

A Bayesian Multiple Technological Frontier Estimation in the Probabilistic Induced Technical Change Model: OECD Countries from 1968-2009

Jangho Yang

January 2018

OMPTEC Working Paper No. 2018-1



A Bayesian Multiple Technological Frontier Estimation in the Probabilistic Induced Technical Change Model: OECD Countries from 1968-2009

Jangho Yang*

Department of Economics, The New School for Social Research

Abstract

The paper develops a Bayesian cluster identification model to estimate the *innovation possibilities frontier* of the OECD economies based on the probabilistic Induced Technical Change (ITC) model, a generalized version of the canonical ITC model stemming from Von Weizsäcker (2010) and Kennedy (1964). The result shows that there are multiple distinctive frontiers in the OECD economies from 1968-2009, each of which is associated with a class of economies whose technical conditions exhibit a distinctive pattern. It is observed that advanced economies with a high “level” of labor productivity and low capital productivity along with a high labor cost tend to have a low rate of cost reduction, the result from a low growth rate of labor and capital productivity. The result explains a stylized fact of economic development of advanced economies from a low-wage, labor-intensive economy to a high-wage, capital-intensive economy.

Keywords: Frontier estimation, induced technical change, technological frontier, Bayesian inference

JEL codes: C11, C15, D24, D80, O33, O4

1 Introduction

The model of Induced Technical Change (ITC) (Von Weizsäcker, 2010; Kennedy, 1964; Samuelson, 1965; Drandakis and Phelps, 1966; Shah and Desai, 1981; Drandakis and Phelps, 1966; Foley, 2003) has received a fair amount of attention in the economics literature as an alternative to the models of exogenous technical change. The ITC model assumes an endogenous dynamic of technical progress where changes in productivity are induced by changes in unit factor prices.¹ One fundamental feature

*Address: 6 E 16th Street, New York, NY 10003; E-mail: yangj994@newschool.edu; Telephone: 347-276-4512

¹For a survey on ITC models, see Brugger and Gehrke (2016).

of the ITC model that makes it distinctive from other models of technical change lies in its assumption about the trade-off between rates of change in productivities, often expressed in the form of *innovation possibilities frontier* (IPF) (Kennedy, 1964). The IPF is usually assumed to be a concave function so that an increase in productivity of one input is made possible at the cost of a decrease in productivity of other inputs.

Despite its important theoretical implication, there have been few attempts to empirically estimate the concave IPF in the economics literature.² To fill this gap, this paper develops an econometric model to estimate the IPF with two production inputs, labor and capital. The proposed model is flexible enough and can be extended to estimate multiple frontiers, through which the characteristics of each frontier class can be studied. The paper finds that there are multiple distinctive IPFs in the OECD economies from 1968 - 2009. One of the results of the multiple frontier estimation is displayed below:

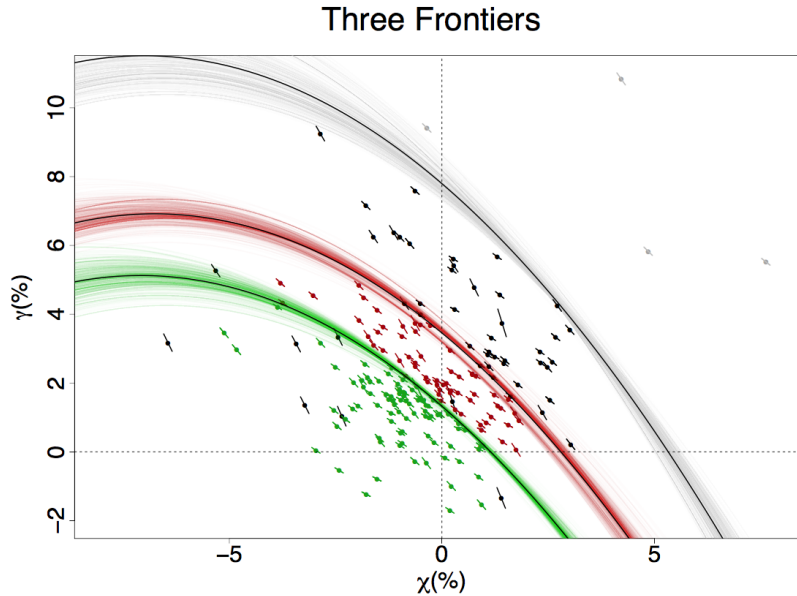


Figure 1: Three recovered frontiers. The black lines represent the frontiers recovered from the mean of the posterior distribution of the parameters, each of which is displayed along with 5,000 frontiers randomly drawn from the posterior distributions. Those data points belonging to the outer, the middle and the inner frontier are colored in black, red and green, respectively. Those four points in gray are invalid data points identified by the noise distribution. The negative ratio of the unit capital to labor cost, $-(1 - \omega)/\omega$ is drawn at each data point. A steeper slope has a longer line. Data source: Expanded Penn World Table Database (EPWT).

The figure displays the *innovation possibilities space* defined by the growth rate of labor and capital productivity, γ (y-axis) and χ (x-axis), with each point in the space representing the realization

²There are a few studies in Environment economics on the operationalization of the ITC model in estimating the potential for environmental policies such as a carbon tax to induce R&D on renewable energy sector (Grubb et al., 2002; Lee et al., 2011; Wing, 2003). However, they do not directly estimate the innovation possibility frontier of production inputs.

of productivity growth. A higher χ and γ implies a faster technological growth. With regard to the frontier, there are three recovered frontiers in black line. The tinted lines around the black line represent the estimation uncertainty. Each economy belongs to one of these three frontiers and its membership is displayed by using different color. The economy in black, red and green belong to the outer, middle and inner frontier.

The slope of the bar on each point represents the negative ratio of the capital to labor cost. The steeper the slope is, the longer the line is drawn for visualization purpose. This slope is used for measuring the technological inefficiency of the economy. For example, those economies with a steep slope have a relatively higher capital cost than unit labor cost and, therefore, adopting a capital-saving technology rather than a labor-saving one is better choice for them. For this reason, if they are in the second quadrant where $\chi < 0$ and $\gamma > 0$, they would be far away from the frontier. For example, the four black data points underneath the innermost frontier in the second quadrant have a very high ratio of the unit capital to labor cost, and they represent the most poorly performing economies in the first frontier class.

One of the most noticeable findings of the paper is that the recovered multiple frontier classes are characterized by distinctive *levels* of important economic variables, the level of labor and capital productivity and the unit labor cost. The outermost frontier class, which is associated with the highest rate of cost reduction, has the lowest labor productivity, the highest capital productivity, and the lowest unit labor cost. In contrast, those economies in the innermost frontier associated with the lowest rate of cost reduction, tend to have the highest level of labor productivity, the lowest level of capital productivity, and the highest unit labor cost. This suggests a well-documented path of economic growth and technical progress during the second half of the 20th century: the economic transition from a low-wage, labor intensive economy to a high-wage, capital-intensive economy.

The remainder of this paper is organized as follows. Section 2 discusses the probabilistic model of ITC and derives the technical inefficiency function, a probability distribution representing non-optimal technical change. In doing so, we will rely on Information Theory and adopt the notion of entropy to model a bounded rational behavior in which entrepreneurs have a limited capacity to process market signals regarding the rate of cost reduction. A non-trivial result of this model is that maximizing expected payoff of the entrepreneurs leads to a non-degenerate distribution of heterogeneous actions even without making any arbitrary assumptions as to the biases of individual entrepreneur.

Section 3 sets up a Bayesian clustering model to estimate multiple technological frontiers in the OECD economies. The clusters identified by the technological inefficiency function in our model are inherently based on an asymmetrical distance between observations because those distances are calculated based on the frontier (the outermost boundary) as a reference point. This marks a stark difference from the conventional cluster models such as K-means algorithms that rely on the symmetric distance of observations, usually around the mean of data in each cluster. Utilizing Metropolis-Hastings simulation algorithm for the posterior distribution sampling, the section recovers multiple technological frontiers of the OECD economies.

Section 4 discusses the distinctive characteristics of the recovered clusters in the 3-frontier model as an example. The key technological variables such as the level and the rate of input productivities and unit input costs are considered. Section 5 concludes the paper.

2 Probabilistic Model of ITC

2.1 Canonical ITC model: optimal technical progress

Consider a production model, in which output Y requires employed labor L and capital K . Denoting labor and capital productivity as $x = Y/L$ and $\rho = Y/K$, technical change is defined as the growth rate of x and ρ , $\gamma = \frac{dx/dt}{x}$ and $\chi = \frac{d\rho/dt}{\rho}$, respectively. Total production cost C consists of labor cost wL and capital cost rK where w and r are respectively the wage rate and the profit rate. Therefore, the unit factor cost, i.e. the factor cost per output, is defined as $\omega = wL/Y$ for unit labor cost and $\pi = rK/Y$ for unit capital cost. Assuming that the unit total cost in the initial state is unity, $\pi = 1 - \omega$.

The ITC model is predicated on the idea that changes in productivities respond to changes in unit cost ω in a way that a new set of techniques decreases the production cost. We define the instantaneous decrease in unit cost, $\zeta = -d(C/Y)/(C/Y)$, which can be expressed in term of the weighted average of γ and χ with respective weights given by unit labor cost and unit capital cost under the assumption that w and r are kept constant.(Kennedy, 1964):

$$\zeta = \omega\gamma + (1 - \omega)\chi, \quad (1)$$

Under the behavioral assumption of the ITC, an increase in unit cost of a certain input prompts the entrepreneur to introduce technology that saves that input. An increase in unit labor cost ω , for example, leads to a labor saving technology that increases the growth rate of labor productivity, γ .

The key assumption of the ITC model is that the course of technological progress is not unbounded but is constrained by the IPF to have a trade-off between χ and γ . The IPF suggests that an increase in productivity of one input is made possible at the cost of a decrease in productivity of the other input. Higher labor productivity growth, for example, is coupled with lower capital productivity growth on the innovation possibilities frontier. This distinctive feature of the IPF can be represented by a concave function:

$$\gamma = f(\chi), \quad \text{with } f' < 0, \quad f'' < 0. \quad (2)$$

Given the IPF, the cost minimizing agent that seeks to maximize the rate of cost reduction ζ solves the following optimization problem:

$$\begin{aligned} \max \quad & \zeta = \omega\gamma + (1 - \omega)\chi, \\ \text{s. t} \quad & \gamma = f(\chi). \end{aligned} \quad (3)$$

The solution to this problem yields the following relationship between technical progress and the unit cost:

$$f'(\chi) = -\frac{1-\omega}{\omega}. \quad (4)$$

Equation 4 states that the slope of the IPF is the negative ratio of the unit capital to unit labor cost. Since the IPF represents the limit of technical change, a point on the IPF has the *maximum rate of cost reduction* and corresponds to the optimal combination of γ and χ according to the actual cost structure given by ω . We denote the χ and γ on the frontier, and the corresponding maximum ζ , as χ_f , γ_f , and ζ_f .

2.2 Probabilistic ITC model

The probabilistic ITC model generalizes the canonical model and assumes that the maximum rate of cost reduction is achieved probabilistically. Let us consider a typical agent whose realized rate of cost reduction, which I denote as ζ_r , is different from its maximum rate on the frontier ζ_f . Let $\zeta_d \geq 0$ be the distance, $\zeta_f - \zeta_r$. To further derive the functional form of ζ_d , let us suppose that the concave IPF takes the form of a simple negative half-quadratic function:³

$$\gamma_f = a\chi_f^2 + b\chi_f + c, \quad (5)$$

where $a, b < 0$ to fulfill the concavity conditions. Using the canonical relation between the unit cost ω and the chosen technology γ and χ on the IPF in Equation 4, we have $-\omega/(1-\omega) = 2a\chi_f + b$. This implies that:

$$\chi_f = \frac{-\frac{1-\omega}{\omega} - b}{2a} \quad (6)$$

$$\gamma_f = a\left(\frac{-\frac{1-\omega}{\omega} - b}{2a}\right)^2 + b\left(\frac{-\frac{1-\omega}{\omega} - b}{2a}\right) + c. \quad (7)$$

Denoting χ_r and γ_r as the realized rate of labor and capital productivity increase, ζ_d has the following functional form given ω :

$$\begin{aligned} \zeta_d &= \zeta_f - \zeta_r \\ &= (1-\omega)\left(\frac{-\frac{1-\omega}{\omega} - b}{2a}\right) + \omega\left(a\left(\frac{-\frac{1-\omega}{\omega} - b}{2a}\right)^2 + b\left(\frac{-\frac{1-\omega}{\omega} - b}{2a}\right) + c\right) - \\ &\quad -(1-\omega)\chi_r - \omega\gamma_r. \end{aligned} \quad (8)$$

This point can be seen more clearly with the following figure where a hypothetical IPF is drawn

³This is a Taylor approximation around $\chi = 0$.

on the innovation space with χ and γ :

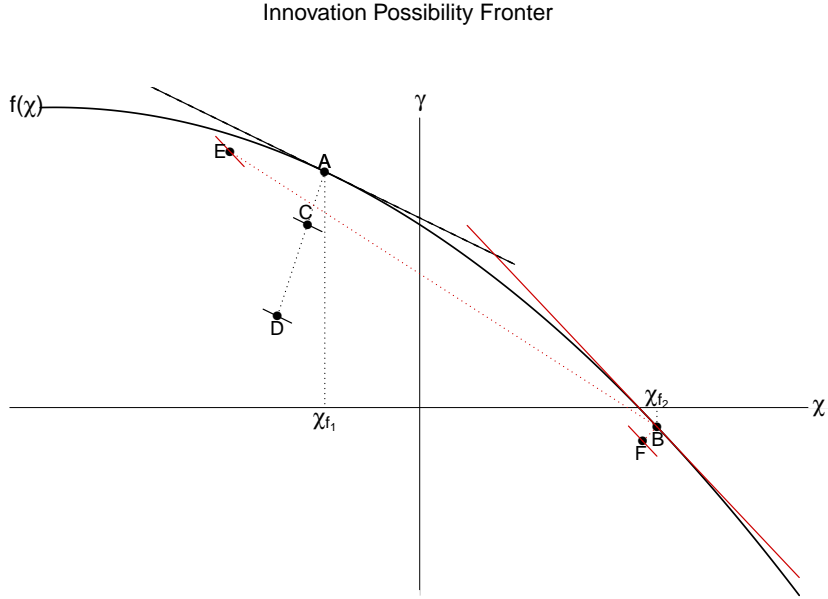


Figure 2: An innovation possibilities frontier. The bold concave function represents the IPF $\gamma = f(\chi)$, a functional relationship that represents the trade-off between the growth rates of capital productivity χ and labor productivity γ . Six points with an alphabet letter represent different realization of χ and γ . The cost structure represented by the negative ratio of unit capital to labor ratio $-(1 - \omega)/\omega$ is drawn as a slope on the point. Economy A, C and D have the same slope which is colored in black while Economy B, E, and F have the same slope in red. Economy A and B are located on the IPF with the optimal technical change in terms of capital and labor productivity growth χ_f and γ_f .

The IPF, $\gamma = f(\chi)$, is drawn for 6 different economies (A-F), with the realized χ and γ along with a different cost structure represented by the slope $-(1 - \omega)/\omega$. Economies A and B are on the IPF, realizing the maximum rate of cost reduction from the technology and thus $\zeta_{d_A} = \zeta_{d_B} = 0$. Economy A has higher unit labor cost than Economy B, $\omega_A > \omega_B$, and therefore has a flatter slope of the IPF, $-(1 - \omega_A)/\omega_A < -(1 - \omega_B)/\omega_B$, so that the maximum rate of cost reduction is realized with a higher growth rate of labor productivity and a lower growth rate of capital productivity: $\gamma_A > \gamma_B$ and $\chi_A < \chi_B$. In contrast, Economies C-F are apart from the IPF so that $\zeta_{d_C}, \zeta_{d_D}, \zeta_{d_E}, \zeta_{d_F} > 0$. Economies C and D have the same cost structure with Economy A, $\omega_A = \omega_C = \omega_D$. However, Economy C has a lower χ and γ than is needed to reach the frontier while Economy D has even lower χ and γ so that $\zeta_{d_A} = 0 < \zeta_{d_C} < \zeta_{d_D}$. It is important to note that ζ_d is not determined by the geometric distance from the frontier but is determined by the cost structure ω given the IPF as was shown in Equation 9. Economy E and F have similar geometric distance from the IPF but Economy E has lower ζ due to the fact that the cost structure of Economy E, which is the same as that of Economy

B and F, requires a significantly lower γ and a higher χ for the technology to be optimal.

The cost-minimizing behavior of the typical agent in the probabilistic ITC model can be formulated by the mixed strategy setting (von Neumann and Morgenstern, 1944). Consider a payoff function $u[\zeta_d]$ that represents the payoff of adopting a new technology that brings about a particular ζ_d . Lower ζ_d leads to higher payoff, $u'[\zeta_d] < 0$, because it implies that the new technique leads to the rate of cost reduction close to ζ_f . The payoff is maximized when $\zeta_d = 0$, that is when the actual rate of cost reduction is equal to ζ_f . The typical agent with a mixed strategy assigns a probability to each ζ_d , $p[\zeta_d]$, to maximize the *expected payoff* $\sum_{\zeta_d} p[\zeta_d]u[\zeta_d]$, where $\sum_{\zeta_d} p[\zeta_d] = 1$:

$$\begin{aligned} \max \quad & \sum p[\zeta_d]u[\zeta_d], \\ \text{s. t} \quad & \sum_d p[\zeta_d] = 1. \end{aligned} \tag{9}$$

With no further constraint, the solution to this problem is the Dirac Delta function, choosing the technique that minimizes ζ_d and thus maximizes the payoff $u[\zeta_d]$:

$$p[\zeta_d] = \text{DiracDelta}[\zeta_d - \hat{\zeta}_d[u, \zeta_d]] \tag{10}$$

where $\hat{\zeta}_d[u, \zeta_d]$ is the distance that maximizes the payoff, which is equal to zero in our model. The Dirac Delta function has the following property:

$$\text{DiracDelta}[x] = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0 \end{cases} \quad \text{and} \quad \int_{-\infty}^{\infty} \text{DiracDelta}[x] dx = 1. \tag{11}$$

so that the resulting frequency distribution of $p[\zeta_d]$ puts unit weight on the payoff maximizing action while puts zero weight on the others, implying that the technology is always chosen on the IPF. This is basically the result of the canonical ITC model, where the typical entrepreneur has a unique payoff-maximizing set of technology on the frontier, that is $\zeta_d = 0$.

Entropy-constrained behavior and technical inefficiency function

The model can be further extended to allow for non-optimal technical progress, in which the choice of χ and γ can be different from their maximum rate on the frontier. The paper does not propose a specific behavioral model to account for the non-optimality but lays out a general probabilistic model in which the agent is exposed to a positive degree of uncertainty in the choice of technology.

Consider a bounded rational behavior where the optimizing agent has a limited capacity to process market signals. This type of behavior can be formulated by constraining the agent's processing capacity to the minimum entropy H_{min} , based on the fact that entropy is a measure of uncertainty

in Information Theory (Sims, 2003, 2006; Scharfenaker and Foley, 2017).⁴⁵ Using the definition of entropy as the negative expected value of log-probability, $H[p] = -\sum p \text{Log}[p]$, the agent's expected payoff maximization program is written as the following:

$$\begin{aligned} \max \quad & \sum_{\zeta_d} p[\zeta_d] u[\zeta_d], \\ \text{s. t} \quad & H[p[\zeta_d]] = -\sum_{\zeta_d} p[\zeta_d] \text{Log} p[\zeta_d] \geq H_{min}, \\ & \sum_{\zeta_d} p[\zeta_d] = 1, \end{aligned} \tag{12}$$

whose associated Lagrangian with the multiplier κ and T is:

$$L[p, u, \zeta_d, T] = -\sum p[\zeta_d] u[\zeta_d] - \kappa \left(\sum p[\zeta_d] - 1 \right) + T \left(\sum p[\zeta_d] \text{Log} p[\zeta_d] - H_{min} \right).$$

The solution to this problem is the Gibbs distribution with the following form (Borwein and Lewis, 1991; Cover and Thomas, 2006):

$$p[\zeta_d] = \frac{\exp\left(\frac{u[\zeta_d]}{T}\right)}{\sum \exp\left(\frac{u[\zeta_d]}{T}\right)}. \tag{13}$$

This result shows that the solution is not a single point on the frontier, but a probability distribution of the possible states of the technological change. In general, the realized rate of cost reduction ζ_r is different from its maximum, ζ_f . We call this resulting probability distribution of ζ_d a *technical inefficiency function* because it represents the probability of how far each economy's technological condition, measured by the current rate of cost reduction ζ_r , is from its maximum condition on the frontier ζ_f .

The inefficiency function can be derived by further specifying the payoff function $u[\zeta_d]$. I assume in this paper that the payoff is a linear function of $-\zeta_d$, so that the agent's payoff is proportional to a reduction in ζ_d .⁶ Using $u[\zeta_d] = m - n\zeta_d$ with $m, n > 0$, the technical inefficiency probability function is written as:

$$p[\zeta_d] \propto \exp(-\lambda\zeta_d), \tag{14}$$

⁴⁵For detailed discussion on information theory, see Jaynes (2003), MacKay (2005), and Cover and Thomas (2006). For an extensive survey on the information theoretic approaches in economics, see Yang (2017b)

⁵It is worthwhile to note that the Dirac Delta distribution we derived in the previous section has zero entropy. In the Dirac Delta function, one choice is assigned probability 1 while the other choices gets zero probability, $p_j = 1$ while $p_{j-} = 0$. And therefore we have $H[\text{DiracDelta}] = -(p_d \text{Log}[1] + 0 \text{Log}[0] + \dots + 0 \text{Log}[0]) = 0 + 0 + \dots + 0 = 0$ where $0 \text{Log}[0] = 0$ by convention. The zero entropy model implies that the typical entrepreneur has a full capacity to process all relevant market information as to the rate of cost reduction, resulting in a complete certainty about their decision. Therefore, any changes in the input costs will induce an optimal response of technical change so that all potential rate of cost reduction are fully exhausted.

⁶As we will see in Section 3, we find empirical evidence for this linear specification.

where $\lambda = n/T$ represents the intensity of the payoff. The result shows that the inefficiency function is an exponential function on ζ_d with a parameter λ .⁷

This result shows a stark contrast to the canonical ITC model, because the derived solution of the maximization is a probability distribution of the possible states of the technological change. The rates of cost reduction around frontier are still most likely, but the rates of cost reductions below the frontier are also possible. That is, all entrepreneurs should not necessarily be on the frontier.

3 Multiple Frontier Estimation

The purpose of this section is to develop a Bayesian econometric model to estimate multiple IPFs. We first derive the likelihood function using the technical inefficiency function derived in Equation 14. Then, we will discuss the prior distributions of the unknown parameters followed by a discussion on the data and the model selection criteria. Finally, the posterior distributions are analyzed along with the recovered IPFs.

3.1 The model

3.1.1 Likelihood function of Single frontier case

As we discussed above in Equation 5, we assume that the IPF takes the form of a negative half-quadratic function that has the property $f'(\chi_f) < 0$ and $f''(\chi_f) < 0$ with no higher order terms: $\gamma_f = a\chi_f^2 + b\chi_f + c$, where $a, b < 0$. Equation 9 on the functional form of ζ_d , combined with the technical inefficiency function in Equation 14, constitute the likelihood function of the IPF *without noise*:

$$p[\chi_r, \gamma_r | \omega, a, b, c, \lambda] = \lambda \exp(-\lambda \zeta_d), \quad (15)$$

To allow for the outliers in the model estimation, we introduce a noise distribution that penalizes the invalid data. The penalizing distribution is defined on the innovation space “outside” the frontier. As the technical inefficiency function that assigns a probability to the data points “inside” frontier, the penalizing distribution should effectively assign a probability of those data points “outside” the frontier and measure how likely the data is “invalid.” Intuitively, the data points further away from the frontier are less likely to be valid than those closer to the frontier. The exponential distribution on $C + \max[\zeta_r] - \zeta_r$, where $\max[\zeta_r]$ is the maximum value of the realized rate of cost reduction

⁷One influential frontier estimation model, Stochastic Frontier Analysis (SFA) has proposed different specifications of technical inefficiency function such as half-normal (Aigner et al., 1977), truncated normal (R. Stevenson, 1980), exponential (Meeusen and van Den Broecke, 1977), the Gamma (Greene, 1990), and Rayleigh distribution (Hajargasht, 2014). Commonly, these different one-sided distributions are proposed for computational convenience, statistical accuracy, and model identification. It would be interesting to see which form of payoff function would lead to different technical inefficiency functions in the SFA literature.

and C is a positive constant, does effectively this job because it assigns with a constant rate a higher probability to the extreme points while putting lower probability to those data points closer to the frontier (smaller ζ_r). For simplicity, we put C to zero in this model, so that the noise distribution is $\lambda_n \exp(-\lambda_n(\max_r[\zeta] - \zeta_r))$, where λ_n is the parameter of the penalizing distribution. Thus, the likelihood function of the single IPF *with noise* is written as:

$$p[\chi_r, \gamma_r, |\omega, a, b, c, \lambda, \lambda_n] = \mathcal{B}(1)\lambda \exp(-\lambda\zeta_d) + \mathcal{B}(2)\lambda_n \exp(-\lambda_n(\max[\zeta_r] - \zeta_r)), \quad (16)$$

where $\mathcal{B} : \subseteq [1, 0]$ is a boolean variable. When χ and γ are inside the frontier, $\mathcal{B}(1) = 1$ while $\mathcal{B}(2) = 0$ so that the likelihood is $\lambda \exp(-\lambda\zeta_d)$. When χ and γ are outside the frontier, $\mathcal{B}(2) = 1$ while $\mathcal{B}(1) = 0$ so that the likelihood is $\lambda_n \exp(-\lambda_n(\max[\zeta_r] - \zeta_r))$.

It is worthwhile noting that there is a trade-off in Equation 16 between having invalid data points inside and outside the frontier. If they are inside the frontier, they will lower the exponential fit of ζ_d inside the frontier but there is no penalty for this data from the penalizing distribution. In contrast, if they are outside the frontier, the fit of ζ_d will increase while the noise distribution penalizes them. From this perspective, the major task of this estimation is to find the optimal number of those invalid data points.

3.1.2 Likelihood function of Multiple frontiers case

We can generalize the single frontier case to the multiple frontier one. To get a sense of the multiple frontier model, let's first suppose that we have two negative half-quadratic IFSs: $a_I\chi_f^2 + b_I\chi_f + c_I$ and $a_{II}\chi_f^2 + b_{II}\chi_f + c_{II}$. In the multiple frontier models, each observation is included in only one frontier so that a multiple membership is not allowed. These frontiers divide up the data into two clusters of the exponential distribution. Therefore, the likelihood function should properly represent the likelihood of each data point coming only from the corresponding frontier even in the existence of the other frontier. The following constraints are added to the likelihood function for this purpose. First, if the data points are between the outer frontier I and the inner frontier II, the likelihood of ζ_d is calculated using only the frontier I, $p_I[\zeta_d]$, because those data points are identified as invalid from the inner frontier. Second, if data points are inside both frontiers, ζ_d is calculated from the frontier whose likelihood $p[\zeta_d]$ is greater than that of the other. For example, suppose that, for a data point ζ_{d_i} , we have $p_I[\zeta_{d_i}] > p_{II}[\zeta_{d_i}]$. Then, its likelihood is calculated from the frontier I, $p_I[\zeta_{d_i}]$. Finally, if data points are outside both of the frontiers, they are assumed to be noise.

Using the exponential distribution with a parameter λ_I and λ_{II} for $p_I[\zeta_d]$ and $p_{II}[\zeta_d]$ respectively, we have the following likelihood function *with noise*:

$$p[\chi_r, \gamma_r | \omega, a_I, b_I, c_I, \lambda_I, a_{II}, b_{II}, c_{II}, \lambda_{II}, \lambda_n] = \mathcal{B}(1)\lambda_I \exp(-\lambda_I\zeta_d) + \mathcal{B}(2)\lambda_{II} \exp(-\lambda_{II}\zeta_d) + \mathcal{B}(3)\lambda_n \exp(-\lambda_n(\max[\zeta_r] - \zeta_r)). \quad (17)$$

When χ and γ are inside the frontier I and outside the frontier II, $\mathcal{B}(1) = 1$ while $\mathcal{B}(2, 3) = 0$ so that the likelihood is $\lambda_I \exp(-\lambda_I \zeta_d)$. When χ and γ are inside both of the frontiers, $\mathcal{B}(2) = 1$ while $\mathcal{B}(1, 3) = 0$ so that the likelihood is $\lambda_{II} \exp(-\lambda_{II} \zeta_d)$. Finally, when χ and γ are outside all the frontiers, $\mathcal{B}(3) = 1$ while $\mathcal{B}(1, 2) = 0$ so that the likelihood is $\lambda_n \exp(-\lambda_n (\max[\zeta_r] - \zeta_r))$.

This two-frontier model can be easily generalized into K number of frontiers using the same logic as above. For example, if the data point is outside the outermost frontier I so that the observed rate of the cost reduction is greater than the maximum cost reduction, we consider them as noise. If the data point is inside the outermost frontier I, but is outside the rest of frontiers, the likelihood is determined by the distance from the outermost frontier. If the data point is inside the outermost frontier I and the second outer frontier II but is outside the rest of frontiers, the likelihood is determined by comparing $P_I[\zeta_{d_i}]$ and $P_{II}[\zeta_{d_i}]$, and take the larger one. If the data point is inside frontier I, frontier II, frontier III but is outside the rest of frontiers, the likelihood is determined by comparing $P_I[\zeta_{d_i}]$ and $P_{II}[\zeta_{d_i}]$ and $P_{III}[\zeta_{d_i}]$, and take the largest one. This process repeats until the data point is inside all the frontiers, in which the likelihood is determined by comparing $P_I[\zeta_{d_i}]$ and $P_{II}[\zeta_{d_i}] \dots P_K[\zeta_{d_i}]$. Therefore the likelihood for K frontiers with a noise distribution is as follows:

$$\begin{aligned}
p[\chi, \gamma | \omega, A, B, C, \Lambda, \lambda_n] &= \mathcal{B}(1) \lambda_I \exp(-\lambda_I \zeta_d) + \mathcal{B}(2) \lambda_{II} \exp(-\lambda_{II} \zeta_d) + \dots \\
&\quad + \mathcal{B}(K) \lambda_K \exp(-\lambda_n \zeta_d) + \mathcal{B}(K + 1) \lambda_n \exp(-\lambda_n (\max[\zeta] - \zeta)),
\end{aligned} \tag{18}$$

where $A = [a_I, a_{II}, \dots, a_K]$, $B = [b_I, b_{II}, \dots, b_K]$, $C = [c_I, c_{II}, \dots, c_K]$, and $\Lambda = [\lambda_I, \lambda_{II}, \dots, \lambda_K, \lambda_n]$.

3.1.3 Prior distribution

Now that we have discussed the likelihood function, we turn our attention to a general principle as to the prior distribution of the unknown parameters of the likelihood function, $A = [a_I, a_{II}, \dots, a_K]$, $B = [b_I, b_{II}, \dots, b_K]$, $C = [c_I, c_{II}, \dots, c_K]$, $\Lambda = [\lambda_I, \lambda_{II}, \dots, \lambda_K, \lambda_n]$. First, the parameter a determines the curvature of the IPF, the elasticity of input substitution. Since a is the least interesting parameter in estimating the general location and the shape of the IPF, we will simplify our model estimation and reduce the number of parameters by setting $a_k = a$ for $k = 1, \dots, K$. When it comes to the prior on a , we utilize information from the result of the single frontier case (Yang, 2017a) where $-a$ was estimated around 0.1. We set up a gamma prior with the shape parameter α being 2, and the scale parameter β being 0.05 whose mean $\alpha\beta$ is 0.1: $-a \sim \text{Gamma}(2, 0.05)$.

Using Equation 4, we can show that the parameter b is the negative Harrod-neutral steady-state ratio of unit capital to unit labor cost, $-(1 - \omega^*)/\omega^*$, while c is the Harrod-neutral steady state labor productivity growth, γ^* , defined at $\chi = 0$ (Uzawa, 1961; Shah and Desai, 1981; Foley, 2003). Using the prior information that the ratio of unit capital to unit labor ratio is slightly above 1 in the OECD

countries, we use, for all the $-b_k$ in B , a gamma prior distribution, $\text{Gamma}(10, 0.1)$ for the mean of the distribution to be 1. For c , we have prior information that inner frontiers should have a lower steady-state growth rate of labor productivity, which is represented by the lower value of parameter c . Therefore, the mean of the prior on c varies among different frontiers. For the prior on c of the outermost frontier, we use the highest γ around zero χ in the data as the mean of the prior distribution on c , which is around 7%. The mean of the prior on c of the inner frontiers is decided by dividing the mean value of c of the outermost prior by the number of frontiers. For example, if there are two frontiers, the mean of the prior on c of the second frontier is 3.5. With 3 frontiers, the mean of c_1 , c_2 , c_3 would be 7, 4.6, and 2.2, respectively. Assigning the mean of the prior distribution on the inner frontiers by equally dividing the distance between the mean of c of the outermost frontier and 0 with the same prior on a and b guarantees that all frontiers are equally apart from the two adjacent frontiers. Since we are less sure about the location of the inner frontiers, we allow a higher variance for those parameters than that of $p[c_1]$. We set the scale of the gamma distribution for c_1 to 100 and 10 to the rest of c .

For the prior on λ_k , whose inverse is the mean distance between the frontier and the observed data point, $E[\zeta_d]$ we use the gamma prior with the mean 1 for the inner frontiers. This means that the technical condition of the economies are less efficient than its maximum potential on the frontier by 1% on average. We use $\text{Gamma}(50, 0.02)$ for $\lambda_{2, \dots, k}$. For λ_1 , the uniform prior is used to allow for the maximum flexibility for ζ_d of the outermost frontier. This represents our prior belief that the behavior of those economies associated with the outer frontier is less constrained because they tend to have less information about the technology.

Finally, we use another gamma prior on the rate parameter λ_n of our penalizing distribution. The higher the rate parameter, the more the invalid data are penalized. The inverse of λ_n is the average distance of ζ_r outside the frontier from the most extreme ζ in the data. We set the mean of λ_n to be 3 so that the mean distance is below 0.5%. A relatively strong prior is required for a stable estimation because the estimation is indifferent to the high value of λ_n so the MCMC chain does not converge properly. We use a rather strong prior using $\text{Gamma}(200, 0.015)$. It turns out that the estimation result for other parameters is not sensitive to the prior choice on λ_n . This issue will be discussed in more detail in the following section on the posterior analysis.

As we will see, the Bayesian method will allow the data to alter these prior distributions of the parameters.

3.1.4 The posterior distribution

Assuming the independence between the prior distributions, the posterior can be obtained as follows:

$$\begin{aligned}
\text{Log}[p[\omega, A, B, C, \Lambda | \chi, \gamma]] &\propto \text{Log}[p[\chi, \gamma, \omega | A, B, C, \Lambda]] + \text{Log}[p[A, B, C, \Lambda]] \\
&= \text{Log}[\mathcal{B}(1)\lambda_1 \exp(-\lambda_1 \zeta_d) + \mathcal{B}(2)\lambda_2 \exp(-\lambda_2 \zeta_d) + \cdots + \\
&\quad + \mathcal{B}(K)\lambda_K \exp(-\lambda_n \zeta_d) + \mathcal{B}(K+1)\lambda_n \exp(-\lambda_n (\max[\zeta] - \zeta))] + \\
&\quad + \text{Log}[p[A]] + \text{Log}[p[B]] + \text{Log}[p[C]] + \text{Log}[p[\Lambda]], \tag{19}
\end{aligned}$$

where the prior distributions of $p[A]$, $p[B]$, $p[C]$, and $p[\Lambda]$ with $k = 1, \dots, K$ frontiers are defined as follows:

$$\begin{aligned}
a_k &\sim \text{Gamma}(2, 0.05) \\
b_k &\sim \text{Gamma}(10, 0.1) \\
c_1 &\sim \text{Gamma}(100, 0.07) \\
c_k &\sim \text{Gamma}(10, 0.7 - 0.7k/K) \\
\lambda_1 &\sim \text{Uniform}(0, \infty). \\
\lambda_{2, \dots, K} &\sim \text{Gamma}(50, 0.02). \\
\lambda_n &\sim \text{Gamma}(200, 0.015).
\end{aligned}$$

3.2 Estimation of the Model

3.2.1 Data and sampling method

We use the Extended Penn World Table database 4.0 (EPWT) assembled by Adalmir Marquetti and Duncan Foley to obtain the country-level data on technical progress and the unit cost.⁸ Four variables are extracted: capital productivity ρ , labor productivity x , unit labor cost ω , and the GDP, all these adjusted by the PPP for the OECD countries from 1968 - 2009. For technological variables - such as the growth rate of labor and capital productivity, γ and χ - we need to remove the business cycle effect in the measurement. For this purpose, we employ a peak-to-peak measure in this paper with the minimum length of each cycle for two years. There are 200 data points in total.

In obtaining a sequence of random samples from the posterior distribution, the paper uses the Metropolis-Hastings algorithm, one of Markov Chain Monte Carlo (MCMC) methods. The MCMC is often used when we have a density function $p[\theta]$ that is not analytically tractable so that direct sampling is difficult. What we do instead is to simulate the random variable from the given density based on a "Markov chain" which has the target distribution as its equilibrium distribution. Then, we recover the probability distribution of the parameters from the simulated random numbers of them θ^* ,

⁸The database is available at <https://sites.google.com/a/newschool.edu/duncan-foley-homepage/home>

the method that is often called the “Monte Carlo” approach. For detailed discussions on the MCMC method, see Gamerman and Lopes (2008); Minh and Minh (2015); Gelman et al. (2013).

3.2.2 Estimation results

The estimation of parametric multiple frontier models depends on the predetermined number of frontiers. If we assume that there is only one frontier, the exponential model of ζ_d might not fit the data well even though the model itself is the simplest. In contrast, if we assume that there are a large number of frontiers, say the extreme case where the same number as that of the data points, the model will not have any predictive power for the out-of-sample data even though this model perfectly fits the observed data. Instead, we need to find the optimal number of frontiers which can make the model as parsimonious as possible while simultaneously fitting the data.

In this paper, I report the estimation results of three representative models with two, three, and four frontiers. Some model selection criteria will be discussed in the following section. The following table in the next page summarizes the recovered coefficients for the three models based on the Metropolis-Hastings simulation with 100,000 iterations and 3 chains after 25,000 burn-in periods.⁹

First, the index for the convergence diagnostic \hat{R} of almost all parameters is very close to 1, meaning that all chains have properly converged.¹⁰ This is not the case for b_1 and c_1 of the four-frontier model whose value is far greater than 1 with 1.72 and 1.99, respectively. As is shown in Appendix B.1, the posterior distributions of b_1 and c_1 appear to have two modes so that the chains move back and fourth between modes with no convergence.¹¹ The multi-modality of the two parameters, however, does not affect the posterior distributions of the other parameters. As we will see in the recovered frontiers, the dual modes arise due to the inclusion or exclusion of one data point in the outermost frontier (Frontier I) of the four-frontier model.

Second, a significant shrinkage is observed in the posterior distribution from the prior distribution so that standard deviation of the posterior distribution is far smaller than that of the prior distribution for almost all parameters.¹² The standard deviation of the posterior distributions of $\lambda_1, \lambda_2, \lambda_3$ and λ_n is similar to their prior value. This result does not pose a problem as long as the location of the distribution has effectively updated. The 95% credible interval of $\lambda_1, \lambda_2,$ and λ_3 does not include its prior mean value meaning that the data altered the location of the prior distribution. However, the posterior distribution of λ_n , which is the rate parameter of the penalizing distribution of the invalid

⁹Appendix B.1 displays the traceplots of the Metropolis-Hastings simulation. Codes in R (RStudio Team, 2016) are available upon request.

¹⁰ \hat{R} is computed by comparing the estimated between-chains and within-chain variances. For detailed discussion on the convergence diagnostic, see Gelman et al. (2013).

¹¹The simple Metropolis-Hastings simulation is not always suitable for multi-modal distributions because the chain can be stuck at one mode and can fail to move to another mode. Different MCMC techniques have been proposed to address the multi-modality in the MCMC sampling. For example, see Neal (1996); Laskey and Myers (2003); Sminchisescu and Welling (2011); Lan et al. (2014)

¹²The posterior distribution of b_1 and c_1 have a higher standard deviation than that of the prior distributions due to the multi-modality.

Parameters	Two Frontier				Three Frontiers				Four Frontiers			
	Mean	SD	95% CI	\hat{R}	Mean	SD	95% CI	\hat{R}	Mean	SD	95% CI	\hat{R}
a_1	-0.07(-0.1)	0.00(0.1)	[-0.07,-0.06]	1.01	-0.08(-0.1)	0.00(0.1)	[-0.08,-0.07]	1.02	-0.07(-0.1)	0.00(0.1)	[-0.07,-0.06]	1.00
b_1	-0.97(-1.00)	0.05(0.32)	[-1.01,-0.90]	1.01	-1.04(-1.00)	0.07(0.32)	[-1.09,-0.91]	1.00	-1.13(-1.00)	0.20(0.32)	[-1.41, 0.94]	1.72
c_1	7.48(7)	0.15(0.7)	[7.39, 7.83]	1.01	7.73(7)	0.23(0.7)	[7.56, 8.23]	1.00	8.02(7)	0.67(0.7)	[7.68, 9.27]	1.99
λ_1	0.38	0.05	[0.35, 0.48]	1.01	0.39	0.06	[0.35, 0.53]	1.01	0.33	0.06	[0.30, 0.50]	1.02
b_2	-1.02(-1.00)	0.02(0.32)	[-1.03,-0.97]	1.00	-1.02(-1.00)	0.05(0.32)	[-1.05,-0.91]	1.02	-0.96(-1.00)	0.05(0.32)	[-0.99,-0.84]	1.00
c_2	2.14(3.5)	0.05(1.11)	[2.13, 2.29]	1.00	3.44(4.6)	0.18(3.03)	[3.25,3.80]	1.00	4.18(5.25)	0.19(3.24)	[4.09, 4.62]	1.07
λ_2	1.01(1.00)	0.09(0.14)	[0.95, 1.21]	1.01	1.12(1.00)	0.11(0.14)	[1.05,1.33]	1.02	1.15(1.00)	0.13(0.14)	[1.07, 1.42]	1.09
b_3	-	-	-	-	-1.07(-1.00)	0.05(0.32)	[-1.10,-0.95]	1.02	-1.02(-1.00)	0.04(0.32)	[-1.04,-0.93]	1.00
c_3	-	-	-	-	1.32(2.2)	0.06(2.10)	[1.30, 1.43]	1.01	2.21(3.5)	0.14(2.65)	[2.15, 2.63]	1.02
λ_3	-	-	-	-	1.13(1.00)	0.12(0.14)	[1.06, 1.35]	1.02	1.28(1.00)	0.13(0.14)	[1.20, 1.57]	1.02
b_4	-	-	-	-	-	-	-	-	-1.01(-1.00)	0.06(0.32)	[-1.06,-0.92]	1.00
c_4	-	-	-	-	-	-	-	-	1.10(1.75)	0.13(1.88)	[1.04, 1.28]	1.01
λ_4	-	-	-	-	-	-	-	-	1.18(1.00)	0.13(0.14)	[1.09, 1.46]	1.02
λ_n	2.86(3)	0.21(0.21)	[2.72, 3.27]	1.01	2.92(3)	0.19(0.21)	[2.79, 3.29]	1.07	2.86(3)	0.20(0.21)	[2.74, 3.34]	1.08

Table 1: Summary statistics of the estimated coefficients a, b, c, λ for the three models. The mean, standard deviation, 95% credible interval (uncertainty interval), and the convergence statistics \hat{R} are reported. The value in the parenthesis in Mean and Standard deviation represents their value of the prior distribution.

data, shows that the posterior distribution has not really been updated from the prior distribution. It turns out that the identified invalid data is not greatly sensitive to λ_n . In our estimation, only 3-4 data points are always identified as noise for a wide domain of λ_n . This implies that the posterior distribution of λ_n does not affect the posterior of the other parameters and therefore the pattern of recovered frontiers and the frequency distribution of ζ_d stay the same with a different λ_n .

Third, the summary result for the posterior distributions of a for all three models shows that the 95% uncertainty interval does not include zero with high certainty, confirming our prior knowledge on the negative quadratic model. The result for the posterior distributions b shows with high certainty that the estimated steady-state ratio of unit capital to labor cost is around 1 for all the frontiers in all three models. The posterior distributions of c , which represents the steady-state labor productivity growth, show that the outermost frontiers of the three models have the highest c (around 7-8%) while the inner frontiers have distinctive level of c depending on the number of parameters. $\lambda_1, \lambda_2, \lambda_3$, and λ_4 , whose inverse is the mean of the distance of the observed cost reduction from the corresponding frontier, varies from 0.33 to 1.13. The outermost frontier has the smallest λ , implying that the actual cost reduction of the economies is further away from the maximum cost reduction in the cluster compared to the other frontiers.

3.2.3 *Estimated frontiers and the frequency distribution of ζ_d*

Based on the posterior distribution of the estimated parameters, the following figure plots the recovered frontiers on the innovation space and the frequency distribution of ζ_d associated with each frontier:

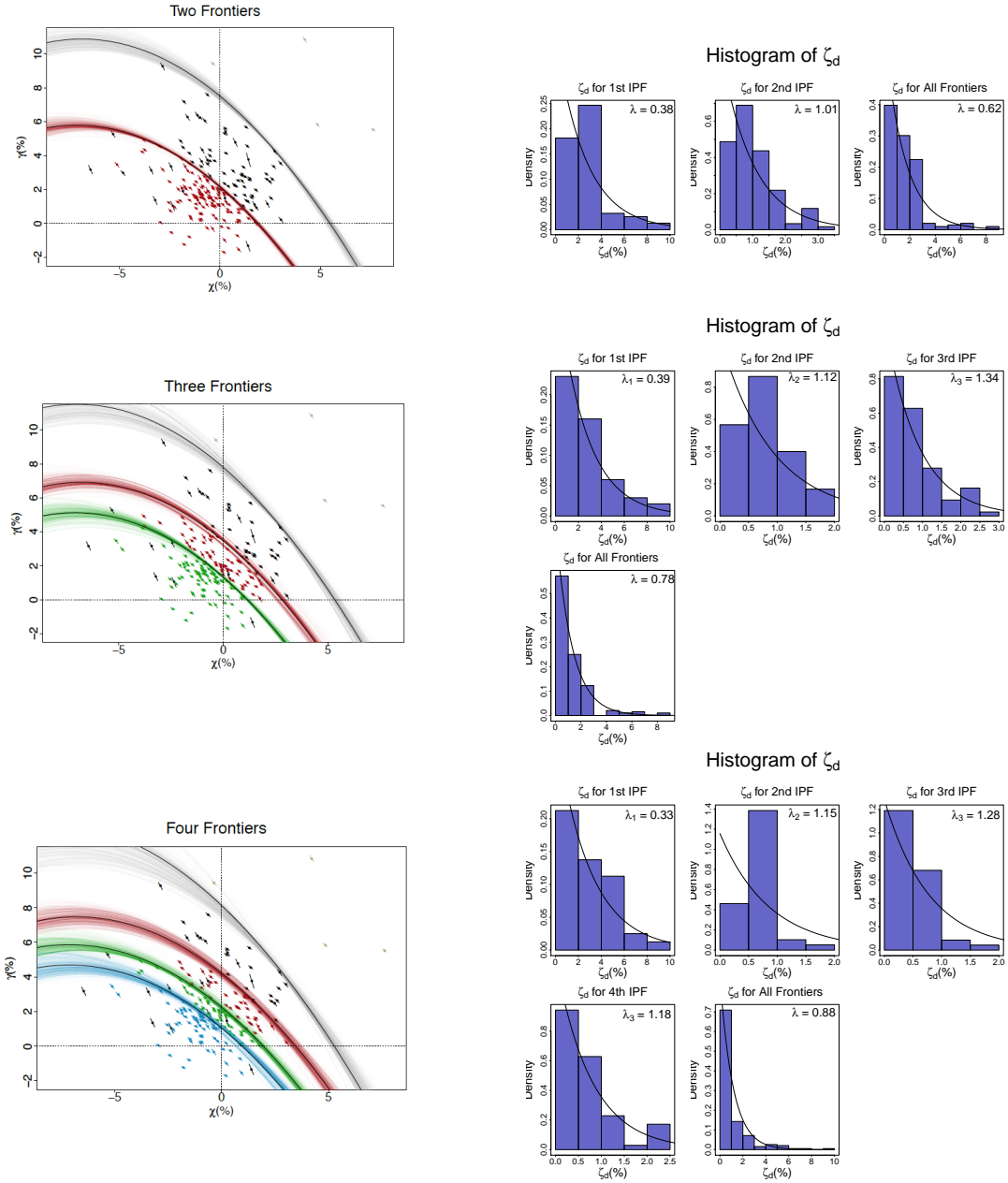


Figure 3: The recovered frontiers on the innovation space and the frequency distribution of the realized and maximum rate of cost reduction on the frontier ζ_d associated with each frontier. In the left panels, the black line is the frontier recovered from the mean value of the posterior distribution of a , b , and c while the colored lines around the frontier are randomly drawn from the posterior distributions. Each data point is colored to specify which frontier it belongs to. The slope of the bar is derived by the negative ratio of the unit labor to capital cost. In the right panels, the histograms of ζ_d are drawn along with the exponential fitted line with the mean value of λ as the rate parameter.

The left side panels represent the recovered frontiers for each model. Black lines are the frontiers recovered from the mean value of each parameter, a , b , and c . Those lines with slight tint around the

black line represent 1000 simulated frontiers drawn from the posterior distributions for each frontier. The data points associated with each frontier are displayed in different colors. The realizations of χ and γ associated with Frontier I (Model 1,2,3), Frontier II (Model 1,2,3), Frontier III (Model 2,3), and Frontier IV (Model 3) are colored in black, red, green, and blue respectively. The four data points outside the outermost frontier are identified as the invalid data points whose ζ value is far higher than the others. The bar drawn at each data point represents the negative ratio of the unit capital to labor cost, $-(1 - \omega)/\omega$. The steeper the slope is the longer the line is drawn for visualization purposes.

The right side panels represent the frequency distribution of the recovered ζ_d associated with each frontier. Each frequency distribution is displayed with the fitted line of the exponential distribution using the estimated mean value of λ as the rate parameter.

There are three important results that deserve our attention. First, the penalizing function identified four invalid data points outside the frontier in the two and three-frontier models. If included beneath the frontier, the four invalid data points would have made the distribution of ζ_d for the outer frontier far less organized and would also have greatly increased the steady-state labor productivity growth $\gamma^* = c$ in the estimation. In contrast, the gray uncertainty lines of the outermost frontier in the four-frontier model sometimes includes a data point that is identified as invalid in the first two models. Due to this problem, the posterior distribution of b_1 and c_1 has dual modes.

Second, some black data points below the inner frontiers suggest that there are some realizations of ζ seemingly close to the inner frontiers that are actually poorly performing economies in the outermost frontier class. The first frontier class with λ_1 around 0.35 is by far less organized than the other frontiers so that ζ_d spans from 0 to 10% as is shown in the histogram of ζ_d . This result makes economic sense because the economies in the outer frontier class have less information about the technology and therefore they are less constrained as to the adoption of a new technology. In contrast, the economies in the inner frontiers are the result of catching up to already well-established technologies with more information and thus, they are more organized.¹³

Third, the negative ratio of the unit capital to labor cost, $-(1 - \omega)/\omega$, expressed as the slope of each data point, explains how some economies are close to the frontier while some are not. For example, those economies with a steep $-(1 - \omega)/\omega$ have a relatively higher unit capital cost than the unit labor cost and, therefore, adopting a capital-saving technology rather than a labor-saving one is better choice for them. For this reason, if they are in the second quadrant where $\chi < 0$ and $\gamma > 0$, they would be far away from the frontier. For example, in the three-frontier model, the four black data points underneath the innermost frontier in the second quadrant have a very high ratio of the unit capital to labor cost, and they represent the most poorly performing economies in the first frontier class.

¹³I am grateful to Duncan Foley for this valuable insight.

3.2.4 Model evaluation and comparison

As an important part of the posterior analysis, we carry out a model evaluation and comparison for our models. For this purpose, I will use Widely Applicable Information Criteria (WAIC) (Watanabe, 2010, 2013). WAIC is a fully Bayesian approach because it utilizes the entire posterior distribution rather than a point estimation. For detailed survey on different model selection criteria, see Gelman et al. (2014).

First, the fit of model to data or the posterior predictive accuracy is assessed using the log pointwise predictive density (lppd). For each data point $i = 1, \dots, n$, the expected log-likelihood of the data is calculated given the posterior distribution and is summed over the entire data points:

$$\text{lppd} = \sum_{i=1}^n \text{Log} \left[\int p[y_i|\theta]p[\theta|y]d\theta \right]. \quad (20)$$

Since the analytical form of the posterior is not always available, we can evaluate the lppd using draws from the posterior simulations. Denoting the simulated θ to be θ_s , $s = 1, \dots, S$, the computed lppd, which I denote with an asterisk, lppd^* , is written as:

$$\text{lppd}^* = \sum_{i=1}^n \text{Log} \left[\frac{1}{S} \sum_{s=1}^S p[y_i|\theta_s] \right] \quad (21)$$

The higher the lppd is, the more accurately the model fit the data.

As other measures based on the log-likelihood, the lppd of observed data y does not account for the number of parameters in the model and therefore is prone to an overestimation of the predictive accuracy. One way to deal with this problem is to augment the lppd with a measure of bias correction. In WAIC, the correction is made by subtracting the “effective number of parameters” of the model ep_{WAIC} , which is a measure of model complexity. It is defined as the difference between the model deviance given the posterior distribution, $2 \text{Log}[E[p[y_i|\theta]]]$, and the expectation of the model deviance, $E[2 \text{Log}[p[y_i|\theta]]]$:

$$\text{ep}_{\text{WAIC}} = \sum_{i=1}^n (2 \text{Log}[E[p[y_i|\theta]]] - E[2 \text{Log}[p[y_i|\theta]]]), \quad (22)$$

which can be computed by:

$$\text{ep}_{\text{WAIC}}^* = 2 \sum_{i=1}^n \left(\text{Log} \left[\frac{1}{S} \sum_{i=1}^S p[y_i|\theta_s] \right] - \frac{1}{S} \sum_{s=1}^S \text{Log}[p[y_i|\theta_s]] \right). \quad (23)$$

Finally, WAIC is obtained by subtracting ep_{WAIC} from the model fit, lppd. In deviation scale, the

measure is multiplied by -2 so that WAIC is written as follows:

$$\text{WAIC} = -2(\text{lppd} - \text{ep}_{\text{WAIC}}). \quad (24)$$

The lower WAIC, the more accurate the posterior prediction is for new data and more parsimonious the model is. The following table summarizes the lppd, ep_{WAIC} , and WAIC for 5 different frontier models with the different number of frontier:

Models	n	lppd	ep_{WAIC}	WAIC
One Frontier	200	-283.31	1.75	570.12
Two Frontiers	200	-264.76	3.44	536.41
Three Frontiers	200	-203.41	9.10	425.00
Four Frontiers	200	-168.98	9.78	357.52
Five Frontiers	200	-160.45	10.71	342.33

Table 2: AICc of the three different frontier models.

The result shows that WAIC decreases as the number of frontiers increases. However, the increase gets less significant as we introduce more frontiers, so that the five-frontier model has a similar predictive accuracy with the four-frontier model.

4 Discussion: Characteristics of Recovered Frontier Class

This section looks into the distinctive characteristics of the frontier class using the recovered frontiers and the associated data points on the innovation possibility space. For illustration, we look at the result of the three-frontier case. Similar results are observed in the two- and four-frontier models as well. The following table shows the countries and years of each frontier class:

Countries	Outer Frontier	Middle Frontier	Inner Frontier
Australia	–	1978, 1984, 1998	1973, 1994, 2005
Austria	1970,1990	1979	1985, 2000, 2007
Belgium	1973, 1988	1980, 2000	1976, 1984, 1994, 1997, 2004, 2007
Brazil	1975, 2007	2004	2000
Canada	–	1973, 1988, 1999	1976, 1984, 1994, 2004
Chile	1989, 1992, 1995	1977	1974, 2000, 2004
Denmark	1986	2000	1973, 1976, 1979, 1994, 1997, 2006
Finland	1989, 2007	1979, 1997	–
France	1988	1973, 1982, 2000	1976, 1994, 2004, 2007
Germany	–	1990	1994, 2000, 2006
Greece	1972, 1988	1978, 1997, 2003, 2006	1985, 2000
Hungary	–	1998, 2002	1993
Iceland	–	1987, 1998	1990
Ireland	1990, 1997	1972, 1978, 2005	1984
Israel	1972, 1987	1981, 1991	1978, 1995, 2000, 2005
Italy	1973, 1979, 1988	1976, 1995	1984, 2000, 2006
Japan	1988	1985, 2004	1979, 1996, 2000
Korea	1973, 1978, 1987, 1990	1994	1983, 1999
Luxembourg	1986, 1989	1973, 2006	1978, 1999
Mexico	1972, 1984, 1990, 1997	1979, 2000	2005
Netherlands	1970, 1976	1973	1986, 1990, 1999, 2006
New Zealand	–	1973, 1999	1984, 1993, 2002, 2005
Norway	1997	1971, 1980, 1984, 2004	1976, 2000
Poland	2007	–	–
Portugal	1976, 1980	1970, 1988	1994, 1998, 2005
Spain	–	1972, 1987	1980, 2000, 2006
Sweden	1970, 1987, 1999	1984, 2005	1973, 1979, 1995
Switzerland	1970	2000, 2006	1980, 1984, 1989, 1994
Turkey	1972, 1976, 1983, 1987,1997,2000	1993	1990
United Kingdom	1988	1973, 1983, 2000	1978, 1994, 2005
United States	–	1988	1973, 1976, 1984, 1994, 1997, 2005

Table 3: Countries and Years of Each Cluster. There are 50, 60, 86 data points in the outer, middle, and inner frontier. The recovered peak-to-peak measures span from 1970 to 2007.

The table shows that advanced OECD countries such as the US, the UK, Belgium Canada, France tend to fall into the middle and the inner frontier while developing countries such as Mexico, Korea, and Turkey tend to belong to the outer and the middle frontier.

To take a further look at the characteristics of each cluster, the following bar charts summarize the labor productivity, capital productivity, unit labor cost, labor productivity growth, capital productivity growth, and the rate of cost reduction for each frontier in the three-frontier model:

Characteristics of the Frontier Class

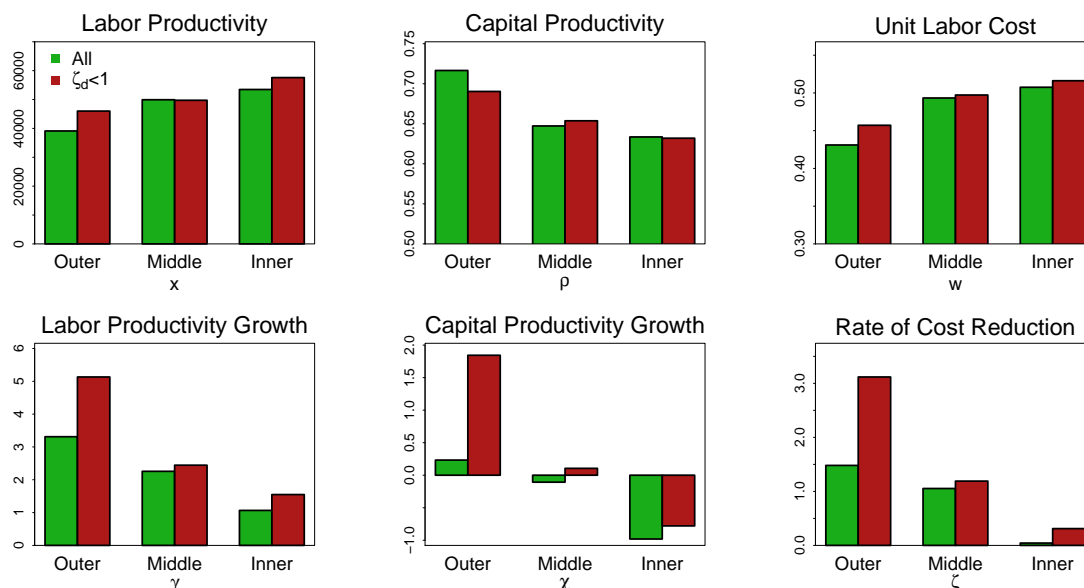


Figure 4: The Characteristics of each frontier by the labor productivity x , capital productivity ρ , unit labor cost ω , labor productivity growth γ , capital productivity growth χ , and the rate of cost reduction ζ for each frontier. The bar in green shows the mean of all the data in the cluster while the bar in red shows the mean of those data points whose $\zeta_d < 1\%$.

The green bar shows the average value of the target variable for each frontier class while the red bar shows the average for those data points whose $\zeta_d < 1\%$ which, therefore, are the most efficient group of economies. The result shows that the level of our target variables, x , ρ , ω , γ , χ , ζ , gradually changes as the frontier moves toward the origin. First and foremost, there are noticeable changes in the *level* of labor and capital productivity. The inner frontier has the highest level of labor productivity, \$53,474 (or \$57,591 for those data points $\zeta_d < 1\%$) per worker, while the middle and the outer frontier have \$49,924 (\$49,741) and \$39,108 (\$45,996) per worker, respectively. In contrast, the level of capital productivity is the highest with 0.72 (0.69) in the outer frontier and 0.65 (0.65) and 0.63 (0.63) in the middle and the inner frontier. Meanwhile, the unit labor cost is the lowest in the outer frontier class with 0.43 (0.45) and is the highest in the inner frontier class with 0.51 (0.52). The middle frontier has the unit labor cost of 0.49 (0.50).

This striking pattern explains a stylized fact of economic development of advanced economies from a low-wage, labor-intensive economy to a high-wage, capital-intensive economy. An economy starts with a relatively low unit labor cost and a low degree of technical progress that involves a low stage of mechanization and low labor productivity. In the course of economic development, the economy experiences a rise in wages and begins to adopt the technology that saves more labor but consumes more capital. Consequently, the level of labor productivity increases while the level of

capital productivity tends to decrease on average.

As is shown in the bottom three rows in Figure 4, where the average growth rates of γ , χ , and ζ are displayed, the outer frontier predictably has the highest ζ coming from the highest γ and χ while the inner frontier class has the lowest ζ whose value is nearly zero. The overall result supports the “catch-up effect” in technological progress, which points out the phenomenon that those economies with relatively backward technology have more room for technological improvement (Baumol, 1986). In other words, those economies which have already achieved a higher level of productivity have difficulty further increasing productivity and end up having a lower rate of growth.¹⁴

5 Conclusion

The paper develops a new econometric model to estimate multiple IPFs within the probabilistic ITC model. Our method is particularly useful for a pooled data where multiple innovation possibilities frontiers characterize different sub-groups of the economies. The paper has found that the innovation possibilities cluster defined by the rate of cost reduction has a distinctive level of important economic variables such as the levels of labor and capital productivity and the unit labor cost. The evidence shows that the lower the rate of cost reduction is, the higher the level of labor productivity, the lower the level of capital productivity and the higher level of unit labor cost. This result implies that the technological frontier is associated with the well-documented economic development from a low-wage labor intensive economy to a high-wage capital-intensive economy.

The paper constitutes one of the first attempts to empirically test the ITC model by demonstrating that the concave IPFs can be a good model to explain the data on the growth rates of capital and labor productivity, and the cost structure, the key variables of the ITC theory. The paper also shows that an entropy constraint in the mixed strategy in the choice of technology provides the general form of the technical inefficiency function, from which its different specifications can be further derived. A possible future research agenda in this line of the entropy-constrained model in the ITC literature will be to study what behavioral assumptions are required to derive the different technical inefficiency functions.

¹⁴The result explains the cross-sectional movement of economies through the “level” effect of productivity on its speed of growth, but does not directly address the movement of frontiers over time. Introducing a time trend in the location parameter of the frontier model will enable us to track the movement of technological frontier in a pooled data. This research would constitute an extension of the multiple frontier estimation model employed in this paper and could provide a potential answer to the question regarding “convergence or divergence” of the world economy, which has been one of the most debated topics in economics. I’m grateful to Mark Setterfield for suggesting to further link my findings in this paper to a larger literature on the convergence.

References

- Aigner, D., Lovell, C., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1):21–37.
- Baumol, W. J. (1986). Productivity growth, convergence, and welfare: What the long-run data show. *The American Economic Review*, 76(5):1072–1085.
- Borwein, J. M. and Lewis, A. S. (1991). On the convergence of moment problems. *Transactions of the American Mathematical Society*, 325(1):249–271.
- Brugger, F. and Gehrke, C. (2016). The neoclassical approach to induced technical change: From hicks to acemoglu. *Metroeconomica*, pages n/a–n/a.
- Cover, T. M. and Thomas, J. A. (2006). *Elements of information theory (2nd Edition)*. Wiley-Interscience.
- Drandakis, E. M. and Phelps, E. S. (1966). A model of induced invention, growth and distribution. *The Economic Journal*, 76(304):823–840.
- Foley, D. K. (2003). Endogenous technical change with externalities in a classical growth model. *Journal of Economic Behavior and Organization*, 52(2):167 – 189.
- Gamerman, D. and Lopes, H. F. (2008). *Markov Chain Monte Carlo: stochastic simulation for Bayesian inference (2nd edn)*. CRC Press.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis (Third Edition)*. CRC Press.
- Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and Computing*, 24(6):997–1016.
- Greene, W. H. (1990). A gamma-distributed stochastic frontier model. *Journal of Econometrics*, 46:141–163.
- Grubb, M., Köhler, J., and Anderson, D. (2002). Induced technical change in energy and environmental modeling: Analytic approaches and policy implications. *Annual Review of Energy and the Environment*, 27(1):271–308.
- Hajargasht, G. (2014). Stochastic frontiers with a rayleigh distribution. *Journal of Productivity Analysis*, 44:199–208.
- Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

- Kennedy, C. (1964). Induced bias in innovation and the theory of distribution. *The Economic Journal*, 74:541–547.
- Lan, S., Streets, J., and Shahbaba, B. (2014). Wormhole hamiltonian monte carlo. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI'14, pages 1953–1959. AAAI Press.
- Laskey, K. B. and Myers, J. W. (2003). Population markov chain monte carlo. *Machine Learning*, 50(1):175–196.
- Lee, J., Veloso, F. M., and Hounshell, D. A. (2011). Linking induced technological change, and environmental regulation: Evidence from patenting in the u.s. auto industry. *Research Policy*, 40(9):1240 – 1252.
- MacKay, D. J. (2005). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.
- Meeusen, W. and van Den Broecke, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International Econometric Review*, 18:435–444.
- Minh, D. D. L. and Minh, D. L. P. (2015). Understanding the hastings algorithm. *Communications in Statistics - Simulation and Computation*, 44(2):332–349.
- Neal, R. M. (1996). Sampling from multimodal distributions using tempered transitions. *Statistics and Computing*, 6(4):353–366.
- R.Stevenson (1980). Likelihood functions for generalized stochastic frontier functions. *Journal of Econometrics*, 13:57–66.
- RStudio Team (2016). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.
- Samuelson, P. A. (1965). A theory of induced innovation along kennedy-weizsäcker lines. *The Review of Economics and Statistics*, 47(4):343–356.
- Scharfenaker, E. and Foley, D. K. (2017). Quantal response statistical equilibrium in economic interactions: Theory and estimation. *Entropy*, 19(9).
- Shah, A. and Desai, M. (1981). Growth cycles with induced technical change. *The Economic Journal*, 91(364):1006–1010.
- Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665 – 690. Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information.

- Sims, C. A. (2006). Rational inattention: A research agenda. Available at <http://sims.princeton.edu/yftp/RIplus/RatInattPlus.pdf>.
- Sminchisescu, C. and Welling, M. (2011). Generalized darting monte carlo. *Pattern Recognition*, 44(10):2738 – 2748. Semi-Supervised Learning for Visual Content Analysis and Understanding.
- Uzawa, H. (1961). Neutral inventions and the stability of growth equilibrium. *The Review of Economic Studies*, 28(2):117–124.
- von Neumann and Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton University Press.
- Von Weizsäcker, C. C. (2010). A new technical progress function (1962). *German Economic Review*, 11(3):248–265.
- Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal Machine Learning Research*, 11:3571–3594.
- Watanabe, S. (2013). A widely applicable bayesian information criterion. *Journal Machine Learning Research*, 14(1):867–897.
- Wing, I. S. (2003). Induced technical change and the cost of climate policy. *MIT Joint Program on the Science and Policy of Global Change Report*, 102.
- Yang, J. (2017a). An entropy-constrained model of induced technical change with a single innovation possibility frontier. *New School for Social Research working paper*, 2017.
- Yang, J. (2017b). Information theoretic approaches in economics. *Journal of Economic Surveys*, pages n/a–n/a. doi = 10.1111/joes.12226.

A A functional form of technical inefficiency probability function

Using a simple payoff function in which the payoff is a linear function of $u[\zeta_d] = m - n(\zeta_d)$, the technical inefficiency function becomes:

$$p[\zeta_d] \propto e^{\frac{m-n(\zeta_d)}{T}} \quad (25)$$

where $m, n > 0$. Since the constant term is washed out when integrating:

$$\frac{e^{\frac{m}{T}} e^{-\frac{n(\zeta_d)}{T}}}{e^{\frac{m}{T}} \int e^{-\frac{n(\zeta_d)}{T}}} = \frac{e^{\frac{m}{T}} e^{-\frac{n(\zeta_d)}{T}}}{e^{\frac{m}{T}} \int e^{-\frac{n(\zeta_d)}{T}}} = \frac{e^{-\frac{n(\zeta_d)}{T}}}{\int e^{-\frac{n(\zeta_d)}{T}}}.$$

Therefore, the technical inefficiency function is written as:

$$p[\zeta_d] \propto e^{-\frac{\zeta_d}{\beta}} \quad (26)$$

where $\beta = T/n$ represents the intensity of the payoff.

B Trace Plot of Metropolis-Hastings simulation

In modeling the Metropolis-Hastings algorithm for a target distribution $p[\theta]$, we take the following steps:

1. Draw from a uniform distribution a random number θ_0^* for the initial value of the parameter.
2. Choose a ‘‘proposal’’ density $g(\theta)$ to propose a next value of the parameter. It is chosen to be a normal distribution with the mean being the previous value of the parameter θ_{t-1}^* .
3. For the t^{th} iteration,
 - (a) Propose a new value θ_t from $g(\theta_t|\theta_{t-1}^*)$
 - (b) Calculate the accept probability α by taking the ratio of $p[\theta_t]/p[\theta_{t-1}^*]$
 - (c) Draw a random number u from the uniform distribution between 0 and 1
 - (d) If $\alpha \geq u$, $\theta_t^* = \theta_t$. If $\alpha < u$, $\theta_t^* = \theta_{t-1}^*$
4. Repeat 3 until the predetermined number of iterations n so that there are n simulated numbers in the sample θ^* .

The following figure shows the convergence plot of Metropolis-Hastings simulation for three posterior distributions.

Convergence Plot for Two Frontier Model

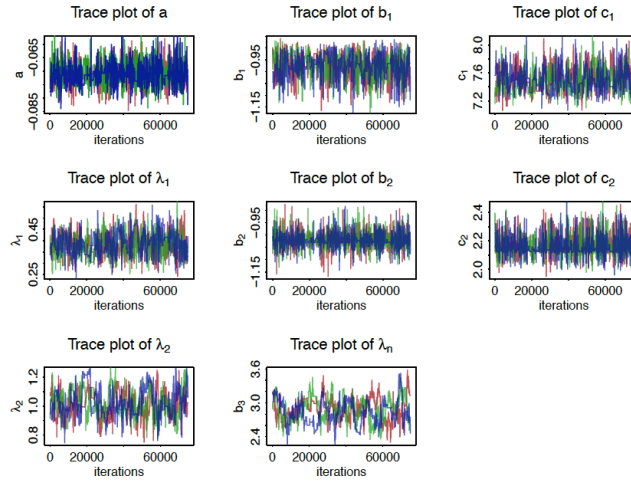


Figure 5: The traceplot of Metropolis-Hastings simulation for the two-frontier model.

Convergence Plot for Three Frontier Model

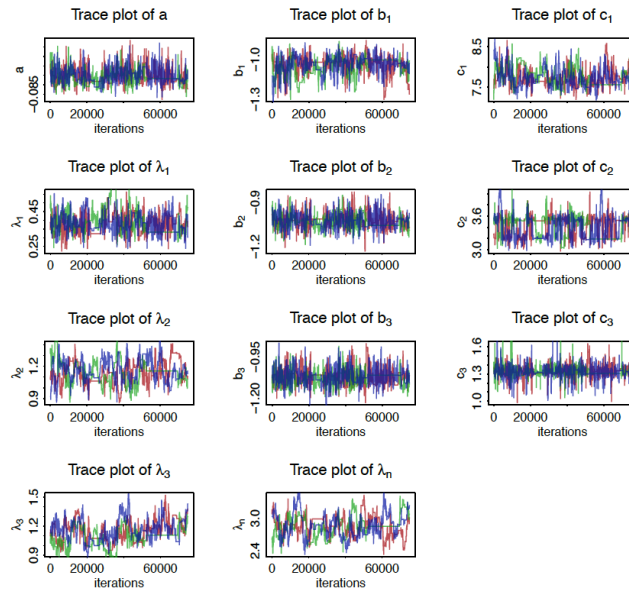


Figure 6: The traceplot of Metropolis-Hastings simulation for the three frontier model.

Convergence Plot for Four Frontier Model

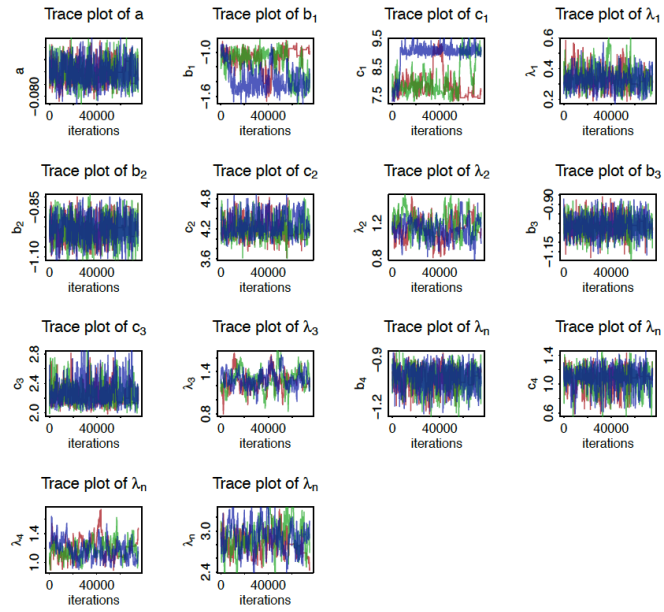


Figure 7: The traceplot of Metropolis-Hastings simulation for the four-frontier model.