RICE UNIVERSITY

Essays on the Use of Duality, Robust Empirical Methods, Panel Treatments, and Model Averaging with Applications to Housing Price Index Construction and World Productivity Growth

by

Chenjun Shang

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Robin C. Sickles Chair
Reginald Henry Hargrove Chair of Economics
Rice University

Mahmoud El-Gamal
Professor of Economics
Chair in Islamic Economics, Finance, and Management
Rice University

Katherine B. Ensor
Professor of Statistics
Rice University

Houston, Texas
April, 2015

RICE UNIVERSITY

Essays on the Use of Duality, Robust Empirical Methods, Panel Treatments, and Model Averaging with Applications to Housing Price Index Construction and World Productivity Growth

by

Chenjun Shang

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Robin C. Sickles

Robin C. Sickles Chair
Reginald Henry Hargrove Chair of Economics
Rice University

Mahmoud El-Gamal

Mahmoud El-Gamal
Professor of Economics
Chair in Islamic Economics, Finance, and Management
Rice University

Katherine B. Ensor

Katherine B. Ensor
Professor of Statistics
Rice University

Houston, Texas
April, 2015

RICE UNIVERSITY

Essays on the Use of Duality, Robust Empirical Methods, Panel Treatments, and Model Averaging with Applications to Housing Price Index Construction and World Productivity Growth

by

Chenjun Shang

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Approved, Thesis Committee:

Robin C. Sickles Chair
Reginald Henry Hargrove Chair of Economics
Rice University

Mahmoud El-Gamal
Professor of Economics
Chair in Islamic Economics, Finance, and Management
Rice University

Katherine B. Ensor
Professor of Statistics
Rice University

Houston, Texas
April, 2015

ABSTRACT

Essays on the Use of Duality, Robust Empirical Methods, Panel Treatments, and Model Averaging with Applications to Housing Price Index Construction and World Productivity Growth

by

Chenjun Shang

This dissertation focuses on analyzing the production side of the economy, and aims to provide robust estimates of the parameters of interest. In a production process, the output level is mainly determined by two parts: inputs and productivity. Compared with the inputs, which are concrete and measurable, productivity is an unobservable factor that relies on economic models for estimation. An appropriate and robust modeling method is essential if we want to accurately capture the productivity term. Chapter 1 reviews the research on productivity with a focus on stochastic frontier analysis, which is a classic framework in the productivity literature. This chapter starts with the definition and decomposition of productivity. Measured as a ratio of the outputs to the inputs, productivity can be divided into two main parts: innovations and technical efficiencies. The growth of technologies and innovations depends heavily on education and research, and the technical efficiencies of firms vary with their administration, management skills, and allocation of inputs etc. In studies analyzing these two components, stochastic frontier models have gradually become the standard method. This chapter briefly introduces the development of stochastic frontier models, with an emphasis on the panel data setting. Twelve specifications, as

well as their implementation methods, are then discussed in detail. These representative models make different assumptions about the efficiency term, aiming to provide better approximations of the underlying data generating process without adding too many constraints. Comparing all these models, we expect different estimates of productivity from different specifications. The evaluation and selection of a suitable model for empirical analysis become a problem. Standard information criteria provide measures of the performance of each candidate model, but multiple criteria can lead to contradicting conclusions about which model is the best one. In addition, the model selection approach itself ignores the risk of model uncertainty. This issue of dealing with multiple competing models will be addressed in Chapter 3.

While Chapter 1 concentrates on the methods of estimating productivity, Chapter 2 focuses on the role of proper specification of the inputs used in generating the output. Though the inputs of a production process are usually observable, their effects on the outputs are often not clear and straightforward. The allocation of different inputs are affected by both the production technology and market prices. Chapter 2 utilizes the duality between the production maximization problem and cost minimization problem to uncover the shadow prices of inputs, and constructs corresponding price indexes for further analysis. This chapter is motivated by recent housing bubbles and considers the housing market for the empirical application. The housing market is an important component of the economy, and constantly attracts interests of researchers. Diewert (2010) for example has provided a comparison of various methods of constructing property price indexes using index number and hedonic regression methods, which he illustrates using data over a number of quarters from a small Dutch town. Chapter 2 provides an alternative approach based on Shephard's dual lemma and I apply it to the same data used by Diewert. This method avoids the multicollinearity problem associated with traditional hedonic regression, and the resulting prices of property characteristics show smoother trends than Diewert's results. The chapter also revisits the Diewert and Shimizu (2013) study that employed hedonic regressions to decompose the price of residential property in Tokyo into land and structure components and that constructed constant quality indexes for land and structure prices respectively. I use three models from Diewert and Shimizu (2013) to fit our real estate data from town A in Netherlands, and also construct the price indices for land and structure, which are compared with results derived using the duality theory.

Again, we have multiple models in the study of housing market. As in the case of productivity, the shadow prices of property characteristics are unobservable (due to the nature of the input or intermediate good, there may not exist an explicit market.) Thus, we rely on certain methods for estimation, and there are a set of candidate models. Chapters 1 and 2 leave us in a dilemma. Which model is correct? Which model do we choose? Is any model actually the correct one or are we choosing among misspecified models? Do we simply choose one model and ignore results from the others? These issues are addressed in Chapter 3 wherein a model averaging approach is explored to provide estimates that are robust to various model specifications. Model averaging methods can be used to provide robust estimates by combining a set of competing models through certain optimization mechanisms.

Chapter 3 pursues robust estimates of world productivity levels as well as its growth rates. Various structural and reduced form models of productivity growth have been proposed in the literature. In either class of models, reduced form measurements of productivity and efficiency are obtained. As the true data generating process of productivity cannot be observed, this chapter examines model averaging approaches that can provide a vehicle to weight predictions (in the form of productivity and efficiency measurements) from different reduced form methods. The reduced form models, typically stochastic frontier methods, have a variety of different settings, which have been discussed in Chapter 1. This chapter considers the jackknife model averaging estimator proposed by Hansen and Racine (2012) and illustrates how to apply the technique to a set of competing stochastic frontier estimators. The derived

method is employed to analyze the productivity and efficiency development in three country groups worldwide. The results of the empirical application show that the model averaging method provides more stable estimates. The model selection method, on the other hand, tends to select a model with superficially high goodness of fit, which results from the match between some specific model setting and the data set. A brief discussion of alternative structural models from which a reduced form forecast can be derived is provided to illustrate a different perspective for productivity analysis.

Acknowledgments

I would like to dedicate this work to my parents, for their love and support since the very beginning of my life. They encourage me to explore the world, to pursue my dream and not to fear failure. I owe them everything I have done.

I would like to express my most sincere gratitude to my adviser, Dr. Sickles. He is the one that led me through the graduate study. He provided invaluable expertise and guidance in my study and research. I could not have accomplished this work without his patience, care, and support.

I would like to thank my committee members, Dr. Muhmoud El-Gamal and Dr. Katherine Ensor, for their advice and assistance. I would like to thank all my friends, for making the life so memorable.

Contents

Illustrations

Tables

Chapter 1

Stochastic Frontier Models Review

1.1 Productivity measurement and the frontier models

When describing a production process, it is essential that key elements such as the number and the nature of inputs and outputs, the form of innovation, and the potential that the decision making unit, or firm, fails to utilize all inputs efficiently given market prices, be considered and correctly modeled. If a firm is operating efficiently, then in order to generate more output, a firm can use more inputs, but if technically inefficient, the firm can increase its level of production with inputs kept fixed. Though in many cases, the two words "productivity" and "efficiency" are interchangably used to express the same concept, they are not the same concept in production theory. Productivity is a broader term and depends on various factors, such as technology, productive efficiency, environmental influences, unexpected incidents, and so on. Productive efficiency can be decomposed into many components, but its two most important, given a constant returns to scale technology, are: 1) technical efficiency, which measures the capability to produce as much output as possible with given inputs or utilize as little inputs as possible to obtain a certain output level; 2) allocative efficiency, which refers to the attempt to adjust the proportion of inputs and outputs according to the price faced, in order to achieve a certain maximum or minimun objective, such as revenue or profit maximization, or cost minimization.

This chapter is focused on the study of technical efficiency, from an econometric approach. A formal definition of technical efficiency was first provided by Koopmans (1951), according to which, a firm is technically efficient if it cannot increase any output without increasing at least one of its inputs or reducing at least one of its other outputs. Alternatively, a firm is technically efficient if it cannot reduce any input without reducing at least one of its outputs or by increasing at least one of its other inputs. As the production unit is usually studied from both output-oriented and input-oriented perspectives, the efficiency problem is also described from these two perspectives. An alternative definition is the Debreu-Farrell measure introduced by Debreu (1951) and Farrell (1957). To clearly state this definition, some notation and terminology will be needed. Denote $x = (x_1, \ldots, x_n) \in \mathbb{R}_+^n$ to be the input vector, and $y = (y_1, \ldots, y_m) \in \mathbb{R}_+^m$ to be the output vector. The production technology can be described by an input set:

$$L(y) = \{x : (x, y) \text{ is feasible}\} \tag{1.1}$$

or equivalently, an output set:

$$P(x) = \{y : (x, y) \text{ is feasible}\} \tag{1.2}$$

The Debreu and Farrell measure of technical efficiency(TE) then can be stated as:

$$\text{Input-oriented } TE_I = \min\{\lambda : \lambda x \in L(x)\} \tag{1.3}$$

$$\text{Output-oriented } TE_O = \max\{\theta : \theta y \in P(x)\} \tag{1.4}$$

By definition, $TE_I \leq 1$ and $TE_O \geq 1$. We can see that while Koopmans' definition considers simultaneous change in both input and output sets, the Debreu-Farrell definition takes only one of them into account and measures the equiproportionate change.

Shephard (1953) and Shephard, Gale, and Kuhn (1970) introduced the distance function to represent the production technology

$$\text{Input distance function:} D_I(x, y) = \max\{\lambda : (x/\lambda) \in L(y)\} \tag{1.5}$$

$$\text{Onput distance function:} D_O(x, y) = \min\{\theta : (y/\theta) \in P(x)\} \tag{1.6}$$

As we can see from the definition above, the Debreu-Farrell measure is just the reciprocal of the corresponding distance function.

Not all production units operate at the technically efficient level. For any input vector, the maximum output that can be obtained from the technology consititutes the "production frontier," which provides an optimal output level for comparison. Technical efficiency of each unit is usually measured by the ratio of realized output to potential optimal output at a centain input level. The input-oriented comparison measure can be defined in a similar manner.

Thinking of a firm as a black box transforming a vector of inputs into a vector of outputs, we take a production, or a transformation function to represent the production technology. Basically, all models that will be discussed below focus on the single-output case. For many multiple-output cases, we can aggregate the outputs into a single output index by some criterion. For those that cannot be aggregated, other treatments will be employed. Let $TE_i \leq 1$ represents the technical efficiency of firm i and $f(\cdot)$ be the production function that identifies the frontier. Realizing that all observed output levels must lie within this frontier, the model to be estimated can be written as:

$$Y_i = f(X_i, \beta) \cdot TE_i \tag{1.7}$$

where Y_i is the output of firm i, X_i is the input vector, and β is the associated parameter vector. In practice, the production function usually takes a logarithmic form in its components. Taking the log of (1.7) gives,

$$\ln Y_i = \ln f(X_i, \beta) - u_i, \quad -u_i = \ln TE_i \tag{1.8}$$

where $u_i \geq 0$ represents the distance to full efficiency, and is thus interpreted as the inefficiency level. Such a specification is labelled as a deterministic frontier model, since the frontier $f(X_i, \beta)$ is known with certainty, and technical inefficiency is the sole cause of any shortfall. One widely-cited example is one proposed by Aigner and Chu (1968) with a Cobb-Douglas production function. The log-transformed formulation is,

$$\begin{aligned} \ln Y_i &= \ln A + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} - u_i \\ &= a + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \epsilon_i \end{aligned} \tag{1.9}$$

Two programming methods (linear and quadratic) were used by Aigner and Chu to calculate the parameters, and both were subject to the constraint that technical inefficiency could be present, i.e., $u_i \geq 0$. It is not possible (without bootstrapping) to conduct inference on the parameters with the goal programming method unless one specifies the distribution of the inefficiency term . Ordinary least squares (OLS) method can be used to consistently estimate the slope parameters if we rewrite the model as

$$\begin{aligned} \ln Y_i &= (\ln A + E[\epsilon_i]) + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \epsilon_i - E[\epsilon_i] \\ &= a^* + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \epsilon^*. \end{aligned} \tag{1.10}$$

Here the new residual term ϵ^* has a zero conditional mean, and thus (a^*, β_1, β_2) can be consistently estimated. However, what is also of interest to productivity analysis is an estimate of $E[u_i|X_i]$. Some extra procedures are needed to obtain an unbiased and consistent estimate of the (in)efficiency term.

One method to identify and estimate inefficiency is to correct the residual by shifting them to nonnegative domain:

$$\tilde{e}_i = \hat{e}_i - \max_i e_i \tag{1.11}$$

where $\hat{e}_i$ is the residual from OLS regression, and $\tilde{e}_i$ is the corrected residual. Thus all residual terms satisfy the constraint that $u_i \geq 0$. This method is referred to as corrected OLS (COLS). We can also impose assumptions on the distribution of u_i and employ the moments method to consistently estimate the parameters of $E[u_i]$. This method is called the modified OLS (MOLS).

As noted before, all the differences between realized and potential optimal output levels are attributed to technical inefficiency, which is not the case in real circumstances. Shocks from random external events will also influence the production and needed to be taken into consideration. The inclusion of random shocks leads to stochastic frontier models.

1.2 A brief history of stochastic frontier thoeries

The stochastic frontier model was first introduced by Aigner, Lovell, and Schmidt (1977) and Meeusen and Van den Broeck (1977), and has become a standard framework for efficiency analysis. The general model that incorporates the firm-specific random noise can be written as:

$$Y_i = f(X_i, \beta) \cdot \exp[v_i] \cdot TE_i \tag{1.12}$$

where $f(X_i, \beta) \cdot \exp[v_i]$ represents the stochastic frontier for firm i. In most applications, the production technology is assumed to be linear in the logged inputs and logged outputs. Thus we make log transformation of the model,

$$\begin{aligned} y_i &= \alpha + \sum_{i=1}^{n} \beta_i x_i + v_i - u_i \\ &= \alpha + \sum_{i=1}^{n} \beta_i x_i + \epsilon_i \end{aligned} \tag{1.13}$$

where $y_i = \ln Y_i$ and $x_i = \ln X_i$, $i = 1, \ldots, n$ are the log-transformed variables. While the inefficiency term $u_i \geq 0$ is assumed to have bounded support, the random shock, v_i, is not restricted and is usually assumed to follow a normal distribution. The mean of the composite error is some constant under such assumptions, and OLS method can be used to obtain consistent estimates of the slope parameters as in the deterministic frontier case. We can also utilize maximum likelihood (ML) methods once some parametric distributions are imposed on the two error terms. Aigner, Lovell, and Schmidt (1977) considered the normal-half-normal case. Here, v_i is assumed to be i.i.d. $N(0, \sigma_v)$ and $u_i = |U_i|$, where U_i is i.i.d. $N(0, \sigma_u)$. It is not difficult to derive the distribution of the composite error ϵ_i, and then formulate the likelihood function. Distribution parameters (σ_v, σ_u) are estimated simultaneously with the inputs coefficients.

A variety of distribution assumptions have been considered. Aigner, Lovell, and Schmidt (1977) and Meeusen and Van den Broeck (1977) discussed the normal-exponential specification; Greene (1980a,b) and Stevenson (1980) proposed the normal-gamma distribution. Arguing that the mean of the inefficiency term doesn't have to be zero, Stevenson (1980) considered the normal-truncated-normal case, in which $u_i = |U_i|$, $U_i \sim N(\mu, \sigma_u^2)$ with $\mu \neq 0$. To evaluate different one-sided distributional assumptions on the inefficiency term, Lee (1983) constructed Lagrangean Multiplier tests using the Pearson family of truncated distributions, since it contains all the distributions that have been mentioned.

Although productivity and efficiency researchers are always interested in the parameters of the underlying technology, stochastic frontier methods are primarily distinguished and motivated based on their ability to providing an estimate of the level and ranking of individual firm's efficiencies. With cross-sectional data, none of those early stochastic frontier models were able to provide a firm specific estimate of inefficiency. Rather they could only identify and consistently estimate the average efficiency of the firms in the sample. A solution to this was provided by Jondrow, Lovell, Materov, and Schmidt (1982). They derived the distribution of u_i conditioning on $v_i - u_i$ in both the half-normal and the exponential cases. Thus, the conditional mean or mode can be used as an estimate of u_i.

Extensions of the stochastic frontier model to a panel data setting would be expected to provide the productivity analyst with more degrees of freedom with which to model technical inefficiency and its heterogeneity over time and among firms. Since firms are observed over multiple periods and this reveals more information about their characristics. The panel stochastic frontier model was first proposed by Pitt and Lee (1981) and Schmidt and Sickles (1984) and specified as,

$$y_{it} = \alpha + x_{it}'\beta + v_{it} - u_i. \tag{1.14}$$

While the random noise differs both over time and cross firms, the inefficiency term is firm-specific and does not change over time. We can combine the inefficiency term with the intercept, and this model becomes a reparameterized, but standard, panel data model. The regression methods for panel data can then be used to estimate the parameters. If there are no time-invariant regressors included, the fixed-effects method, which allows u_i to be correlated with regressors and arbitrarily distributed, can consistently estimate the slope parameters. The basic stochastic frontier model assumes u_i to be independent and relies on the distributional assumption for estimation. The random-effects method also makes no assumption about the distribution of u_i, and allows for the existence of time-invariant variables. For identification, u_i is required to be uncorrelated with the regressors. Greene (2005a,b) revisited the standard panel data stochastic frontier model, and proposed the "true" fixed effects and random effects model, by labeling the constant part of the fixed effects as the "true" fixed effects and changes over time these effects as "technical inefficiency".

With consistent estimates of input coefficients, the inefficiency term for each firm is estimated as $(\max_i \hat{u}_i - \hat{u}_i)$, just as in the deterministic frontier model. The ML method was also considered by Pitt and Lee (1981) and Schmidt and Sickles (1984). Imposing distributional assumptions on the two error terms and assuming u_i is independent, we can estimate the coefficients by maximizing the likelihood function and the firm's inefficiency is thus estimated by $E[u_i|\epsilon_i]$ rather than $(\max_i \hat{u}_i - \hat{u}_i)$, which is a relative value compared to the best performance.

It has been shown that estimated inefficiency is consistent as T grows large. However, in the real world, it is unlikely that a firm's technical efficiency stays the same over a long time period. The stochastic frontier model is then modified to allow for time-varying inefficiencies. Kumbhakar (1990) specified the inefficiency term as $u_{it} = (1+exp(bt+ct^2))^{-1}\tau_i$, where $(1+exp(bt+ct^2))^{-1}$ represents the changing trend over time, and τ_i is the firm-specific effects. A ML method is proposed to estimate the coefficients, and the estimation of the inefficiency term is based on the conditional distribution of $\tau_i|\epsilon_i$. Cornwell, Schmidt, and Sickles (1990) modeled the firm effects as $\alpha_{it} = \theta_{i1} + \theta_{i2}t + \theta_{i3}t^2$, and discussed three cases corresponding to different levels of correlation between the regressors and the inefficiency term. Without distributional assumptions, the firm inefficiency is estimated as $(\max_i \hat{\alpha}_{it} - \hat{\alpha}_{it})$. Battese and Coelli (1992) modeled the effects as $u_{it} = \eta_{it}u_i = exp[-\eta(t-T)]u_i$ and utilized the ML method for estimation. In the model proposed by Lee and Schmidt (1993), the firm effects, consisting of the inefficiency term and some unobserved time-varying effects, was modeled as $\alpha_{it} = \alpha_t - u_{it} = \theta_t \delta_i$. This model incorporates Kumbhakar (1990) and Battese and Coelli (1992) as special cases and assumes no specific changing pattern of the time-varying factor θ_t. They considered fixed-effects and random-effects approaches and estimated u_{it} as $\max_i \alpha_{it} - \alpha_{it}$. Ahn, Lee, and Schmidt (2007, 2013) extended Lee and Schmidt's model in the case of large N and fixed T to let α_{it} be affected by multiple time-varying shocks, i.e., $\alpha_{it} = \sum_{j=1}^{p} \theta_{tj} \delta_{ij}$. This extended model nests all the models mentioned here. Corresponding to the more general setting, the estimation method is more complex and will be explained in the next section.

Most stochastic frontier models being studied and used in application are fully parametric, in both the setting of the production function and the specification of the error terms. Park and Simar (1994) and Park, Sickles, and Simar (1998, 2003, 2007) relaxed the distributional assumptions of the error terms and utilized a semi-parametric approach. In their settings, the firm effects α_i is fixed over time, and (α_i, X_i) is assumed to follow some joint distribution. Kneip, Sickles, and Song (2012) considered a semiparametric model with time-varying firm effects and allowed for arbitrarily temporal changing pattern.

A fully nonparametric case was proposed by Kumbhakar, Park, Simar, and Tsionas (2007), which relied on the local likelihood method for estimation. Being nonparametric, the model is very flexible and avoids the problem of misspecification. The simulation results show that the estimates are robust and numerically stable.

Aside from classic regression techniques, Bayesian methods are also widely studied in the stochastic frontier model literature. They were first applied to the standard stochastic frontier model by Van den Broeck, Koop, Osiewalski, and Steel (1994) and

Koop, Osiewalski, and Steel (1994a,b) with cross-sectional data. In general, a prior joint distribution is assumed for all the parameters $\Theta = (\alpha, \beta, \sigma_v, \sigma_u)$ (σ_v and σ_u are the standard deviations of v_i and u_i respectively), and based on the dataset, we can update the information to form the joint posterior. Different prior distributions are considered for the inefficiency term, and in most cases there are no closed-form expressions for the joint posteriors. To estimate the parameters and the inefficiency term, reseachers usually employ Monte Carlo methods with Gibbs sampling (or other sampling techniques). Koop, Osiewalski, and Steel (1997) discussed Bayesian methods in a panel data setting, corrsponding to both fixed-effects and random-effects apporaches. Osiewalski and Steel (1998) considered the problem of making posterior inference in a fully parametric Bayesian setting with numerical integration methods. Tsionas (2002) proposed a model with random coefficients, which represented different technologies cross firms, and the Bayesian technique was used to separate firms' inefficiency levels from their own technologies. Tsionas (2006) considered Bayesian inference in a dynamic stochastic frontier model. Liu, Sickles, and Tsionas (2013) considered the Bayesian method in a panel data model with unobserved heterogeneous individual effects. The basic setting follows Kneip, Sickles, and Song (2012).

Note that most studies mentioned here, and will be discussed in details in next section, are focused on the production frontier with a single output. One approach to incorporate multiple outputs is to aggregate the outputs into some index under certain aggregating rules (Fernandez, Koop, and Steel (2000)). We can also consider a dual problem of production, such as the stochastic frontier for a cost function or profit function.

1.3 Stochasic frontier models

1.3.1 Cornwell, Schmidt and Sickles (1990)

Model

In Cornwell, Schmidt and Sickles (1990), the basic panel data model was extended to allow for heterogeneity in slopes as well as intercepts. The model can be specified as:

$$y_{it} = x'_{it}\beta + z'_i\gamma + w'_{it}\delta_i + \epsilon_{it} \qquad n = 1, \cdots, N; t = 1, \cdots, T. \tag{1.15}$$

where x_{it}, z_i and w_{it} are $K \times 1$, $J \times 1$ and $L \times 1$ vectors, respectively, and the parameter vectors β, γ and δ_i are dimensioned conformably. Note that 1) variables in z do not vary over time; 2) coefficients of w, δ_i, changes accross different units, thus representing heterogeneity in slopes.

A common construction can relate this model to stardard panel data model. Let $\delta_0 = E[\delta_i]$, and $\delta_i = \delta_0 + u_i$. Then the model can be written as:

$$y_{it} = x'_{it}\beta + z'_{it}\gamma + w'_{it}\delta_0 + v_{it}, \tag{1.16}$$

$$v_{it} = w'_{it}u_i + \epsilon_{it}. \tag{1.17}$$

Here u_i is assumed to be i.i.d., zero mean random variables with covariance matrix Δ. The disturbances ϵ_{it} are taken to be i.i.d. random variable with a zero mean and constant variance σ^2, and uncorrelated with regressors and u_i. Recall the standard panel data stochastic frontier model is,

$$y_{it} = \alpha + x'_{it}\beta + v_{it} - u_i = \alpha_i + x'_{it}\beta + v_{it}$$

To relax the assumption that the firm effects are time-invariant, Cornwell, Schmidt, and Sickles (1990) proposed a quadratic time-varying path for all firms,

$$\alpha_{it} = \theta_{i1} + \theta_{i2}t + \theta_{i3}t^2$$

Let $w'_{it} = (1, t, t^2)$ and $\delta'_i = (\theta_{i1}, \theta_{i2}, \theta_{i3})$, the extended model can be written in the same form as in above setting,

$$y_{it} = x'_{it}\beta + w'_{it}\delta_i + v_{it} \tag{1.18}$$

Put these terms in matrix form, we have:

$$y = X\beta + Z\gamma + W\delta_0 + v, \tag{1.19}$$

$$v = Qu + \epsilon. \tag{1.20}$$

where $Q = diag(W_i)$, $i = 1, \cdots, N$ is a $NT \times NL$ matrix, and u is the associated $NL \times 1$ coefficients vector.

To identify each δ_i, we require Q to be of full column rank, i.e., $L \leq T$.

Implementation

Depending on the level of correlation between the error term and regressors, they considered three cases with different assumptions, and the implementation of (1.19) and (1.20) were discussed respectively. The estimation methods can be seen as modified fixed-effects, random-effects and Hausman-Taylor approaches, respectively.

"Within" Estimator: allow for correlation

Let $P_Q = Q(Q'Q)^{-1}Q$, $M_Q = I - P_Q$. Then P_Q projects matrices onto the column space of Q, and M_Q projects matrices onto the null space of Q. The within estimator of β is given by:

$$\hat{\beta}_w = (X'M_QX)^{-1}X'M_Qy \tag{1.21}$$

Recall that Z do not vary with time, then $M_QZ = 0$, and γ cannot be estimated. The strength is that $\hat{\beta}_w$ is consistent even if (X, Z) and Qu are correlated.

GLS: no correlation

The GLS estimator of $(\beta, \gamma, \delta_0)$ is given by:

$$[(X, Z, W)'\Omega^{-1}(X, Z, W)]^{-1}(X, Z, W)'\Omega^{-1}y \tag{1.22}$$

where $\Omega = \sigma^2 I_{NT} + Q(I_N \otimes \Delta)Q'$. Equivalently, we can first make a transformation of the equation such that covariance of the error term satisfy the assumption of standard linear regression model, and then apply the standard OLS method. The transformed equation can be written as

$$\Omega^{-1/2}y = \Omega^{-1/2}X\beta + \Omega^{-1/2}Z\gamma + \Omega^{-1/2}W\delta_0 + \Omega^{-1/2}v \tag{1.23}$$

It has been shown that

$$\Omega^{-1/2} = \frac{1}{\sigma}M_Q + F$$

with

$$F = Q(Q'Q)^{-1/2}[\sigma^2 I_{NL} + (Q'Q)^{1/2}(I_N \otimes \Delta)(Q'Q)^{1/2}]^{-1/2}(Q'Q)^{-1/2}Q'$$

In this case, γ can be estimated and the GLS estimator is more efficient that within estimator for fixed T. One drawback is that the consistency requires (X, Z, W) and Qu to be uncorrelated.

Extended Hausman-Taylor Estimator: partial correlation

Consider the case in which some of the regressors are correlated with the effects. Following Hausman and Taylor (1981), we partition the regressors and assume that (X_1, Z_1, W_1) are uncorrelated with the effects in the sense that sample covariance converge to zero in probability, while (X_2, Z_2, W_2) are correlated with the effects. Dimensions are denoted k_1, j_1, l_1, k_2, j_2, and l_2 respectively. The general idea is to use uncorrelated variables as instruments and we can estimate all the coefficients if dimensional requirements are met.

We start from the within estimator. Recall that the within method only identifies β, which left residual as

$$(y - X\hat{\beta}_w) = Z\gamma + W\delta_0 + [Qu + \epsilon + X(\beta - \hat{\beta}_w)] \tag{1.24}$$

Premultiplying this equation by $\Omega^{-1/2}$, we get

$$\Omega^{-1/2}(y - X\hat{\beta_w}) = \Omega^{-1/2}Z\gamma + \Omega^{-1/2}W\delta_0 + \Omega^{-1/2}[Qu + \epsilon + X(\beta - \hat{\beta}_w)] \tag{1.25}$$

We use instruments $B^* = \Omega^{-1/2}B = \Omega^{-1/2}(X_1, W_1, Z_1)$ instead of B to prevent possible collinearity between (Z_2, W_2) and (X_1, W_1, Z_1). This gives the estimator:

$$\begin{bmatrix} \hat{\gamma}_w \\ \hat{\delta}_{0w} \end{bmatrix} = [(Z,W)'\Omega^{-1/2}P_{B^*}\Omega^{-1/2}(Z,W)^{-1}] \times (Z,W)'\Omega^{-1/2}P_{B^*}\Omega^{-1/2}(y - X\hat{B}_w) \tag{1.26}$$

Enough instruments are needed to identify all the coefficients, that is, $k_1 + j_1 + l_1 \geq J = L$, or equivalently $k_1 \geq j_2 + l_2$.

The efficient instrumental variables estimator can be obtained by using instruments $A^* = \Omega^{-1/2}A = \Omega^{-1/2}(M_Q, X_1, Z_1, W_1)$ to estimate (1.23). Denoting $G = (X, Z, W)$, we can get

$$\begin{bmatrix} \tilde{\beta}^* \\ \tilde{\gamma}^* \\ \tilde{\delta}_0^* \end{bmatrix} = (G'\Omega^{-1/2}P_{A^*}\Omega^{-1/2}G)^{-1}G'\Omega^{-1/2}P_{A^*}\Omega^{-1/2}y \tag{1.27}$$

The existence condition are summarised in the table below.

$k_1 < j_2 + l_2$	$\tilde{\beta}^* = \hat{\beta}_w$, and $(\tilde{\gamma}^*, \tilde{\delta}_0^*)$ does not exist
$k_1 = j_2 + l_2$	$\tilde{\beta}^* = \hat{\beta}_w$, and $(\tilde{\gamma}^*, \tilde{\delta}_0^*) = (\hat{\gamma}_w, \hat{\delta}_{0w})$
$k_1 > j_2 + l_2$	$(\tilde{\beta}^*, \tilde{\gamma}^*, \tilde{\delta}_0^*) \neq (\hat{\beta}_w, \hat{\gamma}_w, \hat{\beta}_{0w})$, and is more efficient

Table 1.1 : Existence condition

1.3.2 Kumbhakar (1990)

Model

Kumbhakar (1990) considered a translog production function.

$$lny_{it} = \alpha_0 + \alpha' lnx_{it} + \beta' lnz_{it} + lnx'_{it}\Delta lnz_{it} + \theta_{it}, \tag{1.28}$$

$$\theta_{it} = u_{it} + v_{it} = \gamma(t)\tau_i + v_{it}. \tag{1.29}$$

where x_{it} is a $m \times 1$ vector of inputs, z_{it} is a $n \times 1$ vector of shift variables, α, β and Δ are parameter vectors and matrix with conformable dimensions.

The error term θ consists of two components: v_{it} reflects white noise and is assumed to be i.i.d with distributioin $N(0, \sigma_v^2)$; u_{it} is the inefficiency term with time-varying factor $\gamma(t)$ and time-invariant characteristics τ_i. τ_i is assumed i.i.d half-normal distributed and $\gamma(t)$ is specified as following:

$$\gamma(t) = (1 + exp(bt + ct^2))^{-1} \tag{1.30}$$

We can see that $\gamma(t)$ is bounded between $(0, 1)$, and it accommodates increasing, decreasing or time-invariant behavior when b and c vary.

Let f_s be the marginal product of input x_s and let w_i be its price, $s = 2, \ldots, m$, then a cost-minimizing firm is allocatively inefficient if $f_s/f_1 \neq w_s/w_1$. This allocative inefficiency can be estimated following Schmidt and Lovell (1979):

$$\begin{aligned}\xi_s &= lnx_1 - lnx_s - ln(w_s/w_1) + ln(\alpha_s + \sum_j \delta_{sj} lnz_j) \\ & \quad -ln(\alpha_1 + \sum_j \delta_{1j} z_j), \quad s = 1, \ldots, m; \, j = 1, \ldots, n \end{aligned} \tag{1.31}$$

Implementation

The ML method is used for estimation. Denote the inefficiency term as $(\theta_{it}, \xi_{it})'$, $\theta_i = (\theta_{i1}, \ldots, \theta_{iT})'$ and $\xi_i = (\xi_{i1}, \ldots, \xi_{iT})'$. Then the joint distribution of all the residuals is $f(\theta_i, \xi_i, \tau_i) = f(\theta_i, \tau_i) \cdot g(\xi_i)$. Since both τ_{it} and v_{it} are i.i.d distributed and they are independent of each other, the joint pdf of θ_i and τ_i can be expressed as $f(\theta_i, \tau_i) = f(\tau_i) \cdot (\prod_t f(v_{it})) = f(\tau_i) \prod_t f(\theta_{it} - \gamma(t)\tau_i)$.

Marginalizing over τ, we can get the distribution of θ as following:

$$\begin{aligned} f(\theta_i) &= \int_{-\infty}^{0} f(\theta_i, \tau_i) d\tau_i \\ &= \frac{2\tau_* exp(-a_i^*/2)}{(2\pi)^{T/2} \sigma_v^T \sigma_\tau} \cdot \Phi(-\mu_i^*/\tau_*) \end{aligned}$$

where

$$\begin{aligned} \sigma_* &= \frac{\sigma_v \sigma_\tau}{(\sigma_v^2 + \sigma_\tau \sum_t \gamma^2(t))^{1/2}}, \\ \mu_i^* &= \frac{\sigma_\tau^2 \sum_t \gamma(t)\theta_{it}}{\sigma_v^2 + \sigma_\tau^2 \sum_t \gamma^2(t)}, \\ a_i^* &= \frac{1}{\sigma_v^2} \{ \sum_t \theta_{it}^2 - (\sigma_\tau^2 (\sum_t \gamma(t)\theta_{it})^2)(\sigma_v^2 + \sigma_\tau^2 \sum_t \gamma^2(t))^{-1} \}. \end{aligned}$$

The log-likelihood function is then defined as

$$\mathcal{L} = \sum_i lnf(\theta_i) + \sum_i lng(\xi_i) + \sum_i \sum_t ln|J_{it}| \tag{1.32}$$

where

$$|J_{it}| = \sum_s \alpha_s + \sum_s \sum_j \delta_{sj} lnz_{jit}$$

is the Jacobian obtained when we transform $(\theta_{it}, \xi_{it})'$ to $(lnx_{1it}, \ldots, lnx_{mit})'$. And

$$g(\xi_i) = \frac{1}{(2\pi)^{T(m-1)/2}|\Sigma|^{1/2}} exp(-\frac{1}{2}\sum_t (\xi_{it}'\Sigma^{-1}\xi_{it}))$$

The parameter vector $(\alpha_0, \alpha, \beta, \Sigma, b, c, \sigma_v^2, \sigma_\tau^2)$ can be estimated by maximizing (1.32).

1.3.3 Battese and Coelli (1992)

Model

Battese and Coelli (1992) considered a different time-varying path of the firm effects.

$$y_{it} = f(x_{it}; \beta) exp(v_{it} - u_{it}), \tag{1.33}$$

$$u_{it} = \eta_{it} u_i = \{exp[-\eta(t-T)]\} u_i, \qquad t = 1, \ldots, T; i = 1, \ldots, N \tag{1.34}$$

where y_{it} is the production level, x_{it} is a vector of inputs, including the firm-specific factors, and $f(x_{it}; \beta)$ is some production function with unknown parameters β. The panel data can be unbalanced, and in this case $t \in I(i)$, where $I(i)$ is the set of periods that observations of firm i are available.

v_{it} is assumed to be i.i.d $N(0, \sigma_v^2)$ random variables; u_{it} is assumed to be i.i.d. following the non-negatively truncated $N(\mu, \sigma^2)$ distribution.

η is set to be a scaler parameter. The movement of firm-specific effects u_{it} depends on the sign of η. The time-invariant case correponds to $\eta = 0$. To allow for more flexible temporal changing patterns of the firm effects, we can specify η as following

$$\eta_{it} = 1 + \eta_1(t-T) + \eta_2(t-T)$$

This specification permits the effects to be convex or concave rather than simply increasing or decreasing with a constant rate.

Implementation

Denote $e_{it} = v_{it} - uit$ as the efficiency term, and assume a linear production function, the model can be rewritten as:

$$y_{it} = x_{it}\beta + e_{it}, \tag{1.35}$$

$$e_{it} = v_{it} - \eta_{it}u_i, \tag{1.36}$$

$$\eta_{it} = e^{-\eta(t-T)}, \qquad t \in I(i); i = 1, \ldots, N \tag{1.37}$$

u_i's are assumed to follow the non-negative truncations of $N(\mu, \sigma^2)$ distribution, and the density of u_i is

$$f_{U_i}(u_i) = \frac{exp[-\frac{1}{2}(u_i - \mu)^2/\sigma^2]}{(2\pi)^{1/2}\sigma[1 - \Phi(-\mu/\sigma)]}, \qquad u_i \geq 0 \tag{1.38}$$

Where Φ is the cumulative distribution function of the standard normal random variable.

v_{it}'s are assumed i.i.d. with $N(0, \sigma_v^2)$ distribution and are independent of u_i's. Let $v_i = (v_{i1}, \ldots, v_{iT_i})'$ be the vector observations of firm i, and the joint density function of v_i and u_i is

$$f_{U_i,V_i}(u_i, v_i) = \frac{exp\ -\frac{1}{2}\{[\frac{(u_i-\mu)^2}{\sigma^2}] + (v_i'v_i)/\sigma_v^2\}}{(2\pi)^{(T_i+1)/2}\sigma\ \sigma_v^{T_i}[1 - \Phi(-\mu/\sigma)]} \tag{1.39}$$

Denote e_i to be the corresponding vector with $e_{it} = v_{it} - \eta_{it}u_i$. The joint density of u_i and e_i can be derived from (1.39):

$$f_{U_i,E_i}(u_i, e_i) = \frac{exp\ -\frac{1}{2}\{[\frac{(u_i-\mu)^2}{\sigma^2}] + [(e_i + \eta_i u_i)'(e_i + \eta_i u_i)/\sigma_v^2]\}}{(2\pi)^{(T_i+1)/2}\sigma\ \sigma_v^{T_i}[1 - \Phi(-\mu/\sigma)]} \tag{1.40}$$

Integrating the joint density with respect ot u_i, we get the density function of e_i as following:

$$f_{E_i}(e_i) = \frac{[1 - \Phi(-\mu_i^*/\sigma_i^*)]exp\ -\frac{1}{2}\{(e_i'e_i/\sigma_v^2) + (\mu/\sigma)^2 - \mu^*/\sigma^{*2}\}}{(2\pi)^{T_i/2}\sigma_v^{(T_i-1)}[\sigma_v^2 + \eta_i'\eta_i\sigma^2]^{1/2}[1 - \Phi(-\mu/\sigma)]} \tag{1.41}$$

where

$$\mu_i^* = \frac{\mu\sigma_v^2 - \eta_i' e_i \sigma^2}{\sigma_v^2 + \eta_i'\eta_i\sigma^2}, \tag{1.42}$$

$$\sigma_i^* = \frac{\sigma^2\sigma_v^2}{\sigma_v^2 + \eta_i'\eta_i\sigma^2}. \tag{1.43}$$

Let y_i be the $(T_i \times 1)$ vector of production level of firm i, and denote $y = (y_1', y_2', \ldots, y_N')$. The density function of y_i can be derived from (9), and thus the log-likelihood function of y is

$$\begin{aligned} L(\beta, \sigma_v^2, \sigma^2, \mu, \eta; y) &= -\frac{1}{2}(\sum_{i=1}^{N} T_i) ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}(T_i - 1) ln(\sigma_v^2) - \frac{1}{2}\sum_{i=1}^{N} ln(\sigma_v^2 + \eta_i'\eta_i\sigma^2) \\ &\quad - N ln[1 - \Phi(-\mu/\sigma)] + \sum_{i=1}^{N} ln[1 - \Phi(-\mu_i^*/\sigma_i^*)] \qquad (1.44) \\ &\quad - \frac{1}{2}\sum_{i=1}^{N}[(y_i - x_i\beta)'(y_i - x_i\beta)/\sigma_v^2] - \frac{1}{2}N(\mu/\sigma)^2 + \frac{1}{2}\sum_{i=1}^{N}(\mu_i^*/\sigma_i^*)^2 \end{aligned}$$

The parameters can then be estimated by ML method.

From above derivations, we can get the minimum-mean-squared-error predictor of the efficiency for firm i as

$$E[exp(-u_i t)|e_i] = \{\frac{1 - \Phi[\eta_{it}\sigma_i^* - (\mu_i^*)/\sigma_i^*]}{1 - \Phi(-\mu_i^*/\sigma_i^*)}\} exp[-\eta_{it}\mu_i^* + \frac{1}{2}\eta_{it}^2\sigma_i^{*2}] \tag{1.45}$$

1.3.4 Lee and Schmidt (1993)

Model

The model proposed in Lee and Schmidt (1993) incorporates Kumbhakar (1990) and Battese and Coelli (1992) as special cases and is specified as:

$$y_{it} = \alpha_t + x_{it}'\beta + v_{it} + \epsilon_{it} \qquad for \quad i = 1, \ldots, n; t = 1, \ldots, T \tag{1.46}$$

where y_{it} is the output level of firm i at period t, x_{it} is a $k \times 1$ vector of explanatory variables, and β is the corresponding coefficients vector. Denote $\alpha_{it} = \alpha_t + v_{it}$ to be the firm specific effects that change over time. The model can then be rewritten as

$$y_{it} = x_{it}'\beta + \alpha_{it} + \epsilon_{it} \qquad for \quad i = 1, \ldots, n; t = 1, \ldots, T \tag{1.47}$$

Here we assume $\alpha_{it} = \theta_t \delta_i$, where θ_t is the time-varying effects to be estimated and δ_i is the individual multiplier. The scale of δ cannot be determined until the scale of θ is set. Thus we normalize the time-varying effects at period 1 to be 1, that is, $\theta_1 = 1$.

Let $y_i = (y_{i1}, \ldots, y_{iT})'$, $X_i = (x_{i1}, \ldots, x_{iT})'$, $\theta = (1, \theta_2, \ldots, \theta_T)$, and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})'$. ϵ_{it} is assumed to be i.i.d. with mean 0 and variance σ^2, not necessarily normal. The model for firm i over all observed periods is

$$y_i = X_i\beta + \theta\delta_i + \epsilon_i \tag{1.48}$$

And the model for all observations can be expressed in the following matrix form

$$y = X\beta + (I_n \otimes \theta)\delta + \epsilon \tag{1.49}$$

Implementation

Fixed-Effects Approach

In the fixed-effects case, δ is assumed to be the vector of parameters to be determined. If θ were known, we can apply the standard least squares method after making some transformations of the model. Following usual notations for projection matrix, we define $P_\theta = \theta(\theta'\theta)^{-1}\theta$ and $M_\theta = I_T - P_\theta$. Both P_θ and M_θ are idempotent, and $M_\theta\theta = 0$. Utilizing this property, we transform the standard model as below

$$\begin{aligned} M_\theta y_i &= M_\theta X_i \beta + M_\theta \epsilon_i \\ (I_n \otimes M_\theta) y &= (I_n \otimes M_\theta) X\beta + (I_n \otimes M_\theta)\epsilon \end{aligned} \tag{1.50}$$

Now without the individual effects, the standard least squares can be applied to get the estimate of β, $\hat{\beta}$, and δ can then be estimated by $\hat{\delta}_i = \theta'(y_i - X_i\hat{\beta})$. The sum of square errors of model (1.50) is

$$\begin{aligned} SSE_f(\beta,\theta) &\equiv \sum_i (M_\theta y_i - M_\theta X_i\beta)'(M_\theta y_i - M_\theta X_i\beta) \\ &= [(I_n \otimes M_\theta)y - (I_n \otimes M_\theta)X\beta]'[(I_n \otimes M_\theta)y - (I_n \otimes M_\theta)X\beta] \end{aligned} \tag{1.51}$$

Given θ, we know SSE_f is minimized by $\hat{\beta}$ from the properties of OLS estimator. Now since θ is not known in practice, we will determine the estimates of β and θ simultaneously by minimizing the SSE_f defined in (1.52). The minimization problem can only be solved numerically. A relation between $\hat{\beta}$ and $\hat{\theta}$ was given in Lee (1991),

$$\begin{aligned} \hat{\beta} &= (\sum_i X_i'\hat{M}_\theta X_i)^{-1} \sum_i X_i'\hat{M}_\theta y_i \\ &= [X'(I_n \otimes \hat{M}_\theta)X]^{-1}X'(I_n \otimes \hat{M}_\theta)y \end{aligned} \tag{1.52}$$

where $\hat{M}_\theta = I_T - \hat{\theta}(\hat{\theta}'\hat{\theta})^{-1}\hat{\theta}'$. Using this relation, $\hat{\beta}$ and $\hat{\theta}$ can be calculated iteratively. It's shown that the estimates obtained are consistent and asympotitically normal given some regularity conditions. Note that the asympototic properties are proven in the case that $n \to \infty$ and T fixed.

After determining the estimates of β and θ, we can estimate the remaining parameters: $\hat{\delta}_i = \hat{\theta}'(y_i - X_i\hat{\beta})/\hat{\theta}'\hat{\theta}$ and $\hat{\alpha}_{it} = \hat{\theta}_t\hat{\delta}_i$.

Random-Effects Approach

In the random-effects case, δ's are assumed to be random variables instead of fixed parameters, and are uncorrelated with other explanatory variables.

Denote $\mu = E(\delta)_i$, and let $\delta_i^* = \delta_i - \mu$ to be the centered variable. Here we assume δ_i^*'s are i.i.d. with mean zero and variance σ_δ^2. μ is some constant, so δ_i^*'s are also uncorrelated with other explanatory variables. And the model can be rewritten as

$$y_i = X_i\beta + \theta\mu + v_i, \qquad v_i = \theta\delta_i^* + \epsilon_i \tag{1.53}$$

$$y = X\beta + (\mathbf{1} \otimes \theta)\mu + v \tag{1.54}$$

where **1** is an $n \times 1$ vector that all entries equal to one. Now the new error vector v doesn't satisfy the standard assumption that all entries are independently and identicaly distributed. The covariance matrix is

$$\Omega := Cov(v) = \sigma^2 I_{nT} + \theta'\theta\sigma_\delta^2(I_n \otimes P_\theta) \tag{1.55}$$

When the covariance matrix is nonscalar, we would use GLS to get efficient estimates. Parallel to the fixed-effects case, we determine the estimates of (β, θ, μ) by minimizing the corresponding sum of squared errors

$$\begin{aligned} SSE_G(\beta,\theta,\mu) &= \sum_i (y_i - X_i\beta - \theta\mu)'(M_\theta + q^2 P_\theta)(y_i - X_i\beta - \theta\mu) \\ &= [y - X\beta - (\mathbf{1}\otimes\theta)\mu]'[(I_n \otimes M_\theta) + q^2(I_n \otimes P_\theta)] \\ &\quad [y - X\beta - (\mathbf{1}\otimes\theta)\mu] \end{aligned} \tag{1.56}$$

where $\Omega^{-1} = \sigma^2(M_\theta + q^2 P_\theta)$ and $q^2 = \sigma^2/(\sigma^2 + \theta'\theta\sigma_\delta^2)$.

Denote the estimates $\tilde\beta$, $\tilde\theta$ and $\tilde\mu$ respectively, and we can derive a relation between the estimates similar to that in the fixed-effects case.

$$\tilde\beta = \{X'[(I_n \otimes \tilde M_\theta) + q^2(M_{\mathbf{1}} \otimes \tilde P_\theta)]X\}^{-1}X'\{X'[(I_n \otimes \tilde M_\theta) + q^2(M_{\mathbf{1}} \otimes \tilde P_\theta)]X\}y \tag{1.57}$$

where $\tilde P_\theta = \tilde\theta(\tilde\theta'\tilde\theta)^{-1}\tilde\theta'$, $\tilde M_\theta = I_T - \tilde\theta(\tilde\theta'\tilde\theta)^{-1}\tilde\theta'$ and $M_{\mathbf{1}} = I_n - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$. The GLS estimator is more efficient than fixed effects estimator, but only when all the effects are not correlated with the explanatory variables.

The remaining unknown parameters can then be estimated as

$$\tilde\delta_i^* = \frac{\tilde\theta'(y_i - X_i\beta - \tilde\theta\tilde\mu)}{\tilde\theta'\tilde\theta} \tag{1.58}$$

$$\tilde\delta_i = \tilde\delta_i^* + \tilde\mu = \frac{\tilde\theta'(y_i - X_i\beta)}{\tilde\theta'\tilde\theta} \tag{1.59}$$

$$\tilde\alpha_{it} = \tilde\theta_t\tilde\delta_i \tag{1.60}$$

Since the covariance matrix of error term is usually unknown in practice, we must first estimate the covarince matrix, or the value of q^2, which is essentically used in estimation, to obtain the feasible GLS estimator. In the case of nonscalar covariance matrix, the simple OLS method can give consistent estimators of β and θ, which can then be used to estimate q^2.

Following above convention, the consistent estimates can be obtained by minimizing the sum of squared errors

$$SSE_O(\beta, \theta, \mu) = \sum_i (y_i - X_i\beta - \theta\mu)'(y_i - X_i\beta - \theta\mu) \tag{1.61}$$

A closed-form solution can be derived for this simple case

$$\hat{\beta}_O = [\sum_i (X_i - \bar{X})'(X_i - \bar{X})]^{-1} \sum_i (X_i - \bar{X})'(y_i - \bar{y}) \tag{1.62}$$

where $\bar{X} = \frac{1}{n}\sum_i X_i$ and $\bar{y} = \frac{1}{n}\sum_i y_i$.

Other consistent estimator can also be used. And q^2 is consistently estimated by $\hat{q}^2 = SSE_f/[n(T-1)]$, where SSE_f is defined in (1.52).

1.3.5 Park, Sickles and Simar (1998, 2003, 2007)

Model

A series papers by Park, Sickles and Simar (1998, 2003, 2007) considered a stochastic frontier model in which the firm inefficiency effects are correlated with other explanatory variables. Three different patterns of correlation are discussed, so are the respective efficient estimators. The basic setting is

$$y_{it} = X_{it}'\beta + \alpha_i + \epsilon_{it} \qquad for \quad i = 1, \ldots, N; t = 1, \ldots, T \tag{1.63}$$

where y_{it} represents output level of firm i at time t, x_{it} is a $k \times 1$ vector of regressors, and ϵ_{it} is the statistical noise that independently and is identically distributed with

$N(0,\sigma^2)$. α_i is the firm specific effects that is bounded above (or below for the cost frontier model). Let $X_i = (X'_{it}, \ldots, X'_{iT})' \in \mathbb{R}^{kT}$. (α_i, X_i)'s are assumed to be independently and identically distributed with some joint density $h(\cdot,\cdot)$.

With different levels of dependency between the firm effects, α, and other regressors, X, three cases were discussed. Case 1 does not assume any specific pattern of dependency between α and X, which leads to an efficient estimator similar to that of a standard fixed effects model. Case 2 assumes the firm effects are correlated with a subset of the explanatory variables, $Z \in X$. And case 3 assumes that α affects Z only through its long run changes (average movements) $\bar{Z}$. The semiparametric efficient estimators of case 2 and 3 are analogous to those proposed in Hausman and Taylor (1981).

Implementation

In the parametric model, an estimator is efficient when its covariance attains the asymptotic lower bound. The lower bound can be calculated as the inverse of the Fisher information, $\mathcal{I} = E(S^* S^{*\prime})$, where S^* is the efficient score function. With $\mathbb{P}$ represent the class of semiparametric models, we can calculate Fisher like information, $\mathcal{I}(\mathbb{P})$, in a similar way, using adjusted efficient score functions. And the semiparametric efficient estimator would be the one attains the correspongding asymptotic lower bound.

Model 1: no specific dependency structure

We first introduce some additional notations. Denote $y = (y_1, \ldots, y_T)'$, $X = (X'_1, \ldots, X'_T)$. Let $R_t(\beta) = y_t - X'_t(\beta)$, $R\bar{(}\beta) = \frac{1}{T}\sum_{t=1}^{T} R_t(\beta)$ and $U_t(\beta) = R_t(\beta) - R\bar{(}\beta)$. Let Σ_w be the covariance matrix of the regressors, $\Sigma_w = E[\frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X})X_t - \bar{X})']$, where $\bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$, and similarly, $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$. The covariance matrix is assumed to be nonsingular.

We can then derive the efficient score function and the information bound for estimates of β. (For detailed proof, see Park, Sickles, and Simar (1998))

$$S^* = \sigma^{-2}\sum_{t=1}^{T} U_t(\beta)X_t \tag{1.64}$$

$$\mathcal{I}(\mathbb{P}) = \sigma^{-2}T^2\Sigma_w \tag{1.65}$$

Using this information bound, it is easy to show that the within estimator given below is efficient

$$\hat{\beta} = \frac{1}{NT}\hat{\Sigma}_w^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(y_{it}-\bar{y}_i) \tag{1.66}$$

$$\hat{\Sigma}_w = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(X_{it}-\bar{X}_i)' \tag{1.67}$$

Model 2: dependency between firm effects and a subset of regressors

To stress the dependency, we rewrite the model as below

$$y_{it} = X_{it}'\theta + Z_{it}'\gamma + \alpha + \epsilon_{it} \tag{1.68}$$

Now X_{it} is a $p \times 1$ vector, Z_{it} is a $q \times 1$ vector and $p+q=k$. Assume α_i and X_i are conditionally independent and the joint density of (α_i, Z_i, X_i) can be written as

$$f(\alpha, z, x) = h(\alpha, z)g(x|z) \tag{1.69}$$

where $g(\cdot|\cdot)$ represents the density of X_i conditioned on Z_i. Let $\beta = (\theta', \gamma')'$. We can then similarly define $R_t(\beta) = y_t - X_t'\theta - Z_t'\gamma$, $U_t(\beta) = R_t(\beta) - \bar{R(\beta)}$, $\bar{Z} = \frac{1}{T}\sum_{t=1}^{T} Z_t$, and $\bar{\sigma} = \sigma/\sqrt{T}$ The joint density of $(\bar{R}(\beta), Z)$ is then denoted as

$$w(r,z) = \int \phi_{\bar{\sigma}}(r-u)h(u,z)du = \phi_{\bar{\sigma}} * h(\cdot, z)(r) \tag{1.70}$$

where $\phi_{\bar{\sigma}}$ is the density of $N(0, \bar{\sigma}^2)$ distribution.

Define

$$I_0 = \int \frac{(w')^2}{w}(r, z)drdz, \qquad w'(r, z) = \frac{\partial}{\partial r}w(r, z) \tag{1.71}$$

And

$$\Sigma_w(X) = E(\frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X})(X_t - \bar{X})') \tag{1.72}$$

$$\Sigma_w(Z) = E(\frac{1}{T}\sum_{t=1}^{T}(Z_t - \bar{Z})(Z_t - \bar{Z})') \tag{1.73}$$

$$\Sigma_w(X, Z) = E(\frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X})(Z_t - \bar{Z})') \tag{1.74}$$

$$\Sigma_B(X|Z) = E[(\bar{X} - E(\bar{X}|Z))(\bar{X} - E(\bar{X}|Z))'] \tag{1.75}$$

Let $S^{*\prime} = ({S^*}'_\theta, {S^*}'_\gamma)$. Under some regularity conditions, we can derive the efficient score function and the information bound

$$S^*_\theta = \sigma^{-2}\sum_{t=1}^{T}U_t(\beta)X_t - (\bar{X} - E(\bar{X}|Z))\frac{w'}{w}(\bar{R}(\beta), Z) \tag{1.76}$$

$$S^*_\gamma = \sigma^{-2}\sum_{t=1}^{T}U_t(\beta)Z_t \tag{1.77}$$

$$\mathcal{I}(\mathbb{P}) = \begin{pmatrix} \bar{\sigma}^{-2}\Sigma_w(X) + I_0\Sigma_B(X|Z) & \bar{\sigma}^{-2}\Sigma_w(X, Z) \\ \bar{\sigma}^{-2}\Sigma'_w(X, Z) & \bar{\sigma}^{2}\Sigma_w(Z) \end{pmatrix} \tag{1.78}$$

The efficient estimator of β can then be determined in the following steps:

(a) Set $\tilde{S} = I^{-1}S^*$, which is called efficient influence function;

(b) Obtain a consistent estimator $\tilde{\beta}$ of β (for example, the within estimator);

(c) Treat $\tilde{S}$ as a function of $(x, z, y, \beta, \sigma^2, h, g)$. Construct a suitable estimator $\tilde{S}(\cdot, \cdot, \cdot, \beta; X_1,$

$Z_1, y_1, \ldots, X_N, Z_N, y_N)$ of $\tilde{S}(\cdot, \cdot, \cdot, \beta, \sigma^2, h, g)$;

(d) Construct the efficient estimator as

$$\hat{S} = \tilde{S} + \frac{1}{N}\sum_{i=1}^{N} \tilde{S}(X_i, Z_i, y_i, \tilde{\beta}; X_1, Z_1, y_1, \ldots, X_N, Z_N, y_N) \tag{1.79}$$

Model 3: dependency between firm effects and average movements of a subset of regressors

In this case, the firm inefficiency effects only influence the correlated regressors through their long run movements, thus the joint density of (α, z) is:

$$h(\alpha, z) = h_M(\alpha, \bar{z})p(z) \tag{1.80}$$

The method used here is basically the same as in Model 2 except for a few modifications. Instead of $w(r, z)$, now we consider the joint density of $(\bar{R}(\beta), \bar{Z})$,

$$w(r, \bar{z}) = \int \phi_{\bar{\sigma}}(r-u)h_M(u, \bar{z})du \tag{1.81}$$

And define

$$\begin{aligned} \Sigma_B(X|\bar{Z}) &= E[\frac{1}{T}(\bar{X} - E(\bar{X}|\bar{Z}))(\bar{X} - E(\bar{X}|\bar{Z}))'] \quad (1.82)\\ I_0 &= \int [\frac{w'(r,\bar{z})}{w(r,\bar{z})}]^2 dr d\bar{z} \quad (1.83) \end{aligned}$$

Similar results of the efficient score function and the information bound are

$$\begin{aligned} S^*_\theta &= \sigma^{-2}\sum_{t=1}^{T} U_t(\beta)X_t - (\bar{X} - E(\bar{X}|\bar{Z}))\frac{w'}{w}(\bar{R}(\beta), \bar{Z}) \quad (1.84)\\ S^*_\gamma &= \sigma^{-2}\sum_{t=1}^{T} U_t(\beta)Z_t \quad (1.85)\\ \mathcal{I}(\mathbb{P}) &= \begin{pmatrix} \bar{\sigma}^{-2}\Sigma_w(X) + I_0\Sigma_B(X|\bar{Z}) & \bar{\sigma}^{-2}\Sigma_w(X,Z) \\ \bar{\sigma}^{-2}\Sigma'_w(X,Z) & \bar{\sigma}^2\Sigma_w(Z) \end{pmatrix} \quad (1.86) \end{aligned}$$

And the efficient estimators in this case can be determined accordingly.

1.3.6 Latent Class Model

Model

In all the stochastic frontier models we mentioned above, the production function is set uniformly for all units, which implies that all the firms share the same technology and only differ in the level of efficiency. This assumption may not be true in practice. The heterogeneities between firm like different sizes, innovation abilities, targeting groups etc. will lead to different operating strategies and different technologies. Imposing the same functional form in the model may mis-identify some difference in production techonology as the individual inefficiency.

One straightforward way to deal with this problem is to group the firms into different categories in the first place, and then do the regression separately for each group. We can group the observations according to exogenous information about certain characteristics of firms (e.g. size, location etc.), or results from some clustering analysis. A shortcoming of this method is that we cannot utilize information represented by correlation between different groups, since regressions are done separately. Moreover, sometimes we cannot collect enough exogenous information in practice to identify distinct categories and it is hard to evaluate the suitability of the grouping criteria.

By combining the latent structure and stochastic frontier model, Greene (2002) proposed a latent class model. It is assumed that there exist J unobserved classes in the panel data, and the observed dependent variable is characterized by some conditional density function:

$$g(y_{it}|x_{it}, classj) = f(\Theta_j, y_{it}, x_{it}) \tag{1.87}$$

Here the functional form $f(\cdot)$ is the same over the entire sample, while the parameter vector Θ_j is class-specified.

The inefficiency term is assumed with a half-normal distribution, and the likelihood coming for unit i at time period t is defined as

$$P(i,t|j) = f(y_{it}|x_{it};\beta_j,\sigma_j,\lambda_j) = \frac{\Phi(\lambda_j \epsilon_{it|j})}{\Phi(0)}\frac{1}{\sigma_j}\phi(\frac{\epsilon_{it|j}}{\sigma_j}) \tag{1.88}$$

where $\epsilon_{it|j} = y_{it} - x'_{it}\beta_j$. We further assume the inefficiency terms are independently distributed over time, thus the conditional likelihood for individual i is

$$P(i|j) = \prod_{t=1}^{T} P(i,t|j) \tag{1.89}$$

And the unconditional likelihood function can be derived as below

$$P(i) = \sum_{j=1}^{J} \Pi(i,j)P(i|j) = \sum_{j=1}^{J} \Pi(i,j)\prod_{t=1}^{T} P(i,t|j) \tag{1.90}$$

$\Pi(i,j)$ is a prior probability describing distribution of firms in different classes. The simplest specification would be a uniform probability set for all individuals, that is, $\Pi(i,j) = \Pi(j)$, for $i = 1,\ldots,N$. Here we allow heterogeneity in the mixing probabities by adopting a multinomial logit form,

$$\Pi(i,j) = \frac{exp(\theta'_i\pi_j)}{\sum_{m=1}^{J} exp(\theta'_i\pi_m)}, \quad \pi_J = 0 \tag{1.91}$$

And the log likelihood is

$$logL = \sum_{i=1}^{N} logP(i) \tag{1.92}$$

In the specification above, the inefficiency term is assumed to be independent over time. The latent class model proposed by Orea and Kumbhakar (2004) allows the inefficiency term to change over time by following a deterministic path: $\epsilon_{it|j} = \gamma_{it}(\cdot)\cdot\epsilon_{i|j}$, where $\gamma_{it}(\cdot)$ is assumed to be an exponential function to incorporate a vector of firm specific characteristics that might change over time.

$$\epsilon_{it|j} = \gamma_{it}(\eta_j)\cdot\epsilon_{i|j} = exp(z'_{it}\eta_j)\cdot\epsilon_{i|j} \tag{1.93}$$

where $z_{it} = (z_{1it}, \ldots, z_{Hit})'$ is a vector of time-varying variables and $\eta_j = (\eta_{1j}, \ldots, \eta_{Hj})'$ the associated parameter vector. With such a temporal changing path, the individual likelihood in their model is defined directly over all the time periods.

Implementation

We maximize the log likelihood to solve for parameter vector Θ_j and probability π_j simultaneously, and the maximization can be implemented through some numerical opimization method. Greene (2005a) utilized an Expectation-Maximization (EM) algorithm.

Using Bayesian rule, we can derive the posterior probabilities of firms being in some class as

$$w(j|i) = \frac{P(i|j)\Pi(i,j)}{\sum_{j=1}^{J} P(i|j)\Pi(i,j)} \tag{1.94}$$

Two optimization problems are defined using these posterior probabilities to solve for the parameter vector $(\theta_j)_j$ and probability set π_j respectively.

$$\text{O1} \quad \max_{\theta_j} [\sum_{i=1}^{N} w(j|i) logP(i|j)], \quad j = 1, \ldots, J \tag{1.95}$$

$$\text{O2} \quad \max_{(\pi_j, \ldots, \pi_J)} [\sum_{i=1}^{N} \sum_{j=1}^{J} w(j|i) log\Pi(i,j)], \quad \Pi_J = 0 \tag{1.96}$$

Now iteratively optimizing the two problems back and forth and updating the posterior probability $w(j|i)$, we can get the estimates of θ_j and π_j. Here $\Pi(i,j)$ uses the multinomial logit specification. An alternative estimation method using Bayesian approach can be referred to in Tsionas (2002).

Denote the results from the optimization as $((\hat{\Theta}_j), (\hat{\pi}_j), (\hat{w}(j|i)))$. $\hat{w}(j|i)$ best estimates the probability of individual i being in class j, and thus we estimate class of individual i to be the one with largest probability, i.e., $j^* \triangleq \arg\max w(j|i)$. After classifying firms into different groups, firm-specific parameters can be estimated by $\hat{\Theta}_{j^*}$ or $\mathbb{E}[\hat{\Theta}|i] = \sum_{j=1}^{J} \hat{w}(j|i)\hat{\Theta}_j$.

1.3.7 "True" fixed effects model

Model

Greene (2005a,b) discussed the characteristics of fixed-effects method, and proposed a "true" fixed effects model. The basic setting is just the classic panel data stochastic frontier model,

$$y_{it} = \alpha_i + x_{it}\beta + \epsilon_{it} \tag{1.97}$$

$$\epsilon_{it} = v_{it} - u_{it} \tag{1.98}$$

where y_{it} is the dependent variable, x_{it} is the vector of regressors with associated coefficients β. α_i represents the individual fixed effects that is constant over time. ϵ_{it} is the composite error consisting of inefficiency term u_{it} and other unaccounted noise v_{it}. u_{it} is assumed to be one-sided distributed with a nonnegative mean. We can see that the standard fixed effects estimates cannot distinguish the individual time-invariant effects and the inefficiency. Another problem is that the estimators are consistent when $T \to \infty$, but there may exist persistent biases when T is small due to incidental parameter problem.

To deal with these problems and examine the seriousness of the biases in small T case, Greene (2005a,b) proposed the "true" fixed effects (TFE) model and considered an alternative method for estimation In the TFE model, v_{it} is assumed to be i.i.d. with normal distribution $N(0, \sigma_v^2)$ and u_{it} is i.i.d. with distribution $\mathcal{F}_u(\sigma)$. $\mathcal{F}_u$ represents some one-parameter distribution defined on $\mathbb{R}^+$ and σ is its scale parameter. Half-normal and Exponential are two commonly used distributions in this family. Utilizing the ML technique and estimating the individual fixed effects along with other parameters, Greene developed a maximum likelihood dummy variable estimator (MLDVE). One drawback of MLDVE is that the covariance parameter cannot be consistently estimated with a short panel, that is, the estimator biases still exist in small T case. The detailed examination of the incidental problem can be referred to Greene (2005a).

Belotti and Ilardi (2012) proposed two new approches for this TFE model, and both gave consistent estimates with fixed T and $n \to \infty$. Their methods utilize the first difference transformation of the model

$$\begin{aligned} \Delta y_i &= \Delta X_i' \beta + \Delta \epsilon_i && (1.99) \\ \Delta \epsilon_i &= \Delta v_i - \Delta_i u_i && (1.100) \end{aligned}$$

It is easy to derive that Δv_i is i.i.d. following distribution $N_{T-1}(0, \Sigma_v)$. The covariance matrix $\Sigma_v = \sigma_v^4 \Lambda_{T-1}$, and

$$\Lambda_{T-1} = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & 0 & \dots & -1 & 2 \end{pmatrix} \quad (1.101)$$

Denote the unknown conditional distribution of Δu_i as $f(\Delta u_i|\sigma)$, and we can define the marginal likelihood contribution as below

$$\begin{aligned} L_i^*(\theta) &= \int f(\Delta v_i, \Delta u_i|\theta) d\Delta u_i = \int f(\Delta v_i|\theta) f(\Delta u_i|\sigma) d\Delta u_i \\ &= \int f(\Delta y_i|\beta, \sigma_v^2, \Delta X_i, \Delta u_i) f(\Delta u_i|\sigma) d\Delta u_i && (1.102) \end{aligned}$$

The second equality comes from the independence of Δv_i and Δu_i. And θ represents all the parameters $(\beta', \sigma, \sigma_v^2)'$.

Implementation

One approach is to maximize the marginal likelihood function, which can be obtained by simulation. We can see that the marginal likelihood function is the conditional expectation of some function of differenced inefficiency vector Δu_i, which can be consistently estimated by its simulated counterpart under some regularity conditions for distribution of u_i (both half-normal and exponential distributions satisfy these regularity conditions).

$$\begin{aligned} L_i^*(\theta) &= \int f(\Delta y_i|\theta, \Delta X_i, \Delta u_i) f(\Delta u|\sigma) d\Delta u_i & (1.103) \\ &= \mathbb{E}_{\Delta \tilde{u}}[\phi_{T-1}(\Delta y_i - \Delta X_i \beta + \sigma \Delta \tilde{u}_i; 0, \Sigma_v)] & (1.104) \\ &\approx \frac{1}{M} \sum_{m=1}^{M} [\phi_{T-1}(\Delta y_i - \Delta X_i \beta + \sigma \Delta \tilde{u}_{im}; 0, \Sigma_v)] & (1.105) \end{aligned}$$

Here $\phi_{T-1}(\cdot; \mu, \Sigma)$ is the $T-1$ dimensional Guassian density with mean μ and covariance matrix Σ. M denotes the number of draws we take from the distrbution of inefficiency vector $\Delta\tilde{u}$. Instead of simulating from the multivariate distribution, we can simply take draws from the univariate distribution of each element, and the differenced inefficiency is

$$\Delta \tilde{u}_{im} = (\tilde{u}_{i2m} - \tilde{u}_{i1m}, \ldots, \tilde{u}_{iTm} - \tilde{u}_{i(T-1)m})' \quad (1.106)$$

It is proved that when the number of draws $M \to \infty$ and sample size $n \to \infty$, the simulated marginal maximum likelihood estimator converges to marginal maximum likhlihood estimator. Thus we need M to be large enough to ensure the simulated expectation to be a good approximation. And the requirement for large number of draws results in a tradeoff between the computational complexity of this new approach and the inconsistency of MLDVE developed in Greene (2005a,b).

The second approach derives a closed form of the marginal likelihood function under the assumption that inefficiency term u_i has an exponential distribution and

$T = 2$.

$$\begin{aligned} L^*(\theta) &= \prod_{i=1}^{n} f(\Delta y_{it}|\theta, \Delta x_{it}) \\ &= \prod_{i=1}^{n}\{\frac{1}{2\sigma}[exp(\frac{\sigma_v^4}{\sigma^2} - \frac{\Delta\mu_{it}}{\sigma})\Phi(\frac{\Delta\mu_{it}}{\sqrt{2}\sigma_v^2} - \frac{\sqrt{2}\sigma_v^2}{\sigma}) \\ &\quad +exp(\frac{\sigma_v^4}{\sigma^2} + \frac{\Delta\mu_{it}}{\sigma})\Phi(-\frac{\Delta\mu_{it}}{\sqrt{2}\sigma_v^2} - \frac{\sqrt{2}\sigma_v^2}{\sigma})]\} \end{aligned} \tag{1.107}$$

where $\Delta\mu_{it} = \Delta y_{it} - \Delta X_{it}\beta$ and $\Phi(\cdot)$ is the cdf of standard normal.

Function (1.107) can be seen as a marginal likelihood function of a subsample subtracted from the whole panel data with only two periods. There are, in total, $B = \begin{pmatrix} T \\ 2 \end{pmatrix}$ such subsamples, and each provides a consistent subsample estimator. To exploit the information of the entire sample, we can combine the B marginal likelihood functions into one single objective function.

$$U_n(\theta) = n^{-1}\begin{pmatrix} T \\ 2 \end{pmatrix}^{-1} \sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s<t} logf(\Delta_t^s y_i|\theta, \Delta_t^s x_i) \tag{1.108}$$

where $\Delta_t^s y_i = y_{it} - y_{is}$ and $\Delta_t^s x_i = x_{it} - x_{is}$. The resulted estimator from maximizing (1.108) is named Pairwise Difference Estimator by the authors, and it is shown that under condition either $n \to \infty$ with T fixed or $T \to \infty$ with n fixed, the estimator is consistent.

1.3.8 Ahn, Lee and Schmidt (2007)

Model

Ahn, Lee and Schmidt (2007) considered a “multiple time-varying individual effects” model, in which firm-specific efficiencies are described by multiple factors that change over time, and the number of factors can be obtained in the estimation. The production frontier model is specified as below:

$$\begin{aligned} y_{it} &= \delta_t + x_{it}'\beta + \epsilon_{it} - u_{it} \\ &= x_{it}'\beta + \eta_{it} + \epsilon_{it},\ i = 1, 2, \ldots, N, t = 1, 2, \ldots, T \end{aligned} \tag{1.109}$$

where ϵ_{it} is the stochastic noise, and $u_{it} \geq 0$ is the inefficiency term of firm i at time t. δ_t represents the time-varying intercept, and it is assumed that there is no time-invariant intercept term contained in $x_{it}'\beta$. Put together, $\eta \equiv \delta_t - u_{it}$ is interpreted as firm i's efficiency level at time t.

The efficiency parameters η_{it} is assumed to be explained by p unresticted components,

$$\eta_{it} = \theta_{1t}\alpha_{1i} + \theta_{2t}\alpha_{2i} + \cdots + \theta_{pt}\alpha_{pi} = \sum_{j=1}^{p} \theta_{jt}\alpha_{ji} \tag{1.110}$$

Note that the model reduces to Lee and Schmidt (1993) model when we set $p = 1$. And this model also nests the model by Cornwell, Schmidt, and Sickles (1990) in that we can set $p = 3$ and $\theta_{1t} = 1, \theta_{2t} = t$, and $\theta_{3t} = t^2$. The correct number of efficiency components, denote it p_0, is needed for the consistent estimation. If $p > p_0$, the estimators of θ_{jt} are inconsistent; if $p < p_0$, the estimators of both β and θ_{jt} are inconsistent. The detailed procedure of estimating p as well as all other parameters will be described below.

Implementation

Restate the model for firm i in matrix form:

$$y_i = X_i\beta + \eta_i + \epsilon_i, \quad \eta_i = \Theta\alpha_i \tag{1.111}$$

where $y_i = (y_{i1}, \ldots, y_{iT})'$ is the dependent variable vector, $X_i = (x_{i1}, \ldots, x_{iT}')$ is the $T \times k$ matrix of regressors, and Θ is a $T \times p$ matrix on p time-varying factors. The parameters to be estimated are β, Θ and the true number of efficiency components p, and α's are regarded as random variables. A nonlinear least squares method can be used to consistently estimate the model if both N and T are large. Here a GMM method is utilized to analyze data with large N and small T.

It is assumed that Θ is of full column rank. And for the identification purpose, further normalization is made in a way that $\Theta = (\Theta_1', \Theta_2')'$, with $\Theta_2 = -I_p$ and Θ_1 is a $(T-p) \times p$ matrix consisting of unrestricted parameters. In the case of $p = 1$, i.e., the Lee and Schmidt (1993) model, it is equal to normalize θ_T to -1.

In this model specification, all regressors vary across both firms and time and there are no time-invariant regressors included. Though not contained in the model, the firm-specific time-invariant characteristics will be used as instruments in the estimation. Denote $z_i = (x_{i1}', \ldots, x_{iT}', f_i')'$ to be the set of instruments, where f_i is a vector of time-invariant factors. It is assumed that z_i and ϵ_i are uncorrelated. More formally, $E(\epsilon_{it}|z_i, \alpha_i') = 0$, $t = 1, \ldots, T$ and $E(\epsilon_i \epsilon_i'|z_i, \alpha_i') = \Sigma_\epsilon$, $i = 1, 2, \ldots, N$. This assumption will be used to build the moment conditions in GMM. Some other regularity conditions are specified. Details can be referred to Ahn, Lee, and Schmidt (2007).

More notations are needed before explaining the estimation method. Denote $H_0 = (H_{1,0}', H_{2,0}')'$ to be a $T \times (T-p)$ matrix of full column rank. It is constructed in a manner that $H_{1,0}$ is a $(T-p) \times (T-p)$ matrix and $H_0'\Theta_0 = 0_{(T-p)\times p}$. Θ_0 is the normalized matrix we described before, i.e. $\Theta_0 = (\Theta_{1,0}', -I_p)$. Note H_0 is not unique in the sense that $(H_0 B)'\Theta_0 = 0_{(T-p)\times p}$ for any nonsingular matrix B. Thus further normalization is made that $H_{1,0} = I_{T-p}$. Then $H_{2,0}' = \Theta_{1,0}$, resulting from the condition that $H_0'\Theta_0 = 0_{(T-p)\times p}$.

The model is then transformed by pre-multiplying H_0' to (1.111)

$$H_0' y_i = H_0' X_i \beta_0 + H_0' \epsilon_i \tag{1.112}$$

here β_0 is the true value of the parameter. Now α_i is removed through the transformation while Θ_0 remains (in H_0). Let w_i be a $q \times 1$ vector of instruments (it may contain only some of the variables in z_i). Using the uncorrelation assumption, the moment conditions are stated as

$$E(H'u(\beta) \otimes w_i) \equiv E(H'(y - X\beta) \otimes w_i) = 0_{(T-p)q\times 1} \tag{1.113}$$

For a given p, the parameters can be obtained by the continuous-updating GMM estimator proposed by Ahn, Lee, and Schmidt (2001). Note that for identification, the number of moment conditions must be greater than or equal to the number of parameters in θ and β, which requires $(T-p)q - [(T-p)p + k] \geq 0$.

The true value of p can be estimated by utilizing a χ^2 statistic developed in Ahn, Lee, and Schmidt (2001). For a given p, denote $(\tilde{\beta}_p', \tilde{\theta}_p')$ to be the GMM estimator, and define

$$\tilde{\Sigma}_u = \frac{1}{N}\sum_{i=1}^{N} u_i(\tilde{\beta}_p) u_i(\tilde{\beta}_p)', \quad \tilde{\Omega}_{ww} = \frac{1}{N}\sum_{i=1}^{N} w_i w_i', \quad \tilde{\Omega}_{wu} = \frac{1}{N}\sum_{i=1}^{N} w_i u_i(\tilde{\beta}_p)'$$

Then we can calculate the eigenvalues of matrix $\tilde{\Sigma}_u^{-1}\tilde{\Omega}_{wu}'\tilde{\Omega}_{ww}^{-1}\tilde{\Omega}_{wu}$. It is shown that the sum of first $(T-p)$ smallest eigenvalues is asymptotically χ^2 distributed with ((T-p)(q-p)-k) degree of freedom under the null hypothesis that $p = p_0$. And it goes to infinity in the case of $p < p_0$. Thus starting from assuming $p_0 = 0$ and sequentially increasing the number for null hypothesis, we can get the estimate of the true number of efficiency conponents.

1.3.9 Bounded Inefficiency Model

Model

Consider a general setup as below

$$y_i = \alpha_0 + x_i'\beta + \epsilon_i \tag{1.114}$$

$$\epsilon_i = v_i - u_i \tag{1.115}$$

where y_i is the output and x_i is the $K \times 1$ vector of inputs. ϵ_i is the composite error, with u_i representing the inefficiencies and v_i representing the general statistical noise. In the classical model, u_i is assumed to follow some distribution that is bouned below at zero, like half normal, exponential and gamma etc. Qian and Sickles (2008) proposed a model that set an unknown upper bound for the ineffiency term besides the zero lower bound. A natural interpretation for this specification is that the extreme inefficient firms in an industry will be driven out by competition. And the upper bound can serve as an indicator for the extent of competitiveness or supervisory oversight.

Another justification is that the bounded inefficiency specification provides an explanation to the skewness problem in empirical works using the classical model. In the traditional setting, the statistical noise v_i is assumed normally distributed with mean zero and variance σ_v^2, and inefficiency term u_i has a one-sided distribution on nonnegative domain. Thus the composite errors should be negatively skewed by the theory. However, researchers often find positive skewness of the error term in applied work, which leads to questions about the appropriateness of the stochastic frontier model. It has been proved that, with the doubly truncated inefficiency term, the distribution of composite errors can be justified to be positively skewed or the wrong skewness is just a consequence of finite samples. It is easy to see that if the upper bound of the inefficiency term is closer to its mean, then the truncated distribution is negatively skewed, resulting in positive skewness of the composite errors. And even if this is not the case for the inefficiency term, the truncation from right side will reduce the extent of positive skewness, and the "wrong" direction of skewness may simply results from the small sample size.

Denoting the upper bound as $B > 0$ and assuming a doubly truncated normal distribution for the inefficiency, we can derive the density of the inefficiency term to be

$$f(u) = \frac{\frac{1}{\sigma_u}\phi(\frac{u-\mu}{\sigma_u})}{\Phi(\frac{B-\mu}{\sigma_u}) - \Phi(\frac{-\mu}{\sigma_u})}\mathbf{1}_{[0,B]}(u),\ \sigma_u > 0 \tag{1.116}$$

Here $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the pdf and cdf of the standard normal distribution, and $\mathbf{1}_{[0,B]}(\cdot)$ is an indicator function. With $B \to \infty$ and $\mu = 0$, the model reduces to the standard half normal specification.

We can also consider a truncated exponential assumption, and the density function is

$$f(u) = \frac{1}{\sigma_u(1 - e^{-B/\sigma_u})}e^{-\frac{u}{\sigma_u}}\mathbf{1}_{[0,B]}(u) \tag{1.117}$$

This model can be extended to the panel data setting:

$$y_{it} = \alpha_0 + x'_{it}\beta + \epsilon_{it} \tag{1.118}$$

$$\epsilon_{it} = v_{it} - u_{it} \tag{1.119}$$

Where u_{it} is assumed to be the positive inefficiency term and follow a time-varying distribution with upper bound B_t. We can impose different structures for the changing path of B_t to fit different circumstances. A generic setup can be

$$B_t = \sum_{i=1}^{K} b_i(t/T)^i,\ t = 1, \ldots, T \tag{1.120}$$

where b_i represents individual effects, and B_t would be the weighted sum of influences from all firms at period t.

Implementation

The inefficiency u_i and random noise v_i are assumed independently distributed and independent with other regressors, as in standard model. We first estimate coefficients of the independent variables. If there is no intercept specified in the model, the coefficients can be consistently estimated by simple OLS method. And for models with intercept, we can rewrite the production function as below

$$y_i = (\alpha_0 - \mathbb{E}u_i) + x_i'\beta + \epsilon_i^* \tag{1.121}$$

We can see that some bias correction is needed for the intercept since $\mathbb{E}\epsilon = -\mathbb{E}u_i \neq 0$.

In the second step, we can use method of moments to identify the parameters of assumed distribution of the error term. Only the first and second order moments are not enough to identify this model. Detailed examination of higher order moments can be found in Qian and Sickles (2008) for both doubly truncated normal and truncated exponential distribution. It is shown that the truncated half normal model ($\mu = 0$) and the truncated exponential model can be globally identified, and the doubly truncated normal model is locally identified. It is not sure whether doubly truncated normal model is globally identifiable, but we can regard it as a collection of a series of submodels with parameters defined in different domains.

A more efficient method is the maximum likelihood estimation. Though the distribution of inefficiency term is bounded, the statistical noise has an unbounded domain and is independent of the regressors, thus the standard regularity condition is met. Some reparametrization is employed to derive the log-likelihood function. One often used method is the γ-parametrization:

$$\sigma = \sqrt{\sigma_u^2 + \sigma_v^2},\ \gamma = \sigma_u^2/\sigma^2 \tag{1.122}$$

And the log-likelihood function for the doubly truncated normal model using γ-parametri-
zation is:

$$\begin{aligned} lnL &= -nln\Big[\Phi(\frac{-ln\tilde{B}-\mu}{\sigma_u(\sigma,\gamma)}) - \Phi(\frac{-\mu}{\sigma_u(\sigma,\gamma)})\Big] - nln\sigma - \frac{n}{2}ln(2\pi) - \sum_{i=1}^{n}\frac{(\epsilon_i+\mu)^2}{2\sigma^2} \\ &+ \sum_{i=1}^{n} ln\Big[\Phi\big(\frac{(-ln\tilde{B}+\epsilon_i)\sqrt{\gamma/(1-\gamma)} - (ln\tilde{B}+\mu)\sqrt{(1-\gamma)/\gamma}}{\sigma}\big) \\ &-\Phi\big(\frac{\epsilon_i\sqrt{\gamma/(1-\gamma)} - \mu\sqrt{(1-\gamma)/\gamma}}{\sigma}\big)\Big] \end{aligned} \tag{1.123}$$

where $\tilde{B} = exp(-B)$ and $\sigma_u(\sigma,\gamma) = \sigma\sqrt{\gamma}$.

Another parametrization is λ-parametrization brought up by Aigner, Lovell, and Schmidt (1977):

$$\sigma = \sqrt{\sigma_u^2 + \sigma_v^2},\ \lambda = \sigma_u/\sigma_v \tag{1.124}$$

And above log-likelihood function can be equivalently expressed as below using λ-parame-
trization

$$\begin{aligned} lnL &= -nln\big(\Phi(\frac{-ln\tilde{B}}{\sigma_u(\sigma,\gamma)}) - \frac{1}{2}\big) - nln\sigma - \frac{n}{2}ln(2\pi) - \sum_{i=1}^{n}\frac{\epsilon_i^2}{2\sigma^2} \\ &+ \sum_{i=1}^{n} ln\Big[\Phi\big(\frac{(-ln\tilde{B}+\epsilon_i)\sqrt{\gamma/(1-\gamma)} - ln\tilde{B}\sqrt{(1-\gamma)/\gamma}}{\sigma}\big) - \Phi\big(\frac{\epsilon_i\sqrt{\gamma/(1-\gamma)}}{\sigma}\big)\Big] \end{aligned} \tag{1.125}$$

The composite error is then estimated by the sample residual once we obtain the estimates of the coefficients by maximizing the log-likelihood function. Inefficiency term u_i can be estimated by $\mathbf{E}(u_i|\epsilon_i)$, which can be derived from the assumed distribution.

1.3.10 Kneip, Sickles and Song (2012)

Model

Similar to the setting in Ahn, Lee, and Schmidt (2007), the model proposed in Kneip, Sickles, and Song (2012) assumed that the individual effects are influenced by a set of time-varying factors, which are represented by some smooth functions of time t. There are no assumptions imposed on the functional forms, so the model allows for arbitrary patterns of temporal change.

$$y_{it} = \beta_0(t) + \sum_{j=1}^{p} \beta_j x_{itj} + v_i(t) + \epsilon_{it}, \qquad i = 1, \ldots, n; t = 1, \ldots, T \tag{1.126}$$

where $v_i(t)$'s are assumed to be smooth time-varying individual effects, and it is required for identifiability that $\sum_i v_i(t) = 0$. $\beta_0(t)$ is some average function, and can be eliminated by transforming the model to the centered form.

$$y_{it} - \bar{y}_t = \sum_{j=1}^{p} \beta_j(x_{itj} - \bar{x}_t j) + v_i(t) + \epsilon_{it} - \bar{\epsilon}_i, \qquad i = 1, \ldots, n; t = 1, \ldots, T \tag{1.127}$$

where $\bar{y}_t = \frac{1}{n}\sum_i y_{it}$, $\bar{x}_{tj} = \frac{1}{n}\sum_i x_{itj}$ and $\bar{\epsilon}_i = \frac{1}{n}\sum_i \epsilon_{it}$. We can see that all the time-invariant effects are incorporated in $v_i(t)$.

Here $v_i(t)$ is assumed to be a linear combination of some basis functions.

$$v_i(t) = \sum_{r=1}^{L} \theta_{ir} g_r(t) \tag{1.128}$$

This semiparametric construction is more flexible and realistic than parametric models, which presume a fixed functional form for the changing path of the individual effects. And comparing to a total nonparametric model, the existance of some common structure leads to much faster rates of convergence in estimation, and makes it easier to make socioeconomic interpretations about underlying influencing factors.

Now the model can be rewritten as below

$$y_{it} - \bar{y}_t = \sum_{j=1}^{p} \beta_j (x_{itj} - \bar{x}_t j) + \sum_{r=1}^{L} \theta_{ir} g_r(t) + \epsilon_{it} - \bar{\epsilon}_i, \qquad i = 1, \ldots, n; t = 1, \ldots, T \tag{1.129}$$

Parameters β_j, θ_{ir}, the basis functions $g_1, \ldots, g_L$ and the dimension of the space spanned by these functions are unknown and remain to be estimated.

There exists the identifiability problem with such a specification for the individual effects. Given a set of basis functions $g_1, \ldots, g_L$, only the spanned linear space $\mathcal{L}_T := span\{g_1, \ldots, g_L\}$ can be identified. We can obtain another set of basis functions by multiplying some regular $L \times L$ matrix A, $(h_1(t), \ldots, h_L(t))' := A \cdot (g_1(t), \ldots, g_L(t))'$, and the corresponding coefficients for $v_i(t)$ will be $(\vartheta_{1i}, \ldots, \vartheta_{Li})' := A^{-1} \cdot (\theta_{1i}, \ldots, \theta Li)'$.

To deal with this problem, some standardization is needed to identify a specific basis:

(a) $\frac{1}{n} \sum_i \theta_{i1}^2 \geq \frac{1}{n} \sum_i \theta_{i2}^2 \geq \cdots \geq \frac{1}{n} \sum_i \theta_{iL}^2 > 0$

(b) $\frac{1}{n} \sum_i \theta_{ir} \theta_{is} = 0$ for $r \neq s$

(c) $\frac{1}{T} \sum_{t=1}^{T} g_r(t)^2 = 1$ and $\sum_{t=1}^{T} g_r(t) g_s(t) = 0$ for all $r, s \in \{1, \ldots, L\}, r \neq s$

Under this standardization, g_r's are orthogonal and θ_{ir}'s are empirical uncorrelated with each other. This set of basis will be estimated from the sample of all individual effects, $v_1 = (v_1(1), \ldots,$
$v_1(T))', \ldots, v_n = (v_n(1), \ldots, v_n(T))'$.

Specifically, denote the empirical covariance matrix of $v_1, \ldots, v_n$ as $\Sigma_{n,T} = \frac{1}{n} \sum_i v_i v_i'$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_T$ be the eigenvalues of the matrix, and $\gamma_1, \gamma_2, \cdots, \gamma_T$ be the corresponding eigenvectors. The basis functions will be

$$g_r(t) = \sqrt{T} \cdot \gamma_{rt} \qquad \text{for all} \quad r = 1, \ldots ; t = 1, \ldots, T \tag{1.130}$$

$$\theta_{ir} = \frac{1}{T} \sum_t v_i(t) g_r(t) \qquad \text{for all} \quad r = 1, \ldots ; i = 1, \ldots, n \tag{1.131}$$

$$\gamma_r = \frac{T}{n} \sum_i \theta_{ir}^2 \qquad \text{for all} \quad r = 1, \ldots \tag{1.132}$$

And for all $l = 1, 2, \ldots$

$$\begin{aligned} \sum_{r=l+1}^{T} \gamma_r &= \sum_{i,t} (v_i(t) - \sum_{r=1}^{l} \theta_{ir} g_r(t))^2 \\ &= \min_{\tilde{g}_1, \ldots, \tilde{g}_l} \sum_i \min_{\tilde{\vartheta}_{i1}, \ldots, \tilde{\vartheta}_{il}} \sum_t (v_i(t) - \sum_{r=1}^{l} \vartheta_{ir} \tilde{g}_r(t))^2 \end{aligned} \tag{1.133}$$

From (1.133) we can see that $v_i(t) \approx \sum_{r=1}^{l} \theta_{ir} g_r(t)$ will be the best l-dimensional linear estimate, and the dimension L naturally equals to $rank(\Sigma_{n,T})$.

Implementation

Since v_i's are assumed to be smooth trends, we can always find m-times continuously differentiable functions ν_i's defined on $[1, T]$ such that $\nu_i(t) = v_i(t)$ for all $t = 1, \ldots, T$. ν_i will be used as a approximation of v_i, and the smoothness of v_i depends on the roughness of ν_i. The method first estimates β and obtains the approximations ν_i's by least squares, and then determines the estimates of the basis functions $\hat{g}_r$ through the empirical covariance matrix $\hat{\Sigma}_{n,T}$, which is estimated by the $(\hat{v}_1, \ldots, \hat{v}_n) = (\hat{\nu}_1, \ldots, \hat{\nu}_n)$. The corresponding coefficients of the basis functions will be obtained by least squares. In the last step, we update the estimate of v_i by $\sum_{r=1}^{L} \hat{\theta}_{ir} \hat{g}_r$, which is proved to be more efficient than the approximation ν_i (See Kneip, Sickles, and Song (2012)).

Before describing the detailed procedure, we first specify some notations. Denote $\bar{y}_t = \frac{1}{n} \sum_i y_{it}, \bar{y} = (\bar{y}_1, \ldots, \bar{y}_T)'$, and $y_i = (y_{i1}, \ldots, y_{iT})'$. Similarly, let $x_{ij} =$

$(x_{i1j}, \ldots, x_{iTj})', \bar{x}_{tj} = \frac{1}{n}\sum_i x_{itj}, \bar{x}_j = (\bar{x}_{1j}, \ldots, \bar{x}_{Tj})'$, and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})$. $X_i = (x_{itj})_{T\times p}$ and $\bar{X} = (\bar{x}_{tj})_{T\times p}$.

Step 1: Obtain estimates $\hat{\beta}_1, \ldots, \hat{\beta}_p$ and nonparametric approximations $\hat{\nu}_1, \ldots, \hat{\nu}_n$ by least squares.

$$\min_{\beta,\nu} \sum_i \frac{1}{T}\sum_t (y_{it} - \bar{y}_t - \sum_{j=1}^{p} \beta_j(x_{itj} - \bar{x}_{tj}) - \nu_i(t))^2 + \sum_i \kappa \frac{1}{T}\int_1^T (\nu_i^{(m)}(s))^2 ds \tag{1.134}$$

$\nu_i^{(m)}$ is the m-th derivatives of ν_i, and $\kappa > 0$ is some pre-specified smoothing parameter.

According to spline theory (see Eubank (1988)), any $\hat{\nu}_i$ obtained from minimization problem (1.134) can be expressed as a linear combination of a set of natural spline basis $z_1, \ldots, z_T$ of order $2m$, $\hat{\nu}_i(t) = \sum_j \hat{\zeta}_{ji} z_j(t)$. $m = 2$ is a typical choice, which results in cubic smoothing splines.

Let Z and A be $T\times T$ matrices with entries $\{z_j(t)\}_{j,t=1,\ldots,T}$ and $\{\int_1^T z_j^{(m)}(s) z_k^{(m)}(s)\}_{j,k=1,\ldots,T}$, let $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)'$ and $\hat{\zeta}_i = (\hat{\zeta}_{1i}, \ldots, \hat{\zeta}_{Ti})'$. Then (1.134) is equivalent to the following problem.

$$\min_{\beta,\zeta} \sum_i (\| y_i - \bar{y} - (X_i - \bar{X})\beta - Z\zeta_i \|^2 + \kappa\zeta_i' A \zeta_i) \tag{1.135}$$

where $\| \cdot \|$ is the usual Euclidean norm.

The solutions are then given by

$$\hat{\beta} = (\sum_i (X_i - \bar{X})'(I - Z_\kappa)(X_i - \bar{X}))^{-1} \sum_i (X_i - \bar{X})'(I - Z_\kappa)(y_i - \bar{y}) \tag{1.136}$$

and

$$\hat{\zeta}_i = (Z'Z + \kappa A)^{-1} Z'(y_i - \bar{y} - (X_i - \bar{X}\hat{beta})) \tag{1.137}$$

where

$$Z_\kappa = Z(Z'Z + \kappa A)^{-1}Z' = (I - \kappa(Z')^{-1}AZ^{-1})^{-1} \tag{1.138}$$

From above results, we can get the estimates of v_i

$$\hat{v}_i := \hat{\nu}_i = Z\hat{\zeta}_i = Z_\kappa(y_i - \bar{y} - (X_i - \bar{X})\beta) \tag{1.139}$$

Step 2: Obtain the empirical covariance matrix $\hat{\Sigma}_{n,T}$.

$$\hat{\Sigma}_{n,T} = \frac{1}{n}\sum_i \hat{v}_i\hat{v}_i' \tag{1.140}$$

where $\hat{v}_i = (\hat{v}_i(1), \ldots, \hat{v}_i(T))'$ for $i = 1, \ldots, n$. And then we can calculate the eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_T$ and the corresponding eigenvectors $\hat{\gamma}_1, \hat{\gamma}_2, \cdots, \hat{\gamma}_T$.

Step 3: Determine the basis functions and correponding coefficients.

It follows from the analysis of identifiability problem that $\hat{g}_r(t) := \sqrt{T} \cdot \hat{\gamma}_{rt}$, for $r = 1, 2, \ldots, L$; $t = 1, 2, \ldots, T$. And the coefficients $(\hat{\theta}_{1i}, \ldots, \hat{\theta}_{Li})$ can be obtained from the following minimization problem:

$$\min_\vartheta \sum_t (y_{it} - \bar{y}_t - \sum_{j=1}^{p} \hat{\beta}_j(x_{itj} - \bar{x}_{tj}) - \sum_{r=1}^{L} \vartheta_{ri}\hat{g}_r(t))^2 \tag{1.141}$$

Now we can update the estimate of v_i by $\sum_{r=1}^{L} \hat{\theta}_{ir}\hat{g}_r$.

Returning to the non-centered model (1.126), the general average funstion $\beta_0(t)$ is left unestimated. A non-parametric method similar to step 1 can be applied to get an approximation. An alternative is to assume $\beta_0(t)$ also lies in the space spanned by the set of basis functions, that is, $\beta_0(t) = \sum_{r=1}^{L} \bar{\theta}_r g_r(t)$. The coefficients can then be estimated by a similar minimization problem as step 3 with objective function $\sum_t(\bar{y}_t - \sum_{j=1}^{p} \hat{\beta}_j \bar{x}_{tj} - \sum_{r=1}^{L} \vartheta_r \hat{g}_r(t))^2$.

1.3.11 Ahn, Lee and Schmidt (2013)

Model

The model setting in Ahn, Lee and Schmidt (2013) is basically the same as the one in Ahn, Lee and Schmidt (2007). Instead of focusing on the analysis of individual effects, more interests are put on the consistent estimation of slope coefficients β when there exists correlation between individual effects and the regressors. And they still focus on the case with large N and small T. Recall that the model is specified as

$$y_{it} = x_{it}'\beta + \sum_{j=1}^{p} \xi_{tj}\alpha_{ij} + \epsilon_{it} \tag{1.142}$$

With the correlation between the regressors and individual effects, the motivation is thus similar to that of a fixed-effects model. To emphasize this feature, the model interprets ξ_{tj} as "macro shocks", and α_{ij} as "random coefficients" instead of "factors" and "factor loadings", though the model itself resembles the factor models.

Restate the model for individual i in matrix form:

$$y_i = X_i\beta + u_i, \qquad u_i = \eta_i + \epsilon_i = \Theta\alpha_i + \epsilon_i \tag{1.143}$$

where $y_i = (y_{i1}, \ldots, y_{iT})'$ is the dependent variable vector, $X_i = (x_{i1}, \ldots, x_{iT})'$ is the $T \times K$ matrix of regressors, and β is the dimension-comformable coefficients vector. The error term u_i is composed of the random noise ϵ_i and individual effects $\eta_i = \Theta\alpha_i$. Θ is a $T \times p$ $(T > p)$ matrix containing p macro shocks that vary over time. The random noise ϵ_{it} is usually assumed to be independently distributed to assure consistent estimates of coefficients in the case of large N and small T. This model relaxes this assumption in that it allows any kind of autocorrelation of ϵ_i and only assumes that ϵ_i is uncorrelated with regressors x_{it} while α might be correlated with x_{it}. Then for identification, it is assumed that there exist instrument variables that are correlated with α_{ij} but not with ϵ_{it}.

Implementation

Due to rotation problem, it is impossible to separate effects of Θ and α. For identification, Θ is normalized such that $\Theta = (\Theta_1', -I_p)'$ with Θ_1 a $(T-p) \times p$ matrix. With instruments, the GMM method proposed in Ahn, Lee, and Schmidt (2001) is extended to incorporate multiply time-varying effects and two methods are proposed to estimate the true number of individual effects. We will first obtain consistent estimators of β and Θ assuming the true number of effects p_0 is known, and then estimate p using some test statistic. Detailed assumptions and discussion can be referred to the paper.

Define $H(\theta) = (H_1(\theta_1), \ldots, H_{T-p}(\theta_{T-p})) = (I_{T-p}, \Theta_1)'$, where $\theta = vec(\Theta_1')$, θ_j is the jth column of Θ_1', and $H(\theta_j)$ is the jth column of $H(\theta)$. $H(\theta)$ is constructed in this form so that $H(\theta)'\Theta = 0_{(T-p)\times p}$. Thus, the random individual effects α can be removed by premultiplying model (1.143) by $H(\theta)$:

$$H(\theta_0)'y_i = H(\theta_0)'X_i\beta_0 + H(\theta_0)'\epsilon_i \tag{1.144}$$

here and after the subscript 0 is used to represent the true value of parameters.

Depending on the assumptions of the instrument variables, the model considers two cases described as below.

Case 1: strictly exogenous instruments

Denote z_i to be the column vector with all the distinct elements in $(x_{i1}, \ldots, x_{iT})$, f_i to be the column vector of variables, other than the regressors x_i, that are uncorrelated with ϵ_i , and τ_t to be the common variables cross individuals at time t. Let $w_{S,i} =$

$(z_i', f_i')'$ be a $q \times 1$ vector. The strict exogeneity can be stated as:

$$E(\epsilon_i \otimes w_{S,i}|\tau_0) = 0_{Tq\times 1} \tag{1.145}$$

Assuming $p = p_0$, we can derive the moment conditions used in the GMM method:

$$E[m_{S,i}(\delta_0)|\tau_0] = E[H(\theta_0)'u_i(\beta_0) \otimes w_{S,i}|\tau_0] = 0_{(T-p)q\times 1} \tag{1.146}$$

where $\delta = (\beta', \theta')'$, and $u_i(\beta) = y_i - X_i\beta$. The optimal GMM estimator is then obtained by iteratively solving the following problem:

$$\min_{\delta} J(\delta|p) = N(\frac{1}{N}\sum_{i=1}^{N} m_{S,i}(\delta))'(\frac{1}{N}\sum_{i=1}^{N} m_{S,i}(\hat{\delta})m_{S,i}(\hat{\delta})')^{-1}(\frac{1}{N}\sum_{i=1}^{N} m_{S,i}(\delta)) \tag{1.147}$$

where $(N^{-1}\sum_{i=1}^{N} m_{S,i}(\hat{\delta})m_{S,i}(\hat{\delta})')^{-1}$ is the weighting matrix and $\hat{\delta}$ is the previous consistent estimator of δ.

The identification condition of δ requires $p = p_0$, i.e., p is specified correctly. Two methods are considered to estimate the true value. One utilizes the Hansen-Sargan statistic for over-identification test. It is shown in the paper that this statistic is equivalent to the minimized value $J(\tilde{\delta}|p)$, where $\tilde{\delta}$ is the optimal GMM estimator. The property of the Hansen-Sargen statistic implies that when $p = p_0$, $J(\tilde{\delta}|p)$ follows a χ^2 distribution with degree of freedom being $(T - p_0)(q - p_0) - k$, and when $p < p_0$, $J(\tilde{\delta}|p)$ goes to infinity. Thus, we can start by setting null hypothesis $p_0 = 0$, and sequentially increase the testing number of p until the null is not rejected.

The other method follows Bai and Ng (2002) and Bai (2009). The objective function proposed for estimation of p is:

$$S_N(p) = J(\tilde{\delta}|p) - f(N)g(p) \tag{1.148}$$

With some properly chosen $f(N)$ and $g(p)$, a consistent estimator can be obtained by minimizing the above function. The set of functions $f(N)$ and $g(p)$ proposed in the paper is $f(N) = \ln(N)$ and $g(p) = a[(T-p)(q-p)-k]$ with some positive number a.

Case 2: weakly exogenous instruments

The paper also considered consistent estimation of parameters when some regressors are only weakly exogenous. We use $x_{W,it}$ and $x_{S,it}$ to represent the weakly and strictly exogenous regressors respectively, and let $w_{S,i}$ contain all the elements in the set $\{x_{S,it}\}_{t=1,\ldots,T}$. Then denote $w_{j,i} = (w'_{S,i}, x'_{W,i1}, \ldots, x'_{W,ij})$, $j = 1, \ldots, T-p$. The modified weak exogeneity condition can be stated as:

$$E(\epsilon_{is}|w_{j,i}, \tau_0) = 0,\; s \geq j,\; j = 1, \ldots, T-p \tag{1.149}$$

To avoid further comlexity, it is assumed that τ_t is strictly exogenous, and the moment conditions for the GMM method become:

$$E(m_{W,i}(\delta_0)|\tau_0) = E\left(\begin{array}{c} H_1(\theta_{1,0})'u_i(\beta_0)w_{1,i} \\ H_2(\theta_{2,0})'u_i(\beta_0)w_{2,i} \\ \vdots \\ H_{T-p}(\theta_{T-p,0})'u_i(\beta_0)w_{T-p,i} \end{array} \middle| \tau_0 \right) = 0 \tag{1.150}$$

Note that the moment conditions here are a subset of those stated in the strictly exogenous case, thus most results follow from that case and we can use the same methods to estimate true number of p consistently.

1.3.12 Liu, Sickles and Tsionas (2013)

According the to estimation method, the stochastic frontier models can generally be grouped into two categories. One is the traditional method using least squares estimation or the maximum likelihood techniques (MLE), which relies on certain specification of the distribution of the error term. The other is the Bayesian method that emerged in 1990s, which treats all the parameters randomly and usually make use of Markov Chain Monte Carlo(MCMC) techniques and Gibbs sampling algorithm in estimation.

Liu, Sickles and Tsionas (2013) considered two Bayesian stochastic frontier models that allow individual effects to evolve over time. The time-varying effects in a traditional parametric or semiparametric setup are modeled by pre-specifying a particular functional form. The functional form can be rather general, nonetheless, there exists misspecification problem. Moreover, as argued by Swamy and Tavlas (1995), the time-varying idosyncratic factors may follow different patterns, thus a common functional form is not suitable. By utilizing Bayesian method, we treat the unobserved individual effects nonparametrically and avoid problem of misspecification.

Another feature of Bayesian method is that it is much more straightforward and eaiser to implement than the traditional nonparametric regression. And we do not need to rely on complex asymptotic theory to get the inference of parameters.

Model 1

Consider a non-parametric model with time-varying effects.

$$y_{it} = x_{it}'\beta + \varphi_i(t) + \epsilon_{it}, \qquad i = 1, \ldots, n;\ t = 1, \ldots, T \tag{1.151}$$

where y_{it} is the production level, x_{it} is a $p \times 1$ vector of factors and β is the corresponding coefficients. ϵ_{it} is assumed to be i.i.d. following $N(0, \sigma^2)$ distribution, and $\varphi_i(t)$ is some firm-specific effects that changes over time.

Denote $\gamma_{it} = \varphi_i(t)$ and $\gamma_i = (\gamma_{i1}, \ldots, \gamma_{iT})'$. In classic Bayesian models, the individual effects are often assumed to follow a normal prior such that the posterior distribution falls in the same family. To fit the method into more general cases, here we only impose the normal assumption on first-order or second-order difference of the individual effects. The prior of parameter γ_i is specified as below,

$$p(\gamma) \propto \prod_{i=1}^{n} exp(-\frac{\gamma_i' Q \gamma_i}{2\omega^2}) = exp(-\frac{1}{2\omega^2}\gamma'(I_T \otimes Q)\gamma) \tag{1.152}$$

where ω is some smoothness parameter. $Q = D'D$, and D is the $(T-1) \times T$ matrix with $D_{tt} = 1$ for all $t = 1, \ldots, T-1$, $D_{t-1,t} = -1$ for all $t = 2, \ldots, T$ and zeros otherwise. It implies that $\gamma_{i,t} - \gamma_{i,t-1} \sim N(0, \omega^2)$, or equivalently $D\gamma_i \sim N(0, \omega^2 I_{T-1})$. We can alternatively assume normal prior for the second-order difference: $\gamma_{i,t} - 2\gamma_{i,t-1} + \gamma_{i,t+1} \sim N(0, \omega^2)$ or, in matrix form, $D^{(2)}\gamma_i \sim N(0, \omega^2 I_{T-1})$.

Parameters β and σ^2 are imposed with a noninformative joint prior distribution: $p(\beta, \sigma^2) \propto \sigma^{-2}$. The joint prior of all unknown parameters (β, σ, γ) is then obtained as below

$$p(\beta, \sigma, \gamma) \propto \sigma^{-1} \prod_{i=1}^{n} exp - \frac{\gamma_i' Q \gamma_i}{2\omega^2} = \sigma^{-1} exp(-\frac{1}{2}\gamma'(I_n \otimes Q)\gamma) \tag{1.153}$$

Implementation

For a balanced panel data, the likelihood function of the model (1.151) is

$$\mathcal{L}(y|X, \beta, \sigma, \gamma) \propto \sigma^{-nT} exp[-\frac{1}{2\sigma^2}(y - X\beta - \gamma)'(y - X\beta - \gamma)] \tag{1.154}$$

Using this information, the joint distribution of unknown paramters can then be updated through the Bayesian rule: $p(\theta|x) = \frac{f(x|\theta)p(\theta)}{f(x)} = \frac{f(x|\theta)p(\theta)}{\int f(x|\theta)p(\theta)d\theta}$. The posterior distribution is thus derived as

$$\begin{aligned} p(\beta, \sigma, \gamma|y, X, \omega) \quad \propto \quad & \sigma^{-(nT+1)} exp[-\frac{1}{2\sigma^2}(y - X\beta - \gamma)'(y - X\beta - \gamma)] \\ & \times exp[-\frac{1}{2\omega^2}\gamma'(I_n \otimes Q)\gamma] \end{aligned} \tag{1.155}$$

Gibbs sampling is employed here to estimate the joint distribution. Since drawing the random vector (β, σ, γ) simultaneously from above distribution is not feasible in practice, we first derive the standard conditional distribution for each parameter, and take draws of one parameter per time cyclically from their respective conditional distribution. That is, with the kth draw in hand, the $(k+1)$th draw will be obtained in the following order:

1) Draw $\beta^{(k+1)}$ from $p(\beta|\sigma = \sigma^{(k)}, \gamma = \gamma^{(k)})$;

2) Draw $\sigma^{(k+1)}$ from $p(\sigma|\beta = \beta^{(k+1)}, \gamma = \gamma^{(k)})$;

3) Draw $\gamma^{(k+1)}$ from $p(\gamma|\beta = \beta^{(k+1)}, \sigma = \sigma^{(k+1)})$;

And the derived conditional distributions are listed below

$$\beta|\sigma, \gamma, \omega, y, X \sim N(\hat{\beta}, \sigma^2(X'X)^{-1}) \tag{1.156}$$

$$\tfrac{1}{\sigma^2}(y - X\beta - \gamma)'(y - X\beta - \gamma)|\beta, \gamma, \omega, y, X \sim \chi^2_{nT} \tag{1.157}$$

$$\gamma|\beta, \sigma, \omega, y, X \sim N(\hat{\gamma}_i, \sigma^2\omega^2 V) \qquad for \quad i = 1, \dots, n \tag{1.158}$$

where

$$\hat{\beta} = (X'X)^{-1}X'(y - \gamma)$$

$$\hat{\gamma}_i = \omega^2 V(y_i - x_i\beta) \qquad \text{and} \qquad V = (\sigma^2 Q + \omega^2 I_T)^{-1}$$

The choice of smoothing parameter ω can be determined through cross validation in the Bayesian context

$$CV(\omega) = (nT)^{-1}\sum_{i=1}^{n}(y_i - x_i\bar{\beta}_{-i,\omega} - \bar{\gamma}_{-i,\omega})'(y_i - x_i\bar{\beta}_{-i,\omega} - \bar{\gamma}_{-i,\omega}) \tag{1.159}$$

where $\bar{\beta}_{-i,\omega}$ and $\bar{\gamma}_{-i,\omega}$ represent the coefficients obtained from the sample excluding individual i.

Alternatively, Bayesian method can be used to find ω. Suppose that ω has a prior distribution: $\frac{\bar{q}}{\omega^2} \sim \chi^2_{\bar{n}}$. And the updated posterior conditional distribution is

$$\frac{\bar{q} + \sum_{i=1}^{n}\gamma_i' Q\gamma_i}{\omega^2}|\beta, \sigma, \gamma, y, X \sim \chi^2_{\bar{n}}$$

Model 2

Following the setup of Kneip, Sickles and Song (2012), the individual effects $\varphi_i(t)$ is assumed to be in a space generated by some basis functions:

$$y_{it} = x'_{it}\beta + \phi'_t\gamma_i + v_{it} = x'_{it}\beta + \sum_{g=1}^{G}\phi'_{tg}\gamma_{ig} + v_{it} \quad (1.160)$$

where γ_i represents individual effects, and ϕ_i is the common factor that varies over time. Such a specification makes it more suitable for interpretation about different economic effects that associated with each individual unit. The model can be stacked into matrix form:

$$y = X\beta + (I_n \otimes \Phi)\gamma + v = X\beta + (I_T \otimes \Gamma)\phi \quad (1.161)$$

Implementation

We still employ Gibbs sampling for estimation, and make some modification to incorporate the new settings. In the same spirit, the time-varying common factors are assumed to follow some empirical prior distribution, which is controlled by the smoothing parameter ω.

$$\begin{aligned} p(\phi_1, \phi_2, \ldots, \phi_T|\phi_1, \phi_T) &\propto exp\Big(- \frac{\sum_{t=2}^{T}(\phi_t - \phi_{t-1})'(\phi_t - \phi_{t-1})}{2\omega^2}\Big) \\ &= exp(-\frac{1}{2\omega^2}tr\Phi' Q\Phi) \quad (1.162) \end{aligned}$$

And the individual-specific effects are assumed to follow a G dimensional normal distribution: $\gamma_i \overset{i.i.d.}{\sim} N_G(\bar{\gamma}, \Sigma)$. $\bar{\gamma}$ and Σ are some hyperparameters, whose joint distribution is denoted as $p(\bar{\gamma}, \Sigma)$. The inverted Wishart distribution is often used as the conjugate prior for covariance matrix of a multinormal distribution. Thus we assume

$$\begin{aligned} p(\Sigma) &\propto |\Sigma|^{-(\bar{\nu}+1)/2}exp(-\frac{1}{2}\bar{A}\Sigma^{-1}) \quad (1.163) \\ p(\bar{\gamma}|\Sigma) &\propto const \quad (1.164) \end{aligned}$$

where $A = \bar{A} + \sum_{i=1}^{n}(\gamma_i - \bar{\gamma})(\gamma_i - \bar{\gamma})'$.

Parameter vector (β, σ) is imposed with the same noninformative prior: $p(\beta, \sigma) \propto \sigma^{-1}$. Given all these priors, we can form the joint prior distribution for all unknown parameters, and update through the Bayesian rule using information from certain panel data. And to apply Gibbs sampling, we still need to first derive the standard conditional distribution for each parameter.

It is shown by the Monte Carlo simulations that the Bayesian approach outperforms several representative estimators we have discussed (e.g. Cornwell, Schmidt, and Sickles (1990), Park, Sickles, and Simar (1998, 2007)) under a variety of data generating processes.

1.4 Remarks

There have been a lengthy discussion and a variety of extensions on the stochastic frontier analysis since it was first proposed in 1977. As we can see from the reviews of different specifications in the previous section, each setting has its own advantages and drawbacks, and we need a method to evaluate these models based on the data we have. The standard procedure is to first calculate the score of an information criterion for each model, and then pick the one with the highest score as the "true" model. There are some notable problems concerning this model selection procedure. First, there are multiple information criteria evaluating model performance. These criteria have different emphases, and might lead to different "true" models. Second, the selection is base on the in sample performance, and the out-of-sample explaining power may not be satisfactory. Third, by treating the selected model as the "true" one, the possibility of other specifications is ignored and the variance of the parameter of interest can be substantially underestimated. An alternative to the model selection procedure is the model averaging approach, which can be used to combine a set of competing models and form a more efficient and robust estimator. I will explore the model averaging approach in Chapter 3.

This chapter has reviewed the methods of estimating productivity given the outputs and inputs. In the next chapter, I will explore the relations between the outputs and inputs. As shown in Chapter 2, by utilizing the duality theory, we can decompose the output price into the prices of inputs and gain more insight of the production/value-generating process.

Chapter 2

The Decomposition of Housing Prices with Shephard's Dual Lemma[1]

2.1 Introduction

During the recent international financial distresses that led to the Great Recession the balance sheets of banks were (and continue to be) under severe oversight by banking regulators. Basel II requires that bank holding companies have combined Tier 1 and Tier 2 capital ratios of at least 8%. These capital ratios reflect the percentages of a bank's capital to its risk-weighted assets. One of the major assets of a bank is its portfolio of residential mortgages, which typically make up between 20% and 25% of a bank's total assets. As of June 2012 the total residential loans outstanding in the US were roughly US$ 3.5 trillion.

The precipitous drop in the value of such holdings, measured in the valuations from mark to market accounting instead of the more conventional book valuations, was the key component in causing the Great Recession.

There are a set of reasonable questions to ask about the way in which banks, banking regulators, and rating agencies establish the valuation of a bank's real estate holdings and thus the bank's solvency.

1. What methods do banks use to value their real estate holdings?

[1]Part of this chapter is based on the presentation "Pricing Characteristics: An Application of Shephard's Dual Lemma" in North American Productivity Workshop VIII, coauthored with Dr. Grosskopt Shawna, Dr. Rolf Färe, and Dr. Robin Sickles

Figure 2.1 : Real Estate Loans Holdings

http://research.stlouisfed.org/fred2/series/REALLN?cid=100

2. Are these methods the same across real estate holdings in different regions of a country?

3. Are these methods the same across real estate holdings in different countries?

4. Are these methods transparent and easily calculated?

5. Are these methods consistent with the economic assumptions underlying the value at risk paradigm used to set reserve benchmarks?

This chapter addresses these questions by providing a transparent and easily implementable methodology for constructing real estate price indices based on economic assumptions in keeping with the other economic assumptions underlying many of the regulatory criteria used by banking regulators in, for example, the assessment of the value at risk paradigm that provides banks with the rules for setting reserve benchmarks. We use an input distance function to describe the value generating process of residential properties (also referred to as the "dwelling unit"), which is a euphanism for the output of a production process whose price is the price of the residence. The inputs into the production process are a set of characteristics that a buyer demands, proxied in our empirical analysis by the square footage of the structure, the amount of land on which the structure sits, and the age of the structure. The specific form of the input distance function is translog and the shadow prices are derived based on duality theory. We use these shadow prices to construct an imputed residential properties (dwelling unit) price index and compare it to those generated by more conventional hedonic methods. These methods and their advantages and disadvantages are examined by Good et al. (2008), among many others. We implement our modeling approach using a single-output, multi-input distance function. The standard method to estimate the parameters of such an input distance function is to normalize the regression model by moving one of the input variables to the left hand side and to treat it as the dependent variable and view the unknown distance as a right hand side error that is combined with the normal idiosyncratic error in the regression model. Of course this former source of error is bounded due to the bounded nature of the distance function itself. We utilize several methods to address this aspect of the composed error and compare our imputed residential price index across different approaches. The methods are 1) corrected ordinary least squares with White-type robust standard errors; 2) time dummy least squares regression with White-type robust standard errors; 3) stochastic frontier model (Belotti et al., 2012).

The chapter uses data from Diewert (2010) to construct price indices for residential properties in a small Dutch town using quarterly data from 2005 II to 2008 II. We compare results from our approach, which involves an application of Shephard's Dual Lemma (Shephard, 1953), with methods employed in Diewert and Shimizu (2013) that utilize stratification techniques and various hedonic treatments. Compared with these hedonic regression approaches, our empirical models can simultaneously estimate the shadow prices of the main property characteristics without suffering from typical problems of collinearity amoung the quality charceristics. The residential property (dwelling unit) price indices that we derive from our estimations show similar trends to Diewert's results but appear to be less volatile.

2.2 The Theoretical Model

It is known from mathematics that a gradient vector of a function belongs to the dual space of its variables. In economics, a classic example is Shephard's lemma, which says that the derivative of the cost function with respect to a price is an input quantity, i.e., the derivative takes us from the price space to the quantity space. In this chapter, we use Shephard's dual lemma (Shephard, 1953), which says that the derivative of the input distance function with respect to an input is an input price. Next, we apply the dual lemma and use it to derive shadow prices of property characteristics.

Assume that a good is endowed with $z = (z_1, \ldots, z_N)$ characteristics. These characteristics in turn generate a value of the good equal to $p \geqq 0$.[2] We model this relation with an input correspondence

$$L(p) = \{z \in \Re^N_+ : z \text{ generates value } p\}, p \geqq 0. \tag{2.1}$$

[2]The value of the good is $p = wz$, where $w = (w_1, \ldots, w_N) \in \Re^N_+$ are the unknown prices of the characteristics.

This correspondence can in turn be given a functional representation via Shephard (1953) input distance function

$$D_i(p, z) = \sup\{\lambda : z/\lambda \in L(p)\}, \tag{2.2}$$

which, with some mild assumptions on $L(p)$[3] provides a complete characterization of the input correspondence, i.e.,

$$D_i(p, z) \geqq 1 \Leftrightarrow z \in L(p). \tag{2.3}$$

Dual to the input distance is the cost function

$$C(p, w) = \min\{wz : z \in L(p)\} \tag{2.4}$$

where $w \in \Re^N_+$ are the (unknown) prices of the characteristics. From the duality between $C(p, w)$ and $D(p, z)$ we find the shadow price vector of the characteristics to be

$$w^s = \frac{p \cdot \nabla_z D_i(p, z)}{D_i(p, z)}, \tag{2.5}$$

i.e., the unknown price vector w can be derived from observed data (p, z). (See Appendix for the proof.)

To parameterize the distance function (2.2), we begin by choosing the broad family including generalized quadratic (Chambers, 1988) or transformed quadratic (Diewert, 2002) since these functions are linear in their parameters and provide a second-order Taylor approximation. In addition, if these functions are homogeneous of degree

[3]See Färe and Primont (1995).

+1 (like the input distance function is in inputs), they take two specific functional forms (Färe and Sung, 1986). Either they are quadratic means of order ρ (Denny, 1974; Diewert, 1976) or translog (Christensen et al., 1971). The former function has only second-order parameters, while the translog has both second- and first-order parameters. Having no zeros in our data, we choose to estimate the translog formulation of the distance function (2.2). From these estimates, by applying (2.5), we can derive the desired input-characteristics shadow price vector.

To relate our model to the single-output production models by Thorsnes (1997) and McMillen (2003), assume that the technology $L(p)$ exhibits constant returns to scale, i.e.,

$$L(\lambda \cdot p) = \lambda \cdot L(p),\ \lambda > 0. \tag{2.6}$$

Then and only then can the input distance function be written as

$$D_i(p, z) = \frac{1}{p} D_i(1, z), \tag{2.7}$$

noting that p is a scalar. Assuming that z belongs to the isoquant of $L(p)$ so that

$$D_i(p, z) = \frac{1}{p} D_i(1, z) \geq 1, \tag{2.8}$$

then our distance function formulation takes the traditional single-output production function expression,

$$p \leq D_i(1, z). \tag{2.9}$$

Thus the Cobb-Douglas model by McMillen and the CES model of Thorsness are special cases of (2.9) with a pricing formula (2.5) now given by

$$w^s = \frac{p \cdot \nabla_z D_i(1, z)}{D_i(1, z)}. \tag{2.10}$$

Upon applying(2.9) we find that

$$w^s = \nabla_z D_i(1, z) \tag{2.11}$$

2.3 Data Description

In the empirical application we make use of data kindly provided to us by Professor W. Erwin Diewert and analyzed in Diewert (2010). There are small discrepancies between the data analyzed in Diewert (2010) and the data we use here. The comparison of some key summary statistics are shown in Table 2.2. The data we use in our application consist of 2280 observations on quarterly sales of detached houses (what we label residential property or dwelling unit) over 14 quarters for the town of 'A' in the Netherlands. This is a small city (roughly 60,000 inhabitants) and its exact location and of course the name has been masked by Statistics Netherlands. Transactions on dwelling units begin in the first quarter of 2005 and end the second quarter of 2008.

2.4 Empirical Model

2.4.1 General Specification

To justify our choice of the translog model we use in our empirical analyses below, we first note that the input distance function is homogeneous of degress +1 in inputs. It is known, see Färe and Sung (1986), that a homogeneous generalized quadratic function with arguments x_1, x_2 such as

$$\begin{aligned}\varphi^{-1}(F(x_1,x_2)) = a_0 + a_1 f(x_1) + a_2 f(x_2) + b_1 f(x_1) f(x_1) \\ + b_2 f(x_2) f(x_2) + b_3 f(x_1) f(x_2) \qquad (2.12)\end{aligned}$$

takes either a translog or generalized mean of order ρ functional form. As the latter function has only second-order parameters, the translog is the preferred choice in our empirical analysis. In the application below we specificy the arguments as characteristics of property services, which we have denoted as $z_i, i = 1, ..., n$.

In the case of detached residential properties, we treat each dwelling unit as the output whose price is influenced by a number of characteristics. The main variables used are

p: value of the residential dwelling unit;

L: land area of the property;

S: floor space area of the structure;

A: age of the structure.

The land area L, floor space area S, and structure age A are treated as the three input characteristics ($z_i, i = 1, 2, 3$), and p is the value of output. The translog input distance function is specified as below:

$$\begin{aligned}\ln D_i = \alpha_0 + \alpha_1 \ln p + \frac{1}{2}\alpha_{11}(\ln p)^2 + \beta_1 \ln S + \beta_2 \ln L + \beta_3 \ln A + \frac{1}{2}\beta_{11}(\ln S)^2 \\ + \frac{1}{2}\beta_{22}(\ln L)^2 + \frac{1}{2}\beta_{33}(\ln A)^2 + \beta_{12} \ln S \ln L + \beta_{13} \ln S \ln A \\ + \beta_{23} \ln L \ln A + \gamma_1 \ln S \ln p + \gamma_2 \ln L \ln p + \gamma_3 \ln A \ln p.\end{aligned} \tag{2.13}$$

The assumptions of homogeneity of degree +1 in inputs and symmetry of the cross-product terms imply the following restrictions on the parameters:

$$\beta_1 + \beta_2 + \beta_3 = 1 \tag{2.14}$$

$$\sum_{l=1}^{3} \beta_{kl} = 0,\ k = 1, 2, 3 \tag{2.15}$$

$$\gamma_1 + \gamma_2 + \gamma_3 = 0 \tag{2.16}$$

$$\beta_{kl} = \beta_{lk}, \quad k, l = 1, 2, 3. \tag{2.17}$$

Utilizing these restrictions, the distance function can be rewritten as

$$\begin{aligned}\ln D_i &= \alpha_0 + \alpha_1 \ln p + \frac{1}{2}\alpha_{11}(\ln p)^2 + \beta_1 \ln \frac{S}{A} + \beta_2 \ln \frac{L}{A} + \ln A \\ &+ \frac{1}{2}\beta_{11}(\ln \frac{S}{A})^2 + \frac{1}{2}\beta_{22}(\ln \frac{L}{A})^2 + \beta_{12} \ln(\frac{S}{A}) \ln(\frac{L}{A}) \\ &+ \gamma_1 \ln \frac{S}{A} \ln p + \gamma_2 \ln \frac{L}{A} \ln p\end{aligned} \tag{2.18}$$

From the model, we know that the shadow price vector is

$$w^s = \frac{p \cdot \nabla_z D_i(p, z)}{D_i(p, z)} \tag{2.19}$$

where p is the value of output, and z is the vector of all inputs. Denoting w_S, w_L,and w_A as the shadow prices of the structure, the land, and the age of the structure respectively, we can then derive the explicit expressions for the shadow prices of the input characteristics as:

$$w_S = \frac{p}{S}(\beta_1 + \beta_{11} \ln S + \beta_{12} \ln L + \beta_{13} \ln A + \gamma_1 \ln p) \tag{2.20}$$

$$w_L = \frac{p}{L}(\beta_2 + \beta_{22} \ln L + \beta_{12} \ln S + \beta_{23} \ln A + \gamma_2 \ln p) \tag{2.21}$$

$$w_A = \frac{p}{A}(\beta_3 + \beta_{33} \ln A + \beta_{13} \ln S + \beta_{23} \ln L + \gamma_3 \ln p). \tag{2.22}$$

2.4.2 Specification with Constant-Returns-to-Scale Assumption

As indicated in the section talking about the theoretical model, if the production exhibits constant returns to scale (CRS), our model can be related to the single-output production models by Thorsnes (1997) and McMillen (2003). Plug the translog input distance function into the general expression (2.9), we can obtain the regression equation under the CRS constraint, and the inequality will be captured by the input distance. It is the same as the general specification when one sets $\alpha_{11} = \gamma_1 = \gamma_2 = 0$. Thus the regression equation with the CRS constraint is,

$$\begin{aligned}\ln p \quad &= \alpha_0 + \beta_1 \ln S + \beta_2 \ln L + \beta_3 \ln A + \tfrac{1}{2}\beta_{11}(\ln S)^2 + \tfrac{1}{2}\beta_{22}(\ln L)^2 \\ &+\tfrac{1}{2}\beta_{33}(\ln A)^2 + \beta_{12}\ln S \ln L + \beta_{13}\ln S \ln A + \beta_{23}\ln L \ln + \ln D_i\end{aligned} \tag{2.23}$$

This model with CRS assumption is, in essence, the same as the hedonic regression with S, L and A as the characteristics.

Again, we impose the assumption of homogeneity of degree 1. Compared with the general case, now there are no terms on the right-hand side involving $\ln p$, thus we do not have the γ coefficients any more. The homogeneity condition now implies equations (2.14), (2.15) and (2.17). With these three constraints, we can rewrite the equation (2.24) as

$$\begin{aligned}\tilde{\ln} p \quad &= \alpha_0 + \beta_1 \ln \tfrac{S}{A} + \beta_2 \ln \tfrac{L}{A} + \tfrac{1}{2}\beta_{11}(\ln \tfrac{S}{A})^2 \\ &+\tfrac{1}{2}\beta_{22}(\ln \tfrac{L}{A})^2 + \beta_{12}\ln \tfrac{S}{A} \ln \tfrac{L}{A}\end{aligned} \tag{2.24}$$

where $\tilde{\ln} p = \ln p - \ln A$. Under this new specification, the shadow prices of the three characteristics are

$$w_S \quad = \quad \frac{1}{S}(\beta_1 + \beta_{11}\ln S + \beta_{12}\ln L + \beta_{13}\ln A) \tag{2.25}$$

$$w_L \quad = \quad \frac{1}{L}(\beta_2 + \beta_{22}\ln L + \beta_{12}\ln S + \beta_{23}\ln A) \tag{2.26}$$

$$w_A \quad = \quad \frac{1}{A}(\beta_3 + \beta_{33}\ln A + \beta_{13}\ln S + \beta_{23}\ln L) \tag{2.27}$$

2.4.3 Construction of the price index

For comparison purposes, we use the matched model Fisher index discussed in Diewert (2010). Diewert constructs price indices for land and for structures that make up the dwelling unit of a particular age and we do likewise. Dwelling units are grouped into

45 cells consisting of 3 categories for land area (small, medium, large), 3 categories for structures (small, medium, large) and 5 groups for age. The break points for the size of land and structure are chosen in a way that about 50% of the units fall in the medium group, and roughly 25% units are in small and large group respectively. The break points for land area are $L_1 = 160\,m^2$ and $L_2 = 300\,m^2$, and the break points for structure size are $S_1 = 110\,m^2$ and $S_2 = 140\,m^2$. Age of the structure is identified by when the structure was built and ranges from 1960 to 2008. For houses built in 2000-2008, $A = 2$; $A = 3$ for 1990-1999 and so on[4]. Using the structure (or land) prices derived from the above model, we define w_n^t to be the average structure (land) price for properties in cell n that were sold in period t,

$$w_n^t = \frac{\sum_{i\in n} w_i^t z_i^t}{\sum_{i\in n} z_i^t} = \frac{\sum_{i\in n} w_i^t z_i^t}{z_n^t} \tag{2.28}$$

where z_i^t and w_i^t represent the structure (land) area and its corresponding shadow price in cell n. As there is no transaction in some cells across two compared periods, we define $S(s,t)$ to be the set of cells n that have at least one transaction in each of the quarters s and t. The indices are then computed over these matched components. The Laspeyres (L) and Paasche (P) indices for periods s and t are:

$$w^L(s,t) = \frac{\sum_{n\in S(s,t)} w_n^t z_n^s}{\sum_{n\in S(s,t)} w_n^s z_n^s} \tag{2.29}$$

$$w^P(s,t) = \frac{\sum_{n\in S(s,t)} w_n^t z_n^t}{\sum_{n\in S(s,t)} w_n^s z_n^t}. \tag{2.30}$$

Diewert constructs the (ideal) Fisher index by taking the geometric mean of the above two indices:

$$W^F(s,t) = [w^L(s,t)w^P(s,t)]^{\frac{1}{2}} \tag{2.31}$$

[4]In Diewert (2010), the range of age A is 0 to 4. To accommodate our use of the translog distance function, we shift the range of A to 2 to 6.

Two sets of indices are constructed for both structure and land prices. One is a fixed Fisher index, which uses the first quarter as the base period. The other is a chained index: we construct the Fisher index for every two consecutive periods, and the chained index for period t is computed as:

$$I_F^t = W^F(1,2)W^F(2,3)\cdots W^F(t-1,t)$$

2.5 Regression and Results

2.5.1 General Specification

We utilize three different regression methods to estimate the input distance function (2.13). The methods are 1) corrected ordinary least squares with White-type robust standard errors; 2) time dummy least squares regression with White-type robust standard errors; 3) stochastic frontier model (Belotti et al. (2012)). The results are shown in Table 2.1.

The input distance function has long been utilized in theoretical papers to measure the technical efficiency level of a production process. The input distance is bounded from below by unity, which represents a technically efficient level of production. The use of the distance function in empirical work can be traced back to Färe et al. (1985) wherein linear programming was used to estimate non-parametric distance functions and measure technical efficiency. Starting in the 1990's, researchers also considered parametric functions and used econometric methods for estimation. Lovell, Travers, Richardson, and Wood (1994) specified a translog distance function, and used OLS to estimate the parameters. The translog functional form was also used in Coelli and Perelman (1996); coe (2000) and also was estimated using OLS.

In our application, the land, structure and age of a detached property are regarded as the inputs, which are used to "produce" this property, and input distance D_i gives us an estimate of the (in)efficiency level compared to the efficiency frontier. As discussed in previous sections, we choose translog functional form for the input distance. In most empirical studies using translog input distance functions with m outputs and k inputs, the negative of the logged kth input is treated as the dependent variables and is regressed upon the rest terms, and the negative of the logged input distance is treated as an error term. Employing this method, the equation to be estimated is rearranged as:

$$\begin{aligned} -\ln A &= \alpha_0 + \alpha_1 \ln p + \frac{1}{2}\alpha_{11}(\ln p)^2 + \beta_1 \ln \frac{S}{A} + \beta_2 \ln \frac{L}{A} \\ &+ \frac{1}{2}\beta_{11}(\ln \frac{S}{A})^2 + \frac{1}{2}\beta_{22}(\ln \frac{L}{A})^2 + \beta_{12} \ln(\frac{S}{A}) \ln(\frac{L}{A}) \\ &+ \gamma_1 \ln \frac{S}{A} \ln p + \gamma_2 \ln \frac{L}{A} \ln p - \ln D_i \end{aligned} \tag{2.32}$$

Note that by the definition of the input distcance function, $D_i \geq 1$, thus $-\ln D_i \leq 0$ and can be interpreted as a one-sided error. Now the objective function (2.32) fits into the production frontier models, and we can utilize frontier models for estimation, such as those devleoped in Stata (Belotti et al., 2012).

OLS can be used to estimate the coefficients of the distance function. After obtaining the estimates of all coefficients, we correct the estimated intercept by adding the largest positive residual such that the adjusted function bounds all the observed points from below. This gives us the corrected OLS (COLS) estimates. Some researchers suggest that there exists potential simultaneous equation bias, as one of the input variables (here "$-\ln A$") is assumed to be endogenous and there are ratios of inputs in the right-hand-side ("$\ln(S/A)$" and "$\ln(L/A)$"). However, as shown in Coelli (2000), under the cost minimization assumption, OLS provides consistent estimates of the parameters. Though our observations are detached properties with access to basically the same amenities, we still use robust errors in the regression to account for possible heteroskedasticity. The parameter estimates are shown in Table 2.1, and the R^2 is 0.8726.[5]

The data begins in the first quarter of 2005 and ends in the second quarter of 2008. Considering that property prices might be affected by the date of the transaction due to changing market conditions and other factors proxied by time, we add dummy variables to account for yearly effects and reestimate the model by COLS. The adjusted R^2 improves a bit to 0.878.

As noted in Coelli and Perelman (1996), both the linear programming technique (see, for example, Färe et al. (1993)) and the COLS method assume the distance to full efficiency is due entirely to technical inefficiency. To account for the effect of data noise, we can employ stochastic frontier methods. Adding a pure noise term to equation (2.32), we now have a composite error $\epsilon_i = v_i - lnD_i = v_i - u_i$, where u_i is the idiosynchratic error assumed to be $i.i.d.N(0, \sigma_v)$; u_i is our logged input distance, representing technical inefficiency. Here we assume u_i follows the half-normal distribution, i.e., $u_i = |z_i|$, $z_i \sim N(0, \sigma_u)$, and use standard ML techniques for

[5]We have also utilized a modified IV estimator to address the potential for endogeneity in the log ratio's of the other input characteristics and the age characteristic as well as considered other normalizaitons, for example normalizing with respect to the price of the residential dwelling unit. However, the input distance function we utilize is consistent with cost minimization, holding fixed output and input prices, and thus one may question why price should be considered endogeneous. As one would expect, the renormalization made little difference in our results. Unfortunately, instrumenting the right-hand-side variables that interact with the left-hand-side variable ln(A) was not possible due to a dearth of potential instruments.

these estimates. These and other stochastic frontier models can be estimated using Stata (Belotti et al., 2012; **?**). The stochastic frontier estimates from equation (2.32) are reported in Table 2.1.

2.5.2 Comparisons with the methods of Diewert and Shimizu (2013)

Diewert and Shimizu (2013) employed hedonic regression techniques to decompose the price of residential property in Tokyo into land and structure components, and constructed constant quality indices for land and structure prices respectively. In this section we use three different models from **?** to fit our real estate data from town "A" in Netherlands. We also construct the price indices for structures and land and compared these results with those derived above.

In traditional hedonic regression models, the price of one unit of commodity under study is assumed to depend on a function of its characteristics. Diewert (2003), among others, provides the microeconomic support for this method. If one assumes that an agent can consume an hedonic commodity Z with a set of characteristics $z = (z_1, \ldots, z_k)$ and other commodity X, then the consumption of Z units of the hedonic commodity gives a subutility of $f(z_1, \ldots, z_k)$, and the consumption of Z and X together generates utility $u = U(Z, X)$. Denote p^t and w^t to be the prices of the general commodity and hedonic commodity at period t, respectively. The consumer then faces a standard cost minimization problem:

$$\min_{X,Z}\{p^t X + w^t Z : U(X, Z) = u^t\}$$

Under some regulatory conditions, the price of the hedonic commodity can be expressed as:

$$w^t = \rho_t f(z_1, \ldots, z_t)$$

That is, the price of the hedonic good is the product of some time-dependent effects and the utility a consumer gets from its characteristics. The hedonic regression methods are widely used in real estate studies. It explains the property value based on actual choices of people, and it can be comfortably modified to take into account the interaction between the characteristics of the property itself and the effects of the surrounding environment. Some limitations also exist with the classic hedonic pricing model as they do in our approach that utilizes the input distance function. For example, the model assumes that the price of the residential dwelling unit can change immediately after the change in one or some of its characteristics, whereas in reality, there may be a substantial time lag. The model assumes that there are a variety of properties in the market so that consumers can choose the one with the desired combination of characteristics, which is only possible if the market is deep. Another problem is the multicollinearity among characteristics, which we will encounter below. A consumer may find that some properties have all the good characteristics, while some alternatives are inferior in all aspects.

The basic paradigm in Diewert and Shimizu (2013) is referred to as a builder's model, which is based on the assumption that the value of a residential property is the sum of the value of the land on which it is built and the construction cost of its structure. Considering that the structure's price usually falls as the structure ages, they assume the constant quality structure to be a function of its age and a constant depreciation rate over all time periods. Following the notation used above, we can specify the model as below,

$$p_{it} = w_{L,t}L_{it} + w_{S,t}(1 - \delta A_{it})S_{it} + \epsilon_{it} \qquad i = 1, \ldots, N(t);\ t = 1, \ldots, 14 \tag{2.33}$$

where $N(t)$ is the number of properties sold in period t. One concern about this model is the multicollinearity between the land size and structure size: we would expect a larger house structure to be built with a larger land area. Our data shows that the correlation between land size and structure size is 0.6278. Thus the coefficient estimates of land and structure may not be reliable in the sense that small variations in the data may result in erratic changes in the estimates. To deal with this multicollinearity, Diewert and Shimizu (2013) assumed the price of a constant quality structure was proportional to a property construction cost index published by the relevant authority. This method was also employed in Diewert (2010) as one approach to exploring the price change of residential properties, in which the index used was the New Dwelling Output Price Index (NDOPI) published by the Central Bureau of Statistics of Netherlands. The resulting land price index from Diewert (2010) will be included as part of our comparison.[6] Though Diewert examined the same dataset for town "A" as this chapter, the focus of the research was on the price index construction of the residential properties rather than their characteristics. Thus we use applicable methods from Diewert and Shimizu (2013) to make comparisons, which focused more on the decomposition of property price into land and structure components.

We set the constant quality structure price to be proportional to the New Dwelling Output Price Index (NDOPI) mentioned above, the same index used in Diewert (2010) as we examine the same dataset. As Diewert assumes that the price of the structure is proportional to the construction cost index, we follow this assumption in order to provide results as comparable as possible to the hedonic methods he uses and set $w_{S,t} = w_S P_{C,t}$, where P_{C_t} represents the exogenous cost index. The model can then

[6]Compared to what we describe here, Diewert (2010) used a different method to construct the land price index.

be written as

$$p_{it} = w_{L,t}L_{it} + w_{S,t}P_{C_t}(1 - \delta A_{it})S_{it} + \epsilon_{it} \qquad i = 1, \ldots, N(t);\ t = 1, \ldots, 14 \qquad (2.34)$$

We denote this model as DS0, corresponding to the basic builder's model in Diewert and Shimizu (2013).[7] Coefficients of this nonlinear model are estimated by minimizing the mean squared error of the residual term. $w_{L,t}$ is interpreted as a suitable constant quality land price for all residential properties sold in period t, and the constant quality land price index for quarter t is defined by Diewert and Shimizu to be

$$I_{L,1t} = w_{L,t}/w_{L,1}. \qquad (2.35)$$

The age-adjusted constant quality structure is defined to be $(1 - \delta A_{it})S_{it}$ and the corresponding structure price index for quarter t is defined by Diewert and Shimizu to be

$$I_{S,1t} = (w_S P_{C_t})/(w_S P_{C_1}) = P_{C_t}/P_{C_1}. \qquad (2.36)$$

The second model employs splines on both the land size and structure age. Empirical evidence indicates that the growth rate of the property land prices vary with land size. To model the possible changes in land prices as land area increases, Diewert and Shimizu (2013) divided all observations into 3 groups based on the land size, and assumed that the land price in each group was linear in land size. Diewert (2010), which analyzed the same dataset as we do, also considered the possibility of changing land prices over different land area ranges in one of his approaches to measure

[7]In Diewert and Shimizu (2013), the basic builder's model also included the 21 dummy variables indicating different wards in Tokyo, to account for possible differences in land prices. In our dataset, all observations are detached residential properties with access to basically the same amenities, thus the difference in locations has little effect on property prices. The R^2 is quite satisfactory in our regression.

the property price. The method used in Diewert (2010) was the same as Diewert and Shimizu (2013) wherein all observations were grouped into 3 categories based on their land sizes and land price was assumed to be piecewise linear in land areas. To make our results comparable with that from Diewert (2010), we divide our data into the same 3 groups, with break points at $L_1 = 160$ and $L_2 = 300$. This generates a grouping with approximately 50% of the properties in the middle group and 25% in the lower and upper groups. The piecewise linear relative land value function is thus specified as

$$\begin{aligned} f_L(L_{it}) = \mathrm{DL}_{it,1}\gamma_1 L_{it} + \mathrm{DL}_{it,2}(\gamma_1 L_1 + \gamma_2(L_{it} - L_1)) \\ + \mathrm{DL}_{it,3}(\gamma_1 L_1 + \gamma_2(L_2 - L_1) + \gamma_3(L_{it} - L_2)) \end{aligned} \tag{2.37}$$

where $\mathrm{DL}_{it,j}$, $j = 1, 2, 3$ are the land dummy variables with $\mathrm{DL}_{it,j} = 1$ indicating that the property falls into category j and $\mathrm{DL}_{it,j} = 0$ meaning that observation i is not in category j. γ_k, $k = 1, 2, 3$ are unknown parameters to be estimated.

We also group all the observations into three categories based on the age of the structure in order to be consistent with Diewert as he points out that depreciation rates will not be the same for structures of different ages. The break points are chosen to be $A_1 = 1$ and $A_2 = 2$, and the categorical dummy variables for ages are denoted as $\mathrm{DA}_{it,j}$, $j = 1, 2, 3$, with $\mathrm{DA}_{it,j} = 1$ indicating that the property is in category j. This is intended to ensure that the three groups have roughly the same size. However, due to the categorical nature of the data and the unequal number of properties in each of the five age categories (ages are indicative of the building decade and range from 0 to 4 with structures with ages new to 10 years a much larger number than the other 4 categories) the actual number of properties in the three groups are unequal and set at $1052, 481$, and 474, respectively. The piecewise linear depreciation function of the structure's age is defined as

$$g_A(A_{it}) = 1 - (\mathrm{DA}_{it,1}\delta_1 A_{it} + \mathrm{DA}_{it,2}(\delta_2 A_1 + \delta_2(A_{it} - A_1)) \\ + \mathrm{DA}_{it,3}(\delta_1 A_1 + \delta_2(A_2 - A_1)) + \delta_3(A_{it} - A_2)) \tag{2.38}$$

where δ_k, $k = 1, 2, 3$ are the unknow parameters modeling the depreciation schedule of different structure ages.

Now the new model with generalization on land size and structure age can be defined as

$$p_{it} = w_{L,t} f_L(L_{it}) + w_S P_{C_t} g_A(A_{it}) S_{it} + \epsilon_{it} \tag{2.39}$$

We denote this model as DS1. This is also a nonlinear regression model, and we can see that the 3 land relative value parameters (γ's) and the 14 land time parameters ($w_{L,t}$'s) cannot all be identified unless we impose some normalization condition. Thus, we make the normalization that $\gamma_1 = 1$. Note that in this extended model, the marginal land prices for each category, the γ's, are assumed to be the same over all time periods, while $w_{L,t}$ represents the time change in land price for properties in all groups. Thus the constant quality land price index is again defined as in (2.35) for period t, while the constant quality structure price index is again defined as in (2.36)
.

Besides these main characteristics such as the land and structure size, the price of a residential property is also affected by other factors concerning the design of the structure and use of the land space. The model is further extended to adjust for number of rooms, which affect the quality of the structures. To model the effects from number of rooms, we utilize the same technique that first divides all observations into 3 groups and then define the piecewise linear function of number of rooms. In our data, the number of rooms, denoted as N_{it}, ranges from 2 to 10. We first transform the variable to be $R_{it} = N_{it} - 2$, which ranges from 0 to 8, and then divide the properties into three groups based on R_{it}. The break points for R_{it} are chosen to be $R_1 = 2$ and $R_2 = 3$. Let $\mathrm{DR}_{it,j}$ be the dummy variable for the number of rooms, and the piecewise linear function of R_{it} is defined as:

$$
\begin{aligned}
g_B(R_{it}) = \theta_1 + \mathrm{DR}_{it,1}R_{tn} + \mathrm{DR}_{it,2}(\theta_2 R_1 + \theta_3(R_{it} - R_1)) \\
+ \mathrm{DR}_{it,3}(\theta_2 R_1 + \theta_3(R_2 - R_1) + \theta_4(R_{it} - R_2))
\end{aligned}
\tag{2.40}
$$

Then model incorporating this adjustment of structure quality is specified as:

$$
p_{it} = w_{L,t} f_L(L_{it}) + w_S P_{C_t} g_A(A_{it}) g_B(R_{it}) S_{it} + \epsilon_{it} \tag{2.41}
$$

where of course w_S does not vary with time in order to address the high collinearity between structure and land sizes. As the case in DS1, not all parameters can be identified unless some normalizations are made. We thus add two normalization conditions that $\gamma_1 = 1$ and $\theta_1 = 1$, and denote this model as DS2. The constant quality price indices for land and structure are defined the same as in the previous two models.

The land price indices constructed from different models are plotted in Figure 2.2. We can see that the three indices derived using our model exhibit similar trends, while the indices derived from the DS models show somewhat parallel but shifted temporal patterns. The upturns and downturns basically occur at the same time period, but indices generated from our model appear less variable.

Note that the land price indices derived using the DS models climb in the third quarter in 2007, and then fall sharply in the next two quarters, while indices from our model move rather smoothly and exhibit only a moderate fall in 2008 quarter 1. To further explore the difference, we need to jointly look at the property prices and the implied prices of the land and the structure. Using the method described in section 4.1, we can similarly construct the Fisher fixed-base and chained indices for the properties, which are shown in Table 2.4. We can see that the property price index also peaks in 2007 quarter 3, and then falls back a little. When decomposing this temporal change of property price into components of land price and structure price, the DS models assume the structure prices to be proportional to the exogenous construction cost index (NDOPI), which increases all the way over 2007 to the first quarter in 2008. Thus all the decrease in the property price in 2008, quarter 1 is attributed to change in the land price, resulting in a substantial fall of the land price index as we observed. In our model, this decrease in the residential property price index is explained by both land and structure price. Seen from Figure 2.3, the structure price indices derived from our model also fall moderately in the first quarter in 2008.[8]

The use of NDOPI as the structure price index in the DS models is adopted to deal with multicollinearity problems, and cannot be supported within the model. With the same idea, we now introduce exogenous information on land price and derive the structure prices from the DS models. The time-varying parameter $w_{L,t}$ now is set to be $w_L P_{L_t}$, proportional to some land price index, and the structure price parameter $w_{S,t}$ is allowed to freely change over each period. With all other settings remain the same, we still denote the models as DS0, DS1 and DS2. The exogenous land price index is chosen to be the one generated from our SFA model. The resulting structure price indices are shown in Table 2.7 and plotted in Figure 2.4. We can see that with the land price index derived from our model as exogenous information, the structure price indices generated from the DS models show basically the same trend as our structure

[8]In all DS models, the structure price indices are equal to the construction cost index (NDOPI), thus we only use DS0 as a representation.

price indices. Thus, the reliability of land or structure prices generated using the DS models heavily depends on how well the exogenous information reflects the true changing pattern of the price of the other component. The method we propose here can avoid the problem of multicollinearity in hedonic regression and construct the land and structure price indices at the same time. Our methods also tend to smooth the price fluctuations and are less sensitive to market changes compared to the DS models.

2.5.3 Specification with Constant-Returns-to-Scale Assumption

Note that in the general specification, we lack appropriate instruments for the property prices and we put $\ln A$ on the left-hand side as the dependent variable. Under the assumption of constant returns to scale, there are no terms on the right-hand side containing $\ln p$ and we thus circumvent the problem of finding appropriate instrumental variables.

Assuming the white noises are i.i.d with zero mean and constant variance, and taking into account the one-sided log-distance term $\ln D_i$, again we have composite errors in the regression. From the results of the general specification, we can see that the slope coefficients estimates are very close with each other under different regression methods, and only the intercept needs to be corrected for the one-sided error. As only the estimates of slope coefficients are used in the formula generating the shadow prices of characteristics, the choice of which regression method to use does not lead to much difference. Thus I will use robust OLS method for simplicity. To take into account the possible year effects and seasonality, we add dummies for both years and quarters in our regression. The estimation results are shown in Table 2.3. Using formulas (2.27), we can then construct the price indexes of structure and land, which are shown in

Table 2.8. As we add dummies to pick up the yearly and quarterly effects in the regression, we compare the results under the CRS assumption with those obtained in general specification with time dummy variables, and denote as TD-CRS and TD correspondingly. The plots of the indexes are shown in Figure 2.5 and Figure 2.6. As we can see from the figures, compared with the general specification, the structure price indexes under CRS assumption are almost the same, while the land price indexes are lower in all periods. To uncover where this difference comes from, we go back to check the coefficients estimates with and without the CRS assumption. In the general specification with time dummy variables, the coefficient of the cross term of structure and property price, $\ln S \ln p$, is considerably smaller, and the corresponding t-value is near zero, indicating that the coefficient is insignificant. As a contrast, the coefficient of the cross term of land and property price, $\ln L \ln p$, is much larger and significant. In the specification with the CRS assumption, both cross terms are deleted, and their coefficients are no longer in the calculation of the shadow price of corresponding characteristics. Thus the lack of $\ln S \ln p$ has nearly no influence on the shadow price calculation of structures, while the lack of $\ln L \ln p$ results in obvious change in the calculated shadow prices of lands. Despite the decrease of the levels of the land price indexes, the changing trends over the studied period are basically the same with or without the CRS assumption.

2.6 Conclusions

We are optimistic about the potential usefulness of our new approach to construct residential property price indexes. It has reasonable theoretical underpinnings and is parsimonious in terms of required data. Rather than relying on exogenous information to circumvent problems of multicollinearity between different property characteristics, our method can estimate the shadow price of each characteristic with little computational burden. As this model is less sensitive to actual market fluctuations, it also can be combined with traditional hedonic regression methods to provide bounds on residential property prices based on mark-to-market adjustment and less volatile consumer preferences.

One problem with our approach is that, due to limited date, we are unable to find effective instrumental variables for the property price. Thus in the general specification, logged age is put on the left-hand side as the dependent variable, though age should be exogenous. This problem can be circumvented by imposing the constant-returns-to-scale assumption on the distance function. With the CRS assumption, our input distance function is in essence an alternative representation of the hedonic regression with the same set of characteristics. Lack of the cross terms of land and property price leads to a decrease in the levels of estimated land price index, but the changing trend remains the same.

The housing market has always been a hot topic for research. The prices of the property and its characteristics are usually analyzed using hedonic regression method or the index number theory. The chapter provides an alternative method, which is based on the duality theory, to uncover the relation between the property price and the prices of the characteristics. With these different economic theories, we need to consider how to select or combine these models to obtain a robust method. As discussed at the end of Chapter 1, compared with the model selection procedure, the model averaging approach would be a better way to obtain a robust estimator when the data generating process is unobservable. Thus, I will examine in detail the model averaging techniques in the next chapter.

Appendix

We prove the shadow price expression (2.5) in two ways, both when z adjusts to w and when w adjusts to z.

Define the cost function as

$$C(p, w) = \min_z wz - \mu(D_i(p, z) - 1) \tag{A.1}$$

then

$$F.O.C: \qquad w - \mu^* \nabla_z D_i(p, z) = 0. \tag{A.2}$$

Also note that

$$\mu^* = C(p, w). \tag{A.3}$$

This follows from

$$\begin{aligned} \tilde{C}(p, w, \alpha) &= \min_z wz - \mu(D_i(p, z) - \alpha) \\ &= \alpha C(p, w), \end{aligned} \tag{A.4}$$

and

$$\frac{\partial \tilde{C}}{\partial \alpha} = \mu = C(p, w). \tag{A.5}$$

Thus, by the FOC we have,

$$w = C(p, w) \cdot \nabla_z D_i(p, z) \tag{A.6}$$

and by Euler's theorem we have

$$p = wz = C(p, w) D_i(p, z) \tag{A.7}$$

or

$$C(p, w) = p / D_i(p, z) \tag{A.8}$$

and thus

$$w^s = \frac{p \cdot \nabla_z D_i(p,z)}{D_i(p,z)}. \tag{A.9}$$

Next consider the dual optimization problem

$$\begin{aligned} D_i(p,z) &= \min_w wz - \lambda(C(p,w) - 1) \\ &= w^* z - \lambda^*(C(p,w^*) - 1) \end{aligned} \tag{A.10}$$

As above one can show that

$$D_i(p,z) = \lambda^* \tag{A.11}$$

thus

$$D_i(p,z) = w^* z - D_i(p,z)(C(p,w^*) - 1) \tag{A.12}$$

and

$$D_i(p,z) = \frac{w^* z}{C(p,w^*)} \tag{A.13}$$

Therefore

$$\nabla_z D_i(p,z) = \frac{w^*}{C(p,w^*)} - \frac{w^* \cdot \nabla_w C(p,w^*)}{C(p,w^*)^2}. \tag{A.14}$$

Now since $C(p, w^*) = 1$, a constant function,

$$w^* \cdot \nabla_w C(p,w^*) = 0 \qquad \text{and}$$

$$\nabla_z D_i(p,z) = \frac{w^*}{C(p,w^*)}. \tag{A.15}$$

Again, using the Euler's theorem and noting that $w^* z = p$, we have

$$w^{s,*} = \frac{p \cdot \nabla_z D_i(p,z)}{D_i(p,z)} \tag{A.16}$$

2.7 Tables and Figures

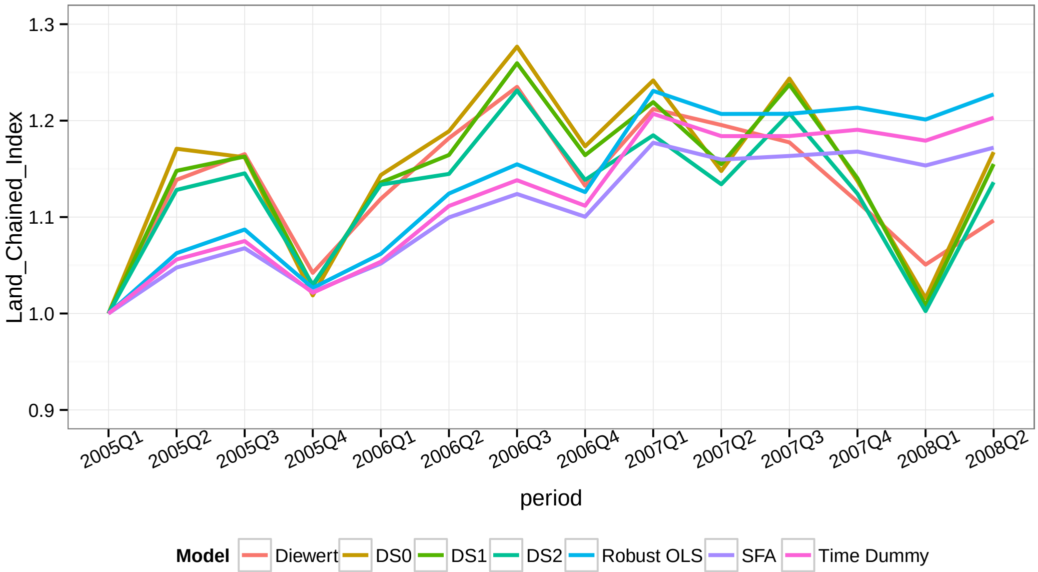


Figure 2.2 : Land Price Index Comparison

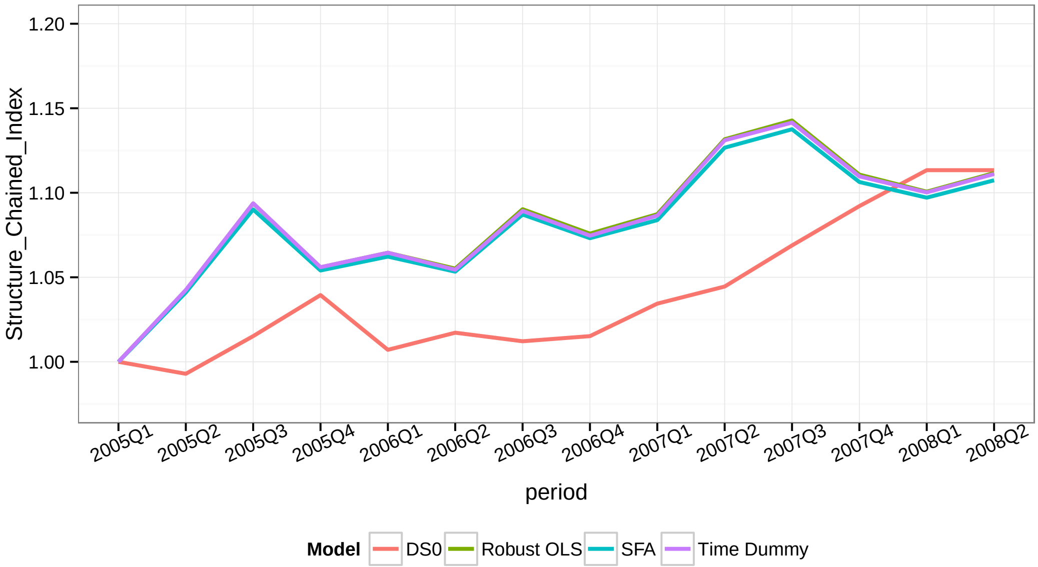


Figure 2.3 : Structure Price Index Comparison

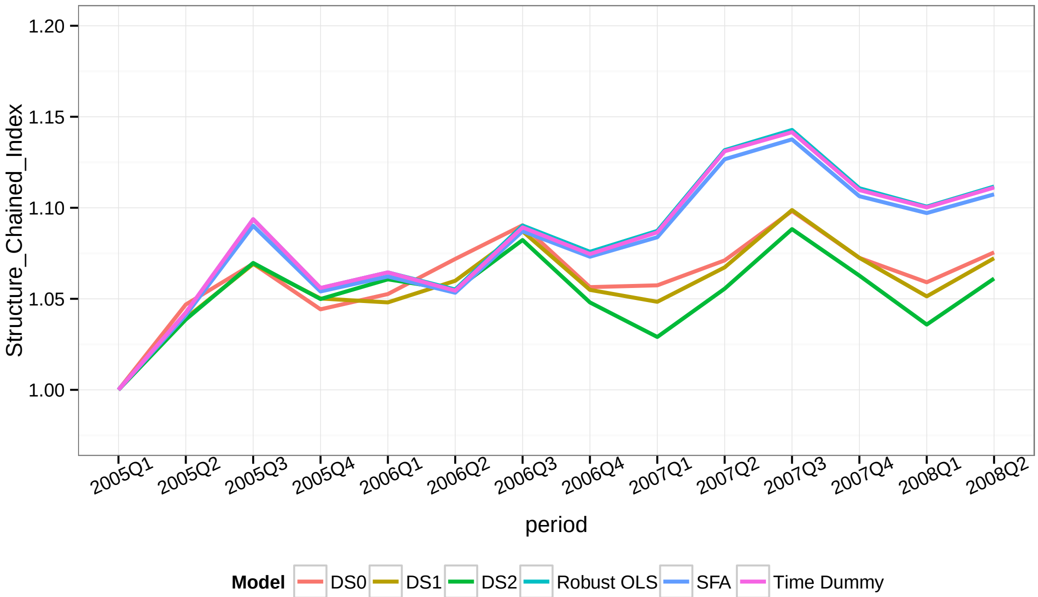


Figure 2.4 : Structure Price Index Comparison with Exogenous Land Price

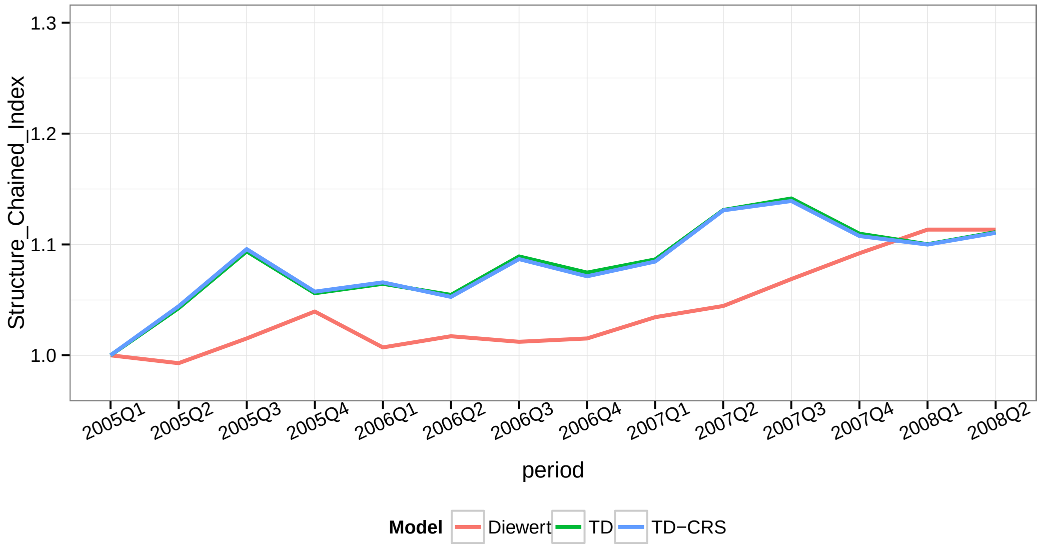


Figure 2.5 : Structure Price Index Comparison with CRS

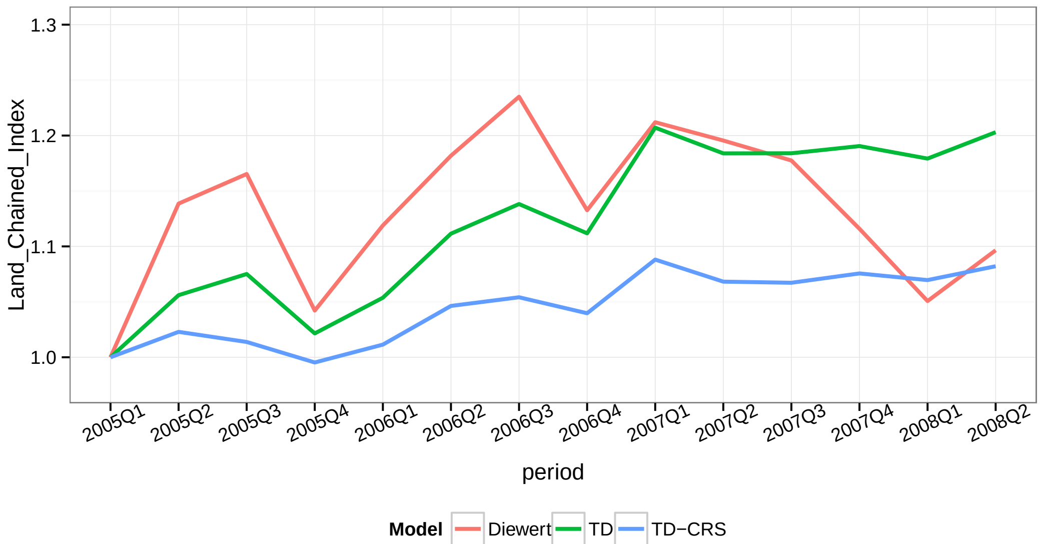


Figure 2.6 : Land Price Index Comparison with CRS

Parameter	COLS			Time Dummy			SFA		
	Coefficient	Std.Err.	t	Coefficient	Std.Err.	t	Coefficient	Std.Err.	t
α_0	-8.654262	0.9184771	-9.42				-7.886361	0.533136	-14.79
α_1	2.999968	0.6205874	4.83	2.519516	0.2859612	8.81	2.506335	0.2969787	8.44
α_{11}	-0.432247	0.1122404	-3.85	-0.3851873	0.042537	-9.06	-0.339389	0.0458892	-7.40
β_1	0.4251001	0.1919702	2.21	0.5175217	0.1231534	4.20	0.5230654	0.1267647	4.13
β_2	-0.7420501	0.2855724	-2.6	-0.6834723	0.1417875	-4.82	-0.5483521	0.1469673	-3.73
$\frac{1}{2}\beta_{11}$	0.0491277	0.0205348	2.39	0.0358318	0.0168336	2.13	0.0550541	0.0168423	3.27
$\frac{1}{2}\beta_{22}$	-0.0471125	0.0296489	-1.59	-0.0423398	0.0150257	-2.82	-0.0146476	0.0160396	-0.91
β_{12}	0.0074992	0.0490074	0.15	0.0073379	0.0249217	0.29	0.0101349	0.0255948	0.40
γ_1	-0.0105265	0.0651377	-0.16	-0.0098862	0.0375809	-0.26	-0.0390997	0.0388763	-1.01
γ_2	0.2456549	0.1031222	2.38	0.2307362	0.0425736	5.42	0.1550929	0.0459129	3.38

Table 2.1 : Regression Results of Different Methods

	Diewert	Ours
Num. of observations	2289	2280
Ave. sale price	190.13	189.76
Ave. land area	257.6	258.98
Ave. structure area	127.2	127.09

Table 2.2 : Summary Statistics Comparison

	Coef.	Std. Err.	t
α_0	-0.257646	0.114862	-2.24
β_1	0.9153355	0.075584	12.11
β_2	0.1219993	0.0539718	2.26
$\frac{1}{2}\beta_{11}$	-0.018214	0.0189258	-0.96
$\frac{1}{2}\beta_{22}$	0.029778	0.0107326	2.77
β_{12}	-0.000395	0.0241278	-0.02

Table 2.3 : Coefficients of Specification with CRS

Time Period	Num. of Obs.	NDOPI	Fisher Fixed-Base House Price Index	Fisher Chained House Price Index
1	157	1.0000	1.0000	1.0000
2	155	0.9929	1.0240	1.0240
3	154	1.0152	1.0682	1.0784
4	155	1.0395	1.0490	1.0408
5	163	1.0071	1.0444	1.0408
6	175	1.0172	1.0668	1.0575
7	157	1.0122	1.0731	1.0734
8	152	1.0152	1.0768	1.0671
9	159	1.0344	1.0683	1.0895
10	194	1.0445	1.1189	1.1148
11	137	1.0688	1.1220	1.1247
12	187	1.0921	1.1132	1.1048
13	148	1.1134	1.1107	1.1045
14	187	1.1134	1.1058	1.1119

Table 2.4 : House Price Index and NDOPI

Period	DS0	DS1	DS2	COLS	SFA	Time Dummy	Diewert
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.17085	1.14797	1.12816	1.06255	1.04764	1.05601	1.13864
3	1.16199	1.16277	1.14536	1.08708	1.06768	1.07512	1.16526
4	1.01873	1.02859	1.02990	1.02655	1.02264	1.02149	1.04214
5	1.14362	1.13589	1.13368	1.06191	1.05178	1.05369	1.11893
6	1.18897	1.16439	1.14474	1.12440	1.09966	1.11148	1.18183
7	1.27672	1.25955	1.23127	1.15472	1.12393	1.13820	1.23501
8	1.17338	1.16416	1.13854	1.12592	1.10034	1.11176	1.13257
9	1.24168	1.21921	1.18481	1.23086	1.17709	1.20706	1.21204
10	1.14775	1.15482	1.13401	1.20693	1.15966	1.18394	1.19545
11	1.24358	1.23731	1.20753	1.20707	1.16338	1.18398	1.17747
12	1.13746	1.14025	1.12421	1.21345	1.16790	1.19053	1.11588
13	1.01596	1.00828	1.00247	1.20116	1.15347	1.17922	1.05070
14	1.16739	1.15499	1.13613	1.22721	1.17198	1.20303	1.09648

Table 2.5 : Land Price Indexes Comparison

Period	DS0	COLS	SFA	Time Dummy
1	1.00000	1.00000	1.00000	1.00000
2	0.99291	1.04218	1.04103	1.04250
3	1.01518	1.09368	1.09011	1.09375
4	1.03947	1.05588	1.05404	1.05597
5	1.00709	1.06450	1.06224	1.06452
6	1.01721	1.05513	1.05333	1.05448
7	1.01215	1.09026	1.08712	1.08924
8	1.01518	1.07582	1.07312	1.07462
9	1.03441	1.08732	1.08382	1.08652
10	1.04453	1.13165	1.12666	1.13105
11	1.06883	1.14276	1.13757	1.14152
12	1.09211	1.11073	1.10636	1.10971
13	1.11336	1.10063	1.09711	1.10022
14	1.11336	1.11168	1.10738	1.11111

Table 2.6 : Structure Price Index Comparison

Period	DS0	DS1	DS2	COLS	SFA	Time Dummy
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.04687	1.03938	1.03864	1.04218	1.04103	1.04250
3	1.06929	1.06910	1.06970	1.09368	1.09011	1.09375
4	1.04422	1.05016	1.04987	1.05588	1.05404	1.05597
5	1.05261	1.04803	1.06071	1.06450	1.06224	1.06452
6	1.07189	1.05990	1.05488	1.05513	1.05333	1.05448
7	1.09040	1.08710	1.08229	1.09026	1.08712	1.08924
8	1.05641	1.05490	1.04801	1.07582	1.07312	1.07462
9	1.05740	1.04835	1.02903	1.08732	1.08382	1.08652
10	1.07110	1.06728	1.05556	1.13165	1.12666	1.13105
11	1.09823	1.09876	1.08828	1.14276	1.13757	1.14152
12	1.07249	1.07250	1.06274	1.11073	1.10636	1.10971
13	1.05908	1.05134	1.03586	1.10063	1.09711	1.10022
14	1.07550	1.07229	1.06113	1.11168	1.10738	1.11111

Table 2.7 : Structure Price Indexes Comparison with Exogenous Land Price

TD-CRS	TD	Diewert
1.0000	1.0000	1.0000
1.0442	1.0425	0.9929
1.0958	1.0938	1.0152
1.0574	1.0560	1.0395
1.0658	1.0645	1.0071
1.0527	1.0545	1.0172
1.0868	1.0892	1.0122
1.0713	1.0746	1.0152
1.0846	1.0865	1.0344
1.1308	1.1311	1.0445
1.1392	1.1415	1.0688
1.1076	1.1097	1.0921
1.0999	1.1002	1.1134
1.1105	1.1111	1.1134

(a) Structure Price Indexes Comparison with CRS

TD-CRS	TD	Diewert
1.0000	1.0000	1.0000
1.0229	1.0560	1.1386
1.0138	1.0751	1.1653
0.9953	1.0215	1.0421
1.0114	1.0537	1.1189
1.0463	1.1115	1.1818
1.0541	1.1382	1.2350
1.0397	1.1118	1.1326
1.0881	1.2071	1.2120
1.0682	1.1839	1.1955
1.0672	1.1840	1.1775
1.0756	1.1905	1.1159
1.0696	1.1792	1.0507
1.0821	1.2030	1.0965

(b) Land Price Indexes Comparison with CRS

Table 2.8 : Price Indexes Comparison with CRS

Chapter 3

Reduced Form Analysis of World Productivity Growth: A Model Averaging Approach

3.1 Introduction

Various structural and reduced form models of productivity growth have been proposed in the literature. The classical structural models, particularly those of Berry, Levinsohn, and Pakes (1995) and OLLEY and PAI (1996) and implemented by Pakes and McGuire (1992), have focused on identifying innovation and efficiency within a dynamic Markov-perfect Nash equilibria modeling framework. Reduced form approaches, typically stochastic frontier methods, often focus on the asymmetry of productivity shocks implied by the presence of a best practice technology. This set of models concentrates on robust methods to identify innovation and efficiency levels, along with their growth rates, without the benefit of overidentifying restrictions implied by structural approaches. In either class of models, however, we can obtain reduced form measurements of productivity and efficiency. For the structural model these are based on the restricted reduced form while in the stochastic frontier approach they are based on the unrestricted reduced form. As the true data generating process cannot be observed, we consider the model averaging approaches in the stochastic frontier setting, trying to reduce model selection risk and obtain more stable and reliable estimates of efficiency and productivity. The model averaging estimates will ultimately be compared with those generated from the structural models discussed above to allow a test of the overidentifying restrictions, which are to be constructed and carried out.

We first consider the stochastic frontier analysis of productivity and efficiency from a model averaging perspective. The stochastic frontier model has become a classic approach to productivity and efficiency analysis since it was introduced by Aigner, Lovell, and Schmidt (1977) and Meeusen and Van den Broeck (1977). The efficiency effect was modeled as an error term aside from the pure white noise. Since then, a huge volume of researchers have extended and modified the stochastic production (cost) frontier to better model the production process and reflect possible circumstances that may contradict the basic assumptions. Most of the variations focused on assumptions of the efficiency errors. Pitt and Lee (1981) and Schmidt and Sickles (1984) extended the stochastic frontier model to panel data. Schmidt and Sickles (1984) discussed the fixed effects, random effects and MLE estimators of the (in)efficiency terms. Later research extended this standard panel data model and allowed the efficiency terms to vary both over time and across sections. The empirical application in Cornwell, Schmidt, and Sickles (1990) modeled the time path of efficiency term with a polynomial function. Meanwhile, Kumbhakar (1990) and Battese and Coelli (1992) used the exponential functions of time t. Lee and Schmidt (1993) separated the efficiency term into time-varying effects that are uniform across sections and a set of individual multipliers. Recent work, such as Kneip, Sickles, and Song (2012) and Ahn, Lee, and Schmidt (2013), further generalized the setting of the efficiency term and assumed the efficiency is determined by a set of underlying factors.

Facing a list of competing models, we want to know which one can best describe the true data genarating process, thus providing a suitable platform on which we can conduct further analysis. This inquiry leads to a general model selection problem. On one hand, we wish to have a model that can explain the relation among variables well, allowing us to identify the influencing factors of certain observed phenomenon or to make predictions about future changes. On the other hand, we want a model that is both parsimonious and simple to implement. The literature discussing model selection is voluminous and various criteria have been proposed, each of which has a different focus. The famous Akaike information criterion (AIC; AKAIKE, 1973) and the Schwarz-Bayes information criterion (BIC; Schwarz et al., 1978) aim to reach a balance between the goodness of fit and the parsimony of the model, but they penalize the number of parameters to a different extent. A more recent concept, the focused information creterion (FIC; Hjort and Claeskens, 2003), considers the estimating quality of the interested parameters instead of the fit of the entire model. The typical procedure of model selection starts with choosing one such criterion and calculating it for all the candidate models. The model of best performance will then be selected and treated as the "true" model that generates the data. Further analysis, such as making inferences or predictions, will be solely based on this "true" model. One problem with the model selection procedure is that it ignores the model uncertainty. That is, we assume the probability that the chosen model is the "true" data generating process (DGP) is one, while in reality it is not. Thus such a procedure leads to over optimistic inferences and higher prediction error. Another problem is that the "true" model can be different based on the criterion employed. Though there can be a general guide directing which criterion to use according to the main interest of our problem, there is no stardard and systematic procedure to follow. It is also difficult to evaluate how well these criteria perform.

The difficulty of selecting an optimal or "true" model lies in the fact that the true DGP is unobservable. All that we can do is to approximate the underlying mechanism as accurately as possible based on the observed data. By assigning a set of weights to all the candidate models instead of treating a single model as the "best" or "true" model, model averaging can be seen as an agnostic method. This method has been shown to be a more favorable approach if the goal of econometric analysis is to approximate the underlying DGP rather than to discover it.

As all the candidate models can reflect the underlying DGP to some extent, it is reasonable to assign each one a weight based on its ability to explain the data. The model selection procedure can be seen as a special case of model averaging in which the "best" model receives a probability of one while others get zero probabilities. Having decided to pursue the model averaging approach, the next thing we consider is how to determine the set of weights. Like other statistical techniques, the model averaging methods can be classified into Frequentist Model Averaging (FMA) and Bayesian Model Averaging (BMA). Early work on model averaging was mainly from the Bayesian perspective, and there has emerged a voluminous literature on both theoretic extensions and empirical applications of BMA. In general, the BMA approach first assigns to each candidate model a prior probability, which will be updated through the observed data, and the updated posterior probability will be used as the weight. A tutorial about the basics of the BMA approach can be found in Hoeting, Madigan, Raftery, and Volinsky (1999).

Choosing the prior probabilities is the first step in the BMA method. This is problematic because we do not know what priors are the most appropriate, and priors can be in conflict with each other. The FMA approach, on the other hand, does not need such ex ante assumptions of probabilities, thus avoiding these associated problems. There has been a growing number of literature discussing the FMA methods in recent years. Buckland, Burnham, and Augustin (1997) proposed to assign weights according to an Information criterion of each candidate model m:

$$I_m = -2log(L_m) + l \tag{3.1}$$

where L_m is the maximized likelihood function of the m-th model, and l is some penalty function. The weights based on this criterion are calculated as

$$w_I = \frac{exp(\frac{1}{2}I_m)}{\sum_{m\in\mathcal{M}} exp(\frac{1}{2}I_m)} \tag{3.2}$$

in which M is the set of all competing models. If $l = 2k$, where k is the number of parameters in the model, the information criterion I_m is just the AIC score. If $l = k \cdot log(n)$, and n represents the number of observations, I_m becomes the BIC score. It is straightforward to base the weights on such model evaluating criteria, but we lack a method to measure the effectiveness of these weighting schemes. It is also difficult to tell by how much they can improve the quality of the estimators.

More weighting schemes have been proposed in recent years. Leung and Barron (2006) considered assigning the weights to a set of least squares estimators based on their risk characteristics. Hansen (2007) proposed to select the weights by minimizing a Mallows criterion, which works in a set of nested linear models with homoskedastic errors. The paper shows that the Mallows model average estimator can asymptotically achieve the lowest squared error among discrete model averaging estimators. Based on the work of Hansen (2007), Wan, Zhang, and Zou (2010) relaxed the assumptions that candidate models should be nested based on certain ordering of regressors. They also provided a proof of the optimality of Mallows criterion in a continous weight set. Wang, Zhang, and Zou (2009) reviewed some important progress in FMA methods.

Hansen and Racine (2012) considered a more general situation in which the candidate models can be non-nested and have heteroskedatic errors. The proposed estimator is termed "jackknife model averaging" (JMA) estimator. It is obtained by minimizing a leave-one-out cross-validation criterion. This JMA estimator is proved to be asymptotically optimal in the sense that it approaches the lowest possible expected squared errors as the sample size approaches infinity (for a panel data of N individuals and T time periods, the sample size is equal to $N \times T$). We will employ the JMA technique in our stochastic frontier analysis.

Though the model averaging methods provide a better approximation of the production process than a single model, it is applied to a set of candidate models in the same class, which, in our case, are stochastic frontier models. We want to pursue the analysis of productivity and efficiency from different perspectives. To measure both the level and changing trend of productivity, we focus on panel data models with time varying effects, which have been of a great interest in recent years. Eckstein and Lifshitz (2011) constructed a dynamic labor supply model to analyze the trends in female employment rates. Bai (2013) considered the estimation of dynamic panel data models with both individual fixed effects and time fixed effects. Dix-Carneiro (2014) studied the Brazilian labor market using a structural dynamic equilibrium model. The structural models have been used extensively to examine the production side of the economy and labor market. Besides employing the reduced form models, we also briefly discuss how to analyze the country-level efficiencies using a structural model. The framework is based on the model proposed by Pakes and McGuire (1992), which provides a description of firms' behaviors in a certain industry and defines a Markov-perfect Nash equilibrium. The single period profit of a firm is decided by its current efficiency level and the efficiencies of its competitiors. This model, as well as those studied in Pakes and Ericson (1998) and Berry, Levinsohn, and Pakes (1995), can be implemented using the algorithm proposed by Pakes and Gowrisankaran (1993). As the behaviors of a country differ to some extent with those of firms, we will modify the structural model to analyze the country-level efficiencies. The implementation method need to be changed accordingly to reflect the simultaneous decision-making process of all the countries. We will leave the detailed structural analysis for future work.

This chapter is organized as follows. In Section 2, we first describe how the stochastic frontier model is developed to model the productivity, and then introduce a set of competing specifications that are widely used in empirical applications. In section 3, we explain the details of the jackknife model averaging technique. Section 4 describes the empirical setting we use for the productivity analysis. Section 5 introduces the data and explains different measurements of the variables. Results are discussed in section 6. In section 7, we briefly show how to examine the country-level efficiencies using a structural model. Section 8 concludes the chapter.

3.2 Productivity Measurement and Stochastic Frontier Models

Productivity is usually measured as a ratio of an output index to an input index, where a higher ratio means more outputs can be produced from a given certain combination of inputs. In a simple one-output production process, total factor productivity (TFP) can be expressed as

$$\text{TFP} = \frac{Y}{\sum a_i X_i} \tag{3.3}$$

where Y is the output, and X_i's are the inputs. When multiple outputs exist, TFP can be measured as a ratio of an index of aggregated outputs to an index of aggregated inputs. The value and properties of such a TFP measure will then depend on the methods of aggregation and index construction. Taking the log of (3.3), we can construct a total factor productivity index as the difference between the logged output and logged input indexes, as stated in Jorgenson and Griliches (1972).

This TFP, also known as the Solow residual, measures the effects in the output that are not explained by the inputs. It can be decomposed into technical efficiency (TE) and technical innovation. In a specific period, the technology may be regarded as fixed, while firms, or countries, might experience different technical efficiency levels. Denote $0 < \mathrm{TE}_i \leq 1$ to be the technical efficiency level for individual i, and let $f(\cdot)$ be the production function that defines the frontier. The production process for each individual can be modeled as

$$y_i = f(X_i, \beta) \cdot \mathrm{TE}_i \tag{3.4}$$

where β is the coefficients vector of inputs X_i. Taking the log of (3.4) gives the following form,

$$\ln y_i = \ln f(X_i, \beta) + \ln \mathrm{TE}_i = \ln f(X_i, \beta) + u_i \tag{3.5}$$

where u_i is the logged efficiency level for each unit. Adding a white noise term v_i to (3.5), we reach the standard stochastic frontier model proposed by Aigner, Lovell, and Schmidt (1977) and Meeusen and Van den Broeck (1977). To identify the efficiency term, the initial stochastic frontier model with cross-sectional data requires a distributional assumption on u_i. With panel data, however, one can allow for a non-parametric form of the efficiency term. In this section, we will focus on panel data models and consider several competing stochastic frontier settings in the study of technical efficiency. After introducing each candidate model, we will show how to apply the model averaging technique to compute the optimal weights and obtain the jackknife averaging estimator.

We start with the most standard stochastic frontier model with panel data, which was first proposed in Schmidt and Sickles (1984).

$$y_{it} = \alpha + X'_{it}\beta + u_i + v_{it} \qquad t = 1, \dots, T; \, i = 1, \dots, N \tag{3.6}$$

where X_{it} is a vector of input variables, y_{it} is the output variable, v_{it} is the usual distrubance term that is assumed to be independently and identically distributed, possible with a parameterized distribution such as $N(0, \sigma_v^2)$, and u_i represents the efficiency level of each individual. If we denote $\alpha_i = \alpha + u_i$, the model can be written as

$$y_{it} = \alpha_i + X'_{it}\beta + v_{it} \tag{3.7}$$

which is just the standard form of the panel data model. The coefficients can be obtained using the within or GLS estimators, depending on whether the error terms are assumed to be correlated with regressors.

In the standard panel data stochastic frontier model, the efficiencies are assumed to be fixed over time, which may not be the case in reality. With the changes in possible influencing factors, such as adjustments of inputs allocation, workers' education levels, firms' administration system, and so on, we expect the efficiency levels to change accordingly from period to period. Many researchers then extended the standard setting to model time-varying inefficiencies. Cornwell, Schmidt, and Sickles (1990) (Hereafter CSS) considered a model allowing for both time-varying individual effects and heterogeneous slope coefficients.

$$y_{it} = X'_{it}\beta + Z'_i\gamma + W'_{it}\delta_i + v_{it} \tag{3.8}$$

where Z_i represents the time invariant variables, and $W_{it}\delta_i$ can be used to model individuals' efficiencies, which vary both over time and across sections. Denote $\delta_0 =$

$E(\delta_i)$, then $\delta_i = \delta_0 + u_i$, with u_i being zero-mean random varibles. The model can be rewritten as

$$y_{it} = X'_{it}\beta + Z'_i\gamma + W'_{it}\delta_0 + \epsilon_{it} \tag{3.9}$$

$$\epsilon_{it} = W'_{it}u_i + v_{it} \tag{3.10}$$

The CSS model can be estimated using extensions of the within and GLS methods. If the effects u_i are assumed to be correlated with the exogenous regressors (X_{it}, Z_i and W_{it}), we can utilize the generalization of the within estimator. One problem of the within method is that γ cannot be identified, since we need to demean the model in the first step, and the time-fixed regressors Z_i disappear from the equation. Assuming no correlation exists between the regressors and the effects u_i, we can use the GLS method to get consistent estimates of $(\beta, \gamma, \delta_0)$.

Another model we consider is the one proposed by Battese and Coelli (1992). The generic setting is

$$y_{it} = X'_{it}\beta + u_{it} + v_{it} \tag{3.11}$$

where v_{it} is the white noise, and u_{it} represents the efficiency effects and can vary across both time and individuals. The time changing path was specified as some exponential function:

$$u_{it} = -\{\exp[-\eta(t - T)]\}u_i \tag{3.12}$$

With such a specification, the individual effects can increase, decrease or remain constant over time, corresponding to $\eta > 0$, $\eta < 0$ or $\eta = 0$. The effects u_i are assumed to be i.i.d. random variables following a non-negative truncation of the normal distribution $N(0, \sigma_u^2)$.

Denote $\epsilon_{it} = u_{it} + v_{it}$ as the composite error. As both u_i and v_{it} are assumed to follow specific distributions, we can derive an explicit expression of the distribution of ϵ_{it}. The ML method can then be applied to estimate the parameters. One problem is that the ML estimator is not linear, while the model averaging method we will employ is proposed for a set of linear estimators. We will show that the Battese and Coelli model can be nested in a more general model setting, and thus estimates of coefficients can be approximated with a linear estimator.

The general setting that nests the BC model is the one proposed by Kneip, Sickles, and Song (2012) (Hereafter KSS) and is specified as

$$y_{it} = \alpha_t + X_{it}'\beta + u_{it} + v_{it} \tag{3.13}$$

The individual effects u_{it} are assumed to be affected by a set of underlying factors and are formulated by linear combinations of some basis functions:

$$u_{it} = \sum_{r=1}^{L} \delta_{ir} g_r(t) \tag{3.14}$$

For identifiability, it is assumed that $\sum_i^n u_{it} = 0$, $t = 1, \ldots, T$. The intercept α_t can be eliminated by transforming the model to the centered form,

$$y_{it} - \bar{y}_t = (X_{it} - \bar{X}_t)'\beta + \sum_{r=1}^{L} \delta_{ir} g_r(t) + v_{it} - \bar{v}_t \tag{3.15}$$

where $\bar{y}_t = \frac{1}{n}\sum_i y_{it}$, $\bar{X}_t = \frac{1}{n}\sum_i X_{it}$ and $\bar{v}_i = \frac{1}{n} v_{it}$. Denote $\tilde{y}_{it} = y_{it} - \bar{y}_t$ and $\tilde{X}_{it} = X_{it} - \bar{X}_t$, we return to the basic stochastic frontier setting

$$\tilde{y}_{it} = \tilde{X}_{it}'\beta + \sum_{r=1}^{L} \delta_{ir} g_r(t) + \tilde{v}_{it} \tag{3.16}$$

We can see that the individual effects u_{it} are assumed to be determined by a number of underlying factors, which are represented by a set of basis functions $(g_1(t), \ldots, g_L(t))$. Denote $\mathcal{L} \equiv span\{g_1, \ldots, g_L\}$ to be the space of the underlying factors. A problem is that the set of basis functions is not unique. Thus certain standardization is imposed for the estimation problem to be well defined. It is assumed that

(a) $\frac{1}{n}\sum_{i=1}^{n}\delta_{i1}^2 \geq \frac{1}{n}\sum_{i=1}^{n}\delta_{i2}^2 \geq \cdots \geq \frac{1}{n}\sum_{i=1}^{n}\delta_{iL}^2 \geq 0$

(b)$\frac{1}{n}\sum_{i=1}^{n}\delta_{ir} = 0$ and $\frac{1}{n}\sum_{i=1}^{n}\delta_{ir}\delta_{is} = 0$ for all $r,\ s \in 1,\ldots,L,\ r \neq s$.

(c) $\frac{1}{T}\sum_{t=1}^{T} g_r(t)^2 = 1$ and $\frac{1}{T}\sum_{t=1}^{T} g_r(t)g_s(t) = 0$ for all $r,\ s \in 1,\ldots,L,\ r \neq s$.

With such standardization, the Battese and Coelli model can be nested after some rescaling. It corresponds to the case that $L = 1$ and $g_1(t) = exp(-\eta(t-T))/\sqrt{\frac{1}{T}\sum_{s=1}^{T} exp(-\eta(s-T))^2}$. Thus the ML estimator Battese and Coelli proposed in their model can be approximated using the KSS estimator.

The CSS specification we described above can also be nested in the KSS model with $L = 3$ and polynomial functions being the basis function for individual effects.

3.3 Jackknife Model Averaging

To illustrate the jackknife model averaging method, we begin with a number of competing models that can be written as

$$y_i = \mu_i + \epsilon_i \qquad i = 1,\ldots,n \tag{3.17}$$

where $\mu_i = E(y_i|X_i)$, and X_i is the vector of input variables. ϵ_i is the error term with zero mean conditional on X_i. The conditional variance of ϵ_i is allowed to vary across observations. That is,

$$E(\epsilon_i|X_i) = 0 \tag{3.18}$$

$$E(\epsilon_i^2|X_i) = \sigma_i^2 \tag{3.19}$$

Suppose we have M candidate models, and for each model m, we have a linear estimator, denoted as $\hat{\mu}^m = P_m y$. The estimator is linear in the sense that P_m is not a function of y. This definition covers a fairly broad class of estimators. As mentioned in Hansen and Racine (2012), the standard OLS, ridge regression, nearest neighbor estimators, series estimators, etc. are all in this class. The jackknife averaging estimator is then calculated as the weighted average of all the candidate models. Next we show how to calculate the desired weights.

For each model m, the jackknife estimator is denoted as $\tilde{\mu}^m = (\tilde{\mu}_1^m, \ldots, \tilde{\mu}_n^m)'$, and $\tilde{\mu}_i^m$ is the estimate of y_i with parameters estimated after the ith observation being deleted (i.e. the leave-one-out cross validation). The jackknife residual is then computed as $\tilde{e}^m = y - \tilde{\mu}^m$. Hansen and Racine (2012) provided a simpler method to calculate the jackknife residual for standard OLS estimator to avoid n times regressions for each model. As the method is not applicable for the stochastic frontier models, we skip the details here.

The weights are assumed to be non-negative and sum to one, thus a weight vector lies on the $\mathbb{R}^M$ unit simplex.

$$\mathcal{H}_M = \left\{ w \in \mathbb{R}^M : w^m \geq 0, \sum_{m=1}^{M} w^m = 1 \right\}$$

Given a specific weight vector w, the averaged estimator is

$$\tilde{\mu}(w) = \sum_{m=1}^{M} w^m \tilde{\mu}^m = \tilde{\mu} w \tag{3.20}$$

and the averaged residual is

$$\tilde{e}(w) = y - \tilde{\mu}(w) = \tilde{e} w \tag{3.21}$$

With the averaging residual, we can state the jackknife or leave-one-out cross-validation criterion as

$$CV_n(w) = \frac{1}{n} \tilde{e}(w)' \tilde{e}(w) \tag{3.22}$$

The jackknife weights can then be obtained by minimizing this criterion over the weight space $\mathcal{H}_M$, i.e.

$$w^* = \underset{w \in \mathcal{H}_M}{\arg\min}\, CV_n(w) \tag{3.23}$$

The jackknife averaging estimator is thus $\hat{\mu}(w^*) = \hat{\mu}w^*$. Depending on different specifications of the model setting, the applicable weight space might be only a subset of $\mathcal{H}_M$.

3.4 Empirical Model

As stated in the above section, the TFP is usually measured as a ratio of an output index to an input index. The most widely used TFP measure at country level is based on the Cobb-Douglas production function

$$Y = AK^{\alpha}L^{\beta} \tag{3.24}$$

where Y is the gross output measured as GDP, and K and L are the capital and labor inputs. A is seen as the level of technology, and α and β are interpreted as the input expenditure share for capital and labor under the assumption of cost minimization. Using the Cobb-Douglas production function, the TFP measure is:

$$\text{TFP} = \frac{Y}{K^{\alpha}L^{\beta}} \tag{3.25}$$

We assume that technology or technical innovations can be accessed by all units in any period, and the differences in productivity are caused by relative efficiencies. The innovations change can be measured by differences in R&D expenditure over time, number of patents appeared, or other similar factors. When building up the empirical model, we can use the time variables to proxy the innovations change, such as the time index method proposed in Baltagi and Griffin (1988), or a polynomial function of time variable. The technical efficiencies can be studied using stochastic frontier models.

We use the generic Cobb-Douglas production function in our empirical analysis of world productivity change. The inputs are capital K and labor L, and the output Y is measured by GDP. To pick up the time-varying productivity change, we add t and t^2 in our model, thus the basic model for regression is

$$lnY_{it} = \alpha_i + \beta_1 lnK_{it} + \beta_2 lnL_{it} + \theta_1 t + \theta_2 t^2 + u_{it} + v_{it} \tag{3.26}$$

As described above, the candidate estimators are the within and GLS estimators from the standard panel data model proposed by Schmidt and Sickles (1984) (denoted as 'Fixed' and 'Random' respectively), the within and GLS estimators from the CSS model (denoted as 'CSSW' and 'CSSG' respectively), and the BC estimator.

Assuming there are no economies of scale, we impose the constant returns to scale constraint on the general case specified above, which means $\alpha + \beta = 1$. Thus the empirical model with constant returns to scale(CRS) assumption is

$$lnY_{it} = \alpha_i + \beta_1 lnK_{it} + (1 - \beta_2) lnL_{it} + \theta_1 t + \theta_2 t^2 + u_{it} + v_{it} \tag{3.27}$$

Denote $ln\tilde{Y}_{it} = lnY_{it} - lnL_{it}$ and $ln\tilde{K}_{it} = lnK_{it} - lnL_{it}$, the regression function for the CRS case is

$$ln\tilde{Y}_{it} = \alpha_i + \beta_1 ln\tilde{K}_{it} + \theta_1 t + \theta_2 t^2 + u_{it} + v_{it} \tag{3.28}$$

As none of these models can be estimated with the standard least squares method, we need to run $N \times T$ times for each candidate estimator, which can be quite time-consuming when the data set gets large. An alternative is to do block cross-validation instead of the leave-one-out cross-validation. With panel data, a natural way is to group the data by time period and delete the data of each period per time. In this way, the regressions needed can be reduced to T times. However, as noted in Zhang, Wan, and Zou (2013), the problem of block cross-validation is that the selection of block length is data dependent, which leads to μ being a non-linear function of y. This non-linearity contradicts the assumption of the jackknife method, thus the asymptotic properties cannot be guaranteed. We will consider the block cross-validation in the general case, and compare the results with the optimal weights obtained using leave-one-out cross-validation to see if the violation of linearity assumption will severely affect the results.

3.5 Data Description

Our data comes from the World Productivity Database (WPD) in UNIDO, which provides information on productivity related variables and statistics of 112 countries over the period 1960-2012. As its name states, WPD focuses on productivity performance at both the level and growth rate. Different measures of TFP are included in the database, as well as some partial measurements, such as labor productivity.

In our empirical model, we use GDP as the output and capital and labor as inputs. GDP in the WPD is the chain-weighted real index, adjusted for purchasing power parity using constant 1996 prices. When the data for one or more of the end years are missing, WPD uses the growth rates of real GDP to impute the value. Information about the real GDP growth rates is obtained from World Development Indicators from World Bank. When the GDP values are missing for the middle of the series, they use interpolated data. Another problem concerns the calculation of capital, which might require GDP and investment data from before 1960. They backcast both GDP and investment to make up for the missing data from before 1960.

The difficulty in measuring capital is largely due to the flow of capital service, which cannot be easily measured for a number of countries covered in WPD. Thus, WPD assumes that the capital services are proportional to the capital stock and provides several measurements of capital stock, denoted as K06, K13, Ks, and Keff, respectively. The differences in these measurements are reflected in how the initial capital stock is computed, the depreciation rate, whether the rate is constant or changes over time, and whether the asset lifetime is explicitly accounted for.

Due to limited data, we will use K06, K13, and Ks as measurements of capital. All three variables are based on the perpetual inventory method and assume a constant depreciation rate. The perpetual inventory method states that the capital stock in any period t equals the remaining capital stock from the previous period after depreciation plus the new investment made in the last period. Dated back to the starting period, the capital stock at period t can be expressed as a function of the initial capital K_0 and the depreciation rate δ,

$$K_t = (1-\delta)K_{t-1} + I_{t-1} = (1-\delta)^t K_0 + \sum_{i=1}^{t}(1-\delta)^{t-1} I_i \tag{3.29}$$

where I_t is the investment in each period, which is also provided in WPD. The three variables then differ in the methods used to estimate the initial capital stock and the value assumed for the depreciation rate. Both K06 and K13 use ten years of investments as the estimate for the initial capital stock K_0, i.e., investments from 1950 to 1959 for the capital series to start from 1960. The only difference between the two variables is the depreciation rate assumed: 6 percent for K06 and 13.3 percent for K13. The rapid depreciation rate used for K13 means this measure places more emphasis on recent investments and the initial capital stock has relatively less impact. The value assumed for δ matches the double-declining balancing method in accounting, implying an asset lifetime of 15 years. K06, by contrast, is more affected by the initial capital stock.

According to the data description given by WPD, the initial capital stock for Ks is computed by assuming the country is at its steady state capital-output ratio (Hence

Ks). The equation used is

$$k = \frac{i}{g + \delta} \tag{3.30}$$

where k is the capital-output ratio (K/Y), g is the growth rate of real GDP, and i is the investment ratio (I/Y). An estimate of the capital-output ratio of the starting year can be obtained by using equation (3.30). The estimate of the initial year's output is then calculated by multiplying this ratio by Y_0. Compared with K06 and K13, Ks does not require the extra ten years of data for the calculation of K_0.

For the labor input, we use the standard measure from the empirical literature: labor force. WPD provides five measurements for the labor inputs: labor force, employment (EMP), derived employment (EDMP), hours worked based on employment (HEMP) and hours worked based on derived employment (HDEMP). Labor force is used as the basis for the other labor measures. Employment is obtained by adjusting the labor force for the population that is employed. A direct measure of employment leads to EMP, and the derived value, which is obtained by applying the unemployment rate to labor force data, leads to DEMP. Further adjustment of EMP and DEMP for the numbers in hours worked results in the last two measurement: HEMP and HDEMP. It is generally considered that productivity estimates based on a finer definition of labor, such as HEMP and HDEPM, are more accurate. The use of labor force may result in underestimation of productivity since not all of the labor force is fully utilized. But labor force data usually has better quality and is available in many more countries over the studied time period compared to the alternatives. The country coverage of EMP is about 50 percent less than that of the labor force. In such a tradeoff, we choose labor force as our labor input to analyze productivity in different regions and income groups.

We divide all the countries into three groups based on their economic development levels and geographic location. Due to limited data, the countries are selected such that we have a balanced panel over the period 1960 to 2012. The countries in each group are listed below:

OECD: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Greece, Iceland, Ireland, Italy, Japan, Republic of Korea, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, UK and USA.

Asia: Bangladesh, China, Hong Kong (SAR of China), India, Indonesia, Israel, Malaysia, Pakistan, Philippines, Singapore, Sri Lanka, Taiwan (Province of China), and Thailand.

Latin America: Argentina, Brazil, Chile, Colombia, Ecuador, Guatemala, Jamaica, Panama, Peru, Trinidad and Tobago, and Venezuela.

3.6 Results

We first examine the results of the specification without constant-returns-to-scale constraint. The coefficients estimates of each model are shown in Tables 3.1 to 3.3. We can see that for OECD countries, the estimated coefficients of labor and capital approximately sum to one, which reflects constant returns to scale. For Asian countries, the coefficients of labor input are much lower than those of the OECD group based on results from standard panel data model and the BC model, while the coefficients of capital range in a similar interval for both country groups. Thus the output growth of Asian countries comes more from the investment of capital. Compared with the within, GLS, and BC estimators, the two CSS estimators give higher estimates for the labor coefficients. Also note that if we just look at the CSS estimators, coefficients estimates differ significantly with the use of different variable measurements.

When K06 or K13 is used as the capital input, coefficients estimated for capital and labor are approximately the same (around 0.4), implying similar contribution to the output from the two inputs. On the other hand, when Ks is used as the capital input, coefficients of capital are more or less twice as large as that of labor. In the Latin America group, the coefficient estimates of capital are fairly stable across all model settings, as is the case in OECD and Asian countries, but the coefficient estimates of labor differ significantly with different model specifications. We can see that the labor coefficients estimated from the standard panel data model range between 0.35 and 0.4, while the cofficients obtained from the CSS model are negative. In summary, the estimation results vary substantially in-between estimators for a given measurement of capital and also across measurements of capital for a given estimator. This is especially true for the two groups consisting of developing coutries, which have experienced large changes in social and economic development during the past fifty years. Thus basing inference and prediction on just one model is risky. The goodness of fit of some models can be superficially high because they happen to fit a specific set of variables from a certain period.

The weights obtained using the jackknife criterion are reported in Tables 3.7 and 3.8. Table 3.7 shows the weights minimizing the squared prediction errors with block cross-validation and Table 3.8 shows the weights minimizing the leave-one-out cross-validation criterion. Looking at the two results, we can see that for OECD countries, of which the data are more stable compared with the other two groups, the weights obtained using two methods are very close. However, for Asian and Latin American countries, the weights obtained using the two criteria differ widely. Using block cross-validation, the CSSW estimator gets zero weight for all three capital measurements, while with leave-one-out cross-validation, the first four estimators – Fixed, Random, CSSW, and CSSG – all get a weight around 0.24-0.25, and the BC estimator gets a weight close to zero. Thus it is necessary to incur the computational burden of running the regression $N \times T$ times to get the optimal weights, especially for data with considerable variations across different blocks. We will use the weights shown in Table 3.8 to average the five candidate models. For all the three country groups, the BC estimators get fairly small weights, even zero. This is probably due to the use of a linear approximation of the ML estimator and needs further analysis.

The technical innovations are proxied by the time variables, i.e. t and t^2, and the technical efficiencies are estimated using stochastic frontier models. Combining the two effects, we have the estimates of total factor productivity. Based on the ending-period productivity levels, we can calculate the annualized productivity growth rates, which are shown in Table 3.10. We can see that the growth rates calculated with different models and capital measurements differ a lot, while the averaged values are much more stable. The Asian countries achieve the highest growth rates at 1.8%, 1.532% and 1.509% for Ks, K06, and K13, followed by the Latin American countries at 1.038%, 0.511% and 0.488% correspondingly. The productivity growth rates of OECD rank last at 0.497%, 0.355% and 0.349% for Ks, K06 and K13. The results exhibit convergence in economics, that is, developing countries grow at faster paces than developed countries. In general, developing countries are able to grow faster by learning the existing technology and experiences from developed countries.

The estimates of total factor productivity for the entire studied period are plotted in Figures 3.1 to 3.3. Figure 3.1 shows the TFP estimates of OECD countries using different stochastic frontier models and capital measurements. Though the estimated values differ, the changing patterns are similar for all the specifications and data measurements. We can see that during the studied period, TFP increases for most of the years, and experiences different levels of drops around 1975, 1988, 1993, and 2008. The TFP level reaches its highest point around 2007, then decreases sharply to the lowest point of the recent 20 years corresponding to the financial crisis. It then increases slightly in subsequent recovering years. For the Asian countries, TFP estimates exhibit a smoother increasing pattern. A notable drop happens around 1997 to 1998, when many countries included in our study experienced severe financial crisis. The recent crisis had a smaller impact on the productivity of Asian countries compared with OECD countries. There is only a slight drop of TFP around 2008 as shown in Figure 3.2. Figure 3.3 shows TFP estimates for Latin American countries. We can observe an obvious drop in productivity from 1980 to 1983, corresponding to the debt crisis at that time. After this drop, the growth of TFP is slow and the TFP plots are comparatively flat. A faster growth in productivity began in 2002 and ended with the crisis in 2008.

From Figures3.1 to 3.3, we find that the Asian group experienced the fastest growth in productivity. We can compare the TFP estimates of the years at the beginning and end of our studied period, and calculate the annualized growth rates. The annualized growth rates of Asian countries are 1.26%, 1.03% and 1.02% for the use of Ks, K06 and K13, respectively. The growth rates of Latin American countries follows at 1.00%, 0.54% and 0.52%, while the OECD grows slowest with the respective rates of 0.49%, 0.29% and 0.28%. Note that TFP estimates from the BC model tends to be negative, especially for the Latin American group. It is unlikely that the TFP estimates are negative for almost the entire studied period. This unrealistic result can be interpreted as an indicator that the BC model is an inadequate approximation of the underlying DGP given the observed data. This poor explaining power can also be seen from the near-to-zero weight obtained from our model averaging process.

Another point to notice is that when we use K06 or K13 as the capital measurements, the TFP estimates are lower than those obtained using Ks. We can see that for the Latin American countries, even the averaged TFP estimates are negative for before 1970 and from 1983 to 1985. As the scale of TFP estimates varies with the use of different model specifications and data measurements, we focus more on the analysis of changing paths and the comparative differences among different groups.

Let us now examine the case with constant-returns-to-scale assumption. The coefficients estimates are shown in Tables 3.4 to 3.6. The coefficients estimates of logged capital stay roughly in the same range: for OECD countries, the estimates range around 0.5 to 0.7; for Asian countries, the estimates are around 0.6 using standard fixed and random estimators, around 0.4 using the two CSS estimators, and are about 0.8 using the BC estimator; for Latin American countries, the standard panel data methods give estimates around 0.6 and the rest estimators give estimates about 0.7 to 0.8. Now with the constant-returns-to-scale assumption, the coefficients estimates of logged labor are just 1 minus the corresponding coefficients estimates of logged capital.

The optimal weights of the candidate estimators under the CRS assumption are shown in Table 3.9. We can see that for all the three country groups and the three measurements of capital, the weights of the random effects estimators are much larger than the rest, ranging from 0.7 to 0.9. The weighted estimates of TFP are thus heavily affected by the results obtained using GLS method.

The annualized growth rates of TFP are shown in Table 3.11. The Asian countries still have the highest growth rates at 0.727%, 0.467% and 0.465% corresponding to Ks, K06 and K13, followed by the OECD countries at 0.398%, 0.013% and 0.015%, while the Latin American countries come at last at 0.26%, −0.12% and −0.109%

for the three capital measurements Ks, K06 and K13. We can see that for all three groups, the estimates of TFP growth rates drop considerably compared with the case without CRS assumption. This decrease may indicate that the growth in the past 50 years largely comes from scale of economics, especially for the two groups consisting of developing countries. The estimates of TFP growth rates of Asian countries drop the most with about 1% difference between the specifications with and without CRS assumption. The growth rates of Latin American countries fall about $0.6 - 0.7\%$ for the three capital measurements, and the growth rates of OECD countries decrease only $0.1 - 0.3\%$. As the developing countries are not as capital intensive as developed countries, they may still in the phase of increasing returns to scale, and benefit from the capital investment.

From both Table 3.10 and Table 3.11, we can see that the estimates of TFP growth rates obtained using K06 and K13 are much lower than those obtained using Ks. For the Latin American group, when we use K06 and K13 as capital measurements and assume constant returns to scale, the growth rates estimates are even negative. We need to further explore how accurately each measurement can represent the variable. Another approach is to weight the results using different measurements as we weight the estimators.

The estimates of TFP over the entire studied period with CRS constrain are shown in Figure 3.4 to 3.6. We can see that, compared with the case without CRS constraint, the TFP estimates are more fluctuated less upward trends.

In summary, based on the results of our averaged stochastic frontier analysis, Asian countries have the highest growth rates of both the technical innovations and total factor productivity. The TFP of OECD countries grows slowest, and that of the Latin American countries falls in between. Generally, the productivities of all three groups increase with fluctuations during the studied period. Notable drops in productivity are more or less correlated with regional or global financial crises. With CRS assumption, the TFP estimates drop considerably for all three groups, showing that the scale of economics plays an important role in the economic growth in past 50 years.

3.7 A Discussion of Structural Model for the Efficiency Analysis

In this section, we briefly show how to employ a structural model to analyze country-level efficiencies. The structural model we use is based on the one proposed in Pakes and McGuire (1992) (PM hereafter), and is modified to be more suitable for country level analysis.

In our model, countries make decisions at each period to maximize the expected discounted value of all future net output. The net output is defined to be gross outputs minus the cost of gross inputs, which is parallel to the profit of a firm. The decisions are based on countries' current information set, which is represented by the following state variables:

$$(\omega, s) \in \Omega \times S. \tag{3.31}$$

Here ω is an index representing the country's efficiency level, and s is a vector recording the number of countries at each efficiency level. s can be seen as an efficiency structure of the countries under study. At each period, each country decides how much to invest for the next period based on its current information set as well as its perceptions of the future efficiency distribution of other countries. These simultaneous decisions then generate the true efficiency structure in the next period. The economy is at equilibrium if, for each country, the perceptions of the future efficiency distribution of other countries are the same as the true efficiency distribution of others. Thus in equilibrium, the expected future efficiency structure will turn out to be the true efficiency structure.

We define a net output function, π, to measure the single-period production process of a country, which resembles the profit function used to describe a firm's production.

$$\pi(\cdot,\cdot) : \Omega \times S \to \mathbb{R} \tag{3.32}$$

where $\pi(\cdot,\cdot)$ is a function of the state variables. We assume that $\pi(\omega, s)$ is increasing in ω for every s.

In the PM model, aside from investment, firms also make decisions about whether to enter or exit the studied industry at each period. The entry decision is made by comparing the opportunity cost and the expected discounted future profits of entering, and the exit decision is made by comparing the expected discounted future profits of staying and a scrap value (the value of investing the firm's resources to its best alternative use). As countries are aggregated entities with a variety of industries and production units, the entry and exit options are not realistic for them. Thus we assume all the countries are active and aim to maximize the expected discounted future net output. The maximized values are assumed to be greater than zero.

As stated above, we assume the investment decisions made by countries will generate the efficiency structure in the next period. For a specific country, we assume its efficiency level in the next period is determined by a Markov process that depends on its current efficiency level, the investment made in the current period and some exogenous factors. With $\omega \in \Omega$ being the current efficiency level, let $\omega' \in \Omega$ be the country's efficiency level of the next period and x be the investment. We assume the country's investment will improve its efficiency by τ, and there exist negative exogenous factors, represented by v, for all countries. Thus the evolution of a country's efficiency can be formulated as

$$\omega' = \omega + \tau - v \tag{3.33}$$

where each τ and v follows a Markov process with the transition probabilities stated below.

$$p(\tau) = \begin{cases} \frac{ax}{1+ax} & \text{if } \tau = 1, \\ \frac{1}{1+ax} & \text{if } \tau = 0. \end{cases}$$

$$p(v) = \begin{cases} \delta & \text{if } v = 1, \\ 1-\delta & \text{if } v = 0. \end{cases}$$

Based on this formulation, the possible incremental change in efficiency level is limited to the set $\{-1, 0, 1\}$ in each period. But we can decide how many decision periods there are in a real time period, for example, a calendar year. If assuming multiple decision periods in a year, we can observe many more possible moves of the efficiency level. Note that in this case, we need to make adjustments to the discount rate according to the length of the decision period.

Economic theory tells us that the efficiency level ω should be in a bounded range. There will be a high enough level that further improvement in efficiency comes at a cost that will reduce the expected discounted net output, such that a country will not invest any more regardless of what the efficiency structure is. In the algorithm described by Pakes and Gowrisankaran (1993), the maximal efficiency level $\bar{\omega}$ was calculated as the highest achievable value for a monopolist. Parallel to the concept of a monopolist in an industry, suppose there exists a central planner that aims to maximize the overall expected discounted value of future net output and is able to make arrangements for all resources. This planner should be capable of achieving the highest possible efficiency level and thus it should be possible in principle to get an estimate for the upper bound of efficiencies over all countries. We denote the upper bound as $\bar{\omega}$.

Remember that, besides the investment, firms need to make decisions about entry or exit in the PM model. The possibility of exit creates a lower bound for the efficiencies. In our country level analysis, we assume there exists a sufficiently low efficiency level at which a country can just feed itself, and has no output left to invest. Without extra capital to improve its efficiency for the next period, the country can do nothing but stay at this surviving level of output. In other words, we may view this as a country exiting when its net output hits zero. The lower bound of the efficiencies is then calculated as the level that induces even a united government to "exit", and is denoted as $\underline{\omega}$.

The cost of investment is c, and the discount rate is denoted as β. We assume output is homogeneous across countries, and these countries have different marginal costs. Thus different efficiency levels are represented by the different marginal costs. The Bellman equation of our maximization problem is stated as:

$$V(\omega, s) = max\{0, sup_{x\geq 0}[\pi(\omega, s) - cx + \beta \sum V(\omega + \tau, s)p(\tau|x, v)p(v)]\}. \qquad (3.34)$$

The efficiency estimates obtained using such a structural model can benefit from the constraint implied by the consumption side of the economy. Compared with the reduced form analysis, the derived (nonlinear) reduced form estimates of productivity based on the structural model estimates might provide better insights on the whole economy. As a full examination of the structural model goes beyond the focus of this chapter, we will leave the detailed discussion and implementation of our structural model for future work.

3.8 Conclusion

This chapter introduces the model averaging method to the analysis of productivity using stochastic frontier models. The optimal weights are calculated by minimizing the jackknife model averaging criterion, which is a leave-one-out cross-validation method. Though each candidate model needs to be estimated $N \times T$ times, the procedure is not complex and the computational burden is moderate in our application. As shown in our productivity analysis of different country groups, the estimated efficiency and productivity level, as well as their growth rates, vary greatly with different model specifications and data measurement methods. The model averaging technique reduces the risk of incurring superficially high goodness of fit of some specific model setting, and provides us with more stable and reliable estimates of efficiencies and productivities. The choice of the frequentist model averaging method avoids the difficulty in choosing proper priors that are required in Bayesian model averaging, and the jackknife model averaging estimator achieves optimality in the sense that it can asymptotically approache the lowest possible expected squared errors.

Based on the our reduced form analysis, the Asian countries experienced the highest growth rate from 1960 to 2012, in both TFP and technical innovations. The OECD countries, on the contrary, have the lowest productivity growth rate, and the Latin American group falls in the middle.

	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
Fixed	Ks		K06		K13	
lnK	0.53880	0.00934	0.56314	0.00926	0.56494	0.00923
lnL	0.38028	0.02377	0.35499	0.02296	0.35120	0.02287
t	0.01030	0.00081	-0.00021	0.00088	-0.00012	0.00088
t^2	-0.00008	0.00001	0.00006	0.00001	0.00006	0.00001
Random	Ks		K06		K13	
lnK	0.54082	0.00926	0.56481	0.00918	0.56652	0.00915
lnL	0.41393	0.01774	0.38742	0.01749	0.38409	0.01754
t	0.00974	0.00078	-0.00076	0.00086	-0.00067	0.00085
t^2	-0.00007	0.00001	0.00006	0.00001	0.00006	0.00001
CSSW	Ks		K06		K13	
lnK	0.61833	0.01962	0.47487	0.01572	0.48109	0.01561
lnL	0.12881	0.04592	0.22057	0.04693	0.21916	0.04652
t	*	*	*	*	*	*
t^2	*	*	*	*	*	*
CSSG	Ks		K06		K13	
lnK	0.62300	0.01732	0.49741	0.01449	0.50296	0.01438
lnL	0.23151	0.03328	0.32925	0.03521	0.32633	0.03483
t	0.00742	0.00202	0.00501	0.00210	0.00479	0.00208
t^2	-0.00005	0.00003	0.00002	0.00003	0.00002	0.00003
BC	Ks		K06		K13	
lnK	0.65211	*	0.66594	*	0.66691	*
lnL	0.35156	*	0.33497	*	0.33399	*
t	0.00416	*	-0.00758	*	-0.00743	*
t^2	-0.00004	*	0.00012	*	0.00011	*

Table 3.1 : Coefficients Estimates-OECD

	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
Fixed	Ks		K06		K13	
lnK	0.61523	0.02323	0.56467	0.02505	0.56578	0.02504
lnL	-0.03732	0.07212	0.07675	0.07727	0.07767	0.07720
t	0.01570	0.00316	0.00706	0.00372	0.00701	0.00372
t^2	0.00003	0.00003	0.00016	0.00003	0.00016	0.00003
Random	Ks		K06		K13	
lnK	0.61664	0.02299	0.57044	0.02455	0.57158	0.02454
lnL	0.07771	0.05458	0.18470	0.04877	0.18518	0.04857
t	0.01205	0.00280	0.00315	0.00314	0.00311	0.00314
t^2	0.00005	0.00003	0.00017	0.00003	0.00017	0.00003
CSSW	Ks		K06		K13	
lnK	0.41294	0.02639	0.40995	0.02595	0.41901	0.02575
lnL	0.19899	0.09231	0.42538	0.08819	0.41810	0.08748
t	*	*	*	*	*	*
t^2	*	*	*	*	*	*
CSSG	Ks		K06		K13	
lnK	0.43497	0.02478	0.43495	0.02437	0.44366	0.02417
lnL	0.23653	0.04871	0.40283	0.04005	0.39651	0.03953
t	0.02297	0.00578	0.01052	0.00610	0.00988	0.00606
t^2	-0.00001	0.00007	0.00011	0.00008	0.00011	0.00007
BC	Ks		K06		K13	
lnK	0.71577	*	0.70603	*	0.70637	*
lnL	0.17821	*	0.18700	*	0.18678	*
t	-0.00665	*	-0.01798	*	-0.01790	*
t^2	0.00017	*	0.00033	*	0.00033	*

Table 3.2 : Coefficients Estimates-Asia

	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
Fixed	Ks		K06		K13	
lnK	0.60869	0.02934	0.63378	0.02583	0.63554	0.02569
lnL	0.37180	0.05186	0.38825	0.04753	0.38621	0.04734
t	0.00286	0.00236	-0.00881	0.00238	-0.00865	0.00236
t^2	0.00000	0.00003	0.00014	0.00003	0.00014	0.00003
Random	Ks		K06		K13	
lnK	0.61720	0.02764	0.63458	0.02451	0.63635	0.02439
lnL	0.36498	0.03326	0.36231	0.03131	0.36061	0.03119
t	0.00264	0.00187	-0.00813	0.00193	-0.00798	0.00192
t^2	0.00000	0.00003	0.00014	0.00003	0.00014	0.00003
CSSW	Ks		K06		K13	
lnK	0.67861	0.04307	0.64647	0.03748	0.65158	0.03706
lnL	-0.47047	0.15100	-0.27746	0.14935	-0.26534	0.14835
t	*	*	*	*	*	*
t^2	*	*	*	*	*	*
CSSG	Ks		K06		K13	
lnK	0.69833	0.03858	0.67029	0.03324	0.67476	0.03286
lnL	-0.21455	0.10933	-0.02502	0.09855	-0.01659	0.09708
t	0.01499	0.00727	0.00046	0.00673	0.00016	0.00669
t^2	0.00000	0.00009	0.00014	0.00009	0.00014	0.00009
BC	Ks		K06		K13	
lnK	0.70338	*	0.69189	*	0.69313	*
lnL	0.28893	*	0.30146	*	0.30035	*
t	0.00058	*	-0.01011	*	-0.00992	*
t^2	0.00002	*	0.00017	*	0.00016	*

Table 3.3 : Coefficients Estimates-Latin America

	Est.	Std. Err	Est.	Std. Err	Est.	Std. Err
Fixed	Ks		K06		K13	
lnK	0.54020	0.00938	0.56454	0.00930	0.56633	0.00928
lnL	0.45980	*	0.43546	*	0.43367	*
t	0.00917	0.00076	-0.00137	0.00083	-0.00131	0.00083
t^2	-0.00007	0.00001	0.00006	0.00001	0.00006	0.00001
Random	Ks		K06		K13	
lnK	0.54422	0.00925	0.56797	0.00918	0.56970	0.00915
lnL	0.45578	*	0.43203	*	0.43030	*
t	0.00899	0.00075	-0.00158	0.00083	-0.00152	0.00082
t^2	-0.00007	0.00001	0.00006	0.00001	0.00006	0.00001
CSSW	Ks		K06		K13	
lnK	0.65818	0.01814	0.51105	0.01473	0.51668	0.01463
lnL	0.34182	*	0.48895	*	0.48332	*
t	0.00394	*	0.00189	*	0.00171	*
t^2	-0.00003	*	0.00003	*	0.00003	*
CSSG	Ks		K06		K13	
lnK	0.65012	0.01598	0.52410	0.01377	0.52916	0.01368
lnL	0.34988	*	0.47590	*	0.47084	*
t	0.00430	0.00205	0.00110	0.00195	0.00095	0.00194
t^2	-0.00004	0.00003	0.00004	0.00003	0.00004	0.00003
BC	Ks		K06		K13	
lnK	0.64852	*	0.66505	*	0.66603	*
lnL	0.35148	*	0.33495	*	0.33397	*
t	0.00437	*	-0.00751	*	-0.00737	*
t^2	-0.00004	*	0.00012	*	0.00011	*

Table 3.4 : Coefficients Estimates with CRS-OECD

	Est.	Std. Err	Est.	Std. Err	Est.	Std. Err
Fixed	Ks		K06		K13	
lnK	0.63748	0.02347	0.58895	0.02483	0.59003	*
lnL	0.36252	*	0.41105	*	0.40997	*
t	0.00148	0.00207	-0.00573	0.00249	-0.00571	*
t^2	0.00009	0.00003	0.00021	0.00003	0.00021	*
Random	Ks		K06		K13	
lnK	0.65378	0.02249	0.60730	0.02383	0.60838	*
lnL	0.34622	*	0.39270	*	0.39162	*
t	0.00056	0.00203	-0.00708	0.00243	-0.00705	*
t^2	0.00009	0.00003	0.00022	0.00003	0.00022	*
CSSW	Ks		K06		K13	
lnK	0.40626	0.02673	0.41516	0.02585	0.42413	*
lnL	0.59374	*	0.58484	*	0.57587	*
t	0.01449	0.00000	0.00698	0.00000	0.00639	*
t^2	0.00002	0.00000	0.00012	0.00000	0.00012	*
CSSG	Ks		K06		K13	
lnK	0.43765	0.02502	0.44340	0.02392	0.45186	*
lnL	0.56235	*	0.55660	*	0.54814	*
t	0.01272	0.00514	0.00491	0.00569	0.00437	*
t^2	0.00003	0.00007	0.00013	0.00007	0.00013	*
BC	Ks		K06		K13	
lnK	0.81461	*	0.79092	*	0.79541	*
lnL	0.18539	*	0.20908	*	0.20459	*
t	-0.01692	*	-0.02605	*	-0.02808	*
t^2	0.00023	*	0.00037	*	0.00039	*

Table 3.5 : Coefficients Estimates with CRS-Asia

	Est.	Std. Err	Est.	Std. Err	Est.	Std. Err
Fixed	Ks		K06		K13	
lnK	0.61212	*	0.63003	0.02442	0.63185	0.02430
lnL	0.38788	*	0.36997	*	0.36815	*
t	0.00224	*	-0.00806	0.00168	-0.00791	0.00167
t^2	0.00000	*	0.00014	0.00003	0.00014	0.00003
Random	Ks		K06		K13	
lnK	0.62157	*	0.63585	0.02360	0.63758	0.02349
lnL	0.37843	*	0.36415	*	0.36242	*
t	0.00204	*	-0.00827	0.00167	-0.00811	0.00165
t^2	0.00001	*	0.00014	0.00003	0.00014	0.00003
CSSW	Ks		K06		K13	
lnK	0.78429	*	0.72357	0.03187	0.72621	0.03151
lnL	0.21571	*	0.27643	*	0.27379	*
t	-0.00129	*	-0.01145	*	-0.01130	*
t^2	0.00004	*	0.00018	*	0.00018	*
CSSG	Ks		K06		K13	
lnK	0.76894	*	0.71264	0.02883	0.71543	0.02856
lnL	0.23106	*	0.28736	*	0.28457	*
t	-0.00098	*	-0.01105	0.00540	-0.01091	0.00540
t^2	0.00004	*	0.00018	0.00009	0.00017	0.00009
BC	Ks	*	K06		K13	
lnK	0.71146	*	0.69864	*	0.69975	*
lnL	0.28854	*	0.30136	*	0.30025	*
t	0.00020	*	-0.01054	*	-0.01035	*
t^2	0.00003	*	0.00017	*	0.00017	*

Table 3.6 : Coefficients Estimates with CRS-Latin America

	Fixed	Random	CSSW	CSSG	BC
Ks, L	0.27265	0.27323	0.22519	0.22893	0.00000
K06, L	0.24872	0.25045	0.23123	0.23795	0.03165
K13, L	0.24845	0.25008	0.23254	0.23871	0.03023

(a) OECD

	Fixed	Random	CSSW	CSSG	BC
Ks, L	0.49664	0.45669	0.00000	0.04304	0.00363
K06, L	0.47975	0.43266	0.00000	0.07231	0.01529
K13, L	0.48083	0.43115	0.00000	0.07382	0.01420

(b) Asia

	Fixed	Random	CSSW	CSSG	BC
Ks, L	0.42554	0.42280	0.00000	0.14115	0.01051
K06, L	0.41339	0.39751	0.00000	0.17578	0.01332
K13, L	0.41311	0.39723	0.00000	0.17637	0.01330

(c) Latin America

Table 3.7 : Weights with block cross-validation

	Fixed	Random	CSSW	CSSG	BC
Ks, L	0.28699	0.28323	0.21864	0.21114	0.00000
K06, L	0.23563	0.23550	0.24607	0.24478	0.03803
K13, L	0.23595	0.23580	0.24628	0.24494	0.03703

(a) OECD

	Fixed	Random	CSSW	CSSG	BC
Ks, L	0.37007	0.41268	0.07097	0.14099	0.00529
K06, L	0.24149	0.24282	0.24820	0.24971	0.01778
K13, L	0.24094	0.24217	0.24891	0.25025	0.01773

(b) Asia

	Fixed	Random	CSSW	CSSG	BC
Ks, L	0.27526	0.26518	0.23203	0.22662	0.00091
K06, L	0.25517	0.24993	0.25016	0.24474	0.00000
K13, L	0.25517	0.25001	0.25009	0.24472	0.00000

(c) Latin America

Table 3.8 : Weights with leave-one-out cross-validation

	Fixed	Random	CSSW	CSSG	BC
Ks	0.00989	0.85410	0.00014	0.00014	0.13572
K06	0.01395	0.86172	0.00031	0.00032	0.12371
K13	0.01348	0.85981	0.00029	0.00030	0.12611

(a) OECD

	Fixed	Random	CSSW	CSSG	BC
Ks	0.03379	0.84643	0.09954	0.01176	0.00848
K06	0.03107	0.95179	0.00042	0.00043	0.01629
K13	0.02391	0.95865	0.00045	0.00046	0.01652

(b) Asia

	Fixed	Random	CSSW	CSSG	BC
Ks	0.01251	0.88118	0.02296	0.02058	0.06277
K06	0.00824	0.70097	0.02492	0.19435	0.07153
K13	0.00834	0.79736	0.02234	0.09854	0.07342

(c) Latin America

Table 3.9 : Weights with CRS Constraint

Tables and Figures

	Fixed	Random	CSSW	CSSG	BC	AVE
Ks	0.00548	0.00500	0.00529	0.00390	0.00134	0.00497
K06	0.00166	0.00120	0.00722	0.00493	-0.00275	0.00355
K13	0.00170	0.00124	0.00702	0.00478	-0.00269	0.00349

(a) OECD

	Fixed	Random	CSSW	CSSG	BC	AVE
Ks	0.01805	0.01516	0.02532	0.02299	0.00396	0.01800
K06	0.01547	0.01244	0.01778	0.01658	0.00038	0.01532
K13	0.01542	0.01239	0.01736	0.01617	0.00042	0.01509

(b) Asia

	Fixed	Random	CSSW	CSSG	BC	AVE
Ks	0.00339	0.00325	0.02230	0.01504	0.00200	0.01038
K06	-0.00137	-0.00074	0.01509	0.00763	-0.00161	0.00511
K13	-0.00137	-0.00075	0.01458	0.00725	-0.00162	0.00488

(c) Latin America

Table 3.10 : Annualized Total Factor Productivity Growth Rate

	Fixed	Random	CSSW	CSSG	BC	AVE
Ks	0.00449	0.00438	0.00121	0.00143	0.00148	0.00398
K06	0.00065	0.00053	0.00243	0.00200	-0.00271	0.00013
K13	0.00066	0.00055	0.00231	0.00190	-0.00265	0.00015

(a) OECD

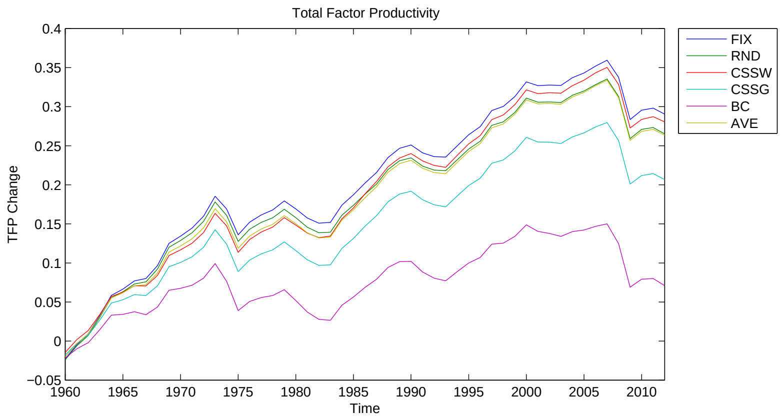
	Fixed	Random	CSSW	CSSG	BC	AVE
Ks	0.00689	0.00624	0.01614	0.01489	-0.00299	0.00727
K06	0.00564	0.00480	0.01353	0.01225	-0.00553	0.00467
K13	0.00564	0.00480	0.01316	0.01190	-0.00618	0.00465

(b) Asia

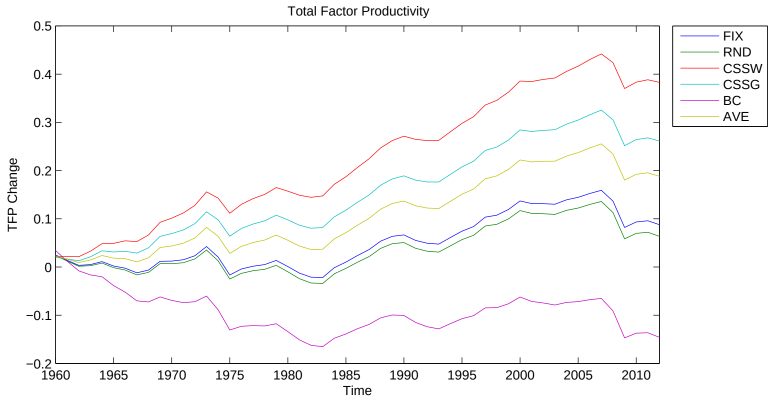
	Fixed	Random	CSSW	CSSG	BC	AVE
Ks	0.00285	0.00274	0.00088	0.00106	0.00171	0.00260
K06	-0.00074	-0.00084	-0.00231	-0.00213	-0.00190	-0.00120
K13	-0.00076	-0.00085	-0.00234	-0.00216	-0.00190	-0.00109

(c) Latin America

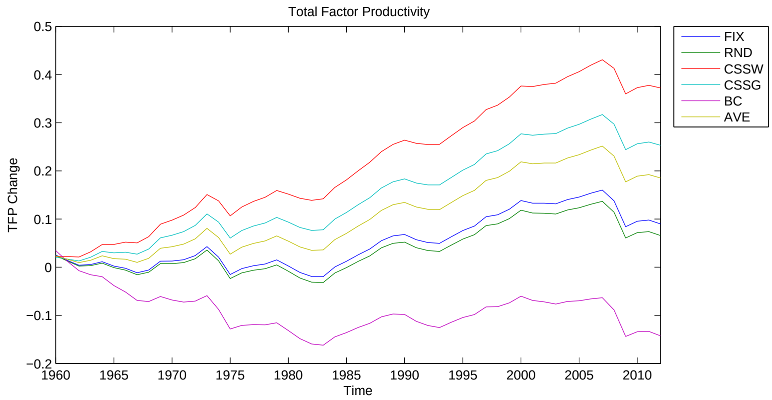
Table 3.11 : Annualized Total Factor Productivity Growth Rate with CRS



(a) Ks

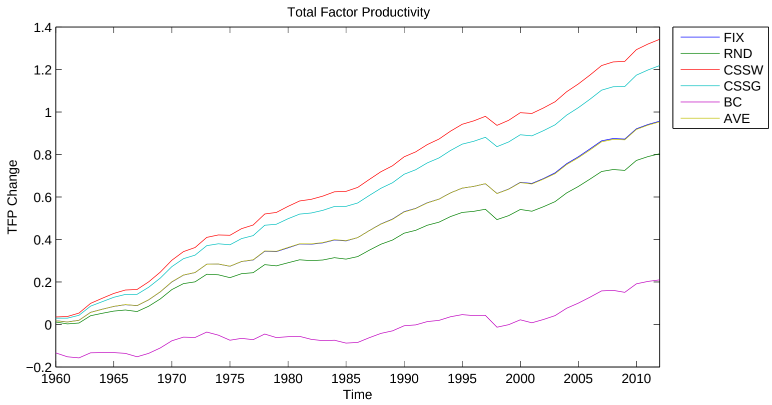


(b) K06

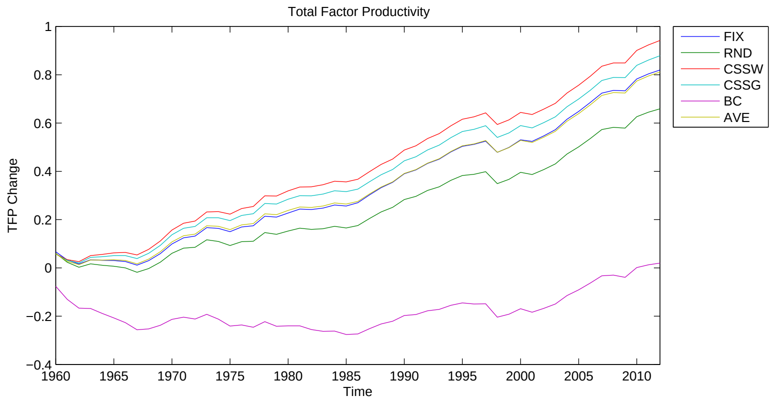


(c) K13

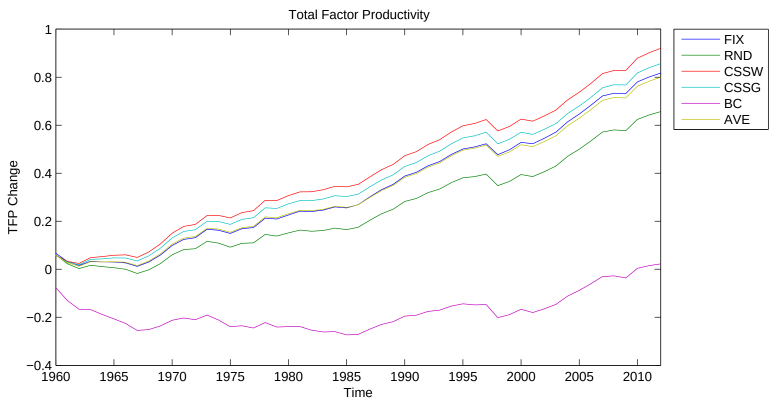
Figure 3.1 : Total Factor Productivity-OECD



(a) Ks

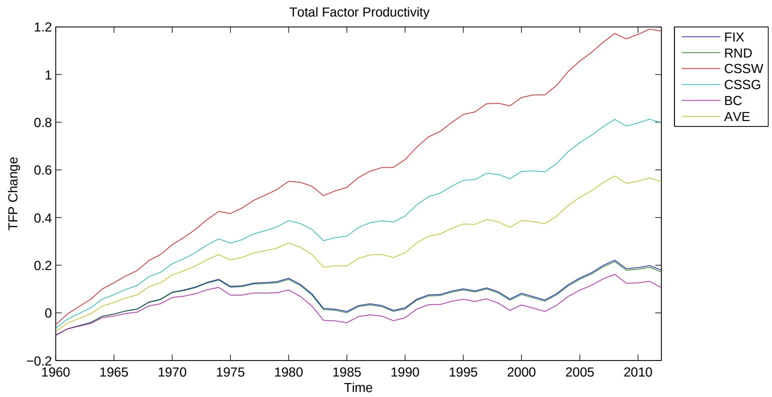


(b) K06

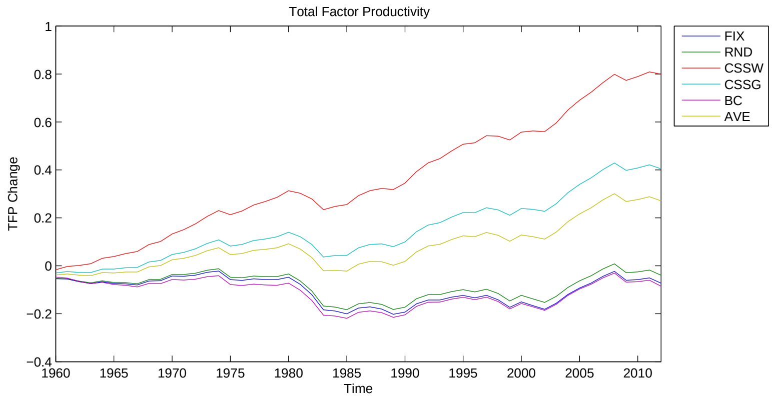


(c) K13

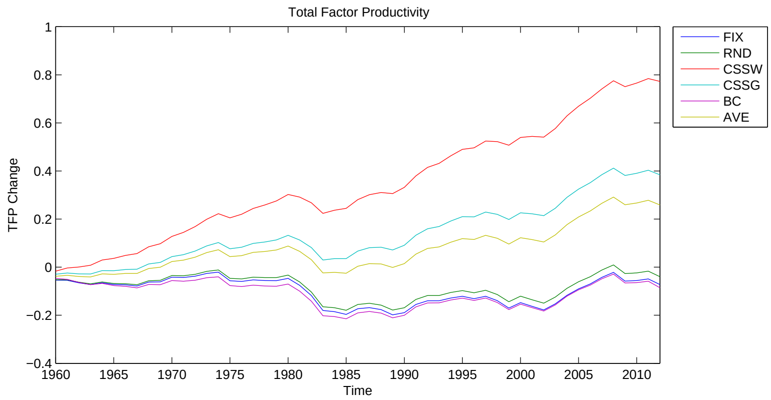
Figure 3.2 : Total Factor Productivity-Asia



(a) Ks

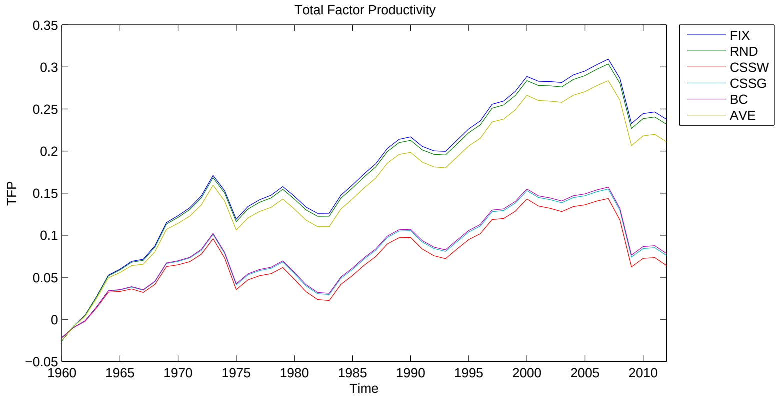


(b) K06

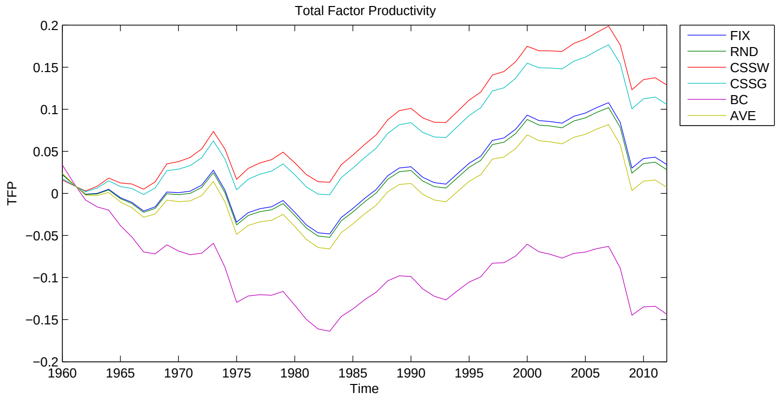


(c) K13

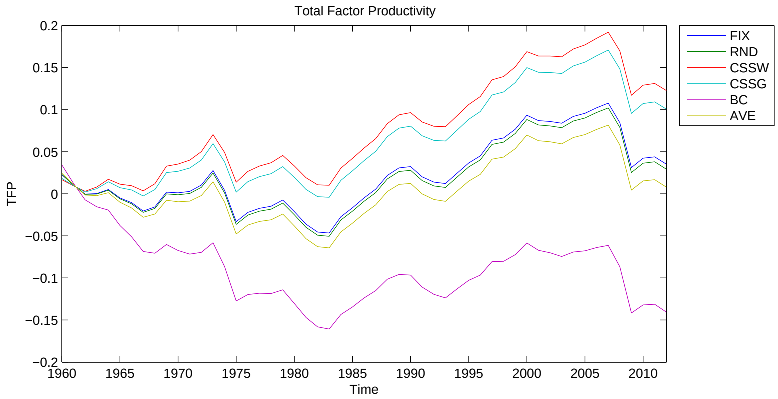
Figure 3.3 : Total Factor Productivity-Latin America



(a) Ks

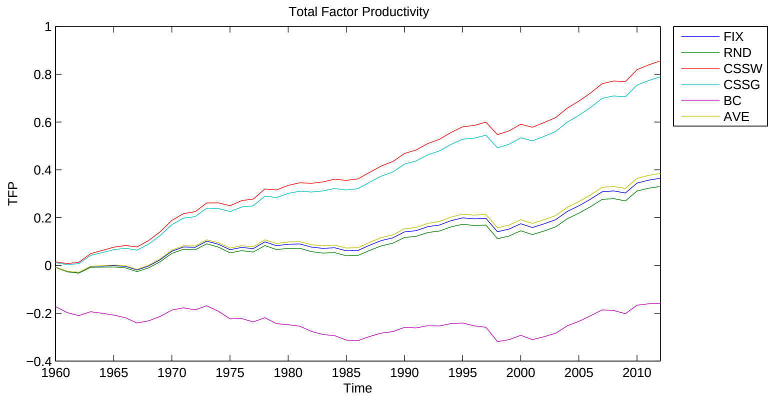


(b) K06

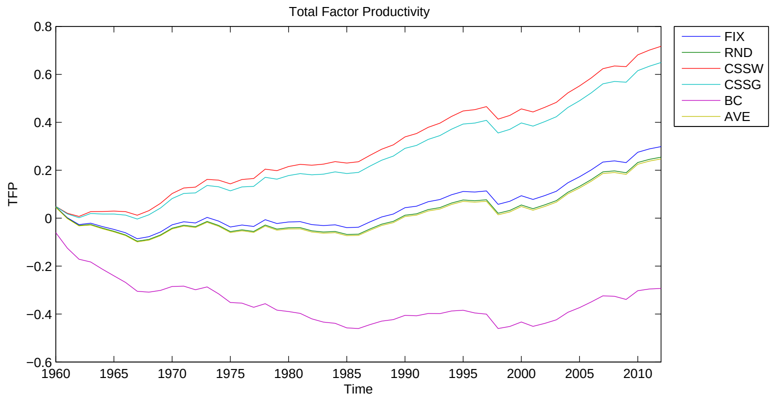


(c) K13

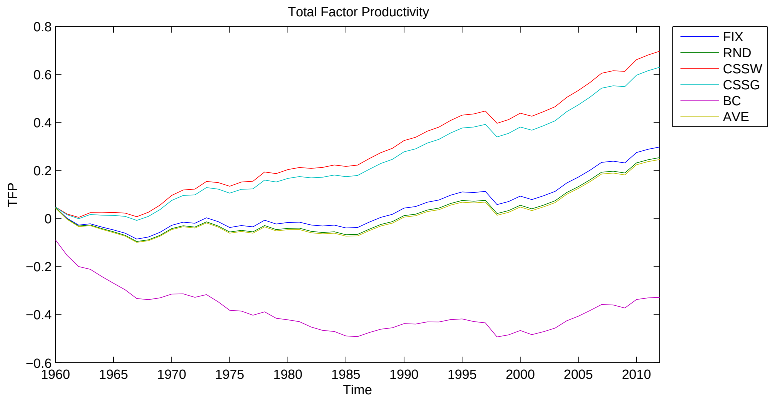
Figure 3.4 : Total Factor Productivity with CRS-OECD



(a) Ks

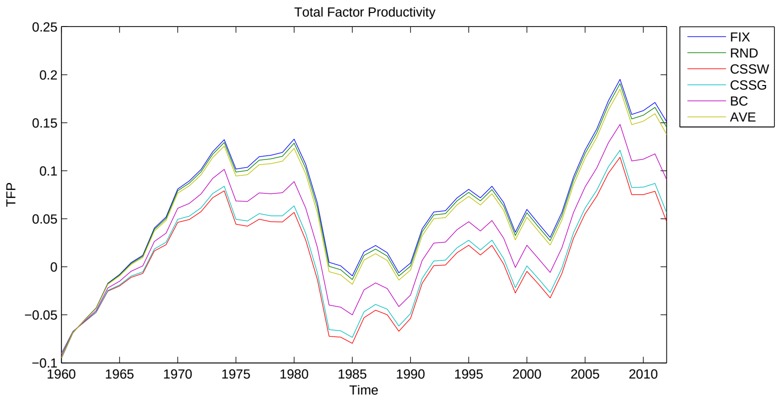


(b) K06

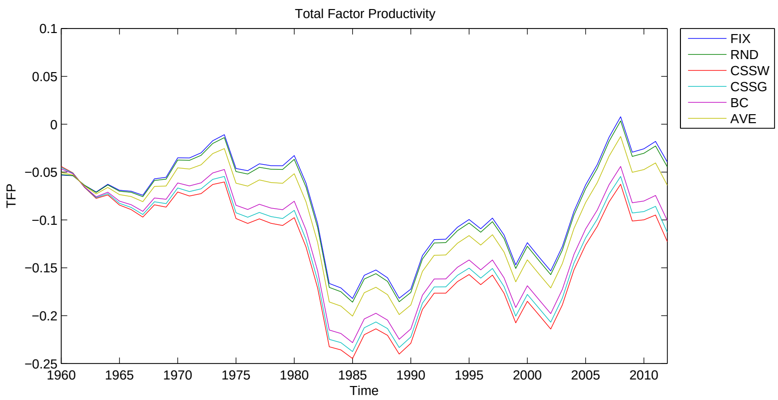


(c) K13

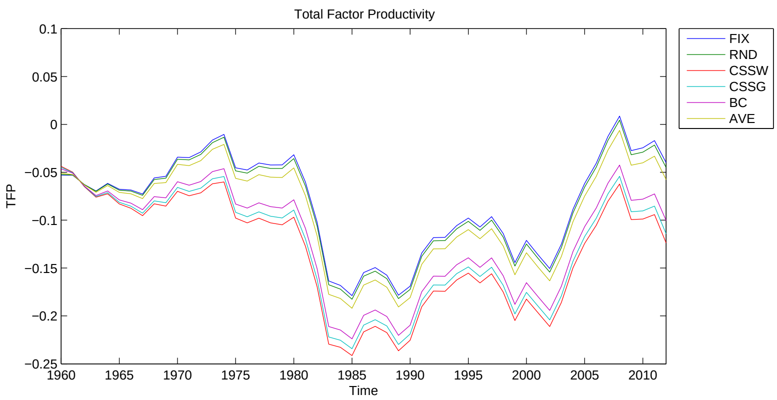
Figure 3.5 : Total Factor Productivity with CRS-Asia



(a) Ks



(b) K06



(c) K13

Figure 3.6 : Total Factor Productivity with CRS-Latin America

Conclusion

This dissertation analyzes the production side of the economy, and explores the model averaging approach to provide robust and efficient estimates of parameters of interest.

Chapter 1 reviews methods of estimating productivity, with a focus on the stochastic frontier models. I make a detailed discussion on twelve representative specifications in the stochastic frontier literature, which constitute a set of candidate models for empirical studies.

Chapter 2 turns to explore the relation between the prices of the outputs and the prices of the inputs. I utilize duality theory in this chapter to decompose the price of the output into the prices of the inputs, and carry out an empirical application on the housing market. I also apply the traditional hedonic regression methods to the same data set and compare the resulting price indexes from these different approaches. The method using duality theory can circumvent the multicollinearity problem associated with hedonic regressions, and the resulting price indexes exhibit smoother trends compared with those obtained using hedonic regressions. Both approaches are based on sound economic theories, and we cannot make a conclusive statement that one method is better than the other.

Both Chapters 1 and 2 focus on panel data models, and face the problem of evaluating different modeling settings for the same topic. I explore the model averaging approach in Chapter 3 to address this issue, and apply a frequentist model averaging method, the jackknife model averaging, to five linear stochastic frontier estimators that are discussed in Chapter 1. The results of the empirical application on world productivity indeed show that the weighted estimator provides more stable and robust estimates. More work is needed in future research on the model averaging approach, such as issues on the limiting behavior of the weighted estimator, the inclusion of non-linear estimators and the computational burden in calculating the optimal weights. Chapter 3 also explores the possibility of a structural approach to the productivity analysis. A more detailed examination of the structural model will be left for future work.

Bibliography

technical efficiency of european railways: a distance function approach. *Applied Economics*, 2000.

S. C. Ahn, Y. H. Lee, and P. Schmidt. Gmm estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, 101(2):219–255, 2001.

S. C. Ahn, Y. H. Lee, and P. Schmidt. Stochastic frontier models with multiple time-varying individual effects. *Journal of Productivity Analysis*, 27(1):1–12, 2007.

S. C. Ahn, Y. H. Lee, and P. Schmidt. Panel data models with multiple time-varying individual effects. *Journal of Econometrics*, 174(1):1–14, 2013.

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

D. J. Aigner and S.-F. Chu. On estimating the industry production function. *The American Economic Review*, pages 826–839, 1968.

H. AKAIKE. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, pages 267–281. Akademinai Kiado, 1973.

J. Bai. Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279, 2009.

J. Bai. Fixed-effects dynamic panel models, a factor analytical method. *Econometrica*, 81(1):285–314, 2013.

J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

B. H. Baltagi and J. M. Griffin. A general index of technical change. *The Journal of Political Economy*, pages 20–41, 1988.

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

F. Belotti, S. Daidone, G. Ilardi, and V. Atella. Stochastic frontier analysis using stata. 2012.

S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

S. T. Buckland, K. P. Burnham, and N. H. Augustin. Model selection: an integral part of inference. *Biometrics*, pages 603–618, 1997.

R. G. Chambers. *Applied production analysis: a dual approach.* Cambridge University Press, 1988.

L. R. Christensen, D. W. Jorgenson, and L. J. Lau. Conjugate duality and the transcedental logarithmic production function. *Econometrica*, (32), 1971.

T. Coelli. On the econometric estimation of the distance function representation of a production technology. Technical report, 2000.

T. Coelli and S. Perelman. Efficiency measurement, multiple-output technologie and distance functions: With application to european railways. 1996.

C. Cornwell, P. Schmidt, and R. C. Sickles. Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of econometrics*, 46(1):185–200, 1990.

G. Debreu. The coefficient of resource utilization. *Econometrica: Journal of the Econometric Society*, pages 273–292, 1951.

M. Denny. The relationship between functional forms for the production system. *Canadian Journal of Economics*, pages 21–31, 1974.

W. E. Diewert. Exact and superlative index numbers. *Journal of econometrics*, 4(2): 115–145, 1976.

W. E. Diewert. The quadratic approximation lemma and decompositions of superlative indexes. *Journal of Economic and Social Measurement*, 28(1):63–88, 2002.

W. E. Diewert. Hedonic regressions. a consumer theory approach. In *Scanner data and price indexes*, pages 317–348. University of Chicago Press, 2003.

W. E. Diewert. Alternative approaches to measuring house price inflation. Technical report, Discussion Paper 10-10, Department of Economics, The University of British Columbia, Vancouver, Canada, V6T 1Z1, 2010.

W. E. Diewert and C. Shimizu. Residential property price indexes for tokyo. *Work. Pap., Sch. Econ., Univ. British Columbia*, 2013.

R. Dix-Carneiro. Trade liberalization and labor market dynamics. *Econometrica*, 82 (3):825–885, 2014.

Z. Eckstein and O. Lifshitz. Dynamic female labor supply. *Econometrica*, 79(6): 1675–1726, 2011.

R. L. Eubank. Spline smoothing and nonparametric regression. 1988.

R. Färe and D. Primont. Multi-output production and duality: theory and applications. 1995.

R. Färe and K. J. Sung. On second-order taylor's-series approximation and linear homogeneity. *Aequationes Mathematicae*, 30(1):180–186, 1986.

R. Färe, S. Grosskopf, and C. K. Lovell. *The measurements of efficiency of production*, volume 6. Springer, 1985.

R. Färe, S. Grosskopf, C. K. Lovell, and S. Yaisawarng. Derivation of shadow prices for undesirable outputs: a distance function approach. *The review of economics and statistics*, pages 374–380, 1993.

M. J. Farrell. The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, pages 253–290, 1957.

C. Fernandez, G. Koop, and M. Steel. A bayesian analysis of multiple-output production frontiers. *Journal of Econometrics*, 98(1):47–79, 2000.

D. H. Good, R. C. Sickles, and J. C. Weiher. A hedonic price index for airline travel. *Review of Income and Wealth*, 54(3):438–465, 2008.

W. Greene. Alternative panel data estimators for stochastic frontier models. *Unpublished manuscript (Septemebr 1, 2002), Department of Economics, New York University*, 2002.

W. Greene. Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis*, 23(1):7–32, 2005a.

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

W. H. Greene. On the estimation of a flexible frontier production model. *Journal of Econometrics*, 13(1):101–115, 1980a.

W. H. Greene. Maximum likelihood estimation of econometric frontier functions. *Journal of econometrics*, 13(1):27–56, 1980b.

B. E. Hansen. Least squares model averaging. *Econometrica*, 75(4):1175–1189, 2007.

B. E. Hansen and J. S. Racine. Jackknife model averaging. *Journal of Econometrics*, 167(1):38–46, 2012.

J. A. Hausman and W. E. Taylor. Panel data and unobservable individual effects. *Econometrica: Journal of the Econometric Society*, pages 1377–1398, 1981.

N. L. Hjort and G. Claeskens. Frequentist model average estimators. *Journal of the American Statistical Association*, 98(464):879–899, 2003.

J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401, 1999.

J. Jondrow, C. K. Lovell, I. S. Materov, and P. Schmidt. On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of econometrics*, 19(2):233–238, 1982.

D. Jorgenson and Z. Griliches. Issues in growth accounting: A reply to edward f. denison. *Survey of Current Business*, 52(5, Part II), 1972.

A. Kneip, R. C. Sickles, and W. Song. A new panel data treatment for heterogeneity in time trends. *Econometric Theory*, 28(03):590–628, 2012.

G. Koop, J. Osiewalski, and M. F. Steel. Bayesian efficiency analysis with a flexible form: The aim cost function. *Journal of Business & Economic Statistics*, 12(3): 339–346, 1994a.

G. Koop, J. Osiewalski, and M. F. Steel. Posterior analysis of stochastic frontier models using gibbs sampling. Technical report, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), 1994b.

G. Koop, J. Osiewalski, and M. F. Steel. Bayesian efficiency analysis through individual effects: Hospital cost frontiers. *Journal of Econometrics*, 76(1):77–105, 1997.

T. C. Koopmans. Analysis of production as an efficient combination of activities. *Activity analysis of production and allocation*, 13:33–37, 1951.

S. C. Kumbhakar. Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics*, 46(1):201–211, 1990.

S. C. Kumbhakar, B. U. Park, L. Simar, and E. G. Tsionas. Nonparametric stochastic frontiers: a local maximum likelihood approach. *Journal of Econometrics*, 137(1): 1–27, 2007.

L.-F. Lee. A test for distributional assumptions for the stochastic frontier functions. *Journal of Econometrics*, 22(3):245–267, 1983.

Y. H. Lee. *Panel data models with multiplicative individual and time effects: appli-*

cations to compensation and frontier production functions. PhD thesis, Michigan State University. Department of Economics, 1991.

Y. H. Lee and P. Schmidt. A production frontier model with flexible temporal variation in technical efficiency. *The measurement of productive efficiency: Techniques and applications*, pages 237–255, 1993.

G. Leung and A. R. Barron. Information theory and mixing least-squares regressions. *Information Theory, IEEE Transactions on*, 52(8):3396–3410, 2006.

J. Liu, R. C. Sickles, and E. G. Tsionas. Bayesian treatment to panel data models with time-varying heterogeneity. 2013.

C. K. Lovell, P. Travers, S. Richardson, and L. Wood. *Resources and functionings: a new view of inequality in Australia*. Springer Berlin Heidelberg, 1994.

D. P. McMillen. the return of centralization to chicago: using repeat sales to identify changes in house price distance gradients. *Regional Science and Urban Economics*, (33):287–304, 2003.

W. Meeusen and J. Van den Broeck. Efficiency estimation from cobb-douglas production functions with composed error. *International economic review*, pages 435–444, 1977.

G. S. OLLEY and A. PAI. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297, 1996.

L. Orea and S. C. Kumbhakar. Efficiency measurement using a latent class stochastic frontier model. *Empirical Economics*, 29(1):169–183, 2004.

J. Osiewalski and M. F. Steel. Numerical tools for the bayesian analysis of stochastic frontier models. *Journal of Productivity Analysis*, 10(1):103–117, 1998.

A. Pakes and R. Ericson. Empirical implications of alternative models of firm dynamics. *Journal of Economic Theory*, 79(1):1–45, 1998.

A. Pakes and G. Gowrisankaran. Implementing the pakes-mcguire algorithm for computing markov perfect equilibria in gauss. Technical report, Working paper, Yale University, New Haven, 1993.

A. Pakes and P. McGuire. Computing markov perfect nash equilibria: Numerical implications of a dynamic differentiated product model, 1992.

B. U. Park and L. Simar. Efficient semiparametric estimation in a stochastic frontier model. *Journal of the American Statistical Association*, 89(427):929–936, 1994.

B. U. Park, R. C. Sickles, and L. Simar. Stochastic panel frontiers: A semiparametric approach. *Journal of Econometrics*, 84(2):273–301, 1998.

B. U. Park, R. C. Sickles, and L. Simar. Semiparametric-efficient estimation of ar (1) panel data models. *Journal of Econometrics*, 117(2):279–309, 2003.

B. U. Park, R. C. Sickles, and L. Simar. Semiparametric efficient estimation of dynamic panel data models. *Journal of Econometrics*, 136(1):281–301, 2007.

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

J. Qian and R. C. Sickles. Stochastic frontiers with bounded inefficiency. *MS, Department of Economics, Rice University*, 2008.

P. Schmidt and C. K. Lovell. Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers. *Journal of econometrics*, 9(3):343–366, 1979.

P. Schmidt and R. C. Sickles. Production frontiers and panel data. *Journal of Business & Economic Statistics*, 2(4):367–374, 1984.

G. Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6 (2):461–464, 1978.

R. W. Shephard. Cost and production functions. Technical report, DTIC Document, 1953.

R. W. Shephard, D. Gale, and H. W. Kuhn. *Theory of cost and production functions.* Number 4. Princeton University Press Princeton, 1970.

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

P. A. Swamy and G. S. Tavlas. Random coefficient models: theory and applications. *Journal of Economic Surveys*, 9(2):165–196, 1995.

P. Thorsnes. Consistent estimates of the elasticity of substitution between land and non-land inputs in the production of housing. *Journal of Urban Economics*, 42(1): 98–108, 1997.

E. G. Tsionas. Stochastic frontier models with random coefficients. *Journal of Applied Econometrics*, 17(2):127–147, 2002.

E. G. Tsionas. Inference in dynamic stochastic frontier models. *Journal of Applied Econometrics*, 21(5):669–676, 2006.

J. Van den Broeck, G. Koop, J. Osiewalski, and M. F. Steel. Stochastic frontier models: A bayesian perspective. *Journal of Econometrics*, 61(2):273–303, 1994.

A. T. Wan, X. Zhang, and G. Zou. Least squares model averaging by mallows criterion. *Journal of Econometrics*, 156(2):277–283, 2010.

H. Wang, X. Zhang, and G. Zou. Frequentist model averaging estimation: a review. *Journal of Systems Science and Complexity*, 22(4):732–748, 2009.

X. Zhang, A. T. Wan, and G. Zou. Model averaging by jackknife criterion in models with dependent data. *Journal of Econometrics*, 174(2):82–94, 2013.