left(y_1 - f(x_1; \beta), \ldots, y_K - f(x_K; \beta); \sigma^2, \lambda\right) \\
&= -\frac{1}{2} K \log\left(\frac{\pi}{2}\right) \\
&\quad -\frac{1}{2} K \log \sigma^2 + \sum_{k=1}^{K} \log \Phi\left(\frac{\lambda(y_k - f(x_k; \beta))}{\sqrt{\sigma^2}}\right) \\
&\quad -\frac{1}{2\sigma^2} \sum_{k=1}^{K} (y_k - f(x_k; \beta))^2.
\end{aligned}$$

The function $l(\beta, \sigma^2, \lambda)$ is the log-likelihood function, which depends on the parameters to be estimated (in this case β, σ^2, and λ) and on the data $(x_1, y_1), \ldots, (x_K, y_K)$. Then, the gradient of $l(\beta, \lambda)$ with respect to β, λ is as follows, with σ^2 defined as $\frac{1}{K} \sum_{k=1}^{K} (y_k - f(x_k; \beta))^2$.

$$\begin{aligned}
l(\beta, \lambda) &= -\frac{k}{2} \log\left(\frac{\pi}{2}\right) - \frac{k}{2} \log\left(\sigma^2\right) + \sum_{k=1}^{K} \log \Phi\left(\frac{\varepsilon_k \lambda}{\sigma}\right) - \frac{k}{2} \\
&= \frac{k}{2} \log\left(\frac{\pi}{2}\right) - k \log \sigma + \sum_{k=1}^{K} \log \Phi\left(\frac{\lambda \varepsilon_k}{\sigma}\right) - \frac{k}{2}.
\end{aligned}$$

Let $\sigma' = \frac{\partial}{\partial \beta_j}(\sigma)$. Then, $\sigma' = \frac{\partial}{\partial \beta_j}\left(\sqrt{\frac{1}{k} \sum_{K=1}^{K} \varepsilon_k^2}\right) = \frac{-2 \sum_{k=1}^{K} \varepsilon_k X_{jk}}{2\sqrt{k}\sqrt{\sum_{k=1}^{K} \varepsilon_k^2}} = -\frac{\sum_{k=1}^{K} \varepsilon_k X_{jk}}{\sqrt{k}\sqrt{\sum_{k=1}^{K} \varepsilon_k^2}}$.

Similarly, $\varepsilon_k' = -X_{jk}$.

$$-k\left(\frac{\sigma'}{\sigma}\right) = -k\left(\frac{\frac{1}{2\sqrt{k}\sqrt{\sum_{k=1}^{K} \varepsilon_k^2}} \sum_{k=1}^{K} 2\varepsilon_k\left(-X_{jk}\right)}{\sqrt{\frac{1}{k} \sum_{k=1}^{K} \varepsilon_k^2}}\right) = k \frac{\sum_{k=1}^{K} \varepsilon_k X_{jk}}{\sum_{k=1}^{K} \varepsilon_k^2},$$

So,

$$\left(\frac{\lambda\varepsilon_k}{\sigma}\right)' = \frac{\lambda\varepsilon_k'\sigma - \lambda\varepsilon_k\sigma'}{\sigma^2} = \frac{-\lambda\sigma X_{jk} + \lambda\varepsilon_k \frac{\sum_{k=1}^{K}\varepsilon_k X_{jk}}{\sqrt{k}\sqrt{\sum_{k=1}^{K}\varepsilon_k^2}}}{\sigma^2}.$$

Now, first, the partial derivative of $l(\beta, \lambda)$ with respect to β is:

$$\begin{aligned}
\frac{\partial}{\partial\beta_j} l(\beta,\lambda) &= -k\left(\frac{\sigma'}{\sigma}\right) + \sum_{k=1}^{K}\frac{\phi\left(\frac{\lambda\varepsilon_k}{\sigma}\right)}{\Phi\left(\frac{\lambda\varepsilon_k}{\sigma}\right)}\left(\frac{\lambda\varepsilon_k}{\sigma}\right) \\
&= k\frac{\sum_{k=1}^{K}\varepsilon_k X_{jk}}{\sum_{k=1}^{K}\varepsilon_k^2} + \sum_{k=1}^{K}\frac{\phi\left(\frac{\lambda\varepsilon_k}{\sigma}\right)}{\Phi\left(\frac{\lambda\varepsilon_k}{\sigma}\right)}\frac{\lambda\left(-\sigma X_{jk} + \varepsilon_K\frac{\sum_{k=1}^{K}\varepsilon_k X_{jk}}{\sqrt{k}\sqrt{\sum_{k=1}^{K}\varepsilon_k^2}}\right)}{\sigma^2} \\
&= -\frac{\lambda}{\sigma}\sum_{k=1}^{K}\frac{\phi\left(\frac{\lambda\varepsilon_k}{\sigma}\right)}{\Phi\left(\frac{\lambda\varepsilon_k}{\sigma}\right)}X_{jk} + \frac{\sum_{k=1}^{K}\varepsilon_k X_{jk}}{\sigma^2}\left(1 + \frac{\lambda}{k\sigma}\sum_{k=1}^{K}\frac{\phi\left(\frac{\lambda\varepsilon_k}{\sigma}\right)}{\Phi\left(\frac{\lambda\varepsilon_k}{\sigma}\right)}\varepsilon_k\right).
\end{aligned}$$

Second, the partial derivative of $l(\beta, \lambda)$ with respect to λ is:

$$\frac{\partial}{\partial\lambda} l(\beta,\lambda) = \sum_{k=1}^{K}\frac{\phi\left(\frac{\lambda\epsilon_k}{\sigma}\right)}{\Phi\left(\frac{\lambda\epsilon_k}{\sigma}\right)}\frac{\epsilon_k}{\sigma}.$$

Appendix B: Getting Started with Excel Solver for SFA and DEA Analyses

Introducing, Getting and Installing Excel Solver

In this chapter and Chap. 4, this book uses Excel Solver to solve optimization problems such as maximum likelihood estimation for SFA, DEA LP problem, and chance-constrained stochastic programming problem. Although the major PC-based spreadsheets provides built-in optimizers, Excel Solver is considered the most widely used optimization software today in the world because of its simple user interface without a need of knowing that the calculations inside the Excel Solver performs are heavily complex in reality. Excel Solver provides three available solving methods such as Simple LP (Linear Programming) method, GRG (Generalized Reduced Gradient) Nonlinear method, and Evolutionary method. An overview of each solving method is discussed in Fig. 2.6.

This Appendix illustrates three examples: (1) SFA parameters estimation (2) DEA LP problem and (3) traveling compressed air expert problem, with an attempt to use GRG method, Simplex LP method and evolutionary method, respectively.

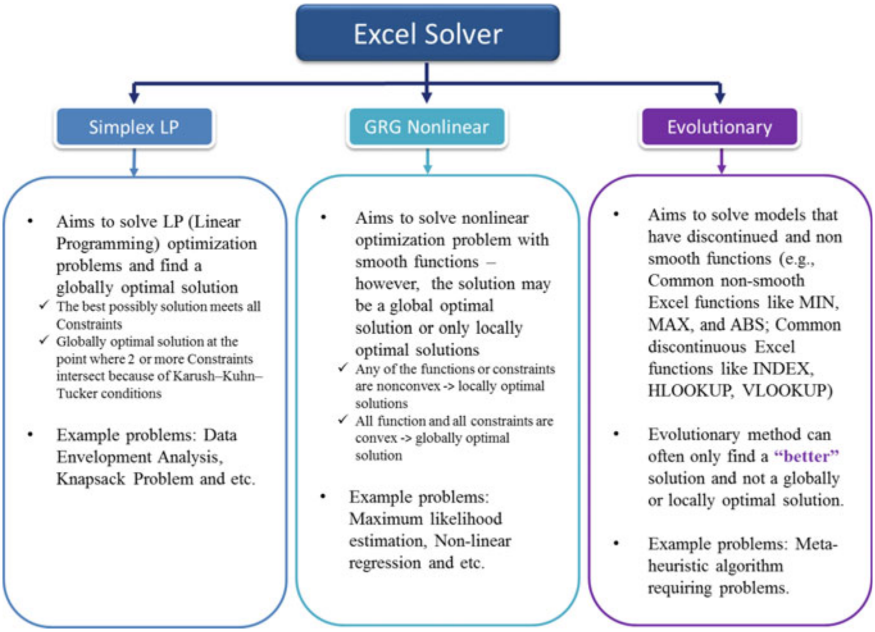


Fig. 2.6 Three methods provided by Excel Solver

However, this Appendix is not an introduction to all of Excel Solver, but to the selected parts of Excel Solver used in the book. For further instructions, see *An Introduction to Spreadsheet Optimization Using Excel Solver*, which is available on http://www.meiss.com/download/Spreadsheet-Optimization-Solver.pdf. Another useful introduction is *Step-By-Step Optimization with Excel Solver*, which is available on http://excelmasterseries.com/D-_Loads/New_Manuals/Step-By-Step_Optimization_S.pdf and provides a detailed introduction for beginners to successfully implement Excel Solver, in particular, by separating a problem solving path into 6 steps such as (1) Step 1—Determine the objective (2) Step 2—Determine the decision variables (3) Step 3—Build the Excel equations that combine the objective with all decision variables (4) Step 4—List all constraints (5) Step 5—Test the Excel Spreadsheet (6) Step 6—Insert all data into the Solver dialog box.

Excel Solver is available in the *Analysis* group on the *Data* tab in Excel as Fig. 2.7. If the *Data* tab does not have the choice *Solver* available, then the following installation procedure should be taken. Note that this guide is based on Microsoft Excel 2013 version.

1. Open *Office Button|Excel Options* to see if the *Add-Ins* option appears
2. If the *Add-Ins* option appears, select *Excel Add-Ins* from the drop down dialog box
3. Click *Go* then *Solver* is enabled.
4. If the *Add-Ins* option does not appear, run the *Setup* program again to install it

Once Excel Solver is installed, it can be used.

Fig. 2.7 Excel Solver in data tab|analysis group

SFA Parameters Estimation Using GRG Method

This example is made to provide an opportunity to learn how to use the generalized reduced gradient (GRG) method of Excel Solver. This method aims to solve nonlinear optimization problem with smooth functions—however, the solution may be a global optimal solution or only locally optimal solutions. In case that any of the functions or constraints are nonconvex, it may obtain locally optimal solutions. If all function and all constraints are convex, it is likely to obtain globally optimal solution. Problems including maximum likelihood estimation or non-linear regression can be solvable with GRG method. This example will use the SFA analysis example discussed in this chapter.

Define Problem, Objective and Decision Variables

This example problem aims to determine and to measure the effectiveness of energy reduction initiatives in terms of a technical improvement that corresponds to a certain structural change in industry-wide energy efficiency between two distinct time periods by proposing a benchmarking model: SFA (stochastic frontier analysis) model based on Hicksian neutral technological change concept.

The proposed SFA model for electricity is:

$$E_i/Y_i = A + \beta_1 WBASE_i + \beta_2 HDD_i + \beta_3 HDD_i^2 + \beta_4 CDD_i + \beta_5 CDD_i^2 + \beta_6 Util_i + \beta_7 Year_i + u_i - v_i$$

where:

- E_i: Total site electricity use at plant i in kWh;
- Y_i: Number of vehicles produced;
- $WBASE_i$: Wheelbase (the distance between its front and rear wheels) of the largest vehicle produced in the plant in inch;
- HDD_i: Thousand heating degree days for the plant location and year;

- HDD_i^2: HDD_i squared;
- CDD_i: Thousand cooling degree days for the plant location and year;
- CDD_i^2: CDD_i squared;
- $Util_i$: Plant utilization rate, defined as output/capacity, where the denominator, capacity is a normalized capacity defined as equal to capacity line rate (or job per hour) × 235 working days × 16 working hours per day;
- $Year_i$: t and $t + \Delta t$ where Δt is the time period at which a significant technical improvement in energy efficiency is observed; and
- β: Vector of parameters to be estimated.

The SFA model above requires several parameters to be estimated, such as β, σ_v^2 and σ_u^2. This work uses the maximum likelihood method for parameter estimation.
The log likelihood function can be expressed as:

$$
\begin{aligned}
l(\beta, \sigma^2, \lambda) = & -\frac{1}{2} N \log\left(\frac{\pi}{2}\right) \\
& -\frac{1}{2} N \log \sigma^2 + \sum_{i=1}^{N} \log \Phi\left(\frac{\lambda(y_i - f(x_i;\beta))}{\sqrt{\sigma^2}}\right) \\
& -\frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - f(x_i;\beta))^2.
\end{aligned}
$$

The objective of this example is to estimate parameters (decision variables) β, σ^2 and λ on the data $(x_1, y_1), \ldots, (x_N, y_N)$ in such a way that maximizes $l(\beta, \sigma^2, \lambda)$ that is the log-likelihood function. The data set are shown below.

	A	B	C	D	E	F	G	H	I	J
1	DMU	Output	Input							
2		kwh/vehicle	Intercept	WheelBase (in)	1000 HDD	$(1000\ HDD)^2$	1000 CDD	$(1000\ CDD)^2$	Util(%)	Year
3		Y	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
4			-174.39	9.35	491.55	-66.06	447.36	-95.38	-205.77	-358.18
5	1	914.64	1	133.50	6.69	44.80	1.22	1.48	1.19	1
6	2	1242.57	1	105.75	6.20	38.48	1.48	2.20	1.27	1
7	3	2098.37	1	155.32	5.17	26.77	2.91	8.44	1.07	1
8	4	1212.36	1	112.01	5.40	29.12	1.71	2.92	1.60	1
9	5	1589.27	1	130.63	3.22	10.37	3.03	9.21	1.78	1
10	6	1336.01	1	133.50	6.47	41.85	1.41	1.99	1.94	1
11	7	1553.32	1	105.87	6.12	37.48	1.43	2.06	1.58	1
12	8	1714.51	1	155.51	5.33	28.44	3.09	9.56	0.77	1
13	9	1548.68	1	112.24	5.85	34.16	2.35	5.52	0.80	1
14	10	1718.41	1	130.63	3.03	9.19	3.43	11.79	1.13	1
15	1	1272.92	1	133.50	5.83	34.01	1.96	3.83	0.73	2
16	2	784.46	1	105.75	5.87	34.46	1.84	3.37	2.09	2
17	3	1950	1	155.32	3.86	14.89	2.82	7.92	1.13	2
18	4	921.72	1	112.01	4.53	20.48	2.52	6.37	2.00	2
19	5	1384.39	1	130.63	2.51	6.31	3.99	15.89	0.50	2
20	6	1008.52	1	133.50	5.87	34.48	1.46	2.13	1.77	2
21	7	789.69	1	105.87	6.17	38.04	1.11	1.24	0.51	2
22	8	1898.7	1	155.51	4.32	18.67	3.23	10.40	0.77	2
23	9	1100.04	1	112.24	4.72	22.24	2.03	4.13	2.22	2
24	10	995.48	1	130.63	2.72	7.38	3.75	14.07	2.29	2
25		SUM	20							

Build Excel Equations Associating the Objective and Decision Variables

Decision variable cells are in cells `C4` to `J4` representing β and `D28` representing λ. The formula of each cell is as follows.

```
>>> K5=SUMPRODUCT($C$4:$J$4,C5:J5)#1st DMU’s best practice level in 1st year.
>>> K6=SUMPRODUCT($C$4:$J$4,C6:J6)#2nd DMU’s best practice level in 1st year.
>>> K7=SUMPRODUCT($C$4:$J$4,C7:J7)#3rd DMU’s best practice level in 1st year.
>>> K8=SUMPRODUCT($C$4:$J$4,C8:J8)#4th DMU’s best practice level in 1st year.
>>> K9=SUMPRODUCT($C$4:$J$4,C9:J9)#5th DMU’s best practice level in 1st year.
>>> K10=SUMPRODUCT($C$4:$J$4,C10:J10)#6th DMU’s best practice level in 1st year.
>>> K11=SUMPRODUCT($C$4:$J$4,C11:J11)#7th DMU’s best practice level in 1st year.
>>> K12=SUMPRODUCT($C$4:$J$4,C12:J12)#8th DMU’s best practice level in 1st year.
>>> K13=SUMPRODUCT($C$4:$J$4,C13:J13)#9th DMU’s best practice level in 1st year.
>>> K14=SUMPRODUCT($C$4:$J$4,C14:J14)#10th DMU’s best practice level in 1st year.
>>> K15=SUMPRODUCT($C$4:$J$4,C15:J15)#1st DMU’s best practice level in 2nd year.
>>> K16=SUMPRODUCT($C$4:$J$4,C16:J16)#2nd DMU’s best practice level in 2nd year.
>>> K17=SUMPRODUCT($C$4:$J$4,C17:J17)#3rd DMU’s best practice level in 2nd year.
>>> K18=SUMPRODUCT($C$4:$J$4,C18:J18)#4th DMU’s best practice level in 2nd year.
>>> K19=SUMPRODUCT($C$4:$J$4,C19:J19)#5th DMU’s best practice level in 2nd year.
>>> K20=SUMPRODUCT($C$4:$J$4,C20:J20)#6th DMU’s best practice level in 2nd year.
>>> K21=SUMPRODUCT($C$4:$J$4,C21:J21)#7th DMU’s best practice level in 2nd year.
>>> K22=SUMPRODUCT($C$4:$J$4,C22:J22)#8th DMU’s best practice level in 2nd year.
>>> K23=SUMPRODUCT($C$4:$J$4,C23:J23)#9th DMU’s best practice level in 2nd year.
>>> K24=SUMPRODUCT($C$4:$J$4,C24:J24)#10th DMU’s best practice level in 2nd year.
```

Similar to formulas in Column `K`, columns `L`, `M`, `L`, `O`, and `P` contains pertinent formulas as follows:

- `L`: $\varepsilon = y_i - f(x_i;\beta)$
- `M`: ε^2
- `N`: $\log \Phi\left(\frac{\lambda(y_i - f(x_i;\beta))}{\sqrt{\sigma^2}}\right)$ where σ^2 replaced with $\frac{1}{N}\sum_{i=1}^{N}(y_i - f(x_i;\beta))^2$
- `O`: $u = y_i - f(x_i;\beta)$
- `P`: $EPI = probability(energy\ inefficiency \geq E_i/Y_i - f(X;\beta) + v_i)$.
  $= 1 - F(E_i/Y_i - f(X;\beta) + v_i)$

	A	K	L	M	N	O	P
1	DMU	Estimates					
2–4		$f(x_K, \beta)$	ε	ε^2	LN(Φ(ελ/σ))	u	EPI(Energy Performance Indicator)
5	1	1204.13	-289.49	83803.87	-3.08	-289.49	198%
6	2	1155.33	87.24	7611.65	-0.36	87.24	47%
7	3	1968.00	130.37	16996.95	-0.25	130.37	28%
8	4	1399.87	-187.51	35159.01	-1.99	-187.51	188%
9	5	1699.03	-109.76	12046.50	-1.34	-109.76	163%
10	6	1171.90	164.11	26930.65	-0.19	164.11	18%
11	7	1110.14	443.18	196404.92	0.00	443.18	0%
12	8	1976.54	-262.03	68661.84	-2.76	-262.03	197%
13	9	1492.99	55.69	3101.21	-0.47	55.69	65%
14	10	1750.21	-31.80	1011.26	-0.85	-31.80	121%
15	1	1336.53	-63.61	4046.11	-1.03	-63.61	140%
16	2	776.32	8.14	66.22	-0.66	8.14	95%
17	3	1744.75	205.25	42128.51	-0.12	205.25	9%
18	4	1137.41	-215.69	46524.00	-2.26	-215.69	192%
19	5	1312.60	71.79	5153.74	-0.41	71.79	55%
20	6	1051.90	-43.38	1881.66	-0.92	-43.38	128%
21	7	892.31	-102.62	10531.77	-1.29	-102.62	160%
22	8	1744.32	154.38	23834.33	-0.20	154.38	20%
23	9	1066.51	33.53	1124.18	-0.55	33.53	78%
24	10	1043.27	-47.79	2283.50	-0.94	-47.79	131%
25		SUM	0.00	589301.87	-19.68		

Other cells used to estimate decision variables are as follows:

```
>>> D28=1 #λ, the initial value is set to 1.
>>> E24=M25/(C25) # σ²
>>> F24=-(1/2)*C25*LN(PI()/2)-(1/2)*C25*LN(E28)
>>>   +N25-(1/2)*(M25/E28)     # l(β,σ²,λ)
>>> G24=(D28^2/(1+D28^2))*E28 # σ_u²
>>> H24=E28-G28    # σ_v²
>>> I24=-T16/(2*I30)-(S11/2)*LN(2*PI())-
>>>    (S11/2)*LN(I30)    # OLS
>>> J24=F28 # MLE
>>> K24=-2*(I28-J28)    # LR
>>> I30=T16/S11    #σ²
```

	C	D	E	F	G	H	I	J	K	L
26										
27		λ	σ^2	l(β,λ,σ)	σ_u^2	σ_v^2	OLS	MLE	LR=-2(OLS-MLE)	
28		1	29465.09	-137.1053419	14732.55	14732.5468	-131.288	-137.11	-11.63391097	
29					121.38	121.38	σ^2 (OLS)			
30							29465.09			
31										

Find Initial Decision Variables Using OLS (Ordinary Least Square)

	Q	R	S	T	U	V	W	X	Y	Z
4		SUMMARY OUTPUT								
5										
6		*Regression Statistics*								
7		Multiple R	0.895573461							
8		R Square	0.802051825							
9		Adjusted R Square	0.686582056							
10		Standard Error	221.6043531							
11		Observations	20							
12										
13		ANOVA								
14			*df*	*SS*	*MS*	*F*	*gnificance F*			
15		Regression	7	[illegible]	341107.1	6.94599	0.001887			
16		Residual	12	589301.9	49108.49					
17		Total	19	2977051						
18										
19			*Coefficients*	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
20		Intercept	-174.3869921	1184.868	[illegible]	0.885435	-2755.99	2407.219	-2755.99	2407.219
21		X Variable 1	9.345086254	3.710522	[illegible]	0.026982	1.260554	17.42962	1.260554	17.42962
22		X Variable 2	491.5494671	[illegible]	0.69119	0.502599	-1057.94	2041.043	-1057.94	2041.043
23		X Variable 3	-66.06286442	[illegible]	-0.87534	0.39857	-230.5	98.37424	-230.5	98.37424
24		X Variable 4	447.3627186	796.0417	0.561984	0.584471	-1287.06	2181.789	-1287.06	2181.789
25		X Variable 5	-95.38169061	164.6779	-0.5792	0.573166	-454.184	263.4207	-454.184	263.4207
26		X Variable 6	-205.7693536	106.545	-1.93129	0.077408	-437.911	26.37216	-437.911	26.37216
27		X Variable 7	-358.1837952	117.4608	-3.04939	0.010096	-614.109	-102.259	-614.109	-102.259

Total observation

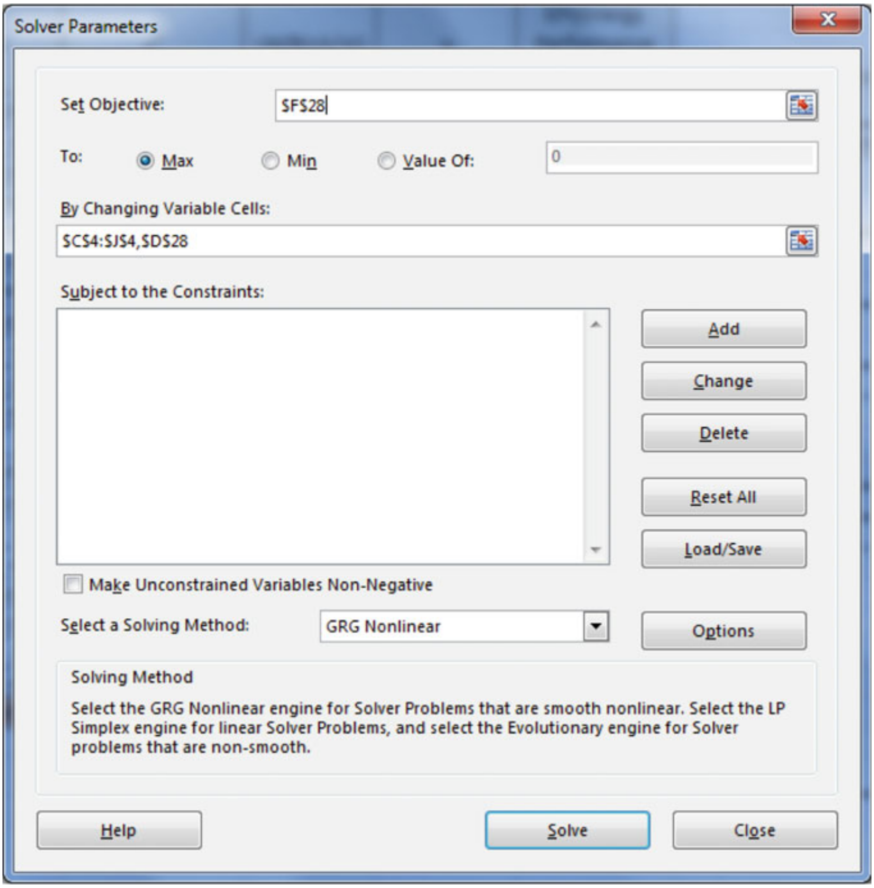
SSE

Initial values of coefficients

Decision variable cells require initial values. Cells C4 to J4 representing β take values found by ordinary linear regression as their initial values while and D28 representing λ is set to 1 as its initial value. Therefore, the formula of each cell is as follows.

```
>>> C4=S20 #initial value for intercept.
>>> D4=S21 #initial value for WheelBase(in).
>>> E4=S22 #initial value for 1000 HDD.
>>> F4=S23 #initial value for (1000 HDD)².
>>> G4=S24 #initial value for 1000 CDD.
>>> H4=S25 #initial value for (1000 CDD)².
>>> I4=S25 #initial value for Util(%).
>>> J4=S26 #initial value for Year.
>>> D28=1 #λ, the initial value is set to 1.
```

Insert All Data into Excel Solver Box



The objective is set to `F28` representing $l(\beta, \sigma^2, \lambda)$ with a maximization option chosen and variable cells are set to `C4` to `J4` representing β and `D28` representing λ. This problem does not need any constraint. Note that s solving method is set to `GRG Nonlinear`.

Understand the Results from Excel Solver

Once the Solve button in the previous Excel Solver Box is hit, the solution appears as shown below. Cells from `C4` to `J4` shows the final parameters for β.

	A	B	C	D	E	F	G	H	I	J
1	DMU	Output	Input							
2		kwh/vehicle	Intercept	WheelBase (in)	1000 HDD	(1000 HDD)^2	1000 CDD	(1000 CDD)^2	Util(%)	Year
3		Y	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
4			636.92	8.39	397.44	-68.83	405.06	-107.60	-331.47	-359.36

Note that the one-sided likelihood-ratio test value (LR) in cell K28 for this model reveals that the model is adequate at the 99.5 % significance level ($8.17 > \chi^2_{1-2\times0.005}(1) = 6.635$) and that the model has very little error attributable to random noise, with most departures attributable to inefficiency. Therefore, the null-hypothesis, $H_0{:}\gamma = \frac{\sigma_u^2}{(\sigma_v^2+\sigma_u^2)} = 0$, is rejected, and the alternative hypothesis $H_1{:}\gamma > 0$ with technical inefficiency effect is accepted for this model.

	C	D	E	F	G	H	I	J	K	L
26										
27		λ	σ^2	l(β,λ,σ)	σ$_u{}^2$	σ$_v{}^2$	OLS	MLE	LR=-2(OLS-MLE)	
28		871.5364051	78017.57	-127.1677954	78017.47	0.102711906	-131.288	-127.17	8.241182104	
29					279.32	0.32	σ^2 (OLS)			
30							29465.09			
31										

The GRG method generates Answer Report. The Answer Report tells an important information that is how long Solver took to solve the problem. In this case, the total run time was just 0.03 s.

	A–K
1	**Microsoft Excel 15.0 Answer Report**
2	**Worksheet: [GRG (SFA).xlsx]SFA**
3	**Report Created: 11/9/2015 8:39:12 AM**
4	**Result: Solver found a solution. All Constraints and optimality conditions are satisfied.**
5	**Solver Engine**
6	Engine: GRG Nonlinear
7	Solution Time: 0.031 Seconds.
8	Iterations: 4 Subproblems: 0
9	**Solver Options**
10	Max Time Unlimited, Iterations Unlimited, Precision 0.000001, Use Automatic Scaling
11	Convergence 0.0001, Population Size 100, Random Seed 0, Derivatives Forward, Require Bounds
12	Max Subproblems Unlimited, Max Integer Sols Unlimited, Integer Tolerance 1%

The Answer Report also provides the information about object, variable cells and constraints as below:

Objective Cell (Max)

Cell	Name	Original Value	Final Value
$F$2	l(β,λ,σ)	-127.1678196	-127.1677954

Variable Cells

Cell	Name	Original Value	Final Value	Integer
$C$4	β_0	636.92	636.92	Contin
$D$4	β_1	8.39	8.39	Contin
$E$4	β_2	397.44	397.44	Contin
$F$4	β_3	-68.83	-68.83	Contin
$G$4	β_4	405.06	405.06	Contin
$H$4	β_5	-107.60	-107.60	Contin
$I$4	β_6	-331.47	-331.47	Contin
$J$4	β_7	-359.36	-359.36	Contin
$D$2	λ	871.5362262	871.5364051	Contin

Constraints

NONE

DEA LP Problem Solving Using Simplex LP

This example is made to provide an opportunity to learn how to use the Simplex method of Excel Solver. This method aims to solve LP (Linear Programming) optimization problems and find a globally optimal solution. The best possibly solution meets all constraints globally to be an optimal solution at the point where 2 or more Constraints intersect because of Karush–Kuhn–Tucker conditions. Data Envelopment Analysis and Knapsack Problem can be solvable with Simplex method. This example will use the DEA example discussed in this chapter.

Define Problem, Objective and Decision Variables

This example problem aims to calculate the efficiency of the t-th time period relative to t-th time period technology, that is, $D_I^t(Z_t, E_t/Y_t)$, which is identical to the following DEA model

$$
\begin{aligned}
& D_I^t(Z_t, E_t/Y_t) = \min_{\phi,\lambda}\phi \\
s.t. \quad & -Z_{it} + Z_t\lambda \geq 0, \\
& -\phi(E_{it}/Y_{it}) + (E_t/Y_t)\lambda \leq 0 \\
& \lambda \geq 0
\end{aligned}
$$

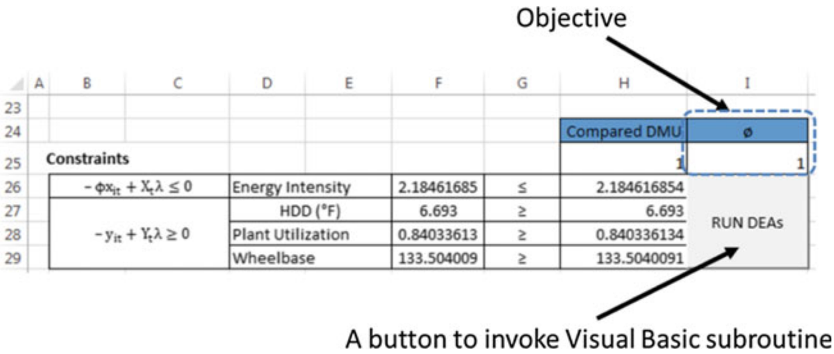
Note that this LP is input-oriented and CRS (Constant Return Scale) is assumed. The dataset for this example is shown below. 10 DMUs (decision making units), each with input (energy intensity) and outputs (HDD, plant utilization, wheelbase) are provided.

	A	B	C	D	E	F
1						
2		DMU	Input	Outputs		
3			Energy Intensity	HDD (°F)	Plant Utilization	Wheelbase
4		1	2.184616854	6.693	0.84	133.504009
5		2	2.967877378	6.203	0.79	105.748089
6		3	5.011954943	5.174	0.93	155.315045
7		4	2.895720819	5.396	0.63	112.007935
8		5	3.79597003	3.22	0.56	130.629992
9		6	3.191058737	6.469	0.51	133.504009
10		7	3.710103486	6.122	0.63	105.866199
11		8	4.095105662	5.333	1.30	155.511895
12		9	3.69902085	5.845	1.25	112.244155
13		10	4.104420809	3.031	0.88	130.629992

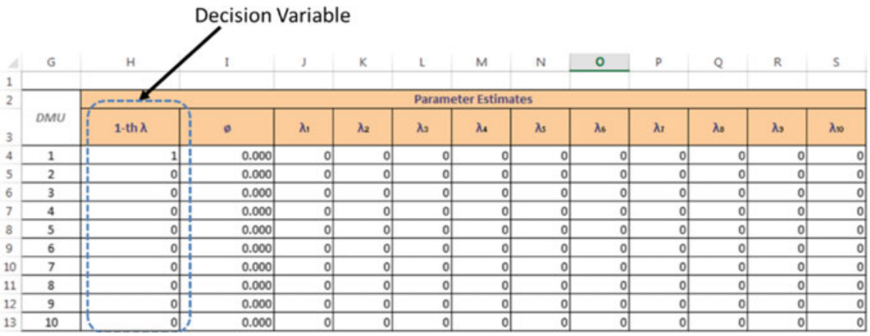
Build Excel Equations Associating the Objective and Decision Variables

The objective of this example is to determine λ and ϕ in such a way to minimize ϕ (technical efficiency). The objective cell is `H25` represenging ϕ. The decision variable cells are from `H4` to `H13` representing λ. Note that this constraint cells use the Excel `INDEX` function which is used to locate the data which corresponds to the pertinent DMU that appears in the column `H25`. The `INDEX` function has a formula, that is = `INDEX (range, row number, column number)`.

```
>>> # $-\phi\left(E_{it}/Y_{it}\right)+\left(E_t/Y_t\right)\lambda \le 0$: Energy Intensity (input)
>>> F26=SUMPRODUCT($C$4:$C$13,$H$4:$H$13)
>>> H26=$I$25*INDEX(C4:C13,H25,1)
>>>
>>> # $-Z_{it}+Z_t\lambda \ge 0$ : HDD (°F) (output)
>>> F27=SUMPRODUCT($D$4:$D$13,$H$4:$H$13)
>>> H27=INDEX(D4:D13,H25,1)
>>>
>>> # $-Z_{it}+Z_t\lambda \ge 0$ : Plant Utilization (output)
>>> F28=SUMPRODUCT($E$4:$E$13,$H$4:$H$13)
>>> H28=INDEX(E4:E13,H25,1)
>>>
>>> # $-Z_{it}+Z_t\lambda \ge 0$ : Wheelbase (output)
>>> F29=SUMPRODUCT($F$4:$F$13,$H$4:$H$13)
>>> H29=INDEX(F4:F13,H25,1)
```



	B	C	D	E	F	G	H	I
23								
24							Compared DMU	ø
25	Constraints						1	1
26	$-\phi x_{it} + X_t\lambda \le 0$		Energy Intensity		2.18461685	≤	2.184616854	RUN DEAs
27	$-y_{it} + Y_t\lambda \ge 0$		HDD (°F)		6.693	≥	6.693	
28			Plant Utilization		0.84033613	≥	0.840336134	
29			Wheelbase		133.504009	≥	133.5040091	



DMU	1-th λ	ø	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
	Parameter Estimates											
1	1	0.000	0	0	0	0	0	0	0	0	0	0
2	0	0.000	0	0	0	0	0	0	0	0	0	0
3	0	0.000	0	0	0	0	0	0	0	0	0	0
4	0	0.000	0	0	0	0	0	0	0	0	0	0
5	0	0.000	0	0	0	0	0	0	0	0	0	0
6	0	0.000	0	0	0	0	0	0	0	0	0	0
7	0	0.000	0	0	0	0	0	0	0	0	0	0
8	0	0.000	0	0	0	0	0	0	0	0	0	0
9	0	0.000	0	0	0	0	0	0	0	0	0	0
10	0	0.000	0	0	0	0	0	0	0	0	0	0

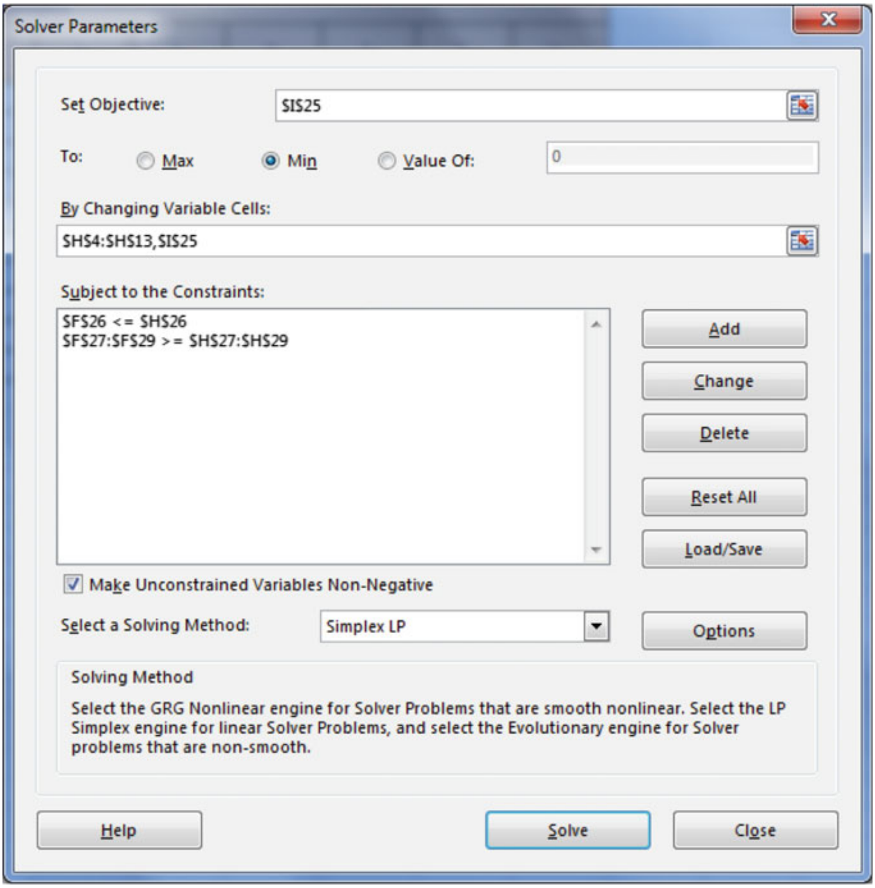
Insert All Data into Excel Solver Box

This example uses VBA in Excel and automates iterations for solving multiple linear programing models with the Simplex method. Briefly, with respect to automation logic, the tool uses "For" loop to automate iterations of solving multiple linear programing models in which Excel Solver with the "Simplex" optimization option calculates the efficiency for each DMU and the results are recorded in a table using the copy/paste function (note: the three major functions used in the loop statement of the VBA programing are as follows: (1) "SolverOk"—defines the objective function and the decision variables; (2) "SolverAdd"—defines model constraints; and (3) "SolverSolv"—runs Solver). The detailed codes are shown below:

```
>>> Sub CommandButton_Click()
>>> Call DEA
>>> End Sub
>>>
>>> Sub DEA()
>>> Dim nDMUs As Long
>>>
>>> nDMUs = Application.InputBox(Prompt:="How many
>>> plants (DMU's) do you want to evaluate?", Type:=1)
>>>
>>> ‘Declare DMU as Integer, DMU iterations 1 through
>>> nDMUs
>>> Dim DMU As Long
>>> For DMU = 1 To nDMUs
>>>
>>> 'Sets the Value of E12 to the DMU Under Evaluation
>>> Range("H25") = DMU
>>>
>>> 'Calls SolverRun Sub
>>> Call SolverRun
>>>
>>> 'Place efficiency value into column J
>>> Range("I" & DMU + 3) = Range("I25")
>>>
>>> 'Select the cells containing the optimal Lambdas
>>> Range("H4:H13").Select
>>>
>>> 'Copy selected lambas and place them in row "DMU +
>>> 2"
>>> Selection.Copy
>>> Range("J" & DMU + 3).Select
>>> Selection.PasteSpecial Paste:=xlPasteValues,
>>> Transpose:=True
>>>
>>> Next DMU
>>> End Sub
>>>
>>> Sub SolverRun()
>>> 'This sub sets up the objective, changing varia-
bles, >>> and constraints of solver
>>>
```

```
>>> 'Reset Solver
>>> SolverReset
>>> 'Defines optimization cell and changing variables
>>>     SolverOk     SetCell:="$I$25",     MaxMinVal:=2,
ValueOf:=0,  >>>  ByChange:=  "$H$4:$H$13,  $I$25",  En-
gine:=2,
>>> EngineDesc:="Simplex LP"
>>> 'Defines the constraints of the model, relation 1
is >>> <=, relation 2 is =, relation 3 is >=
>>> SolverAdd CellRef:="$F$26", Relation:=1, Formu
>>> laText:="$H$26"
>>> SolverAdd CellRef:="$F$27:$F$29", Relation:=3, Fo
>>> rmulaText:="$H$27:$H$29"
>>>
>>> 'Runs Solver model, UserFinish prevents Solver
>>> results dialog box from appearing
>>> SolverSolve UserFinish:=True
>>> End Sub
```

Whenever `SolverRun()` is called, the solver toolbox is filled and run invisibly.



Understand the Results from Excel Solver

Once the `RUN DEAs` button in the spreadsheet is hit, the solution appears as shown below. Cells from `I4` to `I13` shows each DMU's technical efficiency representing ϕ. Meanwhile, cells from `J4` to `S4` shows λ when 1st DMU is considered as a target comparing DMU. Similarly, cells from `J` to `S` associating with each row from 5 to 13 rows shows λ for each DMU.

	G	H	I	J	K	L	M	N	O	P	Q	R	S
1													
2	DMU	Parameter Estimates											
3		10-th λ	ø	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
4	1	1.052671984	1.000	1	0	0	0	0	0	0	0	0	0
5	2	0	0.690	0.937008	0	0	0	0	0	0	0	0	0
6	3	0	0.507	1.163374	0	0	0	0	0	0	0	0	0
7	4	0	0.633	0.838986	0	0	0	0	0	0	0	0	0
8	5	0	0.563	0.978472	0	0	0	0	0	0	0	0	0
9	6	0	0.685	1	0	0	0	0	0	0	0	0	0
10	7	0	0.539	0.914687	0	0	0	0	0	0	0	0	0
11	8	0	0.826	1.54923	0	0	0	0	0	0	0	0	0
12	9	0	0.881	1.492535	0	0	0	0	0	0	0	0	0
13	10	0	0.560	1.052672	0	0	0	0	0	0	0	0	0

Traveling Compressed Air Expert Problem Using Evolutionary Method

This example is made to provide an opportunity to learn how to use the evolutionary method of Excel Solver. The evolutionary method is used the objective contains any cells holding non-smooth or discontinuous formulas. Excel functions such as `INDEX`, `LOOKUP` are common discontinuous functions while `MIN`, `MAX` and `ABS` are common non-smooth Excel functions. This example is modified from the traveling salesman problem located in *Step-By-Step Optimization with Excel Solver* in the context.

Define Problem, Objective and Decision Variables

Assume that a compressed air expert must visit 5 GM engine plants located in US and Canada in order to audit plant's compressed air use practice. He must pick the shortest path that will reach every plants and bring him back to his starting point. In this example, the evolutionary method is used because the objective contains `INDEX` and `LOOKUP` Excel functions, which are discontinuous functions. In addition, this example illustrates how to use `Alldifferent` constrain when solving the problem.

	A	B	C	D	E	F	G	H
1		Distance Chart (mile)						
2			Flint	Spring Hill	St. Catherines	Romulus	Tonawanda	
3		Flint	0	610	244	75	268	
4		Spring Hill	610	0	755	557	753	
5		St. Catherines	244	755	0	257	27	
6		Romulus	75	557	257	0	281	
7		Tonawanda	268	753	27	281	0	
8								

The problem in this example is formally defined as follows: a compressed air expert must make stops in 5 cities where GM engine plants are located: Flint (MI, USA), Spring Hill (TN, USA), St. Catherines (ON, Canada), Romulus (MI, USA), Tonawanda (NY, USA) in such as way that the total length of the trip is minimized. See the distance chart above to refer to the distance between cities in mile.

In this case, the objective is to minimize the total distance travelled when traveling between all 5 cities. The evolutionary method is used to minimize the objective. The decision variables are the order of cities to visit. To specify the order, each city is designated by the row that they appear in the distance chart. For example, Flint appears in the 1st row of the distance chart and therefore, Flint is designated with a "1". Similar to Flint, other cities will be designated by its pertinent row number. At the end, Excel solver will determine the order of cities to visit to minimize the total miles travelled.

Build Excel Equations Associating the Objective and Decision Variables

The decision variables are in cells `B11` to `B15`. The order of the decision variables shown below `(1, 2, 3, 4, 5)` indicates that the expert will visit the cities in this order: Flint (row 1 in the distance chart) → Spring Hill (row 2 in the distance chart) → St. Catherines (row 3 in the distance chart) → Romulus (row 4 in the distance chart) → Tonawanda (row 5 in the distance chart) → Flint.

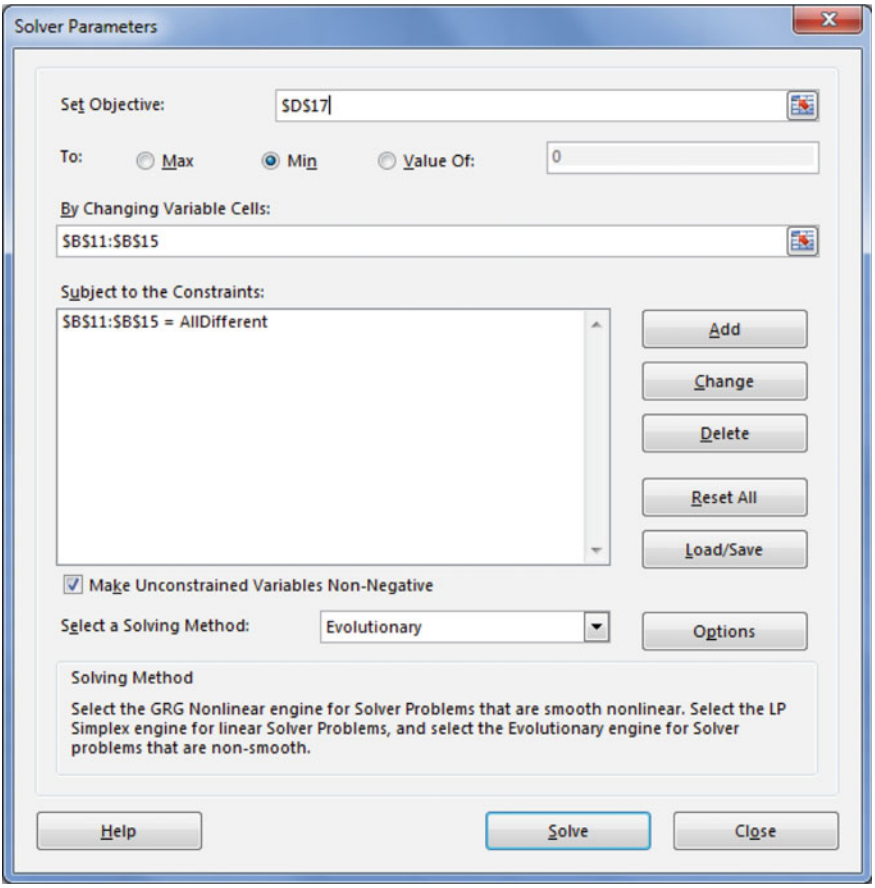
The objective is in cell `D17` where `D17 = SUM(D11:D15)`.

	A	B	C	D	E
9		Decision variables & Objective			
10		visiting order (decision variables)	city	distance (mile)	
11		1	Flint	268	
12		2	Spring Hill	610	
13		3	St. Catherines	755	
14		4	Romulus	257	
15		5	Tonawanda	281	
16				total (objective)	
17				2171	
18					

The formula of each cell is as follows. Note that the Excel `INDEX` function is used to locate the city which corresponds to the Distance Chart row number that appears in the column `B`. The `INDEX` function has a formula, that is `= INDEX (range, row number, column number)`.

```
>>> C11=INDEX($B$3:$B$7,B11,1) #1st visit city name
>>> C12=INDEX($B$3:$B$7,B12,1) #2nd visit city name
>>> C13=INDEX($B$3:$B$7,B13,1) #3rd visit city name
>>> C14=INDEX($B$3:$B$7,B14,1) #4th visit city name
>>> C15=INDEX($B$3:$B$7,B15,1) #5th visit city name
>>>
>>> # Distance between 5th visit city and 1st visit
city
>>> D11=INDEX($C$3:$G$7,B11,B15)
>>>
>>> # Distance between 2nd visit city and 1st visit
city
>>> D12=INDEX($C$3:$G$7,B11,B12)
>>>
>>> # Distance between 3rd visit city and 2nd visit
city
>>> D13=INDEX($C$3:$G$7,B12,B13)
>>>
>>> # Distance between 4th visit city and 3rd visit
city
>>> D14=INDEX($C$3:$G$7,B13,B14)
>>>
>>> # Distance between 5th visit city and 4th visit
city
>>> D15=INDEX($C$3:$G$7,B14,B15)
>>>
>>> # Objective cell
>>> D17=SUM(D11:D15)
```

Insert All Data into Excel Solver Box



This problem uses `Alldifferent` constraint because the problem requires that the expert must visit each city only once without repeating. `Alldifferent` constraint ensures that each city will be visited only once and that all cities will be visited by grouping `B11:B15` cells simultaneously in which the five cells hold the integers 1–5 and no 2 cells in this group will be assigned the same number. Excel Solver Box shown above is set to have `D17` as objective cell and `B11:B15` as decision variables cells. Note that the evolutionary method is set to its solving method.

Understand the Results from Excel Solver

Decision variables & Objective

visiting order (decision variables)	city	distance (mile)
2	Spring Hill	753
4	Romulus	557
1	Flint	75
3	St. Catherines	244
5	Tonawanda	27
		total (objective)
		1656

Once the Solve button in the previous Excel Solver Box is hit, the solution appears as shown above. The solution could be interpreted as follows: The compressed air experts starts in Spring Hill. He then visits Flint, St. Catherine, Tonawanda, and finally back to Spring Hill in that order. The total miles travelled on this route are 1,656 miles. This is the shortest route that will cover all 5 cities starting and ending in Spring Hill. This solution improves the original solution by 23.7 % (=(2,171 − 1,656)/2,171).

Microsoft Excel 15.0 Answer Report
Worksheet: [Evolutionary.xlsx]Sheet1
Report Created: 11/6/2015 1:24:54 PM
Result: Solver cannot improve the current solution. All Constraints are satisfied.
Solver Engine
Engine: Evolutionary
Solution Time: 57.019 Seconds.
Iterations: 0 Subproblems: 165249
Solver Options
Max Time Unlimited, Iterations Unlimited, Precision 0.000001, Use Automatic Scaling
Convergence 0.0001, Population Size 100, Random Seed 0, Mutation Rate 0.075, Time w/o Improve 30 sec, Require Bounds
Max Subproblems Unlimited, Max Integer Sols Unlimited, Integer Tolerance 1%, Assume NonNegative

Objective Cell (Min)

Cell	Name	Original Value	Final Value
$D$17	total (objective)	2171	1656

The evolutionary method generates two reports: Answer Report and Population Report. The Answer Report tells an important information that is how long Solver took to solve the problem. Especially it is important because the evolutionary method can controlled by the Option settings where the options such as the maximum allowable run time, iterations, or subproblems are available for control. In this case, the total run time was 57 s.

	A	B	C	D	E	F
19	Variable Cells					
20		Cell	Name	Original Value	Final Value	Integer
21		$B$11	visiting order (decision variables)	1	2	AllDiff
22		$B$12	visiting order (decision variables)	2	4	AllDiff
23		$B$13	visiting order (decision variables)	3	1	AllDiff
24		$B$14	visiting order (decision variables)	4	3	AllDiff
25		$B$15	visiting order (decision variables)	5	5	AllDiff
26						
27						
28	Constraints					
29			NONE			
30		$B$11:$B$15=AllDiff				

The Answer Report also provides the information about variable cells and constraints. Note the difference of decision variables between their original and final values. Also, note that the unique `Alldifferent` constraint is bound meaning that no slack is still available.

	A	B	C	D	E	F	G	H
1	Microsoft Excel 15.0 Population Report							
2	Worksheet: [Evolutionary.xlsx]Sheet1							
3	Report Created: 11/6/2015 1:24:54 PM							
4								
5								
6	Variable Cells							
7				Best	Mean	Standard	Maximum	Minimum
8		Cell	Name	Value	Value	Deviation	Value	Value
9		$B$11	visiting order (decision variables)	2	3.234693878	1.353125594	5	1
10		$B$12	visiting order (decision variables)	4	2.489795918	1.451944292	5	1
11		$B$13	visiting order (decision variables)	1	3.489795918	1.159851554	5	1
12		$B$14	visiting order (decision variables)	3	2.693877551	1.501944878	5	1
13		$B$15	visiting order (decision variables)	5	3.091836735	1.377931624	5	1
14								
15	Constraints							
16		NONE						

The Population Report gives useful information about the entire population of candidate solutions maintained by the Evolutionary Solving method at the end of the solution process. With the Population Report, a modeler gets some insight into the performance of the Evolutionary method and can decide whether additional runs of the Evolutionary method are likely to yield even better solutions. For each variable and constraint, the Population Report shows the best value found by the Evolutionary method, and the mean (average) value, standard deviation, maximum value, and minimum value of that variable or constraint across the entire population of candidate solutions at the end of the solution process as shown above. These values gives an idea of the diversity of solutions represented by the population. From the sense, the way of interpreting the Population Report is important. For example, if the Best Values are similar from run to run (i.e., the Standard Deviations are small), this may be reason for the high confidence that the final solution is close to the global optimum. However, if the Best Values vary from run to run (i.e., the Standard Deviations are large) might indicate a lack of diversity in the population, suggesting that the modeler should increase the Mutation Rate and run Excel Solver again.

References

Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. J Econ 6:21–37

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manage Sci 30:1078–1092

Battese GE, Corra GS (1977) Estimation of a production frontier model: with application to the pastoral zone of eastern Australia. Aust J Agric Econ 21:169–179

Battese GE, Coelli TJ (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20:325–332

Bogetoft P, Otto L (2011) Benchmarking with DEA, SFA, and R. Springer, Berlin

Boyd GA (2005) Development of a performance-based industrial energy efficiency indicator for automobile assembly plants. Technical report ANL/DIS-05-3. Argonne National Laboratory, DuPage County

Boyd GA (2008) Estimating plant level energy efficiency with a stochastic frontier. Energy J 29:23–43

Boyd GA (2014) Estimating the changes in the distribution of energy efficiency in the U.S. automobile assembly industry. Energy Econ 42:81–87

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making unit. Eur J Oper Res 2:429–444

Coelli TS, Prasada Rao DS, O'Donnell CJ, Battese GE (2005) An introduction to efficiency and productivity analysis, 2nd edn. Springer, Berlin

Färe R, Grosskopt S, Margaritis D (2011) Malmquist productivity indexes and DEA. In: Cooper WW, Seiford LM, Zhe J (eds) Handbook of data envelopment analysis, vol 164, 2nd edn. Springer, Berlin, pp 127–149

Galitsky C, Worrell E (2008) Energy efficiency improvement and cost saving opportunities for the vehicle assembly industry. Lawrence Berkley National Laboratory (LBNL), Orlando

Jurek P, Bras B, Guldberg T, D'Arcy JB, Oh S-C, Biller SR (2012) ABC applied to automotive manufacturing. In: Proceedings of the IEEE power and energy society general meeting, San Diego, CA, USA, 22–26 July

Lin L-C, Tseng L-A (2005) Application of DEA and SFA on the measurement of operating efficiencies for 27 international container ports. In: Proceedings of the Eastern Asia society for transportation studies, Bangkok, Thailand, 21–24

Meeusen W, van den Breoek J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. Int Econ Rev 18:435–444

Nin A, Arndt C, Hertel TW, Preckel PV (2003) Bridging the gap between partial and total factor productivity measures using directional distance functions. Am J Agric Econ 85:928–942

Oh S-C, Hidreth AJ (2013) Decisions on energy demand response option contracts in smart grids based on activity-based costing and stochastic programming. Energies 6:425–443

Oh S-C, Hidreth AJ (2014) Estimating the technical improvement of energy efficiency in the automotive industry—stochastic and deterministic frontier benchmarking approaches. Energies 9:6198–6222

Oh S-C, D'Arcy JB, Arinez JF, Biller SR, Hidreth AJ (2011) Assessment of energy demand response options in smart grid utilizing the stochastic programming approach. In: Proceedings of the IEEE power and energy society general meeting, Detroit, MI, USA, 24–28 July

Productivity Commission (2013) Electricity network regulatory frameworks; inquiry report, 1(62), Australian Government, Productivity Commission, Melbourne, Australia
Sullivan JL, Burnham A, Wang MQ (2010) Energy and carbon emissions analysis of vehicle manufacturing and assembly. Technical report ANL/ESD 10–6. Argonne National Laboratory, DuPage County
Daraio C (2012) The nonparametric approach in efficiency analysis: recent developments and applications. Available online: http://www.siepi.univpm.it/sites/www.siepi.univpm.it/files/siepi/SIEPI%202012/papers/Daraio.pdf. Accessed 22 Sept 2014)
Toshiyuki S, Mika G (2012) Weak and strong disposability vs. natural and managerial disposability in DEA environmental assessment: comparison between Japanese electric power industry and manufacturing industries. Energy Econ 34(3):686–699
Yee JT, Oh S-C (2012) Technology integration to business. Springer, Berlin

Springer

http://www.springer.com/978-3-319-32728-0

Analytics for Smart Energy Management
Tools and Applications for Sustainable Manufacturing
Oh, S.-C.; Hildreth, A.J.
2016, XI, 295 p. 117 illus., 96 illus. in color., Hardcover
ISBN: 978-3-319-32728-0