ESSAYS IN INDUSTRIAL ORGANIZATION

By

Jaemin Ryu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Economics – Doctor of Philosophy

2021

ABSTRACT

ESSAYS IN INDUSTRIAL ORGANIZATION

By

Jaemin Ryu

Chapter 1. Measurement and Decomposition of Cost Inefficiency Using Copulas: An Application to the U.S. Banking Industry

This paper proposes a model and an estimation strategy using copulas in order to measure and decompose technical and allocative inefficiency in the translog cost system. This study adapts the stochastic cost frontier model from Kumbhakar (1997) and employs the APS copulas developed by Amsler et al. (2021) to capture the dependence between technical and allocative inefficiency. The joint density of the system is derived by the probability integral transform and the copula-based version of the Rosenblatt transformation, leading to the method of simulated likelihood estimation. This study also proposes a strategy to estimate individual inefficiency using density deconvolution and conditional distributions. The new methods are then applied to the U.S. banking industry. The results suggest that U.S. bank costs increased by approximately 20% in 2019 and 2020 due to inefficiency, where technical and allocative inefficiency represented 16∼18% and 2.5%, respectively. In addition, ignoring the dependence between technical and allocative inefficiency would produce less plausible results.

Chapter 2. Measurement and Decomposition of Cost Inefficiency Using Copulas: Evidence from Monte Carlo Simulations

The purpose of this paper is to provide methods for copula-based simulations and to demonstrate the performance of the estimation strategy that can measure and decompose cost inefficiency. First, a method to draw random numbers using the APS-3-A copula, which corresponds to the three-input case, is presented. Specifically, copula arguments can be obtained from random numbers distributed independently and uniformly over $[0, 1]$ by applying the inverse Rosenblatt transformation, which needs to derive conditional distributions of cop-

ulas. Then, dependent random numbers can be generated by the inverse transformation method. Second, quasi-Monte Carlo simulations are conducted given the data generating process. Simulation results suggest that the parameters of the translog cost system that accommodates technical and allocative inefficiency can be reliably estimated when the APS copulas are employed. It would also yield biased estimates when the disturbances in the cost function and the cost share equations of the system are regarded as independent.

Chapter 3. Demand Estimation of Deposits: A Case of the Korean Financial Industry

This paper estimates a structural demand model for deposits in the Korean financial sector in order to measure the effects of deregulation in payment and settlement systems in 2009, which caused cash management accounts (CMAs) of securities companies to become close substitutes for traditional deposit services provided by banks. Following the discrete choice literature, depositors choose among differentiated financial institutions, considering their offered interest rates and other attributes. Although it is also assumed that market discipline in banking exists, it depends on the financial stability situation. The results show that consumers respond favorably to deposit rates, the branch staffing, and the number of branches of depository institutions in tranquil times. On the other hand, they consider the financial institution's capital adequacy ratio more important than interest rates during the financial turmoil. This is similar to the phenomenon referred to as the flight to quality in other financial markets. Therefore, although CMAs have the benefit of higher interest rates compared to traditional deposit services, their market share has remained at low levels due to the prolonged financial stress since the global financial crisis, which results in marginal increases in consumer welfare from the deregulation. This implies that the deregulation would not have successfully achieved the purpose of improving consumer welfare by promoting competition.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Kyoo il Kim for his generous support, encouragement, and valuable comments. I would also like to appreciate Peter Schmidt for his deep insights and clear guidance. I additionally thank Jay Pil Choi and Sriram Narayanan for their assistance and helpful comments. Last but not least, I deeply indebted my family and Hyunjung Kim for all their devoted support and love throughout my studies.

I would like to acknowledge financial support from the Bank of Korea and the Michigan State University Department of Economics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

MEASUREMENT AND DECOMPOSITION OF COST INEFFICIENCY USING COPULAS: AN APPLICATION TO THE U.S. BANKING INDUSTRY

1.1 Introduction

Stochastic frontier models (SFMs) developed by Aigner et al. (1977) and Meeusen and van Den Broeck (1977) have been widely used for efficiency analysis. There are two approaches to measure efficiency in the SFMs. The first one is an output-oriented approach, which is used to estimate the production frontier and to measure technical (in)efficiency. The second one is an input-oriented approach, which can be used to estimate the cost frontier and to measure cost (in)efficiency. As Kumbhakar and Lovell (2000) note, there are several differences between these two approaches, an important difference of which is that cost efficiency can be decomposed into input-oriented technical efficiency and input allocative efficiency, whereas output-oriented technical efficiency cannot be decomposed.[1]

Farrell (1957) defines technical and allocative inefficiency as follows. Technical inefficiency occurs when a producer fails to produce the maximum output from a given input bundle. Allocative inefficiency occurs when a producer uses inputs in the wrong proportions, given input prices. As inefficiency can arise by these different causes, it is important to measure and decompose cost inefficiency in order to evaluate the performance of firms.

Schmidt and Lovell (1979) show how to measure both technical and allocative inefficiency, assuming the Cobb-Douglas production technology. Nevertheless, it would be necessary to apply flexible functional forms, such as a translog function[2], when measuring cost inefficiency. Since Christensen et al. (1971) and Christensen and Greene (1976), the translog functional form has played an important role in cost studies, owing to several virtues that overcome the

[1]Hereafter, technical (in)efficiency means input-oriented technical (in)efficiency.

[2]For more details about the translog function, please refer to Kumbhakar and Lovell (2000) and Sickles and Zelenyuk (2019).

limitations of the Cobb-Douglas function. As mentioned in Kumbhakar and Lovell (2000), for example, the translog cost function can accommodate multiple outputs without violating the requisite curvature conditions in output space, while the Cobb-Douglas functional form cannot. In addition, the translog function can provide a second-order approximation to any well-behaved underlying cost frontier, whereas the Cobb-Douglas representation would lead to biased estimates of inefficiency; this is because unmodeled technology complexity could appear in the error term, which contains information about inefficiency, due to its simplicity.

In contrast, as first noted by Greene (1980)[3], econometric issues arise when employing the translog functional form. The problem is characterized as follows. Given the definition in Farrell (1957), the cost system that allows for technical and allocative inefficiency can be written as

$$\begin{aligned} \ln C_i &= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + \epsilon_i \\ &= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + v_i + u_i \\ &= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + v_i + u_i^T + u_i^A \\ s_{ij} &= s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + e_{ij},\ j = 2, \cdots, J, \end{aligned} \tag{1.1.1}$$

where C_i is the actual cost of producer i, $C(\boldsymbol{y}_i, \boldsymbol{w}_i)$ is the deterministic kernel of the stochastic cost frontier, $\boldsymbol{y}_i \in \mathbb{R}_+^M$ is a vector of M outputs produced by producer i, $\boldsymbol{w}_i \in \mathbb{R}_{++}^J$ is a vector of input prices faced by producer i, $v_i \in \mathbb{R}$ is a random disturbance, $u_i^T \in \mathbb{R}_+$ represents a cost increase due to technical inefficiency, $u_i^A \in \mathbb{R}_+$ represents a cost increase due to allocative inefficiency, $u_i = u_i^T + u_i^A$, $\epsilon_i = v_i + u_i$, $s_{ij} \in [0,1]$ is the producer i's actual cost share of input j, $s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) \in [0,1]$ is the optimum cost share of input j derived from Shephard's lemma, and $e_{ij} \in \mathbb{R}$ is the disturbance due to allocative inefficiency of producer i and noise.

The important question is how to model u_i in the cost function and e_{ij} in the cost share equations. As long as e_{ij} represents allocative inefficiency, it cannot be independently distributed of u_i that captures both technical and allocative inefficiency. If it is assumed

[3] It is known as "the Greene Problem." Please refer to Bauer (1990) and Kumbhakar and Lovell (2000) for detailed discussion.

that u_i and e_{ij} are independent, it would lead to inconsistent parameter estimates, and it is impossible to decompose cost inefficiency into two sources (Kumbhakar and Lovell, 2000, p.156).

So far, three types of solutions to "the Greene Problem" have been proposed. The first solution is finding the analytic relationship between u_i^A and $\boldsymbol{e_i} = (e_{i2}, \cdots, e_{iJ})$ (Schmidt and Lovell (1979) and Kumbhakar (1989, 1997)). The second solution is approximating u_i^A as a function of $\boldsymbol{e_i}$, such as in Schmidt (1984), Melfi (1984), Bauer (1985), and Kumbhakar (1991). The third solution proposed by Greene (1980) is ignoring the relation between them and assuming that ϵ_i and $\boldsymbol{e_i}$ are independent. However, as discussed in Bauer (1990), all of these existing solutions are not ideal. The first one can be used when restrictive functional forms, such as a Cobb-Douglas functional form, and/or restrictive assumptions are imposed. The second solution is valid only when the approximation function captures the true relationship between u_i^A and $\boldsymbol{e_i}$. The third approach might fail to use all available information for estimation.

The main contribution of this paper is to propose a translog cost system, which is rigorously developed based on economic theory, providing a solution to "the Greene Problem" that overcomes the limitations of previous studies. In other words, technical and allocative inefficiency can be precisely measured and decomposed by using the proposed model with a flexible functional form. In addition to the main contribution, a novel estimation strategy is proposed that is consistent with the theory behind the stochastic cost frontier model as well as computationally easy. Lastly, this study suggests a method to estimate individual inefficiency when additive noise terms are allowed in the cost share equations.

A stochastic cost frontier model taking a translog cost functional form is constructed based on Kumbhakar (1997) in order to derive the exact relationship between the error terms representing allocative inefficiency in the cost function and the cost share equations. However, as in Schmidt and Lovell (1980), dependence between technical and allocative inefficiency is assumed. This is modeled by the APS copulas developed by Amsler et al.

(2021). Several assumptions are imposed to capture the dependence and to make the model more realistic and estimable. The method of simulated likelihood is applied to estimate the model, where the joint density of the model is derived by the probability integral transform and the copula-based version of the Rosenblatt transformation.

Additionally, given that the model can estimate only average inefficiency of firms, this study also proposes a strategy to measure and decompose individual inefficiency. Jondrow et al. (1982) and Battese and Coelli (1988) propose ways to estimate firm-level inefficiency in production using the conditional distribution $f(u_i^T|\epsilon_i)$, where u_i^T is output-oriented technical inefficiency, $\epsilon_i = v_i - u_i^T$, and v_i is a random disturbance. This method could be adapted to the stochastic cost frontier analysis. However, due to the different environment, we must employ density deconvolution to error terms in the cost share equations of (1.2.1).

The new model and strategies are applied to U.S. depository institutions. Numerous studies on bank efficiency exist for various themes.[4] For instance, innovations in technology, such as telecommunication technologies and information processing, have been intensively adopted in the banking industry (see Feng and Serletis, 2009). In addition, especially in the United States, regulatory changes, such as branching deregulation and enhanced regulatory capital requirements, have affected operation strategies of banks. Since these factors have an impact on technical and allocative inefficiency of banks, respectively, it is necessary to measure and decompose cost inefficiency of banks in order to identify each factor's effect on the performance of banks. Furthermore, it is possible that technical and allocative inefficiency of banks has changed due to outbreak of COVID-19, since the pandemic has affected financial markets and resource allocation of depository institutions.

The remainder of the chapter is organized as follows. Section 1.2 develops the econometric model. Section 1.3 presents the estimation strategy. Section 1.4 develops the strategy to estimate firm-level inefficiency. Section 1.5 presents empirical results applied to the U.S. banking industry. Section 1.6 concludes the chapter.

[4] Please refer to Berger and Humphrey (1997) and Bhatia et al. (2018) for more details.

1.2 Model

1.2.1 Translog Cost System of Kumbhakar (1997) Revisited

Kumbhakar (1997) establishes an exact relationship between the terms representing allocative inefficiency in the cost function and the cost share equations of the stochastic cost frontier model. This subsection summarizes its setup and results.

1.2.1.1 Setup

Let $P(\boldsymbol{y}_i, \boldsymbol{x}_i e^{-u_i^T}) = 0$ be a production possibility frontier[5], where $\boldsymbol{x}_i \in \mathbb{R}_+^J$ is a vector of J inputs used by producer i. Recall that $\boldsymbol{y}_i \in \mathbb{R}_+^M$ is a vector of M outputs, and $u_i^T \in \mathbb{R}_+$ represents technical inefficiency. $P(\cdot)$ is assumed to be differentiable. Then, the cost minimization problem of producer i who is only technically inefficient can be written as

$$\min_{\boldsymbol{x}_i} \ \boldsymbol{w}_i' \boldsymbol{x}_i \quad \text{s.t. } P(\boldsymbol{y}_i, \boldsymbol{x}_i e^{-u_i^T}) = 0.$$

Note that it yields the same solution to the following problem such that

$$\min_{\boldsymbol{x}_i^*} \ \boldsymbol{w}_i' \boldsymbol{x}_i^* \quad \text{s.t. } P(\boldsymbol{y}_i, \boldsymbol{x}_i^*) = 0,$$

where $\boldsymbol{x}_i^* = \boldsymbol{x}_i e^{-u_i^T}$. Its first-order conditions are

$$\frac{P_j(\boldsymbol{y}_i, \boldsymbol{x}_i^*)}{P_1(\boldsymbol{y}_i, \boldsymbol{x}_i^*)} = \frac{w_{ij}}{w_{i1}}, \ j = 2, \cdots, J,$$

where $P_j(\boldsymbol{y}_i, \boldsymbol{x}_i^*), \ j = 1, \cdots, J$, denotes the partial derivative of $P(\boldsymbol{y}_i, \boldsymbol{x}_i^*)$ with respect to x_{ij}^*. Given this result, the first-order conditions of producer i who is both technically and allocatively inefficient can be written as

$$\frac{P_j(\boldsymbol{y}_i, \boldsymbol{x}_i^*)}{P_1(\boldsymbol{y}_i, \boldsymbol{x}_i^*)} = \frac{w_{ij}}{w_{i1}} e^{\xi_j}, \ j = 2, \cdots, J,$$

[5]Kumbhakar (1997) uses a production function to derive the translog cost system, which means he considers a single-output case. However, as noted by Parmeter and Kumbhakar (2014), the derivation and the result of the translog cost system for a multiple-output case are similar to the single-output case.

where $\xi_j \in \mathbb{R},\ j = 2, \cdots, J$, represents producers' allocative inefficiency for the input pair $(j, 1)$. Note that $\xi_1 = 0$ by construction. If $\xi_j = 0$ for $j = 2, \cdots, J$, the input pair $(j, 1)$ is perfectly allocatively efficient.

1.2.1.2 Results

The stochastic cost frontier model can be written as

$$\begin{aligned} \ln C_i &= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + v_i + u_i^T + u_i^A \\ s_{ij} &= s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + \eta_{ij},\ j = 2, \cdots, J, \end{aligned}$$

where $\eta_{ij} \in \mathbb{R}$ is the deviation from the optimum cost share of input j due to only allocative inefficiency of producer i, which does not contain noise. Recall that C_i is the actual cost of producer i, $C(\boldsymbol{y}_i, \boldsymbol{w}_i)$ is the deterministic kernel of the stochastic cost frontier, $v_i \in \mathbb{R}$ is a random disturbance, $u_i^A \in \mathbb{R}_+$ represents a cost increase due to allocative inefficiency, $s_{ij} \in [0, 1]$ is the actual cost share of input j, and $s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) \in [0, 1]$ is the optimum cost share of input j.

Assume that the deterministic kernel of the stochastic cost frontier takes a translog functional form. Then, deterministic components of the system, $\ln C(\boldsymbol{y}_i, \boldsymbol{w}_i)$ and $s_j(\boldsymbol{y}_i, \boldsymbol{w}_i)$, can be written as

$$\begin{aligned} \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) &= \beta_0 + \sum_{m=1}^{M} \beta_m^y (\ln y_{im}) + \frac{1}{2} \sum_{m=1}^{M} \sum_{n=1}^{M} \beta_{mn}^{yy} (\ln y_{im})(\ln y_{in}) \\ &\quad + \sum_{j=1}^{J} \beta_j^w (\ln w_{ij}) + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk}^{ww} (\ln w_{ij})(\ln w_{ik}) \\ &\quad + \sum_{m=1}^{M} \sum_{j=1}^{J} \beta_{mj}^{yw} (\ln y_{im})(\ln w_{ij}) \\ s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) &= \beta_j^w + \sum_{k=1}^{J} \beta_{jk}^{ww} (\ln w_{ik}) + \sum_{m=1}^{M} \beta_{mj}^{yw} (\ln y_{im}),\ j = 2, \cdots, J. \end{aligned}$$

Note that $s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) = \frac{w_{ij} x_{ij}}{C(\boldsymbol{y}_i, \boldsymbol{w}_i)} = \frac{w_{ij}}{C(\boldsymbol{y}_i, \boldsymbol{w}_i)} \frac{\partial C(\boldsymbol{y}_i, \boldsymbol{w}_i)}{\partial w_{ij}} = \frac{\partial \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i)}{\partial \ln w_{ij}}$. The second equality holds because of Shephard's lemma. Also, the terms representing allocative inefficiency can be written as

$$u_i^A = \sum_{j=1}^{J} \beta_j^w \xi_j + \sum_{j=1}^{J}\sum_{k=1}^{J} \beta_{jk}^{ww} (\ln w_{ij}) \xi_k + \frac{1}{2} \sum_{j=1}^{J}\sum_{k=1}^{J} \beta_{jk}^{ww} \xi_j \xi_k$$
$$+ \sum_{m=1}^{M}\sum_{j=1}^{J} \beta_{mj}^{yw} (\ln y_{im}) \xi_j + \ln \sum_{j=1}^{J} (s_{ij}^*/e^{\xi_j})$$
$$\eta_{ij} = \frac{s_j(\boldsymbol{y}_i, \boldsymbol{w}_i)[1 - \{\sum_{k=1}^{J}(s_{ik}^*/e^{\xi_k})\}e^{\xi_j}] + \sum_{k=1}^{J} \beta_{jk}^{ww}\xi_k}{\{\sum_{k=1}^{J}(s_{ik}^*/e^{\xi_k})\}e^{\xi_j}}, \; j = 2, \cdots, J,$$

where $s_{ij}^* = s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + \sum_{k=1}^{J} \beta_{jk}^{ww}\xi_k, \; j = 2, \cdots, J$, is the shadow cost share of input j for producer i who is assumed to be only allocatively inefficient.

To satisfy the properties of the cost function, restrictions on parameters are required such that $\beta_{mn}^{yy} = \beta_{nm}^{yy} \; \forall m \neq n$ and $\beta_{jk}^{ww} = \beta_{kj}^{ww} \; \forall j \neq k$ for symmetry, $\sum_{j=1}^{J} \beta_j^w = 1$, $\sum_{k=1}^{J} \beta_{jk}^{ww} = 0 \; \forall j$, and $\sum_{j=1}^{J} \beta_{mj}^{yw} = 0 \; \forall m$ for linear homogeneity in w_i. Because of these restrictions, $\sum_{j=1}^{J} s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) = 1$ and $\sum_{j=1}^{J} \eta_{ij} = 0 \; \forall i$ are guaranteed.[6]

1.2.2 Modified Model

As noted by Kumbhakar and Lovell (2000), Kumbhakar (1997)'s model treats allocative inefficiency of the translog cost system in a theoretically and econometrically consistent manner and provides a solution to "the Greene Problem." Nevertheless, there is still room for improvement. In Section 1.2.2.1, several limitations of Kumbhakar (1997) are discussed and assumptions of his model are modified. In Section 1.2.2.2, the modified model and distributional assumptions are presented.

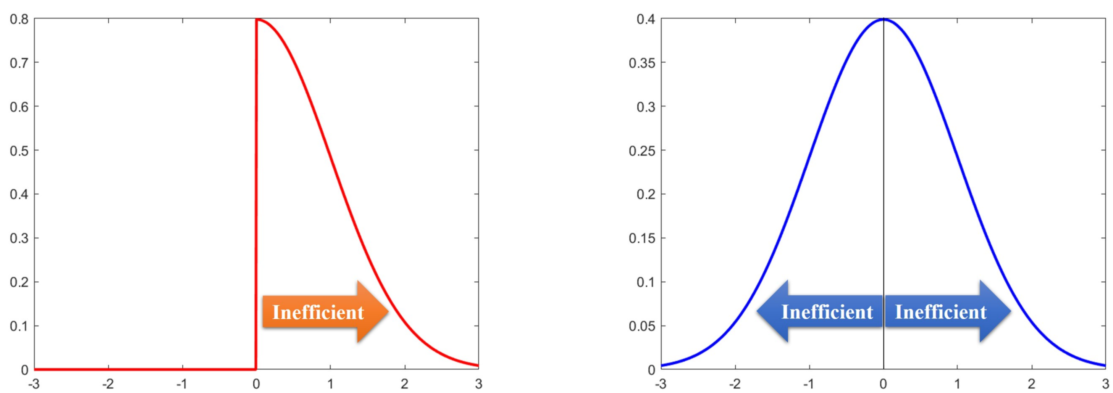
1.2.2.1 Assumptions

First, Kumbhakar (1997)'s model does not capture the relationship between the one-sided term representing technical inefficiency, u_i^T, and the two-sided terms representing allocative inefficiency, ξ_js. It implies that they can be assumed to be independent; this means

[6] Please see to Appendix 1.A.1 for the proof.

technically efficient producers can be allocatively inefficient and vice versa. Rather than imposing such an assumption, it would be more reasonable to assume that they are dependent, which means that technically efficient producers tend to be allocatively efficient, and technically inefficient producers are likely to be allocatively inefficient. However, as Figure 1.2.1 illustrates, the degree of allocative inefficiency does not depend on the value of ξ_j itself but the size of its absolute value instead. Since the degree of technical inefficiency, u_i^T, is non-negative, it is difficult to model the relation between u_i^T and ξ_j. To the best of my knowledge, only two studies, namely Schmidt and Lovell (1980) and Amsler et al. (2021), have modeled the aforementioned relationship between technical and allocative inefficiency.[7] However, both studies consider the single-output cost system assuming Cobb-Douglas production technology, rather than the translog cost system (as in Kumbhakar (1997)) that is most widely used for empirical cost studies and that accommodates multiple-output cases.

Figure 1.2.1: Degree of Technical and Allocative Inefficiency



(a) Technical Inefficiency (u_i^T) (b) Allocative Inefficiency (ξ_j)

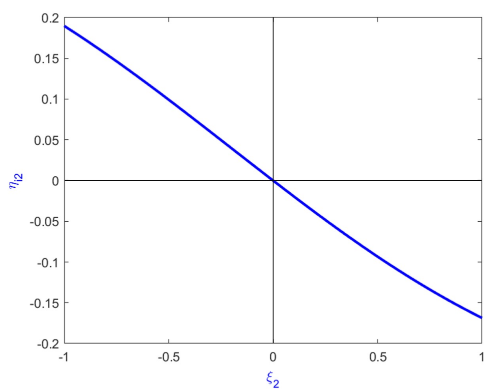
Note: In both figures, the x-axis is the value of u_i^T or ξ_j, and the y-axis is its density value.

Second, Kumbhakar (1997)'s model imposes a restrictive assumption that the magnitudes of allocative inefficiency, ξ_js, are invariant across producers. This implies that it only considers systematic tendency for over- or under-utilization of any input relative to any other

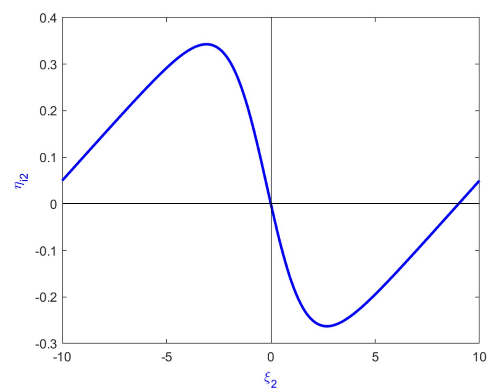
[7]Following the notations in this paper, they assume that u_i^T is positively correlated with $\boldsymbol{\eta}_i = (\eta_{i2}, \cdots, \eta_{iJ})$ since terms like ξ_js in Kumbhakar (1997) are not introduced to their model. Given that there are two terms that can be interpreted as allocative inefficiency in the cost share equations, ξ_js and $\boldsymbol{\eta}_i$, the latter part of this section examines which term will be linked to u_i^T.

inputs. Despite the assumption on ξ_js, their impacts on cost, u_i^A, and on input shares, $\boldsymbol{\eta}_i$, are different across producers, as they are influenced by outputs, y_{im}, and input prices, w_{ij}, by construction. As stated in Kumbhakar and Lovell (2000), the model becomes extremely difficult to estimate without this assumption. To be specific, if the magnitudes of allocative inefficiency are assumed to be random, denoted by $\boldsymbol{\xi}_i = (\xi_{i2}, \cdots, \xi_{iJ})$, it is hard to derive the distribution of $\boldsymbol{\eta}_i$ from the distribution of $\boldsymbol{\xi}_i$. This is because, although they can be one-to-one in the narrow domain, $\boldsymbol{\eta}_i$ is a nonlinear function of $\boldsymbol{\xi}_i$ that is not globally invertible. For example, Figure 1.2.2 shows the relationship between ξ_2 and η_{i2} given parameter values when a firm produces one output using two inputs.[8] Therefore, the change of variables theorem cannot be directly applied. However, this assumption needs to be relaxed to incorporate idiosyncratic deviations from the cost minimizing input ratios. In addition, relaxing this assumption enables researchers to rigorously model the relationship between the terms representing technical and allocative inefficiency.

Figure 1.2.2: Relationship between ξ_2 and η_{i2}



(a) $\xi_2 \in [-1, 1]$



(b) $\xi_2 \in [-10, 10]$

Lastly, stochastic components are not included in the cost share equations. Kumbhakar (1997) analytically derives the optimum cost share, $s_j(\boldsymbol{y}_i, \boldsymbol{w}_i)$, and $\boldsymbol{\eta}_i$ from the use of Shephard's lemma, thereby $\boldsymbol{\eta}_i$ represents pure allocative inefficiency[9] stemmed from the opti-

[8] Although Kumbhakar (1997) does not provide the result, u_i^A and η_{ij} can be simplified when no additive error terms in the cost share equations is assumed. Please refer to Appendix 1.A.2 for the proof. Figures are drawn by the simplified formula when $s_{i2} = 0.55, \beta_{22}^{ww} = 0.05$.

[9] The simplified formula for $\boldsymbol{\eta}_i$ derived in Appendix 1.A.2 clearly shows that $\boldsymbol{\eta}_i$ is a function of ξ_js given the actual cost shares and parameters.

mization error. Moreover, $\boldsymbol{\eta}_i$ is derived based on the assumption that ξ_js do not vary across producers, so it would be unnatural to interpret $\boldsymbol{\eta}_i$ as a stochastic component of the cost share equations. If the assumption about ξ_js is relaxed as described in the previous paragraph, the term representing allocative inefficiency in the cost share equations can be seen as a stochastic component. However, as Reiss and Wolak (2007) point out[10], there are other sources of random components in the cost share equations besides failure in cost minimization. Furthermore, as noted by Brown and Walker (1995), one can easily make distributional assumptions of the system and apply usual estimators by using additive noise terms for the share equations.

The second and third points are related to the issue on the stochastic specification in the models of producers' demand, cost, and production systems. There are contradictory views on how to formulate a stochastic specification for the cost system. The conventional practice is to append additive noise terms to the nonstochastic cost share equations. For example, Christensen and Greene (1976) state that "since the cost share equations are derived by differentiation, they do not contain the disturbance term from the cost function" (p.662), so they add stochastic noise terms following multivariate normal distribution to the cost share equations in an ad hoc fashion. Subsequent research criticizes that such an approach is inconsistent with economic theory and derives stochastic components in the cost share equations in the optimization framework. These studies include Chavas and Segerson (1987), McElroy (1987), Brown and Walker (1995), and Kumbhakar and Tsionas (2011). Note that although they provide theoretical justifications for the stochastic specification of the cost share equations, the sources of stochastic components in the cost share equations vary across studies, such as random environments, measurement errors, and optimization errors.

In order to deal with these issues, I modify the following assumptions to Kumbhakar (1997)'s model.

[10]"The four principal ways in which a researcher can introduce stochastic components into a deterministic economic model are: 1. researcher uncertainty about the economic environment; 2. agent uncertainty about the economic environment; 3. optimization errors on the part of economic agents; and 4. measurement errors in observed variables." (p.4305)

Assumption 1. *The magnitudes of allocative inefficiency vary across producers and are denoted by* $\xi_{i2}, \cdots, \xi_{iJ}$.

Assumption 2. u_i^T *is uncorrelated with* $\xi_{i2}, \cdots, \xi_{iJ}$ *but positively correlated with* $|\xi_{i2}|, \cdots, |\xi_{iJ}|$.

Assumption 3. *Additive noise terms, denoted by* $\boldsymbol{\nu}_i = (\nu_{i2}, \cdots, \nu_{iJ})$, *are allowed in the cost share equations.*

As mentioned, Assumptions 1 and 2 are made to introduce the idiosyncratic disturbance due to optimization errors, as well as to precisely model dependence between technical and allocative inefficiency. Kumbhakar (1997)'s model includes several terms induced by allocative inefficiency, including ξ_js and $\boldsymbol{\eta}_i$. Thus, instead of modifying the assumption on the magnitudes of allocative inefficiency, we can assume that u_i^T is uncorrelated with $\eta_{i2}, \cdots, \eta_{iJ}$ but positively correlated with $|\eta_{i2}|, \cdots, |\eta_{iJ}|$. In this case, ξ_js are not used for estimation. This approach is similar to Schmidt and Lovell (1980) and Amsler et al. (2021), but there are mainly two reasons for imposing Assumption 1 other than ξ_j or ξ_{ij} being the origin of allocative inefficiency in the model.

First of all, the alternative method does not provide a solution to "the Greene Problem." The specification of the two previous studies, Schmidt and Lovell (1980) and Amsler et al. (2021), follows Schmidt and Lovell (1979) as

$$\begin{aligned} y_i &= \alpha + \boldsymbol{x}_i'\beta + v_i - u_i^T \\ x_{i1} - x_{ij} &= \ln\left(\frac{\beta_1 w_{ij}}{\beta_j w_{i1}}\right) + e_{ij},\ j = 2, ..., J, \end{aligned}$$

where y_i is the natural log of output of producer i, $\boldsymbol{x}_i$ is a vector of natural log of inputs, $v_i \in \mathbb{R}$ is a random distrubance, $u_i^T \in \mathbb{R}_+$ represents technical inefficiency, $w_{ij} \in \mathbb{R}_{++}$ is the price of input j, and $e_{ij} \in \mathbb{R}$ is a two-sided term capturing allocative inefficiency and noise. This model does not include an additional term representing allocative inefficiency in the production frontier, therefore issues like "the Greene Problem," which occur in the translog cost system, are not raised. However, an analytic solution, which is proposed by Kumbhakar (1997) and Kumbhakar and Tsionas (2005a,b) that introduce ξ_js and make a distributional assumption on them, is not applicable if ξ_js are not used when estimating the translog cost system. In addition, even though both u_i^A and $\boldsymbol{\eta}_i$ are functions of ξ_js, no closed-form expression for u_i^A in terms of $\boldsymbol{\eta}_i$ exists in the translog cost system. In other words, u_i^A cannot be specified as a function of $\boldsymbol{\eta}_i$. Thus, a method similar to approximation solutions proposed by Schmidt (1984) that specify a functional relationship as $u_i^A = \boldsymbol{e}_i' A \boldsymbol{e}_i$, where A is a positive semi-definite matrix, cannot be applied as well.

Secondly, it should be noted that $\boldsymbol{\eta}_i$ might not correctly measure each firm's degree of allocative inefficiency in the model, although it arises from the fact that a producer uses inputs in an incorrect proportion. Figure 1.2.2(b) shows that the absolute value of the error term in the cost share equation, $|\eta_{i2}|$, can decrease as the size of $|\xi_2|$ increases. This indicates that a producer using inputs fairly inefficiently can have the same cost shares to a producer that allocates inputs efficiently. In addition, it is not guaranteed that $\xi_j = 0$ implies $\eta_{ij} = 0$ by construction, which implies that the input j's actual share deviates from its optimum share even if the input pair $(j, 1)$ is efficiently allocated. For instance, suppose that $J = 3$, $\xi_2 = 0$, but $\xi_3 \neq 0$. Then, although the input pair (2,1) is efficiently allocated, $\eta_{i2} \neq 0$ because of the presence of inefficiency among the input pair (3,1).

Consequently, the magnitudes of allocative inefficiency are allowed to vary and linked to u_i^T, which represents the magnitudes of technical inefficiency, in order to refine Kumbhakar (1997)'s model and to provide a solution to "the Greene Problem." I further discuss Assumption 2, which is about how to model the dependence between u_i^T and ξ_{ij}s. Since the terms representing technical and allocative inefficiency are introduced to the model without any theoretical linkages, u_i^T is assumed to be uncorrelated with $\xi_{i2}, \cdots, \xi_{iJ}$. However, as illustrated in Figure 1.2.1, a producer becomes technically inefficient as the size of u_i^T, which is non-negative, increases and becomes allocatively inefficient as the size of $|\xi_{ij}|$ increases. Hence, u_i^T is assumed to be positively correlated with $|\xi_{i2}|, \cdots, |\xi_{iJ}|$.

Assumption 3 is imposed for two purposes. The first purpose is to capture not only optimization errors stemmed from ξ_{ij}s in the model but also sources of randomness that are not explicitly modeled. For example, Kumbhakar and Tsionas (2005a) use the same specification to take account of measurement errors, and agent and/or researcher uncertainty.[11] The second purpose is to facilitate estimation. As pointed out, it is difficult to derive the distribution of $\boldsymbol{\eta}_i$ from ξ_{ij}s. Moreover, although the assumption on the magnitudes of allocative ineffciency is modified from Kumbhakar (1997)'s model, the cost share equations can be seen as deterministic because η_{ij} is a function of ξ_{ij}s. By appending additive noise, the system is converted to a stochastic model so that one can readily obtain a joint density for estimation. The first and second purposes are related. Reiss and Wolak (2007), for instance, state that one can simply transform a deterministic economic model into an econometric model and justify applying usual estimators by introducing measurement errors.

1.2.2.2 Modified Model and Distributional Assumptions

Because of Assumptions 1 and 3, the stochastic cost frontier model needs to be modified. Although the assumption about the magnitudes of allocative inefficiency has changed from Kumbhakar (1997)'s model, the formula for each component of the system can be identically derived. However, since $\sum_{j=1}^{J} s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) = 1$ and $\sum_{j=1}^{J} \eta_{ij} = 0\ \forall i$, an additional restriction on the sum of ν_{ij} is required so that $\sum_{j=1}^{J} s_{ij} = 1\ \forall i$ is guaranteed.

The modified model can be written as

$$
\begin{aligned}
\ln C_i &= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + \epsilon_i \\
&= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + v_i + u_i \\
&= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + v_i + u_i^T + u_i^A \\
&= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + v_i + u_i^T + g(\boldsymbol{\xi}_i)
\end{aligned}
$$

[11]"These error terms represent measurement error and/or factors that are not under the control of the firm, so they are not modeled explicitly, unlike the $\boldsymbol{\xi}$'s. Alternatively, these errors might not be relevant for the producer (in the sense that they are known to the producer), but nonetheless must be taken into account by the researcher (who does not know them)." (p.739)

$$
\begin{aligned}
s_{ij} &= s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + e_{ij} \\
&= s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + \eta_{ij} + \nu_{ij} \\
&= s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + h_j(\boldsymbol{\xi}_i) + \nu_{ij}, \ j = 2, \cdots, J. \qquad (1.2.1)
\end{aligned}
$$

Recall that C_i is the actual cost of producer i, $C(\boldsymbol{y}_i, \boldsymbol{w}_i)$ is the deterministic kernel of the stochastic cost frontier, $\boldsymbol{y}_i \in \mathbb{R}_+^M$ is a vector of M outputs, $\boldsymbol{w}_i \in \mathbb{R}_{++}^J$ is a vector of input prices, $v_i \in \mathbb{R}$ is a random disturbance, $u_i^T \in \mathbb{R}_+$ represents a cost increase due to technical inefficiency, $u_i^A = g(\boldsymbol{\xi}_i) \in \mathbb{R}_+$ represents a cost increase due to allocative inefficiency, $\boldsymbol{\xi}_i = (\xi_{i2}, \cdots, \xi_{iJ})$, $\xi_{ij}, \ j = 2, \cdots, J$, represents allocative inefficiency for the input pair $(j, 1)$, $u_i = u_i^T + u_i^A$, $\epsilon_i = v_i + u_i$, $s_{ij} \in [0, 1]$ is the actual cost share of input j, $s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) \in [0, 1]$ is the optimum cost share of input j, $\eta_{ij} = h_j(\boldsymbol{\xi}_i) \in \mathbb{R}$ is the disturbance due to allocative inefficiency, $\nu_{ij} \in \mathbb{R}$ is additive noise, and $e_{ij} = \eta_{ij} + \nu_{ij}$ is the disturbance due to both allocative inefficiency and noise. Each component of the system can be written as

$$
\begin{aligned}
\ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) &= \beta_0 + \sum_{m=1}^{M} \beta_m^y (\ln y_{im}) + \frac{1}{2} \sum_{m=1}^{M} \sum_{n=1}^{M} \beta_{mn}^{yy} (\ln y_{im})(\ln y_{in}) \\
&\quad + \sum_{j=1}^{J} \beta_j^w (\ln w_{ij}) + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk}^{ww} (\ln w_{ij})(\ln w_{ik}) \\
&\quad + \sum_{m=1}^{M} \sum_{j=1}^{J} \beta_{mj}^{yw} (\ln y_{im})(\ln w_{ij}) \\
s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) &= \beta_j^w + \sum_{k=1}^{J} \beta_{jk}^{ww} (\ln w_{ik}) + \sum_{m=1}^{M} \beta_{mj}^{yw} (\ln y_{im}), \ j = 2, \cdots, J \\
u_i^A = g(\boldsymbol{\xi}_i) &= \sum_{j=1}^{J} \beta_j^w \xi_{ij} + \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk}^{ww} (\ln w_{ij}) \xi_{ik} + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk}^{ww} \xi_{ij} \xi_{ik} \\
&\quad + \sum_{m=1}^{M} \sum_{j=1}^{J} \beta_{mj}^{yw} (\ln y_{im}) \xi_{ij} + \ln \sum_{j=1}^{J} (s_{ij}^* / e^{\xi_{ij}}) \\
\eta_{ij} = h_j(\boldsymbol{\xi}_i) &= \frac{s_j(\boldsymbol{y}_i, \boldsymbol{w}_i)[1 - \{\sum_{k=1}^{J} (s_{ik}^* / e^{\xi_{ik}})\} e^{\xi_{ij}}] + \sum_{k=1}^{J} \beta_{jk}^{ww} \xi_{ik}}{\{\sum_{k=1}^{J} (s_{ik}^* / e^{\xi_{ik}})\} e^{\xi_{ij}}}, \ j = 2, \cdots, J, \qquad (1.2.2)
\end{aligned}
$$

where $s_{ij}^* = s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + \sum_{k=1}^{J} \beta_{jk}^{ww} \xi_{ik}, \ j = 2, \cdots, J$, is the shadow cost share of input j for producer i who is assumed to be only allocatively inefficient. In addition, $\sum_{j=1}^{J} \nu_{ij} = 0 \ \forall i$

is required to guarantee $\sum_{j=1}^{J} s_{ij} = 1 \ \forall i$. Restrictions on parameters are also necessary in order to satisfy the properties of the cost function such that $\beta_{mn}^{yy} = \beta_{nm}^{yy} \ \forall m \neq n$ and $\beta_{jk}^{ww} = \beta_{kj}^{ww} \ \forall j \neq k$ for symmetry, $\sum_{j=1}^{J} \beta_{j}^{w} = 1$, $\sum_{k=1}^{J} \beta_{jk}^{ww} = 0 \ \forall j$, and $\sum_{j=1}^{J} \beta_{mj}^{yw} = 0 \ \forall m$ for linear homogeneity in w_i.

Regarding the distributional assumptions on the stochastic components of the system, I follow a standard practice, such as in Christensen and Greene (1976), Schmidt and Lovell (1979), and Kumbhakar and Tsionas (2005a,b). Assume that v_i is distributed as $N(0, \sigma_v^2)$, u_i^T is distributed as $|N(0, \sigma_T^2)|$, $\boldsymbol{\xi}_i$ is distributed as $N(0, \boldsymbol{\Sigma}_\xi)$, and $\boldsymbol{\nu}_i$ is distributed as $N(0, \boldsymbol{\Sigma}_\nu)$. For $j = 2, \cdots, J$, assume that u_i^T and ξ_{ij} are distributed independently of v_i and ν_{ij}, and v_i and ν_{ij} are mutually independent.

1.3 Estimation Strategy

1.3.1 Relationship between u_i^T and $\boldsymbol{\xi}_i$: APS Copulas

In Section 1.2.2.1, Assumption 2, which is about the dependence between u_i^T and $\boldsymbol{\xi}_i$, is imposed in order to have the desirable attributes between technical and allocative inefficiency. Based on this assumption, it is required to derive the joint density of u_i^T and $\boldsymbol{\xi}_i$ to estimate the model. One way to obtain the joint distribution of dependent random variables is applying copulas; that is, given specific marginal distributions of u_i^T and ξ_{ij}s, the joint distribution of them can be obtained by employing copulas that capture the dependence.

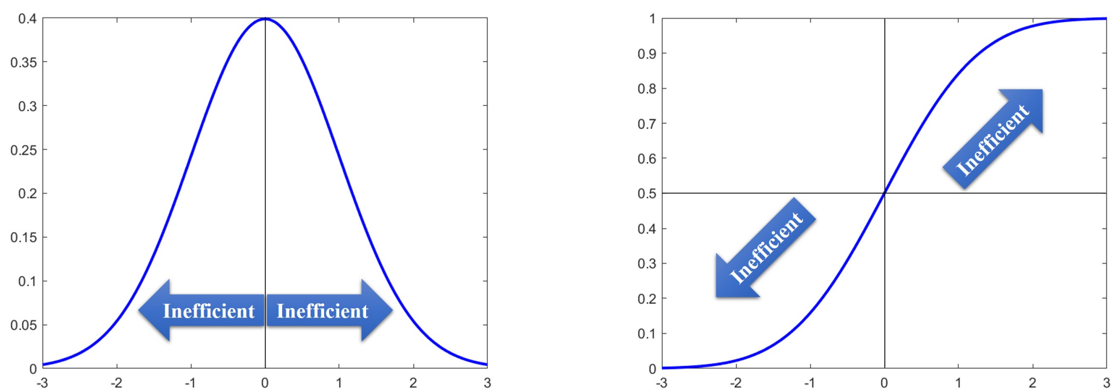
Sklar's theorem states that for every joint cumulative distribution function of random variables $X_1, \cdots, X_J$ with margins $F_1(\cdot), \cdots, F_J(\cdot)$, which are marginal cumulative distribution functions of $X_1, \cdots, X_J$, there exists a copula $C : [0,1]^J \to [0,1]$, which is a cumulative distribution function, such that

$$F(x_1, \cdots, x_J) = C\big(F_1(x_1), \cdots, F_J(x_J)\big)$$

for all $x_i \in \mathbb{R}, \ i = 1, \cdots, J$, where $F(\cdot)$ is a joint cumulative distribution function.

Let $\omega_1 = F_1(u^T)$, $\omega_2 = F_2(\xi_2)$, $\cdots$, $\omega_J = F_J(\xi_J)$, where $u^T, \xi_2, \cdots, \xi_J$ are dummy arguments. In order to have the desired properties between technical and allocative inefficiency, it is required that ω_1 is linked to $\omega_2, \cdots, \omega_J$, for which ω_1 is uncorrelated with $\omega_2, \cdots, \omega_J$ but correlated with $|\omega_2 - 0.5|, \cdots, |\omega_J - 0.5|$. To be specific, ω_1 is uncorrelated with $\omega_2, \cdots, \omega_J$ so that u_i^T is uncorrelated with $\xi_{i2}, \cdots, \xi_{iJ}$. However, as Figure 1.3.1 shows, if we assume that ξ_{ij} is distributed symmetric around zero, like $\xi_{ij} \sim N(0, \sigma_{\xi_j})$, a producer becomes allocatively inefficient when $\omega_j,\ j = 2, \cdots, J$, moves away from 0.5. Therefore, ω_1 needs to be positively correlated with $|\omega_2 - 0.5|, \cdots, |\omega_J - 0.5|$.

Figure 1.3.1: PDF and CDF of ξ_{ij}



(a) PDF of ξ_{ij} (b) CDF of ξ_{ij}

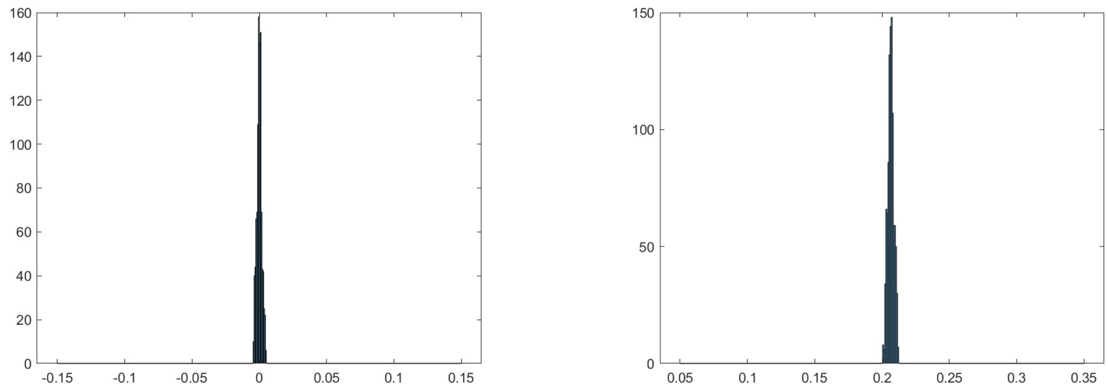
Amsler et al. (2021) propose a new family of copulas (hereafter the APS copulas) that can induce the desired attributes between technical and allocative inefficiency. For example, suppose that two inputs are used for production ($J = 2$). Then, the APS-2 copulas can be applied to capture dependence between u_i^T and ξ_{i2}. The APS-2 copula densities are defined as follows (Amsler et al., 2021, p.4):

$$\text{APS-2-A} : c_{12}(\omega_1, \omega_2) = 1 + \theta_{12}(1 - 2\omega_1)\{1 - 12(\omega_2 - 0.5)^2)\},\ |\theta_{12}| \leq 0.5$$

$$\text{APS-2-B} : c_{12}(\omega_1, \omega_2) = 1 + \theta_{12}(1 - 2\omega_1)(1 - 4|\omega_2 - 0.5|),\ |\theta_{12}| \leq 1,$$

where $c_{12}(\omega_1, \omega_2)$ is the APS-2 copula density, and θ_{12} is the association parameter. Then, $\text{cov}(\omega_1, \omega_2) = 0$, $\text{corr}(\omega_1, (\omega_2 - 0.5)^2) = \frac{2}{\sqrt{15}}\theta$ for the APS-2-A copula, and $\text{corr}(\omega_1, |\omega_2 - 0.5|) = \frac{1}{3}\theta$ for the APS-2-B copula (Amsler et al., 2021, p.3-4); that is ω_1 is uncorrelated with ω_2, but it can be correlated with $|\omega_2 - 0.5|$. Given this, u_i^T and ξ_{i2} can have the desired properties. Figure 1.3.2 illustrates the sample correlation between ω_1, ω_2, and $(\omega_2 - 0.5)^2$ when $\theta_{12} = 0.4$ derived from simulations[12] using the APS-2-A copula. It shows that $\text{corr}(\omega_1, \omega_2) \approx 0$ and $\text{corr}(\omega_1, (\omega_2 - 0.5)^2) \approx 0.2066$ as the theoretical results.

Figure 1.3.2: Sample Correlations (The APS-2-A Copula)



(a) $\text{corr}(\omega_1, \omega_2)$ (b) $\text{corr}(\omega_1, (\omega_2 - 0.5)^2)$

Amsler et al. (2021) also develop a method that the APS-2 copulas can be extended to more dimensions, which is necessary when more than two inputs are used for production. If $J = 3$, for instance, the APS-3 copulas can be applied to capture dependence between u_i^T, ξ_{i2}, and ξ_{i3}. To be specific, ω_2 and ω_3 are allowed to follow any standard bivariate copula but need to be linked to ω_1 as in the APS-2 copulas. Assume that ω_2 and ω_3 follow the bivariate Gaussian copula. Then, the APS-3 copula densities are defined as follows (Amsler et al., 2021, p.6):

$$\text{APS-3-A} : c_{123}(\omega_1, \omega_2, \omega_3) = 1 + (c_{12} - 1) + (c_{13} - 1) + (c_{23} - 1),$$

where

$$c_{12} = c_{12}(\omega_1, \omega_2) = 1 + \theta_{12}(1 - 2\omega_1)\{1 - 12(\omega_2 - 0.5)^2)\},\ |\theta_{12}| \leq 0.5$$
$$c_{13} = c_{13}(\omega_1, \omega_3) = 1 + \theta_{13}(1 - 2\omega_1)\{1 - 12(\omega_3 - 0.5)^2)\},\ |\theta_{13}| \leq 0.5$$
$$c_{23} = c_{23}(\omega_2, \omega_3) = \frac{1}{\sqrt{1 - \rho^2}} \exp\Big[- \frac{\rho^2 \Phi^{-1}(\omega_2)^2 - 2\rho\Phi^{-1}(\omega_2)\Phi^{-1}(\omega_3) + \rho^2\Phi^{-1}(\omega_3)^2}{2(1 - \rho^2)} \Big].$$

[12] The number of replications is 1,000, where the sample size is 1,000 for each replication.

$$\text{APS-3-B} : c_{123}(\omega_1, \omega_2, \omega_3) = 1 + (c_{12} - 1) + (c_{13} - 1) + (c_{23} - 1),$$

where

$$\begin{aligned}
c_{12} &= c_{12}(\omega_1, \omega_2) = 1 + \theta_{12}(1 - 2\omega_1)(1 - 4|\omega_2 - 0.5|),\ |\theta_{12}| \leq 1 \\
c_{13} &= c_{13}(\omega_1, \omega_3) = 1 + \theta_{13}(1 - 2\omega_1)(1 - 4|\omega_3 - 0.5|),\ |\theta_{13}| \leq 1 \\
c_{23} &= c_{23}(\omega_2, \omega_3) = \frac{1}{\sqrt{1-\rho^2}} \exp\Big[- \frac{\rho^2 \Phi^{-1}(\omega_2)^2 - 2\rho \Phi^{-1}(\omega_2)\Phi^{-1}(\omega_3) + \rho^2 \Phi^{-1}(\omega_3)^2}{2(1-\rho^2)} \Big],
\end{aligned}$$

where $c_{123}(\omega_1, \omega_2, \omega_3)$ is the APS-3 copula density, $c_{12}(\omega_1, \omega_2)$ and $c_{13}(\omega_1, \omega_3)$ are the APS-2 copula densities, $c_{23}(\omega_2, \omega_3)$ is the bivariate Gaussian copula density, θ_{12} and θ_{13} are the association parameters, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and $\rho \in (-1, 1)$ is the correlation parameter. By applying the APS-3 copulas, one can capture dependence such that ω_1 is uncorrelated with ω_2 and ω_3 but correlated with $|\omega_2 - 0.5|$ and $|\omega_3 - 0.5|$ as in the case of $J = 2$. Hence, u_i^T and ξ_{ij}s, $j = 2$ and 3, can have the desired properties. Figure 1.B.1 illustrates the sample correlations between ω_1, ω_2, $(\omega_2 - 0.5)^2$, ω_3, and $(\omega_3 - 0.5)^2$ when $\theta_{12} = \theta_{13} = 0.2$ and $\rho = -0.5$[13] obtained by simulations[14] based on the APS-3-A copula. It shows that $\text{corr}(\omega_1, \omega_2)$ and $\text{corr}(\omega_1, \omega_3)$ are approximately zero when $J = 2$, but $\text{corr}(\omega_1, |\omega_2 - 0.5|)$ and $\text{corr}(\omega_1, |\omega_3 - 0.5|)$ are positive.

1.3.2 Derivation of the Joint Density

By employing the APS copulas, we can derive the joint density of the translog cost system that captures the dependence between technical and allocative inefficiency. For notational convenience, rewrite $X = \epsilon_i = v_i + u_i^T + u_i^A = X_1 + Z_1 + g(\boldsymbol{Z}_2)$, $\boldsymbol{Y} = (Y_2, \cdots, Y_J) = (e_{i2}, \cdots, e_{iJ}) = (\eta_{i2} + \nu_{i2}, \cdots, \eta_{iJ} + \nu_{iJ}) = (h_2(\boldsymbol{Z}_2) + W_2, \cdots, h_J(\boldsymbol{Z}_2) + W_J)$, and $\boldsymbol{Z} = (Z_1, \boldsymbol{Z}_2) = (u_i^T, \xi_{i2}, \cdots, \xi_{iJ})$. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$ be a vector of parameters, where $\boldsymbol{\theta}_1 =$

[13] Amsler et al. (2021) show that the allowable range of θ_{12} and θ_{13} depends on ρ. To be specific, if ω_2 and ω_3 are strongly correlated, the range of $|\theta_{12}| + |\theta_{13}|$ decreases. Please refer to Result 10 of Amsler et al. (2021, p.6) in detail.

[14] The number of replications is 1,000, where the sample size is 1,000 for each replication.

$(\beta_0, \boldsymbol{\beta}^y, \boldsymbol{\beta}^{yy}, \boldsymbol{\beta}^w, \boldsymbol{\beta}^{ww}, \boldsymbol{\beta}^{yw}, \sigma_v^2, \boldsymbol{\Sigma}_\nu)$, $\boldsymbol{\theta}_2 = (\boldsymbol{\theta}^{\text{APS}}, \boldsymbol{\theta}^{\text{Gauss}})$, and $\boldsymbol{\theta}_3 = (\sigma_T^2, \boldsymbol{\Sigma}_\xi)$.[15] $\boldsymbol{\theta}^{\text{APS}}$ and $\boldsymbol{\theta}^{\text{Gauss}}$ denote vectors of association or correlation parameters of the APS and Gaussian copulas, respectively.[16] Then, the translog cost system can be rewritten as

$$\begin{aligned}
\ln C_i &= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + X \\
&= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + X_1 + Z_1 + g(\boldsymbol{Z}_2) \\
s_{ij} &= s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + Y_j \\
&= s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + h_j(\boldsymbol{Z}_2) + W_j, \ \ j = 2, \cdots, J.
\end{aligned} \tag{1.3.1}$$

It is required to derive the joint density of $X = \epsilon_i$ and $\boldsymbol{Y} = (Y_2, \cdots, Y_J) = (e_{i2}, \cdots, e_{iJ})$ in order to form a likelihood function of the translog cost system. First, the joint density of X, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ can be written as

$$\begin{aligned}
f_{X,\boldsymbol{Y},\boldsymbol{Z}}(x, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta}) &= f_{X,\boldsymbol{Y}|\boldsymbol{Z}}(x, \boldsymbol{y}|\boldsymbol{z}; \boldsymbol{\theta}) \cdot f_{\boldsymbol{Z}}(\boldsymbol{z}; \boldsymbol{\theta}_2, \boldsymbol{\theta}_3) \\
&= f_{X|\boldsymbol{Y},\boldsymbol{Z}}(x|\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta}) \cdot f_{\boldsymbol{Y}|\boldsymbol{Z}}(\boldsymbol{y}|\boldsymbol{z}; \boldsymbol{\theta}) \cdot f_{\boldsymbol{Z}}(\boldsymbol{z}; \boldsymbol{\theta}_2, \boldsymbol{\theta}_3).
\end{aligned}$$

Then, the joint density of X and $\boldsymbol{Y}$ can be obtained by integrating out $\boldsymbol{Z}$ such that

$$\begin{aligned}
f_{X,\boldsymbol{Y}}(x, \boldsymbol{y}; \boldsymbol{\theta}) &= \int_{\mathbb{R}_+ \times \mathbb{R}^{J-1}} f_{X,\boldsymbol{Y},\boldsymbol{Z}}(x, \boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta}) d\boldsymbol{z} \\
&= \int_{\mathbb{R}_+ \times \mathbb{R}^{J-1}} f_{X|\boldsymbol{Y},\boldsymbol{Z}}(x|\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta}) \cdot f_{\boldsymbol{Y}|\boldsymbol{Z}}(\boldsymbol{y}|\boldsymbol{z}; \boldsymbol{\theta}) \cdot f_{\boldsymbol{Z}}(\boldsymbol{z}; \boldsymbol{\theta}_2, \boldsymbol{\theta}_3) d\boldsymbol{z}.
\end{aligned} \tag{1.3.2}$$

Suppose that the distribution of $\boldsymbol{Z}$ is known and simple. Then, the joint density of $\boldsymbol{X}$ and $\boldsymbol{Y}$ (1.3.2) can be approximated by drawing random numbers from the density of $\boldsymbol{Z}$. However, $\boldsymbol{\theta}_2 = (\boldsymbol{\theta}^{\text{APS}}, \boldsymbol{\theta}^{\text{Gauss}})$, which governs the joint distribution of $\boldsymbol{Z}$, should be estimated. This implies that we need to transform the random vector $\boldsymbol{Z}$ to another random vector in order to estimate $\boldsymbol{\theta}_2$. In addition, a transformed random vector needs to be simple to make the process of drawing random numbers easy.[17]

[15] Notation in bold represents row vectors whose elements are corresponding parameters of the system. For instance, $\boldsymbol{\beta}^y = (\beta_1^y, \beta_2^y, \cdots)$ and $\boldsymbol{\beta}^{yy} = (\beta_{11}^{yy}, \beta_{12}^{yy}, \cdots, \beta_{21}^{yy}, \beta_{22}^{yy}, \cdots)$.

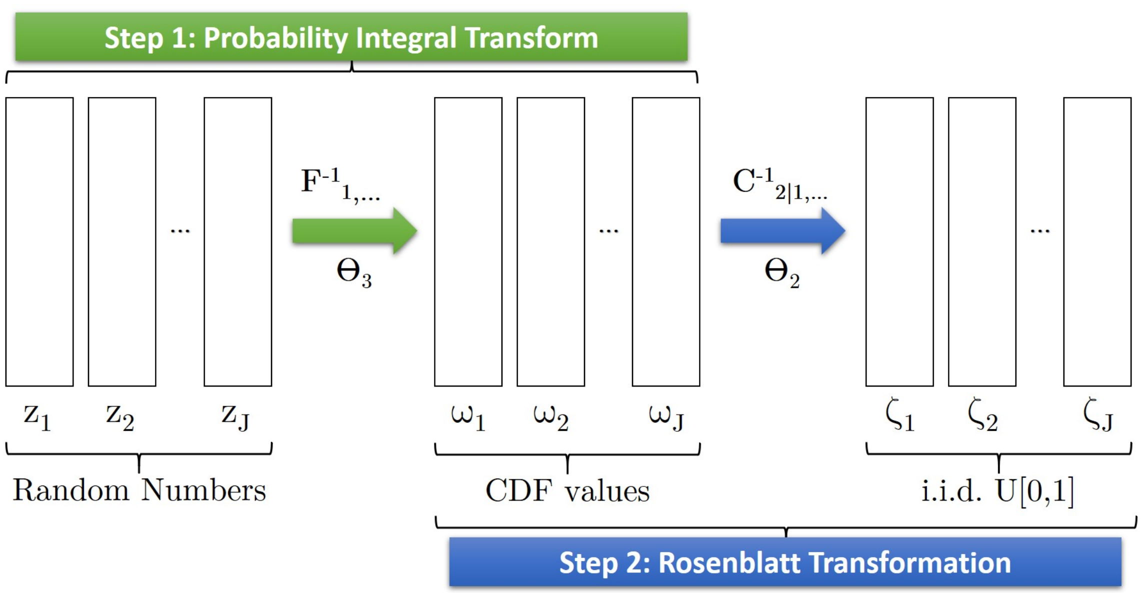
[16] For example, if $J = 3$, $\boldsymbol{\theta}^{\text{APS}} = (\theta_{12}, \theta_{13})$ and $\boldsymbol{\theta}^{\text{Gauss}} = \rho$.

[17] "If the researcher wants to take a draw from a standard normal density (that is, a normal with zero mean and unit variance) or a standard uniform density (uniform between 0 and 1), the process from a programming perspective is very easy."(Train, 2009, p.205-206)

As pointed out in Section 1.2.2.1, it is difficult to apply the change of variables theorem into the Kumbhakar (1997)'s model when the magnitudes of allocative inefficiency are assumed to be random. However, the theorem can be used to derive the joint density of the translog cost system under the assumptions made in Section 1.2.2.1 by the following procedure. It is established on the probability integral transform and the copula-based version of Rosenblatt transformation (see Rosenblatt, 1952) that are monotone.

Rosenblatt (1952) proposes a method using conditional cumulative distribution functions for transforming a dependent random vector to the independent random vector whose components are uniformly distributed on $[0,1]$. Appendix 1.A.3 provides detailed explanations for the Rosenblatt transformation. Based on the both transformations, $\boldsymbol{Z}$ that has dependent components can be replaced with functions of the independent ramdom vector $\boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_J)$, where $\zeta_j \sim U[0,1]$, $j = 1, \cdots, J$ are uniformly and independently distributed over $[0,1]$. The procedure for a change of variables using the probability integral transform and the Rosenblatt transformation is illustrated in Figure 1.3.3.

Figure 1.3.3: Procedure for a Change of Variables



The first step is to replace $\boldsymbol{Z}$ with functions of $\boldsymbol{\omega}$ using the probability integral transform. Given that $\omega_j = F(z_j; \boldsymbol{\theta_3})$ for $j = 1, \cdots, J$, let $J_{\boldsymbol{Z}}$ be the Jacobian matrix whose $(i,j)^{\text{th}}$

element is $\frac{\partial z_i}{\partial \omega_j} = \frac{\partial F_i^{-1}(\omega_i;\theta_3)}{\partial \omega_j}$. Then, by applying the change of variables theorem, the joint density of X and $\boldsymbol{Y}$ can be written as

$$
\begin{aligned}
& f_{X,\boldsymbol{Y}}(x,\boldsymbol{y};\boldsymbol{\theta}) \\
&= \int_{[0,1]^J} f_{X|\boldsymbol{Y},\boldsymbol{Z}}\big(x|\boldsymbol{y},\boldsymbol{z}(\boldsymbol{\omega});\boldsymbol{\theta}\big) \cdot f_{\boldsymbol{Y}|\boldsymbol{Z}}\big(\boldsymbol{y}|\boldsymbol{z}(\boldsymbol{\omega});\boldsymbol{\theta}\big) \cdot f_{\boldsymbol{Z}}\big(\boldsymbol{z}(\boldsymbol{\omega});\boldsymbol{\theta}_2,\boldsymbol{\theta}_3\big) \cdot |J_{\boldsymbol{Z}}| d\boldsymbol{\omega} \\
&= \int_{[0,1]^J} f_{X|\boldsymbol{Y},\boldsymbol{Z}}\big(x|\boldsymbol{y},\boldsymbol{z}(\boldsymbol{\omega});\boldsymbol{\theta}\big) \cdot f_{\boldsymbol{Y}|\boldsymbol{Z}}\big(\boldsymbol{y}|\boldsymbol{z}(\boldsymbol{\omega});\boldsymbol{\theta}\big) \cdot c(\omega_1,\cdots,\omega_J) \cdot \prod_{j=1}^{J} f_j\big(z_j(\omega_j)\big) \cdot |J_{\boldsymbol{Z}}| d\boldsymbol{\omega} \\
&= \int_{[0,1]^J} f_{X|\boldsymbol{Y},\boldsymbol{Z}}\big(x|\boldsymbol{y},\boldsymbol{z}(\boldsymbol{\omega});\boldsymbol{\theta}\big) \cdot f_{\boldsymbol{Y}|\boldsymbol{Z}}\big(\boldsymbol{y}|\boldsymbol{z}(\boldsymbol{\omega});\boldsymbol{\theta}\big) \cdot c(\omega_1,\cdots,\omega_J) d\boldsymbol{\omega},
\end{aligned}
$$

where $\boldsymbol{\omega} = (\omega_1,\cdots,\omega_J)$. The second equality holds as a joint density of dependent random variables equals the product of the copula density $c(\omega_1,\cdots,\omega_J)$ and marginal densities $f_j\big(z_j(\omega_j)\big)$, $j = 2,\cdots,J$, for each random variable. The third equality holds as

$$
J_{\boldsymbol{Z}} = \begin{pmatrix} \frac{\partial z_1}{\partial \omega_1} & \frac{\partial z_1}{\partial \omega_2} & \cdots & \frac{\partial z_1}{\partial \omega_J} \\ \frac{\partial z_2}{\partial \omega_1} & \frac{\partial z_2}{\partial \omega_2} & \cdots & \frac{\partial z_2}{\partial \omega_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial z_J}{\partial \omega_1} & \frac{\partial z_J}{\partial \omega_2} & \cdots & \frac{\partial z_J}{\partial \omega_J} \end{pmatrix} = \begin{pmatrix} \frac{1}{f_1(z_1;\boldsymbol{\theta}_3)} & 0 & \cdots & 0 \\ 0 & \frac{1}{f_2(z_2;\boldsymbol{\theta}_3)} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \frac{1}{f_J(z_J;\boldsymbol{\theta}_3)} \end{pmatrix},
$$

thus $|J_{\boldsymbol{Z}}| = \Big[\prod_{j=1}^J f_j\big(z_j;\boldsymbol{\theta}_3\big)\Big]^{-1}$.

The second step is to replace $\boldsymbol{\omega}$ with functions of $\boldsymbol{\zeta}$ employing the Rosenblatt transformation. Considering that a copula C is also a cumulative distribution function, let $T_R : [0,1]^J \to [0,1]^J$ be the Rosenblatt transformation given by

$$
\begin{aligned}
\zeta_1 &= C_{1|2,\cdots,J}(\omega_1|\omega_2,\cdots,\omega_J) \\
&\vdots \\
\zeta_{J-2} &= C_{J-2|J-1,J}(\omega_{J-2}|\omega_{J-1},\omega_J) \\
\zeta_{J-1} &= C_{J-1|J}(\omega_{J-1}|\omega_J) \\
\zeta_J &= C_J(\omega_J).
\end{aligned}
$$

For example, the conditional APS-2-A copula function $C_{1|2}(\omega_1|\omega_2)$ is

$$
C_{1|2}(\omega_1|\omega_2) = \omega_1 + g_2\omega_1(1-\omega_1),
$$

where $g_2 = \theta_{12}\{1-12(\omega_2-0.5)^2\}$. The conditional APS-3-A copula functions $C_{1|23}(\omega_1|\omega_2,\omega_3)$ and $C_{2|3}(\omega_2|\omega_3)$ are

$$C_{1|23}(\omega_1|\omega_2,\omega_3) = \frac{1}{c_{23}}\{h(\omega_1-\omega_1^2)+c_{23}\omega_1\}$$
$$C_{2|3}(\omega_2|\omega_3) = \Phi\Big(\frac{\Phi^{-1}(\omega_2)-\rho\Phi^{-1}(\omega_3)}{\sqrt{1-\rho^2}}\Big),$$

where $c_{23} = c_{23}(\omega_2,\omega_3)$ is the bivariate Gaussian copula density, $h = g_2 + g_3$, where $g_2 = \theta_{12}\{1-12(\omega_2-0.5)^2\}$ and $g_3 = \theta_{13}\{1-12(\omega_3-0.5)^2\}$, Φ is the cumulative distribution function of the standard normal distribution, and ϕ is the probability density function of the standard normal distribution.[18]

Let T be the inverse function of T_R, and J_T denotes the Jacobian matrix whose $(i,j)^{\text{th}}$ element is $\frac{\partial\omega_i}{\partial\zeta_j}$. Then, by applying the change of variables theorem once more, the joint density of X and $\boldsymbol{Y}$ can be written as

$$\begin{aligned}
&f_{X,\boldsymbol{Y}}(x,\boldsymbol{y};\boldsymbol{\theta})\\
&= \int_{[0,1]^J} f_{X|\boldsymbol{Y},\boldsymbol{Z}}\Big(x\Big|\boldsymbol{y},\boldsymbol{z}\big(T(\boldsymbol{\zeta})\big);\boldsymbol{\theta}\Big)\cdot f_{\boldsymbol{Y}|\boldsymbol{Z}}\Big(\boldsymbol{y}\Big|\boldsymbol{z}\big(T(\boldsymbol{\zeta})\big);\boldsymbol{\theta}\Big)\cdot c\big(T_1(\boldsymbol{\zeta}),\cdots,T_J(\boldsymbol{\zeta});\boldsymbol{\theta_2}\big)|J_T|d\boldsymbol{\zeta}\\
&= \int_{[0,1]^J} f_{X|\boldsymbol{Y},\boldsymbol{Z}}\Big(x\Big|\boldsymbol{y},\boldsymbol{z}\big(T(\boldsymbol{\zeta})\big);\boldsymbol{\theta}\Big)\cdot f_{\boldsymbol{Y}|\boldsymbol{Z}}\Big(\boldsymbol{y}\Big|\boldsymbol{z}\big(T(\boldsymbol{\zeta})\big);\boldsymbol{\theta}\Big)d\boldsymbol{\zeta}
\end{aligned}$$

where $T_j(\boldsymbol{\zeta}) = \omega_j,\ j = 1,\cdots,J$. The second equality holds as

$$J_T = \begin{pmatrix} \frac{\partial\omega_1}{\partial\zeta_1} & \frac{\partial\omega_1}{\partial\zeta_2} & \cdots & \frac{\partial\omega_1}{\partial\zeta_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\omega_{J-1}}{\partial\zeta_1} & \frac{\partial\omega_{J-1}}{\partial\zeta_2} & \cdots & \frac{\partial\omega_{J-1}}{\partial\zeta_J} \\ \frac{\partial\omega_J}{\partial\zeta_1} & \frac{\partial\omega_J}{\partial\zeta_2} & \cdots & \frac{\partial\omega_J}{\partial\zeta_J} \end{pmatrix} = \begin{pmatrix} \frac{1}{c_{1|2,\cdots,J}} & \cdots & \frac{1}{\partial C_{1|2,\cdots,J}/\partial\omega_{J-1}} & \frac{1}{\partial C_{1|2,\cdots,J}/\partial\omega_J} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \frac{1}{c_{J-1|J}} & \frac{1}{\partial C_{J-1|J}/\partial\omega_J} \\ 0 & \cdots & 0 & \frac{1}{c_J} \end{pmatrix},$$

where $c_{1|2,\cdots,J},\cdots,c_{J-1|J}$ are conditional copula densities. Therefore, $|J_T| = [c_{1|2,\cdots,J}\times\cdots\times c_{J-1|J}\times c_J]^{-1} = [c(\omega_1,\cdots,\omega_J)]^{-1}$.[19]

[18] Sections 2.A.1 and 2.2 provide derivation of these conditional copula functions.
[19] Arguments are occasionally dropped for brevity.

Note that

$$
\begin{aligned}
&f_{X|\boldsymbol{Y},\boldsymbol{Z}}\Big(x\Big|\boldsymbol{y},\boldsymbol{z}\big(T(\boldsymbol{\zeta})\big);\boldsymbol{\theta}\Big)\\
&=f_{X|\boldsymbol{Y},\boldsymbol{Z}}\bigg(x_1+z_1\big(T(\boldsymbol{\zeta})\big)+g\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big)\bigg|h_2\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big),\cdots,h_J\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big),\boldsymbol{z}\big(T(\boldsymbol{\zeta})\big);\boldsymbol{\theta}\bigg)\\
&=f_{X_1}\bigg(x-z_1\big(T(\boldsymbol{\zeta})\big)-g\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big);\boldsymbol{\theta}\bigg),
\end{aligned}
$$

where $x=\ln C-\ln C(\boldsymbol{y},\boldsymbol{w})$, and

$$
\begin{aligned}
&f_{\boldsymbol{Y}|\boldsymbol{Z}}\Big(\boldsymbol{y}\Big|\boldsymbol{z}\big(T(\boldsymbol{\zeta})\big);\boldsymbol{\theta}\Big)\\
&=f_{\boldsymbol{Y}|\boldsymbol{Z}}\bigg(h_2\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big)+\nu_2,\cdots,h_J\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big)+\nu_J\bigg|\boldsymbol{z}\big(T(\boldsymbol{\zeta})\big);\boldsymbol{\theta}\bigg)\\
&=f_{\nu}\bigg(e_2-h_2\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big),\cdots,e_J-h_J\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big);\boldsymbol{\theta}\bigg),
\end{aligned}
$$

where $e_j=s_j-s_j(\boldsymbol{y},\boldsymbol{w}),\ j=2,\cdots,J.$

Therefore, the joint density of X and $\boldsymbol{Y}$ can be simplified as

$$
\begin{aligned}
&f_{X,\boldsymbol{Y}}(x,\boldsymbol{y};\boldsymbol{\theta})\\
&=\int_{[0,1]^J}f_{X|\boldsymbol{Y},\boldsymbol{Z}}\Big(x\Big|\boldsymbol{y},\boldsymbol{z}\big(T(\boldsymbol{\zeta})\big);\boldsymbol{\theta}\Big)\cdot f_{\boldsymbol{Y}|\boldsymbol{Z}}\Big(\boldsymbol{y}\Big|\boldsymbol{z}\big(T(\boldsymbol{\zeta})\big);\boldsymbol{\theta}\Big)d\boldsymbol{\zeta}\\
&=\int_{[0,1]^J}f_{X_1}\bigg(x-z_1\big(T(\boldsymbol{\zeta})\big)-g\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big);\boldsymbol{\theta}\bigg)\\
&\qquad\qquad\cdot f_{\nu}\bigg(e_2-h_2\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big),\cdots,e_J-h_J\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big)f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})d\boldsymbol{\zeta}\\
&=\mathbb{E}_{\boldsymbol{\zeta}}\bigg[f_{X_1}\bigg(x-z_1\big(T(\boldsymbol{\zeta})\big)-g\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big);\boldsymbol{\theta}\bigg)\\
&\qquad\quad\cdot f_{\nu}\bigg(e_2-h_2\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big),\cdots,e_J-h_J\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big);\boldsymbol{\theta}\bigg)\bigg],
\end{aligned}
\tag{1.3.3}
$$

where $f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})$ is the joint probability density function of $\boldsymbol{\zeta}$, and $\mathbb{E}_{\boldsymbol{\zeta}}$ represents the expectation with respect to the distribution of $\boldsymbol{\zeta}$. The second equality holds as $\boldsymbol{\zeta}$ is the independent random vector such that each component follows a uniform distribution over $[0,1]$, which implies $f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})=1$.

1.3.3 Maximum Simulated Likelihood Estimator

Since the joint density of X and $\boldsymbol{Y}$ involves an intractable integral, simulation-based methods are necessary to compute the joint density. The direct or crude frequency simulator for the joint density $f_{X,\boldsymbol{Y}}(x,\boldsymbol{y};\boldsymbol{\theta})$ can be written as

$$\begin{aligned}
&\hat{f}_{X,\boldsymbol{Y}}(x,\boldsymbol{y};\boldsymbol{\theta})\\
&=\frac{1}{R}\sum_{r=1}^{R}\Bigg\{f_{X_1}\Big(x-z_1\big(T(\boldsymbol{\zeta}^{(r)})\big)-g\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta}^{(r)})\big)\Big);\boldsymbol{\theta}\Big)\\
&\qquad\cdot f_{\nu}\Big(e_2-h_2\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta}^{(r)})\big)\Big),\cdots,e_J-h_J\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta}^{(r)})\big)\Big);\boldsymbol{\theta}\Big)\Bigg\},
\end{aligned}$$

where $\boldsymbol{\zeta}^{(r)}=(\zeta_1^{(r)},\cdots,\zeta_1^{(r)})$ is the independent r^{th} draw of R draws from multivariate standard uniform distribution. The maximum simulated likelihood estimator $\hat{\boldsymbol{\theta}}^{\text{MSL}}$ maximizes the following simulated log likelihood function:

$$\begin{aligned}
&\ln\hat{\mathcal{L}}(\boldsymbol{\theta})\\
&=\sum_{i=1}^{N}\ln\hat{f}_{X,\boldsymbol{Y}}(x,\boldsymbol{y};\boldsymbol{\theta})\\
&=\sum_{i=1}^{N}\ln\Bigg[\frac{1}{R}\sum_{r=1}^{R}\Bigg\{f_{X_1}\Big(x-z_1\big(T(\boldsymbol{\zeta}^{(r)})\big)-g\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta}^{(r)})\big)\Big);\boldsymbol{\theta}\Big)\\
&\qquad\cdot f_{\nu}\Big(e_2-h_2\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta}^{(r)})\big)\Big),\cdots,e_J-h_J\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta}^{(r)})\big)\Big);\boldsymbol{\theta}\Big)\Bigg\}\Bigg].
\end{aligned}\tag{1.3.4}$$

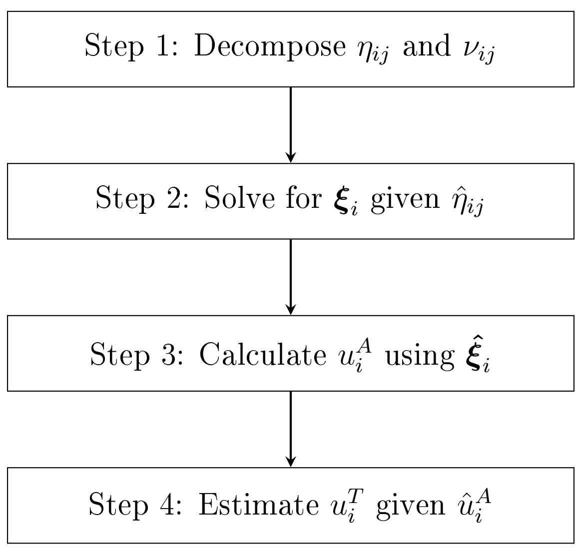
1.4 Firm-level Inefficiency

Researchers can measure and decompose the average inefficiency by estimating the model proposed in previous sections. However, as Jondrow et al. (1982) point out, it is also desirable to estimate inefficiency for each observation to compare (in)efficiency levels across firms, which is the original motivation of Farrell (1957). However, the error term in the cost share equations, e_{ij}, is assumed to contain both the disturbance from the optimal share and the additive noise. Therefore, methods that use conditional distributions to derive firm-level inefficiency developed in previous studies, such as Jondrow et al. (1982) and Battese and

Coelli (1988), cannot be directly employed. In this section, I propose a strategy to measure and decompose firm-level technical and allocative inefficiency for (1.2.1).

The process unfolds in four steps as summarized in Figure 1.4.1. The first step is decomposing the error term in the cost share equations, e_{ij}, into η_{ij} and ν_{ij} to calculate $\hat{\eta}_{ij}$. As a point estimate of $\hat{\eta}_{ij}$, we can use the mode of the conditional distribution $\hat{f}_{\eta|e}(\eta|e)$ as Jondrow et al. (1982), which can be obtained by deconvolution density estimation. The second step is solving for $\boldsymbol{\xi}_i$ given $\hat{\eta}_{ij}$. As $J-1$ cost share equations $\eta_{ij} = h_j(\boldsymbol{\xi}_i)$, $j = 2, \cdots J$, include $J-1$ unknowns, we can find the solution to the system of equations, $\hat{\boldsymbol{\xi}}_i$. The third step is calculating u_i^A using $\hat{\boldsymbol{\xi}}_i$ following (1.2.2). The last step is estimating the conditional expectation of u_i^T given $\hat{u}_i^A$ and $\hat{\epsilon}_i$, where $\hat{u}_i^T = \mathbb{E}[u_i^T|v_i + u_i^T] = \mathbb{E}[u_i^T|\epsilon_i - u_i^A]$ that is in line with Jondrow et al. (1982).

Figure 1.4.1: Process to Measure and Decompose Individual Inefficiency



Details on Step 1 and 4 are as follows. First, since the two terms representing allocative inefficiency, η_{ij} and u_i^A, are functions of $\boldsymbol{\xi}_i$, it is required to estimate $\boldsymbol{\xi}_i$ to measure individual allocative inefficiency, which can be obtained from the system of equations $\eta_{ij} = h_j(\boldsymbol{\xi}_i)$, $j = 2, \cdots, J$. However, because the error term in the cost share equations, e_{ij}, which can be obtained after estimating the model, contains additive noise, ν_{ij}, it is essential to decompose η_{ij} and ν_{ij} that are unobservable. Density deconvolution methods can be used to recover an unknown probability density function that is noise-free, which implies we can recover the density function of η_{ij}. The setup of the density deconvolution problem is as follows. Suppose that one can only observe samples $Y_1, \cdots, Y_n$ given by $Y_i = X_i + U_i,\ i = 1, \cdots, n$, where U_i is noise from a known distribution and independent of X_i. The problem is how to estimate the density function of X, $f_X(x)$, and the conditional density of X given Y, $f_{X|Y}(x|y)$, based on the observations $Y_1, \cdots, Y_n$. For more details about density deconvolution to estimate $f_X(x)$, please refer to Carroll and Hall (1988), Stefanski and Carroll (1990), and Fan (1991). Wang and Ye (2015) propose re-weighted deconvolution kernel methods to estimate the conditional density function $f_{X|Y}(x|y)$ in an additive error model. Their estimator, which is applied to estimate $f_{\eta|e}(\eta|e)$, is

$$\hat{f}_{X|Y}(x|y) = \hat{\tau}_0(x|y) \sum_{j=1}^{n} K_h^*(x - Y_j),$$

where $\hat{\tau}_0(x|y) = \frac{f_U(y-x)}{\sum_{j=1}^{n} L_b(y-Y_j)}$, $L_b(\cdot) = L(\cdot/b)/b$, $L(\cdot)$ is a real non-negative kernel function, b is the bandwidth that associates with the kernel density estimate of f_Y, $K_h^*(\cdot) = K^*(\cdot/h)/h$, $K^*(z) = \frac{1}{2\pi}\int e^{-itz}\frac{\phi_K(t)}{\phi_U(t/h)}dt$ is the deconvoluting kernel, $h \in \mathbb{R}_{++}$ is a smoothing parameter, ϕ_U is the characteristic function of U, and $\phi_K(t) = \int e^{it\chi}K(\chi)d\chi$ is the Fourier transform of $K(\chi)$. Then, η_{ij} can be estimated from the estimates of $f_{\eta|e}(\eta|e)$.

Second, upon solving for $\boldsymbol{\xi}_i$ given $\hat{\eta}_{ij}$ (Step 2) and calculating u_i^A given $\hat{\boldsymbol{\xi}}_i$ by (1.2.2) (Step 3), one can obtain $v_i + u_i^T$ from the cost function. Then, an approach like Jondrow et al. (1982) and Battese and Coelli (1988), which use the conditional expectation, can be applied to estimate technical inefficiency. As it is assumed that $v_i \sim N(0, \sigma_v^2)$, $u_i^T \sim |N(0, \sigma_T^2)|$, conditional expectation of u_i^T given $v_i + u_i^T$ is

$$\mathbb{E}[u_i^T | v_i + u_i^T] = \sigma_* \left[\frac{\phi(\frac{\lambda}{\sigma}\tilde{\epsilon}_i)}{\Phi(\frac{\lambda}{\sigma}\tilde{\epsilon}_i)} + \frac{\lambda}{\sigma}\tilde{\epsilon}_i \right]$$

where $\sigma_*^2 = \sigma_v^2\sigma_T^2/\sigma^2$, $\sigma^2 = \sigma_v^2 + \sigma_T^2$, $\lambda = \sigma_T + \sigma_v$, ϕ is the probability density function of the standard normal distribution, Φ is the cumulative density function of the standard normal distribution, and $\tilde{\epsilon}_i = v_i + u_i^T$ (Kumbhakar and Lovell, 2000, p.141).

1.5 Application: U.S. Banking Industry

1.5.1 Data

The new model and strategies are applied to the U.S. banking industry. The dataset for this empirical study is based on the Reports of Condition and Income (Call Reports) for all FDIC insured U.S. banks. These are retrieved from FDIC's Statistics on Depository Institutions (https://www7.fdic.gov/sdi). Although the sample includes all U.S. depository institutions in the Call Report for the end of 2019 and 2020, it is filtered for the following reasons. First, the observations are dropped if key variables in the model are missing. Second, banks that enter or exit the market during the corresponding year are excluded, as their reported cost does not represent the total yearly cost. To the end, the number of banks in the dataset is 5,105 in 2019 and 4,937 in 2020, whereas the number of reported institutions is 5,177 in 2019 and 5,002 in 2020, respectively.

There is a long-standing disagreement over the input and output of banks.[20] I follow the asset (or intermediation) approach (see Sealey and Lindley, 1977) that banks are regarded as firms that transform various financial and physical resources, such as deposits and labor, into loans and investments. Similar to Altunbas et al. (2007) and Ding and Sickles (2019), it is assumed that banks produce two outputs using three inputs. The output variables are loans (y_{i1}) and other earning assets (y_{i2}), such as securities and trading assets. The input variables are funds that the bank owes (x_{i1}), such as deposits and debentures, the number of full-time employees (x_{i2}), and fixed assets (x_{i3}). Given this definition of inputs, input prices are defined as w_{i1} = interest expenses (C_{i1})/funds that the bank owes (x_{i1}), w_{i2} = salaries (C_{i2})/the number of full-time employees (x_{i2}), and w_{i3} = fixed assets expenses (C_{i3})/fixed assets (x_{i3}). The total cost (C_i) is defined as the sum of interest expenses (C_{i1}), salaries (C_{i2}), and fixed assets expenses (C_{i3}).

Tables 1.B.1 and 1.B.2 summarize descriptive statistics of key variables in 2019 and 2020,

[20]For more details, please refer to Berger and Humphrey (1992) and Guarda et al. (2013).

respectively. While other variables are not changed significantly during the sample period, there are notable changes in two key variables as shown in Table 1.5.1. First, interest expenses decrease significantly in 2020. Although they had had an increasing trend since 2016, they decreased from $158.7 billion in 2019 to $77.1 billion in 2020 for all depository institutions. Second, other earning assets have increased substantially during the pandemic. To be specific, their balance has slightly increased from 2016 to 2019 but increased from $6.9 trillion in 2019 to $9.8 trillion in 2020. It implies that the composition of banks' costs and outputs has considerably changed since the outbreak of COVID-19. This might lead the changes in the cost frontier of the U.S. banking industry between 2019 and 2020.

Table 1.5.1: Interest Expenses and Other Earning Assets of the U.S. Banks

Year	2016	2017	2018	2019	2020
Interest Expenses ($ billion)	54.4	73.3	119.8	158.7	77.1
Other Earning Assets ($ billion)	6,321.8	6,547.6	6,611.1	6,870.8	9,829.8

Source: FDIC

1.5.2 Average Inefficiency

The main purpose of this subsection is to check whether the average cost inefficiency of the U.S. banking industry has varied before and after the pandemic. In addition, the estimation results are compared to those following Greene (1980) to show that more plausible results can be obtained from the model proposed in this paper.

Greene (1980)'s assumptions can be summarized as follows. Given the translog cost system

$$
\begin{aligned}
\ln C_i &= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + \epsilon_i \\
&= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + v_i + u_i \\
s_{ij} &= s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + e_{ij},\ j = 2, 3,
\end{aligned}
$$

it is assumed that ϵ_i is independent of $\boldsymbol{e}_i = (e_{i2}, e_{i3})$, $v_i \overset{\text{i.i.d}}{\sim} N(0, \sigma_v^2)$, $u_i \overset{\text{i.i.d}}{\sim} |N(0, \sigma_u^2)|$, and $\boldsymbol{e}_i \overset{\text{i.i.d}}{\sim} N(0, \boldsymbol{\Sigma}_e)$. Maximum likelihood estimation can be applied to estimate parameters of the model, where the joint density of ϵ_i, e_{i2}, e_{i3} is simply $f_{\epsilon,e_2,e_3}(\epsilon, e_2, e_3) = f_\epsilon(\epsilon) \cdot f_{\boldsymbol{e}}(\boldsymbol{e})$, as ϵ_i and $\boldsymbol{e}_i$ are independent. Note that the probability density function of ϵ_i, which is the sum of two random variables following the normal distribution and the half normal distribution, respectively, is

$$f_\epsilon(\epsilon) = \frac{2}{\sigma} \cdot \phi\left(\frac{\epsilon}{\sigma}\right) \cdot \Phi\left(\frac{\epsilon\lambda}{\sigma}\right),$$

where $\sigma^2 = \sigma_v^2 + \sigma_u^2$, $\lambda = \sigma_u/\sigma_v$, and ϕ and Φ are density and distribution functions of the standard normal distribution.[21] Even though technical and allocative inefficiency cannot be decomposed, overall inefficiency u_i can be estimated using Greene (1980)'s assumptions.

Tables 1.5.2 and 1.5.3 show the estimation results for 2019 and 2020, respectively, where Models I and II stand for the model of this paper and the model following the independence assumption stated by Greene (1980). To estimate the model of this paper (Model I), $R = 10,000$ sets of random numbers are drawn.

Table 1.5.2: Estimation Results for 2019

	Model I	Model II		Model I	Model II		Model I	Model II
β_0	0.5176	0.5529	β_{11}^{ww}	0.1465	0.1488	$\sigma_{\nu_2}/\sigma_{e_2}$	0.0625	0.0755
	(0.1466)	(0.1484)		(0.0023)	(0.0011)		(0.0008)	(0.0006)
β_1^y	0.5130	0.5099	β_{12}^{ww}	-0.1370	-0.1358	$\sigma_{\nu_3}/\sigma_{e_3}$	0.0218	0.0431
	(0.0178)	(0.0187)		(0.0020)	(0.0010)		(0.0010)	(0.0004)
β_2^y	0.4879	0.4883	β_{22}^{ww}	0.1365	0.1325	ρ_ν/ρ_e	0.5233	0.0314
	(0.0231)	(0.0231)		(0.0020)	(0.0014)		(0.0279)	(0.0133)
β_{11}^{yy}	0.1180	0.1147	β_{11}^{yw}	0.0088	0.0044	θ_{12}	0.3229	-
	(0.0026)	(0.0027)		(0.0016)	(0.0018)		(0.1566)	
β_{12}^{yy}	-0.1102	-0.1069	β_{12}^{yw}	-0.0060	-0.0025	θ_{13}	0.0000	-
	(0.0031)	(0.0030)		(0.0012)	(0.0014)		(0.1506)	
β_{22}^{yy}	0.1005	0.0972	β_{21}^{yw}	0.0156	0.0187	ρ_θ	0.8137	-
	(0.0039)	(0.0039)		(0.0014)	(0.0016)		(0.0165)	
β_1^w	0.3704	0.3896	β_{22}^{yw}	-0.0172	-0.0199	σ_T/σ_u	0.2242	0.2239
	(0.0138)	(0.0182)		(0.0012)	(0.0013)		(0.0175)	(0.0128)
β_2^w	0.5368	0.5254	σ_v	0.2868	0.2922	σ_{ξ_2}	0.5941	-
	(0.0111)	(0.0142)		(0.0061)	(0.0049)		(0.0240)	
						σ_{ξ_3}	0.6447	-
							(0.0228)	

Note: BHHH standard errors are in parentheses.

[21]Please refer to Appendix 1.A.4 for the derivation.

Table 1.5.3: Estimation Results for 2020

	Model I	Model II		Model I	Model II		Model I	Model II
β_0	0.8529	0.9687	β_{11}^{ww}	0.1265	0.1276	$\sigma_{\nu_2}/\sigma_{e_2}$	0.0553	0.0724
	(0.1514)	(0.1539)		(0.0021)	(0.0010)		(0.0008)	(0.0006)
β_1^y	0.3599	0.3549	β_{12}^{ww}	-0.1224	-0.1202	$\sigma_{\nu_3}/\sigma_{e_3}$	0.0201	0.0467
	(0.0195)	(0.0179)		(0.0018)	(0.0010)		(0.0009)	(0.0004)
β_2^y	0.5805	0.5748	β_{22}^{ww}	0.1272	0.1222	ρ_ν/ρ_e	0.3968	-0.1413
	(0.0242)	(0.0238)		(0.0018)	(0.0013)		(0.0411)	(0.0122)
β_{11}^{yy}	0.1050	0.1053	β_{11}^{yw}	0.0086	0.0033	θ_{12}	0.3191	-
	(0.0030)	(0.0030)		(0.0012)	(0.0016)		(0.2036)	
β_{12}^{yy}	-0.0841	-0.0846	β_{12}^{yw}	-0.0039	0.0012	θ_{13}	0.0000	-
	(0.0034)	(0.0030)		(0.0011)	(0.0013)		(0.1428)	
β_{22}^{yy}	0.0647	0.0661	β_{21}^{yw}	0.0118	0.0159	ρ_θ	0.8162	-
	(0.0040)	(0.0036)		(0.0012)	(0.0015)		(0.0132)	
β_1^w	0.3633	0.3800	β_{22}^{yw}	-0.0180	-0.0219	σ_T/σ_u	0.2060	0.1690
	(0.0103)	(0.0147)		(0.0011)	(0.0013)		(0.0269)	(0.0227)
β_2^w	0.5569	0.5429	σ_v	0.3193	0.3325	σ_{ξ_2}	0.6299	-
	(0.0088)	(0.0117)		(0.0069)	(0.0058)		(0.0227)	
						σ_{ξ_3}	0.7005	-
							(0.0180)	

Note: BHHH standard errors are in parentheses.

There are three points to be noted. First, the results indicate a dependence between technical and allocative inefficiency, although θ_{13}, which captures dependence between u_i^T and ξ_{i3}, is close to zero in both periods. Second, reflecting large increase in other earning assets (y_{i2}), β_1^y decreased and β_2^y increased in 2020 relative to 2019. Lastly, both σ_T of Model I and σ_u of model II decreased, implying that technical or overall inefficiency decreased during the pandemic. This might be due to the fact that interest expenses decreased in 2020. However, the magnitude of decreases is much higher for Model II than for Model I.

Table 1.5.4 shows the estimation results of average inefficiency. In both models, the mean of u_i^T, u_i^A, or u_i cannot be directly estimated; only their standard deviations and the standard deviation of ξ_{ij}s can be estimated. However, since the mean of a random variable that follows the half normal distribution is a function of its standard deviation[22], the mean of u_i^T for Model I and u_i for Model II, which are assumed to follow the half normal distribution, can be calculated. The average u_i^A of Model I is calculated using the estimation results for

[22] If $X \sim |N(0, \sigma_X^2)|$, $\mathbb{E}[X] = \frac{\sigma_X \sqrt{2}}{\sqrt{\pi}}$.

firm-level inefficiency in the next subsection. For Model I, the average u_i is computed as the sum of the mean of both u_i^T and u_i^A.

Table 1.5.4: Average Inefficiency

	2019			2020		
	u_i	u_i^T	u_i^A	u_i	u_i^T	u_i^A
Model I	0.2040	0.1789	0.0251	0.1887	0.1643	0.0244
Model II	0.1786	-	-	0.1349	-	-

The results suggest that costs of U.S. banks increased by around 16∼18% and 2.5% during the sample period due to technical and allocative inefficiency, respectively. There are two main findings from the results. The first one is that overall inefficiency in 2020 decreased compared to 2019. The effect can be also decomposed. Cost increases due to allocative inefficiency do not differ between 2019 and 2020 (2.5% → 2.4%), whereas changes in technical inefficiency are non-trivial (17.9% → 16.4%). The second main finding is that assuming independence between ϵ_i and $\boldsymbol{e}_i$ would produce unrealistic results. Particularly, the overall inefficiency levels are generally underestimated by $3 \sim 5\%$ when compared to Model I. Furthermore, overall inefficiency decreased to a great extend in 2020 in spite of the pandemic (17.9% → 13.5%), which is seemingly less plausible.

1.5.3 Firm-level Inefficiency

As the first step to estimate firm-level inefficiency, it is suggested in Section 1.4 that the mode of the conditional distribution $\hat{f}_{\eta|e}(\eta|e)$ is used. Figures 1.5.1 and 1.5.2 show density estimates of e, $\hat{f}_e(e)$, and conditional density estimates of η given e, $\hat{f}_{\eta|e}(\eta|e)$, for 2019 obtained by kernel density estimation and density deconvolution. Although both $\hat{f}_{e_2}(e_2)$ and $\hat{f}_{e_3}(e_3)$ are skewed, $\hat{f}_{\eta|e}(\eta|e)$ is fairly symmetric and centered around its mode for various values of e. Along with Figures 1.B.2 and 1.B.3, which illustrate density estimates of e and conditional density estimates of η given e for 2020, it supports the validity of using the mode as the estimate of η.

Figure 1.5.1: $\hat{f}_{e_2}(e_2)$ and $\hat{f}_{\eta_2|e_2}(\eta_2|e_2)$ for 2019

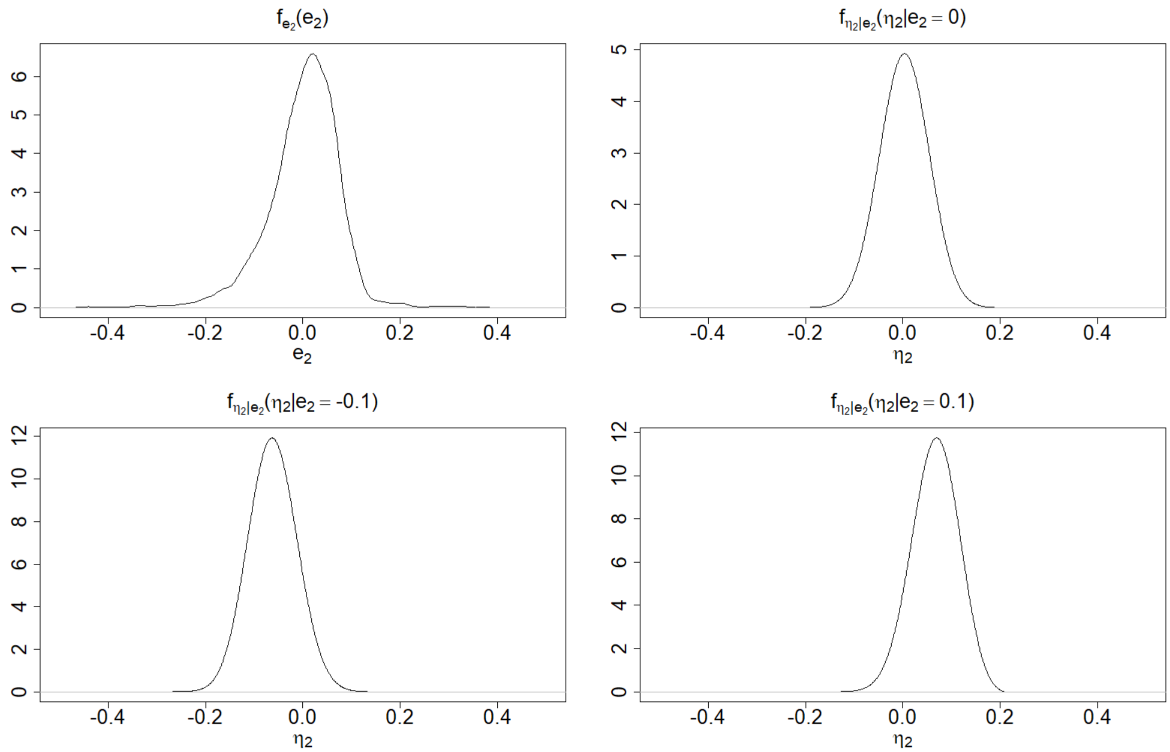


Figure 1.5.2: $\hat{f}_{e_3}(e_3)$ and $\hat{f}_{\eta_3|e_3}(\eta_3|e_3)$ for 2019

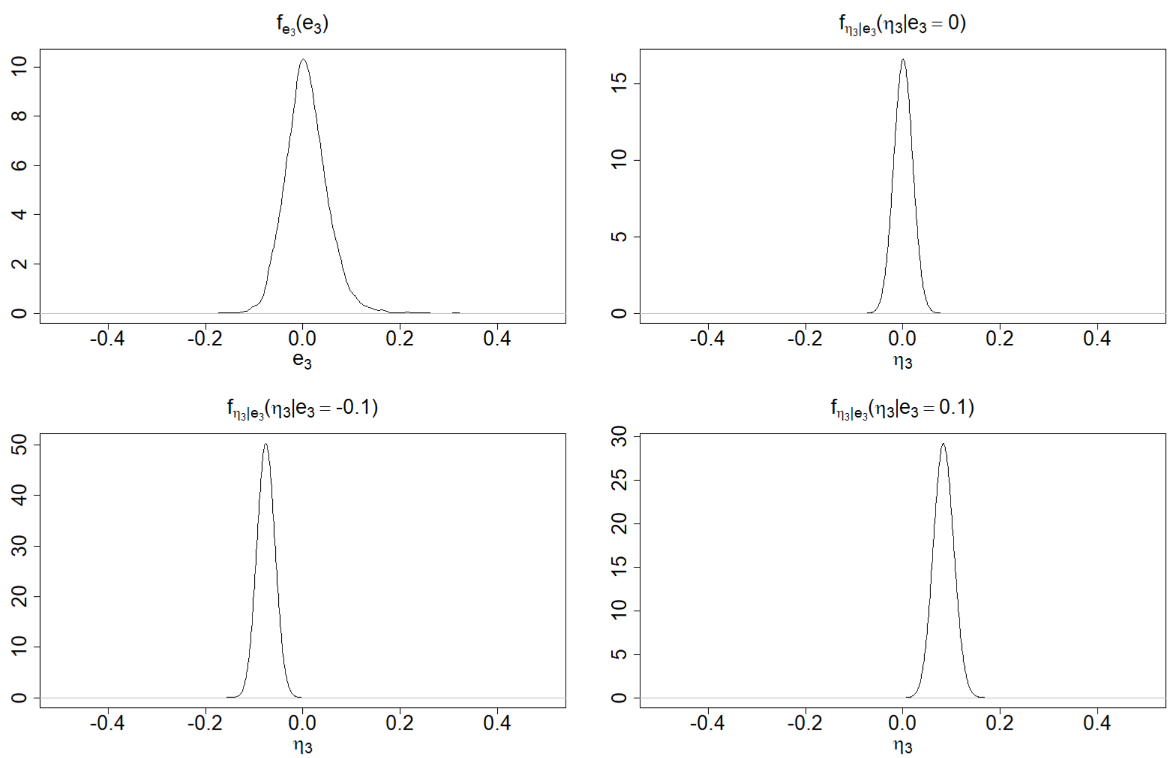


Table 1.5.5 shows the standard deviations of ξ_{i2} and ξ_{i3}, $\hat{\sigma}_{\xi_2}$ and $\hat{\sigma}_{\xi_3}$, using the estimation results of the second step; recall that ξ_{i2} and ξ_{i3} represent the degree of allocative inefficiency for the input pair $(j, 1),\ j = 2, 3$. Compared to the model estimates, the standard deviations of estimates for individual firm's allocative inefficiency tend to be rather large but not too different. In addition, as the estimates of firm-level allocative inefficiency, $\hat{\xi}_{i2}$ and $\hat{\xi}_{i3}$, can be obtained, the results can be compared by the bank's classification, such as the banks' size and the charter class[23]. The results show that the degree of allocative inefficiency of larger banks, such as banks with assets greater than $1 billion or nationally chartered commercial banks, and of thrifts is more dispersed than that of other banks.

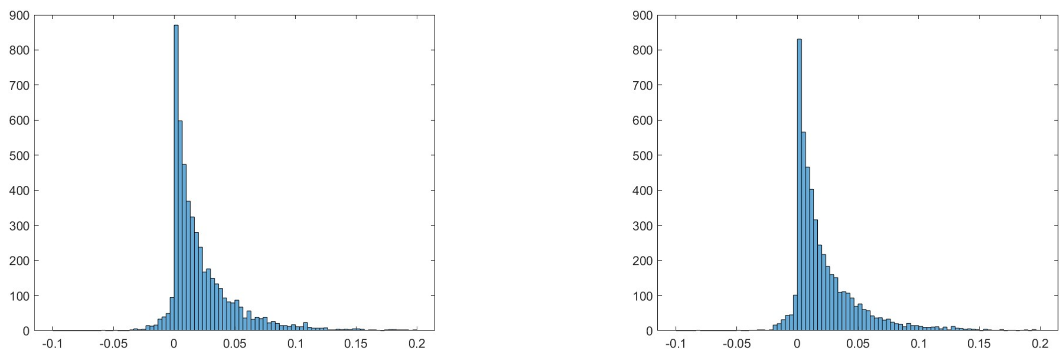
Table 1.5.5: Standard Deviations of $\hat{\xi}_{i2}$ and $\hat{\xi}_{i3}$

	2019		2020	
	$\hat{\sigma}_{\xi_2}$	$\hat{\sigma}_{\xi_3}$	$\hat{\sigma}_{\xi_2}$	$\hat{\sigma}_{\xi_3}$
Model Estimates	0.5941	0.6447	0.6299	0.7005
Full Sample	0.7693	0.7495	0.7313	0.7286
Asset Size				
- Large Banks	0.8805	0.7667	0.8416	0.7617
- Small Banks	0.7476	0.7459	0.7016	0.7176
Charter Class				
- N	0.8024	0.7422	0.7614	0.7185
- NM	0.7571	0.7461	0.7102	0.7232
- SM	0.7354	0.7060	0.7137	0.6975
- SB	0.7732	0.7475	0.7319	0.7040
- SA	0.8287	0.8488	0.8434	0.8465

[23] Banks are classified by its asset size and classification codes assigned by the FDIC, which indicate an institution's charter type, an institution's charter type, its Federal Reserve membership status, and its primary federal regulator. Please refer to Table 1.B.3 for more details.

Figure 1.5.3 shows the distribution of estimates for cost increases due to allocative inefficiency, u_i^A, of individual banks, which is obtained from the third step of the process. The distribution shape is fairly similar to the results in Kumbhakar and Tsionas (2005b) in that it is highly skewed to the right. However, magnitudes are smaller seemingly owing to different definitions of inputs and outputs, sample periods, and the assumption about the variation of the term representing allocative inefficiency, $\boldsymbol{\xi}_i$. Even though a small number of estimates are negative, contrary to the definition of u_i^A, their magnitudes are not considerable.

Figure 1.5.3: Distribution of $\hat{u}_i^A$



(a) 2019 (b) 2020

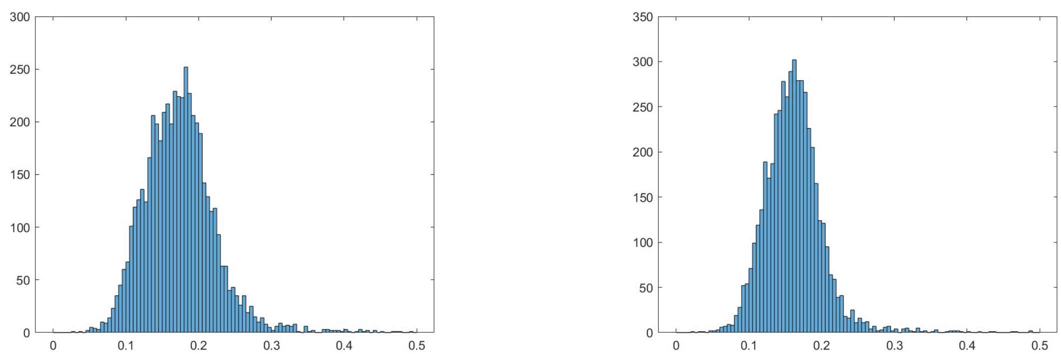
Table 1.5.6 shows the average cost increases due to allocative inefficiency by the banks' classification. The estimates of larger banks, such as banks with assets greater than \$1 billion or nationally chartered commercial banks, and of thrifts are higher than those of other banks. In addition, the average cost increases due to allocative inefficiency has not been changed significantly between 2019 and 2020 by banks' asset sizes and the charter classes.

Table 1.5.6: Average of $\hat{u}_i^A$

	2019	2020
Full Sample	0.0251	0.0244
Asset Size		
- Large Banks	0.0327	0.0332
- Small Banks	0.0237	0.0223
Charter Class		
- N	0.0288	0.0282
- NM	0.0240	0.0229
- SM	0.0239	0.0242
- SB	0.0225	0.0210
- SA	0.0316	0.0326

Figure 1.5.4 shows the distribution of estimates for cost increases due to technical inefficiency, u_i^T, of individual banks, which is obtained from the last step of the process. As opposed to the distribution of $\hat{u}_i^A$, it is slightly skewed to the right. In addition, the estimates for cost increases due to technical inefficiency are less dispersed in 2020 compared to those of the previous year. Its standard deviation is decreased from 0.0567 in 2019 to 0.0452 in 2020; this implies that banks became somewhat homogeneous in terms of technical inefficiency.

Figure 1.5.4: Distribution of $\hat{u}_i^T$



(a) 2019 (b) 2020

Table 1.5.7 shows the average cost increases due to technical inefficiency by the banks' classification. The estimates are fairly similar to those of u_i^T obtained from the model. Also, unlike the results of the third step, the estimates of the large banks are slightly smaller than those of the small banks. By the banks' charter class, nationally chartered commercial banks seem to be technically less efficient than others, but the gap between them decreased in 2020.

Table 1.5.7: Average of $\hat{u}_i^T$

	2019	2020
Model Estimates	0.1789	0.1643
Full Sample	0.1748	0.1620
Asset Size		
- Large Banks	0.1737	0.1619
- Small Banks	0.1750	0.1621
Charter Class		
- N	0.1819	0.1666
- NM	0.1743	0.1610
- SM	0.1751	0.1634
- SB	0.1641	0.1569
- SA	0.1710	0.1626

Table 1.5.8 summarizes the estimation results of individual technical and allocative inefficiency. To sum up, first, the ratio of cost increase due to allocative inefficiency, u_i^A, has not changed between 2019 and 2020. However, the ratio of cost increase due to technical inefficiency, u_i^T, has decreased. It suggests the possibility of changes in the cost frontier during the pandemic, while banks try to maintain resource allocation. Second, cost increase due to inefficiency is generally higher for larger banks. It is in line with previous studies, such as Altunbas et al. (2007) and Ding and Sickles (2019).

Table 1.5.8: Average of $\hat{u}_i^T$ and $\hat{u}_i^A$

	2019		2020	
	$\hat{u}_i^T$	$\hat{u}_i^A$	$\hat{u}_i^T$	$\hat{u}_i^A$
Full Sample	0.1748	0.0251	0.1620	0.0244
Asset Size				
- Large Banks	0.1737	0.0327	0.1619	0.0332
- Small Banks	0.1750	0.0237	0.1621	0.0223
Charter Class				
- N	0.1819	0.0288	0.1666	0.0282
- NM	0.1743	0.0240	0.1610	0.0229
- SM	0.1751	0.0239	0.1634	0.0242
- SB	0.1641	0.0225	0.1569	0.0210
- SA	0.1710	0.0316	0.1626	0.0326

1.6 Conclusion

Cost efficiency analysis has the virtue that it enables researchers to decompose inefficiency into two main sources; input-oriented technical inefficiency, and resource misallocation. However, there is no satisfactory method to measure and decompose both types of inefficiency when flexible functional forms are allowed for.

In this paper, a model and an estimation strategy for the translog cost system are developed to overcome limitations of previous stochastic cost frontier studies. By employing APS copulas, one can model the dependence between technical and allocative inefficiency as well as provide a solution to "the Greene Problem." The model can be estimated by the method of simulated likelihood. The proposed estimation strategy is developed upon economic intuition behind the stochastic frontier model. It is also uncomplicated, as random numbers can be drawn from the simple density, the standard uniform density. In addition, a strategy to estimate individual inefficiency is proposed, which uses not only conditional distributions as in previous studies, but also density deconvolution.

An empirical exercise for the U.S. banking industry in 2019 and 2020 shows that the costs of U.S. banks increased by around 20% during the sample period due to inefficiency, where technical and allocative inefficiency account for around 16∼18% and 2.5%, respectively. During the pandemic, banks' technical inefficiency has slightly decreased seemingly due to changes in the composition of costs and output, while the degree of allocative inefficiency has not changed significantly. Lastly, the results suggest that it would produce less plausible results when ignoring the dependence between technical and allocative inefficiency.

APPENDICES

APPENDIX A

Additional Details

1.A.1 Proof of $\sum_{j=1}^{J} s_{ij}(\boldsymbol{y}_i, \boldsymbol{w}_i) = 1$ and $\sum_{j=1}^{J} \eta_{ij} = 0$

1.A.1.1 $\sum_{j=1}^{J} s_{ij}(\boldsymbol{w}_i, \boldsymbol{y}_i) = 1$

$$
\begin{aligned}
&\sum_{j=1}^{J} s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) \\
&= \sum_{j=1}^{J} \left[\beta_j^w + \sum_{k=1}^{J} \beta_{jk}^{ww} (\ln w_{ik}) + \sum_{m=1}^{M} \beta_{mj}^{yw} (\ln y_{im}) \right] \\
&= \sum_{j=1}^{J} \beta_j^w + \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk}^{ww} (\ln w_{ik}) + \sum_{j=1}^{J} \sum_{m=1}^{M} \beta_{mj}^{yw} (\ln y_{im}) \\
&= \sum_{j=1}^{J} \beta_j^w + \sum_{k=1}^{J} \sum_{j=1}^{J} \beta_{jk}^{ww} (\ln w_{ik}) + \sum_{m=1}^{M} \sum_{j=1}^{J} \beta_{mj}^{yw} (\ln y_{im}) \\
&= \sum_{j=1}^{J} \beta_j^w + \sum_{k=1}^{J} (\ln w_{ik}) \sum_{j=1}^{J} \beta_{jk}^{ww} + \sum_{m=1}^{M} (\ln y_{im}) \sum_{j=1}^{J} \beta_{mj}^{yw} \\
&= 1.
\end{aligned}
$$

The last equality holds because of the restrictions on the parameters, such as $\sum_{j=1}^{J} \beta_j^w = 1$, $\sum_{k=1}^{J} \beta_{jk}^{ww} = \sum_{j=1}^{J} \beta_{jk}^{ww} = 0$ $\forall j$ or $\forall k$, and $\sum_{j=1}^{J} \beta_{mj}^{yw} = 0$ $\forall m$.

1.A.1.2 $\sum_{j=1}^{J} \eta_{ij} = 0$

Since $\sum_{j=1}^{J} s_{ij} = 1$ and $\sum_{j=1}^{J} s_{ij}(\boldsymbol{y}_i, \boldsymbol{w}_i) = 1$, $\sum_{j=1}^{J} \eta_{ij} = 0$ is guaranteed. For $J = 2$ and 3, it can be also shown using the formula of η_{ij}.

(i) $J = 2$

$$
\begin{aligned}
&\eta_{i1} + \eta_{i2} \\
&= \frac{s_1(\boldsymbol{y}_i, \boldsymbol{w}_i)[1 - \{s_{i1}^* + (s_{i2}^*/e^{\xi_{i2}})\}] + \beta_{12}^{ww}\xi_{i2}}{\{s_{i1}^* + (s_{i2}^*/e^{\xi_{i2}})\}} \\
&\quad + \frac{s_2(\boldsymbol{y}_i, \boldsymbol{w}_i)[1 - \{s_{i1}^* + (s_{i2}^*/e^{\xi_{i2}})\}e^{\xi_{i2}}] + \beta_{22}^{ww}\xi_{i2}}{\{s_{i1}^* + (s_{i2}^*/e^{\xi_{i2}})\}e^{\xi_{i2}}}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{s_1(\boldsymbol{y}_i,\boldsymbol{w}_i)e^{\xi_{i2}}[1-\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})\}]+\beta^{ww}_{12}\xi_{i2}e^{\xi_{i2}}}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})\}e^{\xi_{i2}}} \\
&\quad + \frac{s_2(\boldsymbol{y}_i,\boldsymbol{w}_i)[1-\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})\}e^{\xi_{i2}}]+\beta^{ww}_{22}\xi_{i2}}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})\}e^{\xi_{i2}}} \\
&= \frac{s_1(\boldsymbol{y}_i,\boldsymbol{w}_i)e^{\xi_{i2}}+s_2(\boldsymbol{y}_i,\boldsymbol{w}_i)}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})\}e^{\xi_{i2}}} \\
&\quad - \frac{\big(s_1(\boldsymbol{y}_i,\boldsymbol{w}_i)+s_2(\boldsymbol{y}_i,\boldsymbol{w}_i)\big)\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})\}e^{\xi_{i2}}+\beta^{ww}_{12}\xi_{i2}e^{\xi_{i2}}+\beta^{ww}_{22}\xi_{i2}}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})\}e^{\xi_{i2}}} \\
&= \frac{\big(s_1(\boldsymbol{y}_i,\boldsymbol{w}_i)-s^*_{i1}+\beta^{ww}_{12}\xi_{i2}\big)e^{\xi_{i2}}+\big(s_2(\boldsymbol{y}_i,\boldsymbol{w}_i)-s^*_{i2}+\beta^{ww}_{22}\xi_{i2}\big)}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})\}e^{\xi_{i2}}} \\
&= 0.
\end{aligned}
$$

The fourth equality holds as $s_1(\boldsymbol{y}_i,\boldsymbol{w}_i)+s_2(\boldsymbol{y}_i,\boldsymbol{w}_i)=1$, and the last equality holds by the definition of s_{ij}.

(ii) $J=3$

$$
\begin{aligned}
&\eta_{i1}+\eta_{i2}+\eta_{i3} \\
&= \frac{s_1(\boldsymbol{y}_i,\boldsymbol{w}_i)[1-\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}]+\beta^{ww}_{12}\xi_{i2}+\beta^{ww}_{13}\xi_{i3}}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}} \\
&\quad + \frac{s_2(\boldsymbol{y}_i,\boldsymbol{w}_i)[1-\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}e^{\xi_{i2}}]+\beta^{ww}_{22}\xi_{i2}+\beta^{ww}_{23}\xi_{i3}}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}e^{\xi_{i2}}} \\
&\quad + \frac{s_3(\boldsymbol{y}_i,\boldsymbol{w}_i)[1-\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}e^{\xi_{i3}}]+\beta^{ww}_{32}\xi_{i2}+\beta^{ww}_{33}\xi_{i3}}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}e^{\xi_{i3}}} \\
&= \frac{s_1(\boldsymbol{y}_i,\boldsymbol{w}_i)e^{\xi_{i2}}e^{\xi_{i3}}[1-\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}]}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}} \\
&\quad + \frac{\beta^{ww}_{12}\xi_{i2}e^{\xi_{i2}}e^{\xi_{i3}}+\beta^{ww}_{13}\xi_{i3}e^{\xi_{i2}}e^{\xi_{i3}}}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}} \\
&\quad + \frac{s_2(\boldsymbol{y}_i,\boldsymbol{w}_i)e^{\xi_{i3}}[1-\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}e^{\xi_{i2}}]}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}} \\
&\quad + \frac{\beta^{ww}_{22}\xi_{i2}e^{\xi_{i3}}+\beta^{ww}_{23}\xi_{i3}e^{\xi_{i3}}}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}} \\
&\quad + \frac{s_3(\boldsymbol{y}_i,\boldsymbol{w}_i)e^{\xi_{i2}}[1-\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}e^{\xi_{i3}}]}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}} \\
&\quad + \frac{\beta^{ww}_{32}\xi_{i2}e^{\xi_{i2}}+\beta^{ww}_{33}\xi_{i3}e^{\xi_{i2}}}{\{s^*_{i1}+(s^*_{i2}/e^{\xi_{i2}})+(s^*_{i3}/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{s_1(\boldsymbol{y}_i,\boldsymbol{w}_i)e^{\xi_{i2}}e^{\xi_{i3}} + s_2(\boldsymbol{y}_i,\boldsymbol{w}_i)e^{\xi_{i3}} + s_3(\boldsymbol{y}_i,\boldsymbol{w}_i)e^{\xi_{i2}}}{\{s_{i1}^* + (s_{i2}^*/e^{\xi_{i2}}) + (s_{i3}^*/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}} \\
&\quad - \frac{\big(s_1(\boldsymbol{y}_i,\boldsymbol{w}_i) + s_2(\boldsymbol{y}_i,\boldsymbol{w}_i) + s_3(\boldsymbol{y}_i,\boldsymbol{w}_i)\big)e^{\xi_{i2}}e^{\xi_{i3}}\{s_{i1}^* + (s_{i2}^*/e^{\xi_{i2}}) + (s_{i3}^*/e^{\xi_{i3}})\}}{\{s_{i1}^* + (s_{i2}^*/e^{\xi_{i2}}) + (s_{i3}^*/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}} \\
&\quad + \frac{\beta_{12}^{ww}\xi_{i2}e^{\xi_{i2}}e^{\xi_{i3}} + \beta_{13}^{ww}\xi_{i3}e^{\xi_{i2}}e^{\xi_{i3}}}{\{s_{i1}^* + (s_{i2}^*/e^{\xi_{i2}}) + (s_{i3}^*/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}} \\
&\quad + \frac{\beta_{22}^{ww}\xi_{i2}e^{\xi_{i3}} + \beta_{23}^{ww}\xi_{i3}e^{\xi_{i3}}}{\{s_{i1}^* + (s_{i2}^*/e^{\xi_{i2}}) + (s_{i3}^*/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}} \\
&\quad + \frac{\beta_{32}^{ww}\xi_{i2}e^{\xi_{i2}} + \beta_{33}^{ww}\xi_{i3}e^{\xi_{i2}}}{\{s_{i1}^* + (s_{i2}^*/e^{\xi_{i2}}) + (s_{i3}^*/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}} \\
&= \frac{e^{\xi_{i2}}e^{\xi_{i3}}\big(s_1(\boldsymbol{y}_i,\boldsymbol{w}_i) - s_{i1}^* + \beta_{12}^{ww}\xi_{i2} + \beta_{13}^{ww}\xi_{i3}\big)}{\{s_{i1}^* + (s_{i2}^*/e^{\xi_{i2}}) + (s_{i3}^*/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}} \\
&\quad + \frac{e^{\xi_{i3}}\big(s_2(\boldsymbol{y}_i,\boldsymbol{w}_i) - s_{i2}^* + \beta_{22}^{ww}\xi_{i2} + \beta_{23}^{ww}\xi_{i3}\big)}{\{s_{i1}^* + (s_{i2}^*/e^{\xi_{i2}}) + (s_{i3}^*/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}} \\
&\quad + \frac{e^{\xi_{i2}}\big(s_3(\boldsymbol{y}_i,\boldsymbol{w}_i) - s_{i3}^* + \beta_{32}^{ww}\xi_{i2} + \beta_{33}^{ww}\xi_{i3}\big)}{\{s_{i1}^* + (s_{i2}^*/e^{\xi_{i2}}) + (s_{i3}^*/e^{\xi_{i3}})\}e^{\xi_{i2}}e^{\xi_{i3}}} \\
&= 0.
\end{aligned}
$$

The fourth equality holds as $s_1(\boldsymbol{y}_i,\boldsymbol{w}_i) + s_2(\boldsymbol{y}_i,\boldsymbol{w}_i) + s_3(\boldsymbol{y}_i,\boldsymbol{w}_i) = 1$, and the last equality holds by the definition of s_{ij}.

1.A.2 Simplifying u_i^A and η_{ij}

Assume that there is no additive noise term in the cost share equations. Let $G_i = \sum_{k=1}^{J}(s_{ik}^*/e^{\xi_k})$. Then,

$$
\begin{aligned}
\eta_{ij} &= \frac{s_j(\boldsymbol{y}_i,\boldsymbol{w}_i)(1 - G_i e^{\xi_j}) + \sum_{k=1}^{J}\beta_{jk}^{ww}\xi_k}{G_i e^{\xi_j}} \\
&\Rightarrow\ G_i\eta_{ij}e^{\xi_j} = s_j(\boldsymbol{y}_i,\boldsymbol{w}_i)(1 - G_i e^{\xi_j}) + \sum_{k=1}^{J}\beta_{jk}^{ww}\xi_k \\
&\Rightarrow\ G_i\big(s_j(\boldsymbol{y}_i,\boldsymbol{w}_i) + \eta_{ij}\big)e^{\xi_j} = s_j(\boldsymbol{y}_i,\boldsymbol{w}_i) + \sum_{k=1}^{J}\beta_{jk}^{ww}\xi_k \\
&\Rightarrow\ G_i s_{ij}e^{\xi_j} = s_j(\boldsymbol{y}_i,\boldsymbol{w}_i) + \sum_{k=1}^{J}\beta_{jk}^{ww}\xi_k
\end{aligned}
$$

$$
\begin{aligned}
&\Rightarrow \sum_{j=1}^{J} G_i s_{ij} e^{\xi_j} = \sum_{j=1}^{J} \Big(s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + \sum_{k=1}^{J} \beta_{jk}^{ww} \xi_k \Big) \\
&\Rightarrow G_i \sum_{j=1}^{J} s_{ij} e^{\xi_j} = \sum_{j=1}^{J} s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk}^{ww} \xi_k \\
&\therefore \; G_i = \frac{1}{\sum_{j=1}^{J} s_{ij} e^{\xi_j}}
\end{aligned}
$$

The last equality holds because $\sum_{j=1}^{J} s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) = 1$ and $\sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk}^{ww} \xi_k = 0$ by the symmetry and the linear homogeneity conditions of the cost function. Also, note that

$$
\begin{aligned}
&G_i s_{ij} e^{\xi_j} = s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + \sum_{k=1}^{J} \beta_{jk}^{ww} \xi_k \\
&\Rightarrow \; G_i s_{ij} e^{\xi_j} = s_{ij} - \eta_{ij} + \sum_{k=1}^{J} \beta_{jk}^{ww} \xi_k \\
&\therefore \; \eta_{ij} = s_{ij}(1 - G_i e^{\xi_j}) + \sum_{k=1}^{J} \beta_{jk}^{ww} \xi_k
\end{aligned}
$$

Therefore, u_i^A and η_{ij} can be simplified as

$$
\begin{aligned}
u_i^A = &\sum_{j=1}^{J} \beta_j^w \xi_j + \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk}^{ww} (\ln w_{ij}) \xi_k + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk}^{ww} \xi_j \xi_k + \sum_{m=1}^{M} \sum_{j=1}^{J} \beta_{mj}^{yw} (\ln y_{im}) \xi_j \\
&- \ln \Big(\sum_{j=1}^{J} s_{ij} e^{\xi_j} \Big) \\
\eta_{ij} = &\; s_{ij} \Big(1 - \frac{e^{\xi_j}}{\sum_{k=1}^{J} s_{ik} e^{\xi_k}} \Big) + \sum_{k=1}^{J} \beta_{jk}^{ww} \xi_k .
\end{aligned}
$$

1.A.3 Rosenblatt Transformation

This subsection is written based on Rosenblatt (1952), Chapter 6.9.1 of Joe (2014), and Appendix B of Melchers and Beck (2018). Rosenblatt (1952) shows that a dependent random vector $\boldsymbol{X} = (X_1, \cdots, X_k)$ may be transformed to the random vector $\boldsymbol{Z} = (Z_1, \cdots, Z_k)$, where $Z_i \overset{\text{i.i.d}}{\sim} U[0,1] \; \forall \; i = 1, \cdots, k$. This subsection summarizes the procedure for generating a dependent random vector $\boldsymbol{X}$ from the independent random vector.

Let $F_{1:k}(\chi_1, \cdots, \chi_k)$ be a multivariate distribution of $\boldsymbol{X}$ with marginal distributions $F_1, \cdots, F_k$, where the corresponding random variables $X_1, \cdots, X_k$ can be continuous, discrete, or mixed. Rosenblatt (1952) proposes the transformation T_R such that $\boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_k) = T_R(\chi_1, \cdots, \chi_k)$, where

$$\begin{aligned}
\zeta_1 &= \mathbb{P}(X_1 \leq \chi_1) = F_1(\chi_1) \\
\zeta_2 &= \mathbb{P}(X_2 \leq \chi_2 | X_1 = \chi_1) = F_{2|1}(\chi_2|\chi_1) \\
\zeta_3 &= \mathbb{P}(X_3 \leq \chi_3 | X_1 = \chi_1, X_2 = \chi_2) = F_{3|12}(\chi_3|\chi_1, \chi_2) \\
&\vdots \\
\zeta_k &= \mathbb{P}(X_k \leq \chi_k | X_1 = \chi_1, \cdots, X_{k-1} = \chi_{k-1}) = F_{k|1,\cdots,k-1}(\chi_k|\chi_1, \cdots, \chi_{k-1}).
\end{aligned}$$

With all the conditional distributions $F_1, F_{2|1}, \cdots, F_{k|1,\cdots,k-1}$ and their inverse functions, one can successively obtain dependent random numbers $(\chi_1, \cdots, \chi_k)$ from independent uniform random numbers $(\zeta_1, \cdots, \zeta_k)$ on $[0,1]^k$ such that

$$\begin{aligned}
\chi_1 &= F_1^{-1}(\zeta_1) \\
\chi_2 &= F_{2|1}^{-1}(\zeta_2|\chi_1) \\
\chi_3 &= F_{3|12}^{-1}(\zeta_3|\chi_1, \chi_2) \\
&\vdots \\
\chi_k &= F_{k|1,\cdots,k-1}^{-1}(\zeta_k|\chi_1, \cdots, \chi_{k-1}).
\end{aligned}$$

To sum up, the consecutive process to generate $(\chi_1, \cdots, \chi_k)$ is as follows:

(i) Derive conditional distributions $F_{2|1}, F_{3|12}, \cdots, F_{k|1,\cdots,k-1}$

(ii) Draw ζ_1 from $U[0,1]$ and obtain χ_1 from $\chi_1 = F_1^{-1}(\zeta_1)$ or by solving $F_1(\chi_1) - \zeta_1 = 0$

(iii) Given χ_1, draw ζ_2 from $U[0,1]$ and obtain χ_2 from $\chi_2 = F_{2|1}^{-1}(\zeta_2|\chi_1)$ or by solving $F_{2|1}(\chi_2|\chi_1) - \zeta_2 = 0$

(iv) Given χ_1 and χ_2 , draw ζ_3 from $U[0,1]$ and obtain χ_3 from $\chi_3 = F^{-1}_{3|12}(\zeta_3|\chi_1,\chi_2)$ or by solving $F_{3|12}(\chi_3|\chi_1,\chi_2) - \zeta_3 = 0$, and so on.

Note that a permutation of the order $1,\cdots,k$ is possible, so one would choose a permutation in practice where the computations are simplest. For example, one can choose the reverse order to obtain dependent random numbers $(\chi_1,\cdots,\chi_k)$ from independent uniform random numbers $(\zeta_1,\cdots,\zeta_k)$ on $[0,1]^k$ such that

$$\begin{aligned}
\chi_1 &= F^{-1}_{1|2,\cdots,k}(\zeta_1|\chi_2,\cdots,\chi_k)\\
&\vdots\\
\chi_{k-2} &= F^{-1}_{k-2|k-1,k}(\zeta_{k-2}|\chi_{k-1},\chi_k)\\
\chi_{k-1} &= F^{-1}_{k-1|k}(\zeta_{k-1}|\chi_k).\\
\chi_k &= F^{-1}_k(\zeta_k)
\end{aligned}$$

1.A.4 Density of $\epsilon_i = v_i + u_i$

Let $\epsilon_i = v_i + u_i$, where $v_i \sim N(0,\sigma_v^2)$, $u_i \sim |N(0,\sigma_u^2)|$, and v_i and u_i are independent. Aigner et al. (1977) derive the density of a random variable $\tilde{\epsilon}_i = v_i - u_i^T$ as $\frac{2}{\sigma}\cdot\phi\Big(\frac{\tilde{\epsilon}}{\sigma}\Big)\cdot\Big[1-\Phi\Big(\frac{\tilde{\epsilon}\lambda}{\sigma}\Big)\Big]$, where $\sigma^2 = \sigma_v^2+\sigma_u^2$, $\lambda = \sigma_u/\sigma_v$, and ϕ and Φ are density and distribution functions of the standard normal distribution.[24] Similarly, the density of $\epsilon_i = v_i + u_i$ can be derived as follows. Note that the marginal densities of v_i and u_i are

$$\begin{aligned}
f_v(v) &= \frac{1}{\sigma_v\sqrt{2\pi}}\exp\Big(-\frac{v^2}{2\sigma_v^2}\Big),\ v\in\mathbb{R}\\
f_u u &= \frac{2}{\sigma_u\sqrt{2\pi}}\exp\Big(-\frac{u^2}{2\sigma_u^2}\Big),\ v\in\mathbb{R}_+.
\end{aligned}$$

Since v_i and u_i are independent, the joint density of them is

$$\begin{aligned}
f_{v,u}(v,u) &= \frac{1}{\sigma_v\sqrt{2\pi}}\frac{2}{\sigma_u\sqrt{2\pi}}\exp\Big(-\frac{v^2}{2\sigma_v^2}-\frac{u^2}{2\sigma_u^2}\Big)\\
&= \frac{1}{\sigma_v\sqrt{2\pi}}\frac{2}{\sigma_u\sqrt{2\pi}}\exp\Big\{-\frac{1}{2}\Big(\frac{v^2}{\sigma_v^2}+\frac{u^2}{\sigma_u^2}\Big)\Big\},
\end{aligned}$$

[24]Please refer to Ch.11.7.3 of Sickles and Zelenyuk (2019) for more details.

and the joint density of ϵ_i and u_i becomes

$$\begin{aligned} f_{\epsilon,u}(\epsilon, u) &= f_{v,u}(\epsilon - u, u) \\ &= \underbrace{\frac{1}{\sigma_v\sqrt{2\pi}}\frac{2}{\sigma_u\sqrt{2\pi}}}_{\textcircled{1}} \exp\Big\{-\frac{1}{2}\underbrace{\Big(\frac{(\epsilon-u)^2}{\sigma_v^2} + \frac{u^2}{\sigma_u^2}\Big)}_{\textcircled{2}}\Big\}. \end{aligned}$$

$\textcircled{1}$ and $\textcircled{2}$ can be transformed as follows:

$$\begin{aligned} \textcircled{1} &: \frac{\sqrt{\sigma_v^2+\sigma_u^2}}{\sigma_v\sigma_u\sqrt{2\pi}}\frac{2}{\sqrt{\sigma_v^2+\sigma_u^2}\sqrt{2\pi}} \\ \textcircled{2} &: \frac{\epsilon^2 - 2\epsilon u + u^2}{\sigma_v^2} + \frac{u^2}{\sigma_u^2} \\ &= \frac{\epsilon^2(\sigma_v^2+\sigma_u^2)}{\sigma_v^2(\sigma_v^2+\sigma_u^2)} - \frac{(2\epsilon u/\sqrt{\sigma_v^2+\sigma_u^2})(\sqrt{\sigma_v^2+\sigma_u^2}/\sigma_v\sigma_u)}{\sigma_v^2/(\sigma_v\sigma_u)} + \frac{u^2(\sigma_v^2+\sigma_u^2)}{\sigma_v^2\sigma_u^2} \\ &= \frac{\epsilon^2}{\sigma_v^2+\sigma_u^2}\Big(1 + \frac{\sigma_u^2}{\sigma_v^2}\Big) - \frac{2\epsilon u}{\sqrt{\sigma_v^2+\sigma_u^2}}\frac{\sqrt{\sigma_v^2+\sigma_u^2}}{\sigma_v\sigma_u}\frac{\sigma_u}{\sigma_v} + u^2\frac{\sigma_v^2+\sigma_u^2}{\sigma_v^2\sigma_u^2}. \end{aligned}$$

Let $\sigma^2 = \sigma_v^2 + \sigma_u^2$, $\lambda = \sigma_u/\sigma_v$, and $\delta^2 = (\sigma_v^2+\sigma_u^2)/(\sigma_v^2\sigma_u^2)$. Then,

$$\begin{aligned} f_{\epsilon,u}(\epsilon, u) &= \frac{\delta}{\sqrt{2\pi}}\frac{2}{\sigma\sqrt{2\pi}}\exp\Big\{-\frac{1}{2}\Big(\frac{\epsilon^2}{\sigma^2} + \frac{\epsilon^2\lambda^2}{\sigma^2} - 2\frac{\epsilon u}{\sigma}\delta\lambda + u^2\delta^2\Big)\Big\} \\ &= \frac{2}{\sigma\sqrt{2\pi}}\exp\Big(-\frac{\epsilon^2}{2\sigma^2}\Big)\frac{\delta}{\sqrt{2\pi}}\exp\Big\{-\frac{1}{2}\Big(\frac{\epsilon\lambda}{\sigma} - u\delta\Big)^2\Big\}. \end{aligned}$$

As the support of u is $[0, \infty)$, the density of ϵ_i is

$$\begin{aligned} f_\epsilon(\epsilon) &= \int_0^\infty f_{\epsilon,u}(\epsilon, u)du \\ &= \frac{2}{\sigma\sqrt{2\pi}}\exp\Big(-\frac{\epsilon^2}{2\sigma^2}\Big)\int_0^\infty \frac{\delta}{\sqrt{2\pi}}\exp\Big\{-\frac{1}{2}\Big(\frac{\epsilon\lambda}{\sigma} - u\delta\Big)^2\Big\}du. \end{aligned}$$

Let $\gamma = \frac{\epsilon\lambda}{\sigma} - u\delta$. Then, if $u = 0$, $\gamma = \frac{\epsilon\lambda}{\sigma}$, and if $u \to \infty$, $\gamma \to -\infty$. Also, $\frac{d\gamma}{du} = -\delta$, that is $du = -\frac{1}{\delta}d\gamma$. Therefore,

$$\begin{aligned} f_\epsilon(\epsilon) &= \frac{2}{\sigma}\frac{1}{\sqrt{2\pi}}\exp\Big\{-\frac{1}{2}\Big(\frac{\epsilon}{\sigma}\Big)^2\Big\}\int_{\frac{\epsilon\lambda}{\sigma}}^{-\infty}\frac{\delta}{\sqrt{2\pi}}\exp\Big(-\frac{1}{2}\gamma^2\Big) - \frac{1}{\delta}d\gamma \\ &= \frac{2}{\sigma}\cdot\phi\Big(\frac{\epsilon}{\sigma}\Big)\cdot\Big\{-\Phi(\gamma)\Big|_{\frac{\epsilon\lambda}{\sigma}}^{-\infty}\Big\} \\ &= \frac{2}{\sigma}\cdot\phi\Big(\frac{\epsilon}{\sigma}\Big)\cdot\Phi\Big(\frac{\epsilon\lambda}{\sigma}\Big). \end{aligned}$$

APPENDIX B

Tables and Figures

Table 1.B.1: Descriptive Statistics of Key Variables for 2019

	Mean	Std. Dev	Min	Max
Total Cost (C_i, \$ million)	82.7	1,211.5	0.1	51,615.0
Interest Expenses (C_{i1}, \$ million)	30.3	434.2	0.0	17,008.0
Salaries (C_{i2}, \$ million)	43.5	662.2	0.1	28,538.0
Fixed Assets Expenses (C_{i3}, \$ million)	8.9	140.6	0.0	6,069.0
Loans (y_{i1}, \$ million)	1,995.6	26,598.9	0.0	969,383.0
Other Earning Assets (y_{i2}, \$ million)	1,274.8	24,618.5	0.3	1,174,359.0
Funds that Bank owes (x_{i1}, \$ million)	3,054.9	47,236.1	0.5	1,980,733.1
Number of Full-time Employees (x_{i2})	402	5,552	2	232,982
Fixed Assets (x_{i3}, \$ million)	36.6	504.6	0.0	22,432.0
$w_{i1} = C_{i1}/x_{i1}$	0.0090	0.0045	0.0000	0.0381
$w_{i2} = C_{i2}/x_{i2}$	0.0839	0.0283	0.0056	0.3706
$w_{i3} = C_{i3}/x_{i3}$	0.4039	0.8936	0.0075	22.7273

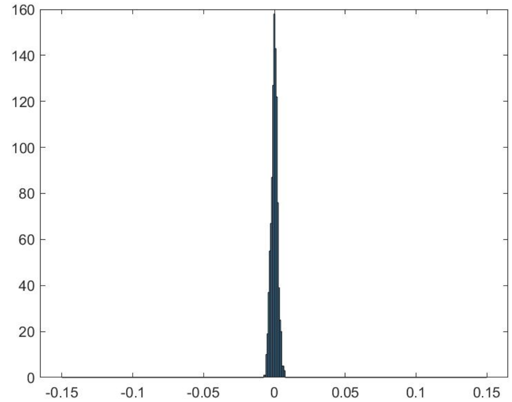
Table 1.B.2: Descriptive Statistics of Key Variables for 2020

	Mean	Std. Dev	Min	Max
Total Cost (C_i, \$ million)	73.0	1,000.1	0.0	40,331.0
Interest Expenses (C_{i1}, \$ million)	15.3	172.6	0.0	7,459.0
Salaries (C_{i2}, \$ million)	47.9	706.8	0.0	28,982.0
Fixed Assets Expenses (C_{i3}, \$ million)	9.8	151.8	0.0	6,362.0
Loans (y_{i1}, \$ million)	2,111.3	26,425.0	0.0	995,415.0
Other Earning Assets (y_{i2}, \$ million)	1,872.9	36,520.7	0.3	1,801,495.0
Funds that Bank owes (x_{i1}, \$ million)	3,737.5	58,393.4	0.5	2,626,377.0
Number of Full-time Employees (x_{i2})	417	5,689	2	233,403
Fixed Assets (x_{i3}, \$ million)	38.1	523.5	0.0	23,184.0
$w_{i1} = C_{i1}/x_{i1}$	0.0062	0.0035	0.0000	0.0285
$w_{i2} = C_{i2}/x_{i2}$	0.0889	0.0306	0.0000	0.3616
$w_{i3} = C_{i3}/x_{i3}$	0.4005	0.7763	0.0025	13.0000

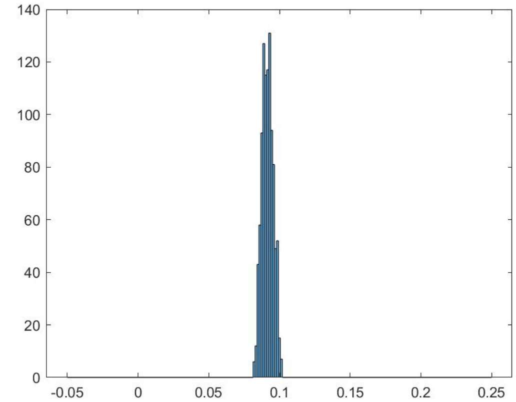
Table 1.B.3: Classification of Banks

Asset Size	
- Large Banks	Banks with assets greater than \$1 billion
- Small Banks	Banks with assets less than \$1 billion
Charter Class	
- N	Commercial banks, federal charter, fed member
- NM	Commercial banks, state charter, fed non-member
- SM	Commercial or savings banks, state charter, fed member
- SB	Savings banks, state charter
- SA	Thrifts, federal or state charter

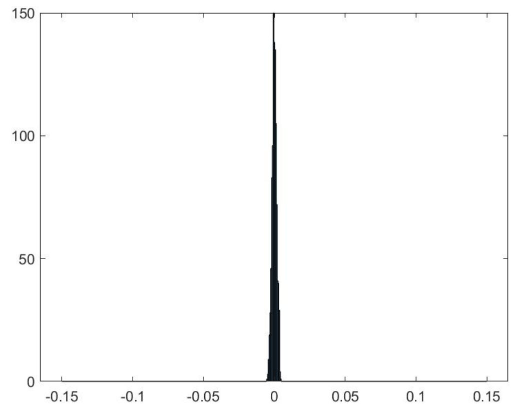
Figure 1.B.1: Sample Correlations (The APS-3-A Copula)



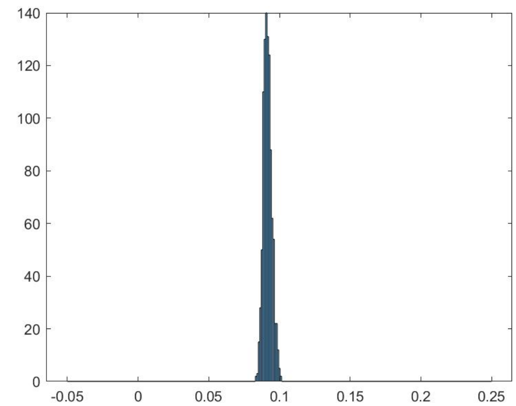
(a) $\text{corr}(\omega_1, \omega_2)$



(b) $\text{corr}(\omega_1, (\omega_2 - 0.5)^2)$



(c) $\text{corr}(\omega_1, \omega_3)$



(d) $\text{corr}(\omega_1, (\omega_3 - 0.5)^2)$

Figure 1.B.2: $\hat{f}_{e_2}(e_2)$ and $\hat{f}_{\eta_2|e_2}(\eta_2|e_2)$ for 2020

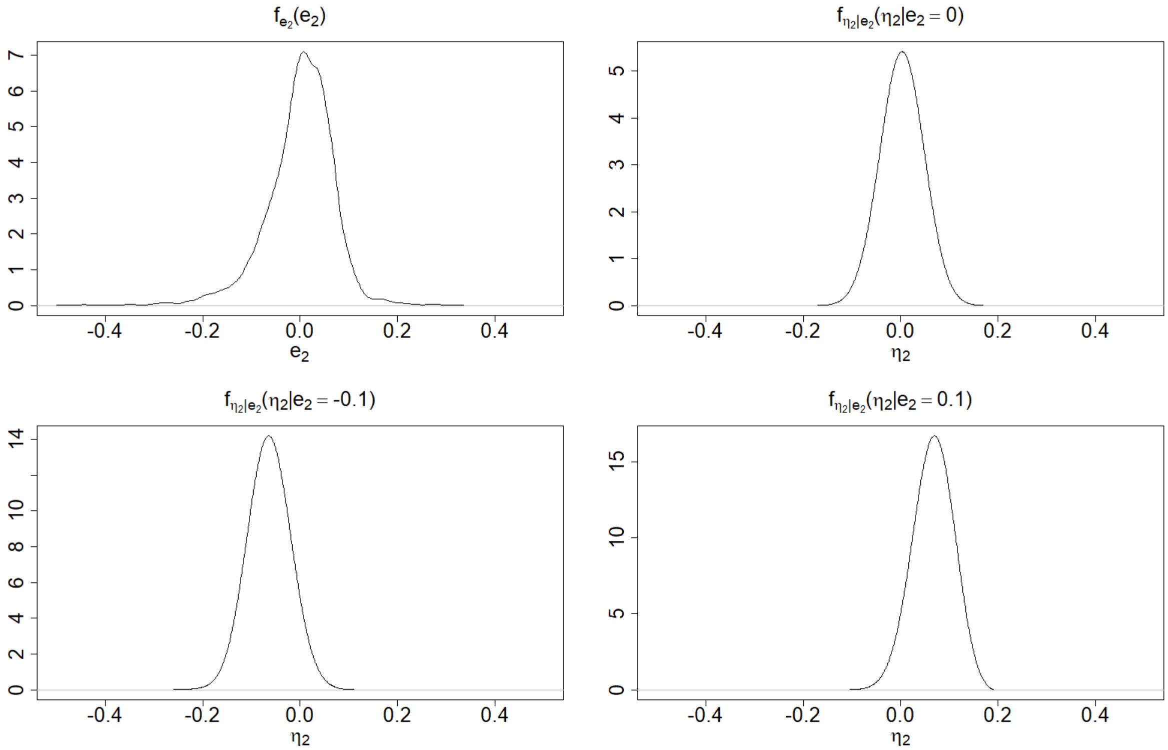
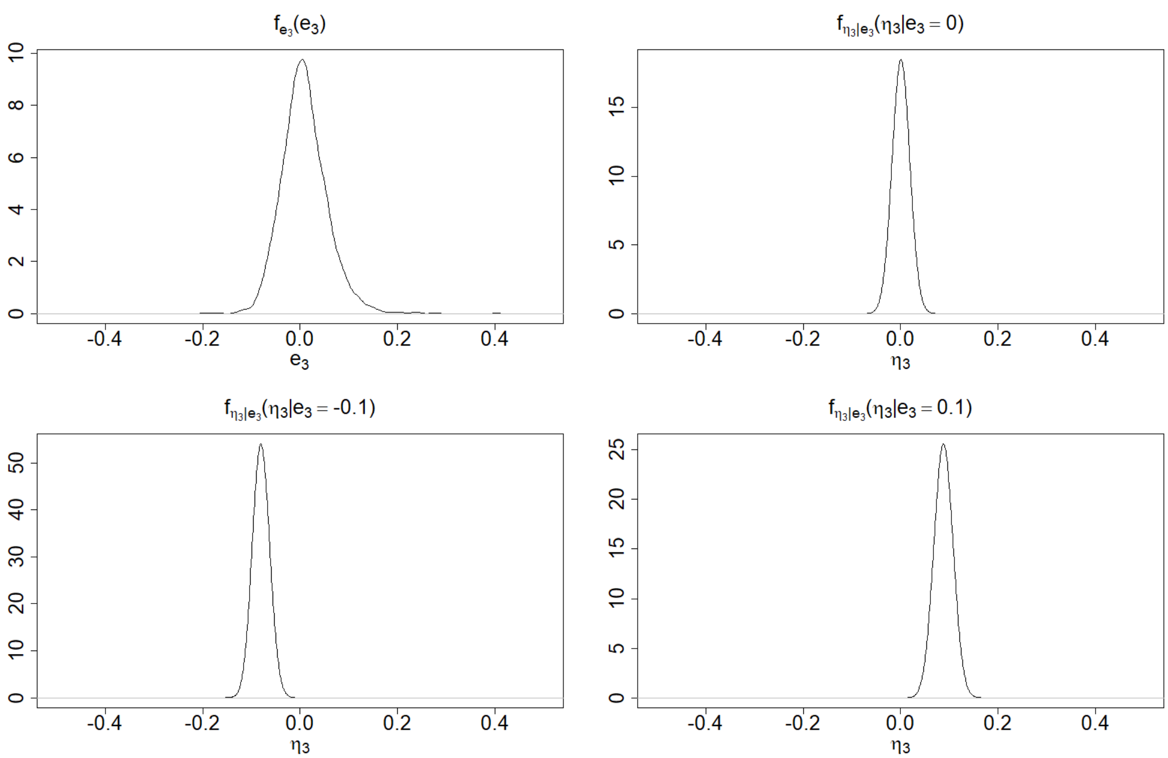


Figure 1.B.3: $\hat{f}_{e_3}(e_3)$ and $\hat{f}_{\eta_3|e_3}(\eta_3|e_3)$ for 2020



CHAPTER 2

MEASUREMENT AND DECOMPOSITION OF COST INEFFICIENCY USING COPULAS: EVIDENCE FROM MONTE CARLO SIMULATIONS

2.1 Introduction

This paper provides methods for copula-based simulations and demonstrates the performance of the estimation strategy proposed by Ryu (2021). First, a method to generate pseudo data using the APS copulas developed by Amsler et al. (2021) is presented; it applies the inverse Rosenblatt transformation and the inverse transformation method. Second, given the data generating process, quasi-Monte Carlo simulations are conducted in order to confirm the validity of the estimation strategy in Ryu (2021) that can measure and decompose technical and allocative inefficiency.

Amsler et al. (2021) conduct Monte Carlo simulations to show that the stochastic production frontier model employing the APS-2-A copula can be reliably estimated. The model used for their simulations can be written as

$$\begin{aligned} y_i &= \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + v_i - u_i^T \\ x_{i1} - x_{i2} &= \ln\left(\frac{\beta_1 w_{i2}}{\beta_2 w_{i1}}\right) + e_{i2}, \end{aligned} \tag{2.1.1}$$

where y_i is the natural log of output of producer i, x_{i1} and x_{i2} are the natural log of inputs, $v_i \in \mathbb{R}$ is a random disturbance, $u_i^T \in \mathbb{R}_+$ represents technical inefficiency, $w_{ij} \in \mathbb{R}_{++},\ j = 1, 2$, are the price of input j, and $e_{i2} \in \mathbb{R}$ is a two-sided term capturing allocative inefficiency and noise. They assume that a firm produces one output given two inputs, and u_i^T and e_{i2} are linked by the APS-2 copulas such that u_i^T is uncorrelated with e_{i2} but positively correlated with $|e_{i2}|$. Since their model uses the method of simulated likelihood for estimation, it requires to draw a $N \times 1$ vector of random numbers, where N is the number of producers, from the distribution of u_i^T, such as the half normal distribution. This is because the joint density of their model is

$$f_{\epsilon,e_2}(\epsilon,e_2) = f_{e_2}(e_2)\cdot \mathbb{E}_{u^T}[c(\omega_1,\omega_2)\cdot f_v(\epsilon+u^T)],$$

where $\epsilon = v - u^T$, $\mathbb{E}_{u^T}$ represents the expectation with respect to the distribution of u_i^T, $\omega_1 = F_1(u^T)$, $\omega_2 = F_2(e_2)$, and $F_1(u^T)$ and $F_2(e_2)$ are marginal cumulative distribution functions of u_i^T and e_{i2}, respectively. Furthermore, if a firm produces one output using three inputs, we can extend the above model, and the joint density in this case becomes

$$f_{\epsilon,e_2,e_3}(\epsilon,e_2,e_3) = f_{e_2}(e_2)\cdot f_{e_3}(e_3)\cdot \mathbb{E}_{u^T}[c(\omega_1,\omega_2,\omega_3)\cdot f_v(\epsilon+u^T)],$$

where $\omega_3 = F_3(e_3)$, and $F_3(e_3)$ is a marginal cumulative distribution function of e_{i3} (Amsler et al., 2021, p.6). It implies that it requires to draw a $N\times 1$ vector of random numbers as well even if the number of inputs increases.

However, numerous studies on efficiency analysis have assumed that producers use more than two inputs. Furthermore, we need to confirm whether an estimation strategy employing copulas would produce reliable estimates in more complex settings. For example, consider a translog cost system that can measure and decompose technical and allocative inefficiency, which can be written as

$$\begin{aligned} \ln C_i &= \ln C(\boldsymbol{y}_i,\boldsymbol{w}_i) + v_i + u_i^T + g(\boldsymbol{\xi}_i) \\ s_{ij} &= s_j(\boldsymbol{y}_i,\boldsymbol{w}_i) + h_j(\boldsymbol{\xi}_i) + \nu_{ij},\ j = 2,\cdots,J, \end{aligned} \tag{2.1.2}$$

where C_i is the actual cost of producer i, $C(\boldsymbol{y}_i,\boldsymbol{w}_i)$ is the deterministic kernel of the stochastic cost frontier, $\boldsymbol{y}_i \in \mathbb{R}_+^M$ is a vector of M outputs, $\boldsymbol{w}_i \in \mathbb{R}_{++}^J$ is a vector of input prices, $v_i \in \mathbb{R}$ is a random disturbance, $u_i^T \in \mathbb{R}_+$ represents a cost increase due to technical inefficiency, $g(\boldsymbol{\xi}_i) \in \mathbb{R}_+$ represents a cost increase due to allocative inefficiency, $\boldsymbol{\xi}_i = (\xi_{i2},\cdots,\xi_{iJ})$, ξ_{ij} represents producers' allocative inefficiency for the input pair $(j,1)$, $s_{ij} \in [0,1]$ is the actual cost share of input j, $s_j(\boldsymbol{y}_i,\boldsymbol{w}_i) \in [0,1]$ is the optimum cost share of input j, $h_j(\boldsymbol{\xi}_i) \in \mathbb{R}$ is the disturbance due to allocative inefficiency, and $\nu_{ij} \in \mathbb{R}$ is additive noise. We can assume that u_i^T and $\boldsymbol{\xi}_i$ are linked by the APS copulas such that u_i^T is uncorrelated with $\xi_{i2}, \cdots, \xi_{iJ}$ but positively correlated with $|\xi_{i2}|, \cdots, |\xi_{iJ}|$. Then, the joint density of this model is

$$f_{\epsilon,\boldsymbol{\nu}}(\epsilon, \boldsymbol{\nu}) = \mathbb{E}_{\boldsymbol{\zeta}}\Bigg[f_v\bigg(\epsilon - z_1\big(T(\boldsymbol{\zeta})\big) - g\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big); \boldsymbol{\theta}\bigg) \cdot f_{\nu}\bigg(e_2 - h_2\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big), \cdots, e_J - h_J\Big(\boldsymbol{z}_2\big(T(\boldsymbol{\zeta})\big)\Big); \boldsymbol{\theta}\bigg)\Bigg],$$

where $\mathbb{E}_{\boldsymbol{\zeta}}$ represents the expectation with respect to the distribution of $\boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_J)$, $\zeta_j \sim U[0,1]$, $\epsilon = v + u^T + g(\boldsymbol{\xi})$, z_1 and $\boldsymbol{z}_2$ are functions that transform CDF values to random numbers, T is the inverse function of the Rosenblatt transformation, and $\boldsymbol{\theta}$ is the parameters. Contrary to Amsler et al. (2021), it requires to draw $N \times J$ uniform random numbers $\boldsymbol{\zeta}$, where J represents the number of inputs, that each columns are uncorrelated. Hence, it is necessary to conduct another set of simulations that allows more inputs in order to check the validity of the economic model applying the APS copulas. In addition, methods of data generating process need to be provided in order to conduct simulations.

Lastly, the plausibility of the assumption in Greene (1980) can be examined by conducting simulations. As discussed in Bauer (1990) and Kumbhakar and Lovell (2000), an econometric issue occurs in a cost system that employs flexible functional forms, such as a translog function. The key question is how to model the relationship between $u_i = u_i^T + g(\boldsymbol{\xi}_i)$ and $e_{ij} = h_j(\boldsymbol{\xi}_i) + \nu_{ij}$ of (2.1.2). Greene (1980) proposes a solution to this problem, which assumes that the disturbance in the cost function, $\epsilon_i = v_i + u_i$, and the disturbance in the cost share equations, e_{ij}, are independent. By conducting simulations using a pseudo-data set that assumes technical and allocative inefficiency are linked by the APS copulas, we can verify the validity of the assumption in Greene (1980).

The remainder of the chapter is organized as follows. Section 2.2 illustrates how to draw observations from the APS-3-A copula that corresponds to a three-input case.[1] Section 2.3 shows the data generating process. Section 2.4 presents the Monte Carlo simulation results. Section 2.5 concludes the chapter.

[1]Amsler et al. (2021) provide the procedure to draw observations from the APS-2-A copula. It is summarized in Appendix A.

2.2 Simulating from the APS-3-A Copula

Since a copula itself is a cumulative distribution function whose marginal distributions follow $U[0,1]$, one needs to find the conditional copula in order to employ the Rosenblatt transformation. Copula arguments can then be obtained by applying the process described in this section that focuses on the APS-3-A copula.

2.2.1 Derivation of Conditional Distributions

Let $c_{123}(\omega_1,\omega_2,\omega_3)$ be a copula density, where $(\omega_1,\omega_2,\omega_3)\in[0,1]^3$ are copula arguments. It is assumed that ω_1 is uncorrelated with ω_2 and ω_3 but correlated with $|\omega_2-0.5|$ and $|\omega_3-0.5|$. In addition, the dependence between ω_2 and ω_3 is captured by any bivariate copula.

For computational ease, ω_3, ω_2, and ω_1 are generated sequentially; this the use of the conditional copulas $C_{1|23}(\omega_1|\omega_2,\omega_3)$ and $C_{2|3}(\omega_2|\omega_3)$ for the Rosenblatt transformation. In this subsection, the conditional copulas, $C_{1|23}(\omega_1|\omega_2,\omega_3)$ and $C_{2|3}(\omega_2|\omega_3)$, are derived, then methods to obtain the copula arguments, ω_3, ω_2, and ω_1, are presented.

2.2.1.1 Conditional Copula $C_{1|23}(\omega_1|\omega_2,\omega_3)$

Assume that ω_2 and ω_3 follow the bivariate normal copula such that

$$
\begin{aligned}
C_{23}(\omega_2,\omega_3) &= \Phi_2(\Phi^{-1}(\omega_2),\Phi^{-1}(\omega_3);\rho)\\
c_{23}(\omega_2,\omega_3) &= \frac{\partial^2}{\partial\omega_2\partial\omega_3}C_{23}(\omega_2,\omega_3)\\
&= \frac{\partial^2\Phi_2(\Phi^{-1}(\omega_2),\Phi^{-1}(\omega_3);\rho)}{\partial\Phi^{-1}(\omega_2)\partial\Phi^{-1}(\omega_3)}\frac{\partial\Phi^{-1}(\omega_2)}{\partial\omega_2}\frac{\partial\Phi^{-1}(\omega_3)}{\partial\omega_3}\\
&= \frac{\phi_2(\Phi^{-1}(\omega_2),\Phi^{-1}(\omega_3);\rho)}{\phi(\Phi^{-1}(\omega_2))\phi(\Phi^{-1}(\omega_3))}\\
&= \frac{\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\Big(-\frac{\Phi^{-1}(\omega_2)^2-2\rho\Phi^{-1}(\omega_2)\Phi^{-1}(\omega_3)+\Phi^{-1}(\omega_3)^2}{2(1-\rho^2)}\Big)}{\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{\Phi^{-1}(\omega_2)^2}{2}\Big)\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{\Phi^{-1}(\omega_3)^2}{2}\Big)}\\
&= \frac{1}{\sqrt{1-\rho^2}}\exp\Big[-\frac{\rho^2\Phi^{-1}(\omega_2)^2-2\rho\Phi^{-1}(\omega_2)\Phi^{-1}(\omega_3)+\rho^2\Phi^{-1}(\omega_3)^2}{2(1-\rho^2)}\Big],
\end{aligned}
$$

where Φ is the cumulative distribution function of the standard normal distribution, ϕ is the probability density function of the standard normal distribution, Φ_2 is the cumulative distribution function of the standardized bivariate normal distribution, ϕ_2 is the probability density function of the standardized bivariate normal distribution, and ρ is the correlation parameter.

Given that $C_{1|23}(\omega_1|\omega_2,\omega_3) = \int c_{1|23}(t_1|\omega_2,\omega_3)dt_1$, it is required to obtain the conditional density $c_{1|23}(\omega_1|\omega_2,\omega_3)$, which is

$$\begin{aligned}
c_{1|23}(\omega_1|\omega_2,\omega_3) &= \frac{c_{123}(\omega_1,\omega_2,\omega_3)}{f_{23}(\omega_2,\omega_3)} \\
&= \frac{c_{123}(\omega_1,\omega_2,\omega_3)}{c_{23}(F_2(\omega_2),F_3(\omega_3))f_2(\omega_2)f_3(\omega_3)} \\
&= \frac{c_{123}(\omega_1,\omega_2,\omega_3)}{c_{23}(\omega_2,\omega_3)}.
\end{aligned}$$

The second inequality holds because

$$\begin{aligned}
F_{XY}(x,y) &= C_{XY}\big(F_X(x),F_Y(y)\big) \\
\Rightarrow\ f_{XY}(x,y) &= \frac{\partial^2 F_{XY}(x,y)}{\partial x\partial y} \\
&= \frac{\partial^2 C_{XY}\big(F_X(x),F_Y(y)\big)}{\partial x\partial y} \\
&= \frac{\partial^2 C_{XY}\big(F_X(x),F_Y(y)\big)}{\partial F_X(x)\partial F_Y(y)}\frac{\partial F_X(x)}{\partial x}\frac{\partial F_Y(y)}{\partial y} \\
&= c_{XY}\big(F_X(x),F_Y(y)\big)f_X(x)f_Y(y)
\end{aligned}$$

and the third equality holds as $\omega_2,\omega_3 \sim U[0,1]$. Note that

$$\begin{aligned}
c_{123}(\omega_1,\omega_2,\omega_3) &= 1 + [1+\theta_{12}(1-2\omega_1)\{1-12(\omega_2-0.5)^2)\}-1] \\
&\qquad + [1+\theta_{13}(1-2\omega_1)\{1-12(\omega_3-0.5)^2)\}-1] + \{c_{23}(\omega_2,\omega_3)-1\} \\
&= g_2(1-2\omega_1) + g_3(1-2\omega_1) + c_{23}(\omega_2,\omega_3) \\
&= h(1-2\omega_1) + c_{23}(\omega_2,\omega_3),
\end{aligned}$$

where $g_2 = \theta_{12}\{1-12(\omega_2-0.5)^2\}$, $g_3 = \theta_{13}\{1-12(\omega_3-0.5)^2\}$, and $h = g_2+g_3$.

Hence, the conditional copula $C_{1|23}(\omega_1|\omega_2,\omega_3)$ is

$$\begin{aligned}
C_{1|23}(\omega_1|\omega_2,\omega_3) &= \int_0^{\omega_1} c_{1|23}(t_1|\omega_2,\omega_3)dt_1 \\
&= \int_0^{\omega_1} \frac{h(1-2t_1)+c_{23}(\omega_2,\omega_3)}{c_{23}(\omega_2,\omega_3)}dt_1 \\
&= [\frac{1}{c_{23}(\omega_2,\omega_3)}\{h(t_1-t_1^2)+c_{23}(\omega_2,\omega_3)t_1\}]\Big|_0^{\omega_1} \\
&= \frac{1}{c_{23}(\omega_2,\omega_3)}\{h(\omega_1-\omega_1^2)+c_{23}(\omega_2,\omega_3)\omega_1\}.
\end{aligned}$$

2.2.1.2 Conditional Copula $C_{2|3}(\omega_2|\omega_3)$

Given that $C_{2|3}(\omega_2|\omega_3) = \int c_{2|3}(t_2|\omega_3)dt_2$, it is required to obtain the conditional density $c_{2|3}(\omega_2|\omega_3)$, which is

$$c_{2|3}(\omega_2|\omega_3) = \frac{c_{23}(\omega_2,\omega_3)}{f_3(\omega_3)} = c_{23}(\omega_2,\omega_3).$$

The second equality holds because $\omega_3 \sim U[0,1]$.

Hence, the conditional copula $C_{2|3}(\omega_2|\omega_3)$ is

$$\begin{aligned}
C_{2|3}(\omega_2|\omega_3) &= \int_0^{\omega_2} c_{2|3}(t_2|\omega_3)dt_2 \\
&= \int_0^{\omega_2} \frac{1}{\sqrt{1-\rho^2}}\exp\Big[-\frac{\rho^2\Phi^{-1}(t_2)^2-2\rho\Phi^{-1}(t_2)\Phi^{-1}(\omega_3)+\rho^2\Phi^{-1}(\omega_3)^2}{2(1-\rho^2)}\Big]dt_2
\end{aligned}$$

Let $y_2 = \Phi^{-1}(t_2)$ and $y_3 = \Phi^{-1}(\omega_3)$. Note that $dt_2 = \phi(y_2)dy_2$. Then, integrate by substitution such as

$$\begin{aligned}
C_{2|3}(\omega_2|\omega_3) &= \int_{-\infty}^{\Phi^{-1}(\omega_2)} \frac{1}{\sqrt{1-\rho^2}}\exp\Big[-\frac{\rho^2y_2^2-2\rho y_2y_3+\rho^2y_3^2}{2(1-\rho^2)}\Big]\phi(y_2)dy_2 \\
&= \int_{-\infty}^{\Phi^{-1}(\omega_2)} \frac{1}{\sqrt{1-\rho^2}}\frac{\frac{1}{\sqrt{2\pi}}\exp\Big[-\frac{y_2^2-2\rho y_2y_3+\rho^2y_3^2}{2(1-\rho^2)}\Big]}{\frac{1}{\sqrt{2\pi}}\exp\Big[-\frac{y_2^2}{2}\Big]}\phi(y_2)dy_2 \\
&= \int_{-\infty}^{\Phi^{-1}(\omega_2)} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\exp\Big[-\frac{(y_2-\rho y_3)^2}{2(1-\rho^2)}\Big]dy_2 \\
&= \Phi\Big(\frac{\Phi^{-1}(\omega_2)-\rho\Phi^{-1}(\omega_3)}{\sqrt{1-\rho^2}}\Big).
\end{aligned}$$

2.2.2 Obtain Copula Arguments

2.2.2.1 Obtain ω_3

As $\omega_3 \sim U[0,1]$, $\zeta_3 = F_3(\omega_3) = \omega_3$. Therefore, draw ζ_3 from $U[0,1]$ and define $\omega_3 = \zeta_3$.

2.2.2.2 Obtain ω_2

First, draw ζ_2 from $U[0,1]$. Then, one can obtain ω_2 by solving the equation $C_{2|3}(\omega_2|\omega_3) - \zeta_2 = 0$. It yields

$$\begin{aligned}
&\Phi\Big(\frac{\Phi^{-1}(\omega_2) - \rho\Phi^{-1}(\omega_3)}{\sqrt{1-\rho^2}}\Big) - \zeta_2 = 0 \\
\Rightarrow\ &\frac{\Phi^{-1}(\omega_2) - \rho\Phi^{-1}(\omega_3)}{\sqrt{1-\rho^2}} = \Phi^{-1}(\zeta_2) \\
\Rightarrow\ &\Phi^{-1}(\omega_2) = \rho\Phi^{-1}(\omega_3) + \sqrt{1-\rho^2}\Phi^{-1}(\zeta_2) \\
\Rightarrow\ &\omega_2 = \Phi\big(\rho\Phi^{-1}(\omega_3) + \sqrt{1-\rho^2}\Phi^{-1}(\zeta_2)\big)
\end{aligned}$$

2.2.2.3 Obtain ω_1

First, draw ζ_1 from $U[0,1]$. Then, one can obtain ω_1 by solving the equation $C_{1|23}(\omega_1|\omega_2,\omega_3) - \zeta_1 = 0$. It yields,

$$\begin{aligned}
&\frac{1}{c_{23}}\{(h(\omega_1 - \omega_1^2) + c_{23}\omega_1\} - \zeta_1 = 0 \\
\Rightarrow\ &h(\omega_1 - \omega_1^2) + c_{23}\omega_1 - c_{23}\zeta_1 = 0 \\
\Rightarrow\ &h\omega_1^2 - (h + c_{23})\omega_1 + c_{23}\zeta_1 = 0 \\
\Rightarrow\ &\omega_1 = \frac{(h + c_{23}) \pm \sqrt{(h + c_{23})^2 - 4hc_{23}\zeta_1}}{2h},
\end{aligned}$$

where $c_{23} = c_{2|3}(\omega_2|\omega_3)$.

It is necessary to check whether the square roots are real numbers. Given that $(\omega_2, \omega_3) \in [0,1]^2$ and $(\theta_{12}, \theta_{13}) \in [-0.5, 0.5]^2$, one can find the upper and lower bounds of g_2, g_3. If $x \in [0,1]$, $1 - 12(x - 0.5)^2 \in [-2, 1]$. Therefore, $(g_2, g_3) \in [-1,1]^2$ and $h \in [-2, 2]$. Since

$c_{23} > 0$ and $\zeta_1 \in [0,1]$, $(h+c_{23})^2 - 4hc_{23}\zeta_1 \geq 0$ when $h \in [-2,0)$. Also, for $h \in [0,2]$, $(h+c_{23})^2 - 4hc_{23}\zeta_1 = (h-c_{23})^2 + 4hc_{23}(1-\zeta_1) \geq 0$. Hence, the solutions are real numbers unless $h = 0$.

The remained question is which solution to take. Rewrite the solutions as

$$\begin{aligned}
\omega_1 &= \frac{(h+c_{23}) \pm \sqrt{(h+c_{23})^2 - 4hc_{23}\zeta_1}}{2h} \\
&= \frac{\{(h+c_{23}) \pm \sqrt{(h+c_{23})^2 - 4hc_{23}\zeta_1}\}\{(h+c_{23}) \mp \sqrt{(h+c_{23})^2 - 4hc_{23}\zeta_1}\}}{2h\{(h+c_{23}) \mp \sqrt{(h+c_{23})^2 - 4hc_{23}\zeta_1}\}} \\
&= \frac{(h+c_{23})^2 - (h+c_{23})^2 + 4hc_{23}\zeta_1}{2h\{(h+c_{23}) \mp \sqrt{(h+c_{23})^2 - 4hc_{23}\zeta_1}\}} \\
&= \frac{2c_{23}\zeta_1}{(h+c_{23}) \mp \sqrt{(h+c_{23})^2 - 4hc_{23}\zeta_1}}.
\end{aligned}$$

Consider the first solution $\omega_1 = \frac{(h+c_{23})+\sqrt{(h+c_{23})^2-4hc_{23}\zeta_1}}{2h} = \frac{2c_{23}\zeta_1}{(h+c_{23})-\sqrt{(h+c_{23})^2-4hc_{23}\zeta_1}}$. Then, the denominator $(h+c_{23}) - \sqrt{(h+c_{23})^2 - 4hc_{23}\zeta_1} < 0$ for $h \in [-2,0)$, which violates the condition that $\omega_1 \in [0,1]$. Now, consider the second solution $\omega_1 = \frac{(h+c_{23})-\sqrt{(h+c_{23})^2-4hc_{23}\zeta_1}}{2h} = \frac{2c_{23}\zeta_1}{(h+c_{23})+\sqrt{(h+c_{23})^2-4hc_{23}\zeta_1}}$. Given that $c_{23} > 0$ and $\zeta_1 \in [0,1]$, it is required to show that $\omega_1 \in [0,1]$. It can be shown as follows:

(i) $\omega_1 \geq 0$

As $c_{23} > 0$ and $\zeta_1 \in [0,1]$, the numerator is positive. So it suffices to show that the denominator is strictly positive. Since $\sqrt{(h+c_{23})^2 - 4hc_{23}\zeta_1} \geq 0$, the denominator is guaranteed to be strictly positive for $h \in (-c_{23}, 2]$; that is, if $c_{23} > 2$, the denominator is strictly positive. If $c_{23} \in (0,2]$ and $h \in [-2, -c_{23})$, one needs to compare the values of $|h+c_{23}|$ and $|\sqrt{(h+c_{23})^2 - 4hc_{23}\zeta_1}|$, which is equivalent to compare their squares. Note that $\{(h+c_{23})^2 - 4hc_{23}\zeta_1\} - (h+c_{23})^2 = -4hc_{23}\zeta_1 > 0$ unless $\zeta_1 = 0$.[2] Hence the denominator is strictly positive except the special case.

(ii) $\omega_1 \leq 1$

To find the maximum value of ω_1 given the arguments in the numerator, it is necessary to find the minimum value of the denominator. Let $\lambda(h) = (h+c_{23}) +$

[2] As $Z_1 \sim U[0,1]$, $\mathbb{P}(Z_1 = 0) = 0$.

$\sqrt{(h+c_{23})^2-4hc_{23}\zeta_1}$. Then,

$$\begin{aligned}\frac{\partial\lambda(h)}{\partial h} &= 1+\frac{1}{2}\{(h+c_{23})^2-4hc_{23}\zeta_1\}^{-\frac{1}{2}}\{2(h+c_{23})-4c_{23}\zeta_1\}\\ &= 1+\frac{(h+c_{23})-2c_{23}\zeta_1}{\sqrt{(h+c_{23})^2-4hc_{23}\zeta_1}}\end{aligned}$$

If $(h+c_{23})-2c_{23}\zeta_1 \geq 0$, $\frac{\partial\lambda(h)}{\partial h} > 0$. For $(h+c_{23})-2c_{23}\zeta_1 < 0$, compare the values of $|(h+c_{23})-2c_{23}\zeta_1|$ and $|\sqrt{(h+c_{23})^2-4hc_{23}\zeta_1}|$, which is equivalent to compare the values of $\{(h+c_{23})-2c_{23}\zeta_1\}^2$ and $(h+c_{23})^2-4hc_{23}\zeta_1$. Note that

$$\begin{aligned}&\{(h+c_{23})-2c_{23}\zeta_1\}^2-(h+c_{23})^2-4hc_{23}\zeta_1\\ &= (h+c_{23})^2-4c_{23}\zeta_1(h+c_{23})-(h+c_{23})^2+4hc_{23}\zeta_1\\ &= -4c_{23}\zeta_1\{(h+c_{23})-h\}\\ &= -4c_{23}^2\zeta_1 \leq 0,\end{aligned}$$

which implies $\frac{(h+c_{23})-2c_{23}\zeta_1}{\sqrt{(h+c_{23})^2-4hc_{23}\zeta_1}} \leq -1$, so $\frac{\partial\lambda(h)}{\partial h} \geq 0$. Therefore, given c_{23} and ζ_1, $\lambda(h)$ has the smallest value when $h=-2$. For $h=-2$, $\omega_1 = \frac{(2-c_{23})+\sqrt{(-2+c_{23})^2+8c_{23}\zeta_1}}{4}$. Note that

$$\begin{aligned}\omega_1 = \frac{(2-c_{23})+\sqrt{(-2+c_{23})^2+8c_{23}\zeta_1}}{4} &\leq 1\\ \Leftrightarrow \sqrt{(-2+c_{23})^2+8c_{23}\zeta_1} &\leq 2+c_{23}\\ \Leftrightarrow (-2+c_{23})^2+8c_{23}\zeta_1 &\leq (2+c_{23})^2\\ \Leftrightarrow 4-4c_{23}+c_{23}^2+8c_{23}\zeta_1 &\leq 4+4c_{23}+c_{23}^2\\ \Leftrightarrow 8c_{23}(1-\zeta_1) &\geq 0.\end{aligned}$$

The last inequality holds because of $c_{23} > 0$ and $\zeta_1 \in [0,1]$.■

Let $A = h+c_{23}$ and $B = c_{23}\zeta_1$, where $c_{23} = c_{2|3}(\omega_2|\omega_3)$. Then the solution is

$$\omega_1 = \frac{2B}{A+\sqrt{A^2-4hB}}.$$

2.3 Generating Pseudo Data

Consider the following stochastic cost frontier model such that

$$\begin{aligned}\ln C_i &= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + v_i + u_i^T + g(\boldsymbol{\xi}_i)\\ s_{ij} &= s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + h_j(\boldsymbol{\xi}_i) + \nu_{ij},\ j = 2, \cdots, J.\end{aligned}$$

Each component of the system defined in Section 2.1 can be written as

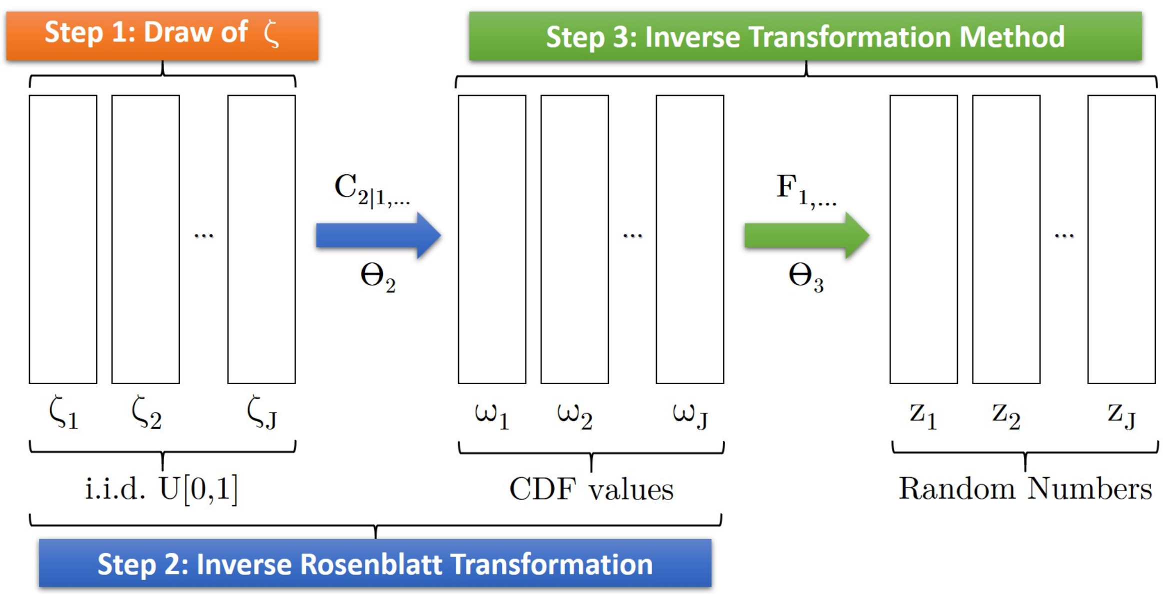
$$\begin{aligned}\ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) &= \beta_0 + \sum_{m=1}^{M} \beta_m^y (\ln y_{im}) + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} \beta_{mn}^{yy} (\ln y_{im})(\ln y_{in})\\ &\quad + \sum_{j=1}^{J} \beta_j^w (\ln w_{ij}) + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} \beta_{jk}^{ww} (\ln w_{ij})(\ln w_{ik})\\ &\quad + \sum_{m=1}^{M}\sum_{j=1}^{J} \beta_{mj}^{yw} (\ln y_{im})(\ln w_{ij})\\ s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) &= \beta_j^w + \sum_{k=1}^{J} \beta_{jk}^{ww} (\ln w_{ik}) + \sum_{m=1}^{M} \beta_{mj}^{yw} (\ln y_{im}),\ j = 2, \cdots, J\\ g(\boldsymbol{\xi}_i) &= \sum_{j=1}^{J} \beta_j^w \xi_{ij} + \sum_{j=1}^{J}\sum_{k=1}^{J} \beta_{jk}^{ww} (\ln w_{ij})\xi_{ik} + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} \beta_{jk}^{ww} \xi_{ij}\xi_{ik}\\ &\quad + \sum_{m=1}^{M}\sum_{j=1}^{J} \beta_{mj}^{yw} (\ln y_{im})\xi_{ij} + \ln \sum_{j=1}^{J} (s_{ij}^*/e^{\xi_{ij}})\\ h_j(\boldsymbol{\xi}_i) &= \frac{s_j(\boldsymbol{y}_i, \boldsymbol{w}_i)[1 - \{\sum_{k=1}^{J}(s_{ik}^*/e^{\xi_{ik}})\}e^{\xi_{ij}}] + \sum_{k=1}^{J} \beta_{jk}^{ww}\xi_{ik}}{\{\sum_{k=1}^{J}(s_{ik}^*/e^{\xi_{ik}})\}e^{\xi_{ij}}},\ j = 2, \cdots, J,\end{aligned}$$

where $s_{ij}^* = \ s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + \sum_{k=1}^{J} \beta_{jk}^{ww}\xi_{ik},\ j = 2, \cdots, J$, is the shadow cost share of input j for producer i who is assumed to be only allocatively inefficient.

Four types of pseudo-data sets are generated by the pair of the numbers of inputs (J) and outputs (M): two inputs - one output, two inputs - two outputs, three inputs - one output, and three inputs - two outputs, where each set includes 1000 producers ($N = 1000$). In order to construct pseudo-data sets, it is necessary to draw three types of variables: (i) variables consisting of deterministic kernels of the translog cost system, $\boldsymbol{y}_i$ and $\boldsymbol{w}_i$; (ii) random components of the model not linked by the APS copulas, v_i and $\boldsymbol{\nu_i}$; and (iii) random components of the model linked by the APS copulas, u_i^T and $\boldsymbol{\xi}_i$. The dependent variables of the translog cost system, C_i and $s_{i2}, \cdots, s_{iJ}$, can be calculated using those variables and parameters.

The first and second types of variables are generated as follows. Note that the underlying production technology of the translog cost function is unknown, as the translog cost function has neither a closed-form dual production nor a transformation function (Kumbhakar and Lovell, 2000, p.154). Hence, outputs, $\ln y_{i1}$ and $\ln y_{i2}$, are drawn from $\Gamma(2,2)$ and $\Gamma(3,2)$ with a correlation coefficient of 0.8. Input prices, $\ln w_{i1}$, $\ln w_{i2}$, and $\ln w_{i3}$, are independently drawn from $N(1, 0.1^2)$, $N(2, 0.1^2)$, and $N(3, 0.1^2)$ distributions, respectively. Stochastic noise terms, v_i and $\boldsymbol{\nu_i}$, are generated from $N(0, \sigma_v^2)$ and $N(0, \boldsymbol{\Sigma}_\nu)$. The third type of variables are generated in the reverse order of the procedure for a change of variables that is used to derive the joint density in Ryu (2021). Figure 2.3.1 illustrates the procedure to simulate $\boldsymbol{Z} = (u_i^T, \xi_{i2}, \cdots, \xi_{iJ})$.

Figure 2.3.1: Procedure to Simulate $\boldsymbol{Z} = (u_i^T, \xi_{i2}, \cdots, \xi_{iJ})$



To be specific, random components of the model linked by the APS copulas are simulated by the following procedure. First, draw independent random numbers, $\boldsymbol{\zeta}$, from the uniform distribution over $[0, 1]$. Second, produce CDF values, $\boldsymbol{\omega}$, given $\boldsymbol{\theta}_2$ and the inverse of conditional APS copula functions employing the inverse Rosenblatt transformation[3]. Third, generate $\boldsymbol{Z} = (Z_1, \boldsymbol{Z}_2) = (u_i^T, \xi_{i2}, \cdots, \xi_{iJ})$ given $\boldsymbol{\theta}_3$ and the inverse of cumulative distribution functions by the inverse transformation method.

There is a practical issue in the first step. In order to apply the inverse Rosenblatt transformation, it is essential to draw an $N \times J$ array of uniformly distributed random numbers in which columns are uncorrelated in order to correctly estimate $\boldsymbol{\theta}_2$. Because of the practical difficulties to use truly random variables in the Monte Carlo methods[4], two methods for generating random numbers are generally used in applications: (i) pseudo-random number generators (PRNGs); and (ii) quasi-random number generators (QRNGs). Two functions provided by MATLAB are considered to draw $\zeta_1, \cdots, \zeta_J$: (i) `rand` that generates uniformly distributed pseudo-random numbers; and (ii) `haltonset` that produces Halton sequences that make up the representative example of quasi-random number sequences.

For illustrative purposes, Figure 2.B.1 shows sample correlations between two uniform random variables, ζ_1 and ζ_2 to compare the performance of the two functions. The number of replications is 1000, where three sample sizes $N \in \{100, 1000, 10000\}$ are considered for each replication. Figures 2.B.1(a), 2.B.1(c), and 2.B.1(e) are obtained by the function `rand`. Figures 2.B.1(b), 2.B.1(d), and 2.B.1(f) are obtained by the function `haltonset`, where several methods are applied to address the inherent issue that the points of a quasi-random sequence are correlated.[5]

As shown in Figure 2.B.1, although the columns of the arrays produced by `rand` are theoretically uncorrelated, some pairs of ζ_1 and ζ_2 are highly correlated, especially when the number of draws are not sufficiently large. Also, even if $N = 10000$, some pairs of ζ_1 and ζ_2 seems to be significantly correlated. By contrast, correlations between ζ_1 and ζ_2 generated

[3]It is known as "conditional distribution method." Please refer to Embrechts et al. (2003) and Cambou et al. (2017) for more details.

[4]Please see Ch.8 of Judd (1998) for more details.

[5]MATLAB provides three methods: (i) omit initial points in the sequence; (ii) set interval between points; and (iii) scramble the sequence. In the simulation, the first 100,000 values of the Halton point set are omitted, the every 100,001st point are retained, and then the Halton point set is scrambled by a reverse-radix operation.

by `haltonset` are mostly negligible. Hence, a QRNG, `haltonset` of MATLAB, is used to generate $\zeta_1, \cdots, \zeta_J$.

In the second step, $\omega_1, \cdots, \omega_J$ are generated through a consecutive process based on the Rosenblatt transformation by solving equations such that

$$\begin{aligned}
\zeta_J &= F_J(\omega_J) \\
\zeta_{J-1} &= F_{J-1|J}(\omega_{J-1}|\omega_J) \\
\zeta_{J-2} &= F_{J-2|J-1,J}(\omega_{J-2}|\omega_{J-1}, \omega_J) \\
&\vdots \\
\zeta_1 &= F_{1|2,\cdots,J}(\omega_1|\omega_2, \cdots, \omega_J).
\end{aligned}$$

As derived in Appendix A, ω_1 and ω_2 for the APS-2-A copula are generated as follows:

$$\begin{aligned}
\omega_2 &= \zeta_2 \\
\omega_1 &= \frac{2\zeta_1}{A + \sqrt{A^2 - 4(A-1)\zeta_1}},
\end{aligned}$$

where $A = 1 + g_2$ and $g_2 = \theta_{12}\{1 - 12(\omega_2 - 0.5)^2\}$. For the association parameter of the APS-2-A copula, three values of $\theta_{12} \in \{0, 0.2, 0.4\}$ are considered. Also, Section 2.2 shows that ω_1, ω_2, and ω_3 for the APS-3-A copula are generated as follows:

$$\begin{aligned}
\omega_3 &= \zeta_3 \\
\omega_2 &= \Phi\big(\rho\Phi^{-1}(\omega_3) + \sqrt{1-\rho^2}\Phi^{-1}(\zeta_2)\big) \\
\omega_1 &= \frac{2B}{A + \sqrt{A^2 - 4hB}},
\end{aligned}$$

where ρ is the correlation parameter of the bivariate Gaussian copula, $A = h + c_{23}(\omega_2, \omega_3)$, $h = g_2 + g_3$, $g_2 = \theta_{12}\{1 - 12(\omega_2 - 0.5)^2\}$, $g_3 = \theta_{13}\{1 - 12(\omega_3 - 0.5)^2\}$, and $B = c_{23}(\omega_2, \omega_3)\zeta_1$. For the association parameter of the APS-3-A copula, two pairs of $(\theta_{12}, \theta_{13}) \in \{(0, 0), (0.2, 0.2)\}$ are considered to generate data sets, and the correlation parameter of the bivariate Gaussian copula ρ is set to -0.5.

2.4 Results of Monte Carlo Simulations

Here, two sets of simulations are conducted. The first set of simulations (Simulation I) is conducted to confirm the validity of the estimation strategy in Ryu (2021). The purpose of the second set of simulations (Simulation II) is to examine the plausibility of the assumption in Greene (1980) when technical and allocative inefficiency are indeed dependent. The number of replications is 1,000 for both sets of simulations.

2.4.1 Simulation I

A set of quasi-Monte Carlo simulations based on quasi-random sequences is conducted. Halton sequences are also used for the simulations as the data generating process but for different reasons. The joint density of X and $\boldsymbol{Y}$ involves a multidimensional integral, in which a quasi-Monte Carlo integration is generally superior to standard Monte Carlo methods in terms of integration error and its convergence rate.[6] For example, Morokoff and Caflisch (1995) show that a quasi-Monte Carlo method using a Halton sequence has the lowest integration error and the fastest convergence rate up to around six dimensions among (quasi-)Monte Carlo methods using Halton, Sobol, Faure, and pseudo-random sequences. For copula sampling, in addition, Cambou et al. (2017) show that replacing PRNGs with QRNGs for integration also improves performance, reducing the variance of the obtained estimators and improving the convergence rate of the variance.

Tables 2.4.1 to 2.4.4 report the results of quasi-Monte Carlo simulations for $\theta_{12} = 0.4$ for two-input cases or $\theta_{12} = \theta_{13} = 0.2$ for three-input cases. Other results are reported in Tables 2.B.1 to 2.B.6 of Appendix B. R sets of $N \times J$ Halton sequences are drawn for estimation, where $R = 10,000$, $N = 1,000$, and $J = 2$ or 3. Although the standard deviations of the association parameters θ_{12} and θ_{13} are somewhat large, the results suggest that the stochastic cost frontier model in Ryu (2021) can be also reliably estimated like the stochastic production frontier model as in Amsler et al. (2021). That is, the modified translog cost system based on

[6] Please refer to Caflisch (1998) for more details.

Kumbhakar (1997) and the APS copulas is estimable by the maximum simulated likelihood estimator established on the probability integral transformation and the copula-based version of the Rosenblatt transformation.

Table 2.4.1: Result of Simulation I ($J = 2,\ M = 1,\ \theta_{12} = 0.4$)

θ	θ^{True}	$\hat{\bar{\theta}}^{\text{MSL}}$	θ	θ^{True}	$\hat{\bar{\theta}}^{\text{MSL}}$	θ	θ^{True}	$\hat{\bar{\theta}}^{\text{MSL}}$
β_0	10.0000	9.9999	β_1^w	0.6000	0.6006	σ_v	0.3162	0.3154
		(0.0306)			(0.0127)			(0.0071)
β_1^y	0.8250	0.8249	β_{11}^{ww}	0.0500	0.0507	σ_{ν_2}	0.0100	0.0062
		(0.0111)			(0.0129)			(0.0085)
β_{11}^{yy}	0.0500	0.0500	β_{11}^{yw}	0.0100	0.0100	θ_{12}	0.4000	0.3765
		(0.0018)			(0.0010)			(0.1219)
						σ_T	0.2236	0.2239
								(0.0170)
						σ_{ξ_2}	0.3162	0.3183
								(0.0246)

Note: Standard deviations are in parentheses.

Table 2.4.2: Result of Simulation I ($J = 2,\ M = 2,\ \theta_{12} = 0.4$)

θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$
β_0	10.0000	9.9984	β_{22}^{yy}	0.0400	0.0402	σ_v	0.3162	0.3146
		(0.0353)			(0.0031)			(0.0072)
β_1^y	0.4000	0.4002	β_1^w	0.6000	0.6000	σ_{ν_2}	0.0100	0.0099
		(0.0140)			(0.0057)			(0.0104)
β_2^y	0.3000	0.2995	β_{11}^{ww}	0.0500	0.0500	θ_{12}	0.4000	0.4014
		(0.0130)			(0.0032)			(0.1113)
β_{11}^{yy}	0.0500	0.0502	β_{11}^{yw}	0.0100	0.0100	σ_T	0.2236	0.2261
		(0.0036)			(0.0017)			(0.0173)
β_{12}^{yy}	-0.0100	-0.0102	β_{21}^{yw}	-0.0150	-0.0150	σ_{ξ_2}	0.3162	0.3112
		(0.0032)			(0.0010)			(0.0162)

Note: Standard deviations are in parentheses.

Table 2.4.3: Result of Simulation I ($J = 3,\ M = 1,\ \theta_{12} = \theta_{13} = 0.2$)

θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$
β_0	10.0000	9.9993	β_{22}^{ww}	0.0250	0.0250	θ_{12}	0.2000	0.1741
		(0.0310)			(0.0041)			(0.1110)
β_1^y	0.8250	0.8247	β_{11}^{yw}	0.0100	0.0100	θ_{13}	0.2000	0.1780
		(0.0113)			(0.0004)			(0.1109)
β_{11}^{yy}	0.0500	0.0500	β_{12}^{yw}	-0.0050	-0.0051	ρ_θ	-0.5000	-0.4996
		(0.0019)			(0.0009)			(0.0176)
β_1^w	0.2500	0.2495	σ_v	0.3162	0.3155	σ_T	0.2236	0.2252
		(0.0027)			(0.0072)			(0.0171)
β_2^w	0.4000	0.4006	σ_{ν_2}	0.0100	0.0095	σ_{ξ_2}	0.3162	0.3170
		(0.0048)			(0.0025)			(0.0121)
β_{11}^{ww}	0.0350	0.0347	σ_{ν_3}	0.0100	0.0100	σ_{ξ_3}	0.3162	0.3167
		(0.0014)			(0.0008)			(0.0047)
β_{12}^{ww}	-0.0150	-0.0148	ρ_ν	0.0000	-0.0409			
		(0.0019)			(0.0796)			

Note: Standard deviations are in parentheses.

Table 2.4.4: Result of Simulation I ($J = 3,\ M = 2,\ \theta_{12} = \theta_{13} = 0.2$)

θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$
β_0	10.0000	10.0000	β_{11}^{ww}	0.0350	0.0348	σ_{ν_2}	0.0100	0.0095
		(0.0348)			(0.0015)			(0.0024)
β_1^y	0.4000	0.3994	β_{12}^{ww}	-0.0150	-0.0149	σ_{ν_3}	0.0100	0.0100
		(0.0139)			(0.0019)			(0.0005)
β_2^y	0.3000	0.3001	β_{22}^{ww}	0.0250	0.0250	ρ_ν	0.0000	-0.0665
		(0.0129)			(0.0033)			(0.1010)
β_{11}^{yy}	0.0500	0.0501	β_{11}^{yw}	0.0100	0.0100	θ_{12}	0.2000	0.1791
		(0.0038)			(0.0007)			(0.1108)
β_{12}^{yy}	-0.0100	-0.0100	β_{12}^{yw}	-0.0050	-0.0050	θ_{13}	0.2000	0.1788
		(0.0033)			(0.0011)			(0.1095)
β_{22}^{yy}	0.0400	0.0400	β_{21}^{yw}	0.0150	0.0150	ρ_θ	-0.5000	-0.5000
		(0.0032)			(0.0004)			(0.0127)
β_1^w	0.2500	0.2496	β_{22}^{yw}	-0.0050	-0.0050	σ_T	0.2236	0.2243
		(0.0026)			(0.0007)			(0.0173)
β_2^w	0.4000	0.4000	σ_v	0.3162	0.3151	σ_{ξ_2}	0.3162	0.3164
		(0.0032)			(0.0071)			(0.0078)
						σ_{ξ_3}	0.3162	0.3165
								(0.0041)

Note: Standard deviations are in parentheses.

2.4.2 Simulation II

I conduct another set of simulations using the pseudo-data set described in Section 2.3, which assumes that technical and allocative inefficiency are linked by the APS copulas. The objective of this simulation is to check the validity of the assumption in Greene (1980) when technical and allocative inefficiency are actually dependent. Consider a stochastic cost frontier model such as

$$\begin{aligned} \ln C_i &= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + \epsilon_i \\ &= \ln C(\boldsymbol{y}_i, \boldsymbol{w}_i) + v_i + u_i \\ s_{ij} &= s_j(\boldsymbol{y}_i, \boldsymbol{w}_i) + e_{ij},\ j = 2, \cdots, J. \end{aligned} \tag{2.4.1}$$

Greene (1980) assumes that ϵ_i is independent of $\boldsymbol{e}_i = (e_{i2}, e_{i3})$. I assume that $v_i \overset{\text{i.i.d}}{\sim} N(0, \sigma_v^2)$, $u_i \overset{\text{i.i.d}}{\sim} |N(0, \sigma_u^2)|$, and $\boldsymbol{e}_i \overset{\text{i.i.d}}{\sim} N(0, \boldsymbol{\Sigma}_e)$. Maximum likelihood estimation can be applied to estimate parameters of the model, where the joint density of ϵ_i, e_{i2}, e_{i3} is simply $f_{\epsilon, e_2, e_3}(\epsilon, e_2, e_3) = f_\epsilon(\epsilon) \cdot f_{\boldsymbol{e}}(\boldsymbol{e})$, as ϵ_i and $\boldsymbol{e}_i$ are assumed to be independent.[7]

Tables 2.4.5 to 2.4.8 show the result of simulations when $\theta_{12} = 0.4$ for $J = 2$ and $\theta_{12} = \theta_{12} = 0.2$ for $J = 3$. The key finding is as follows. As $u_i = u_i^T + g(\boldsymbol{\xi}_i)$, we do not know the true standard deviations of u_i.[8] However, as both u_i^T and $g(\boldsymbol{\xi}_i)$ are positive, the value of u_i is higher than that of u_i^T. It implies that the standard deviation of u_i, σ_u, should be higher than the standard deviation of u_i^T, σ_u^T, as the mean of a random variable from the half normal distribution is an increasing function in its standard deviation. However, the estimates of σ_u for all cases are less than the true value of $\sigma_T = 0.2236$. It suggests that if one ignores the relationship between technical and allocative inefficiency when they are indeed dependent, estimates of a cost increase due to inefficiency would be biased.

[7] Given the assumption about the distribution of v_i and u_i, the probability density function of ϵ_i is

$$f_\epsilon(\epsilon) = \frac{2}{\sigma} \cdot \phi\Big(\frac{\epsilon}{\sigma}\Big) \cdot \Phi\Big(\frac{\epsilon\lambda}{\sigma}\Big),$$

where $\sigma^2 = \sigma_v^2 + \sigma_u^2$, $\lambda = \sigma_u / \sigma_v$, and ϕ and Φ are density and distribution functions of the standard normal distribution.

[8] In addition, as $e_{ij} = h_j(\boldsymbol{\xi}_i) + \nu_{ij},\ j = 2, 3$, we do not know the true standard deviations of e_{ij}.

Table 2.4.5: Result of Simulation II ($J = 2,\ M = 1,\ \theta_{12} = 0.4$)

θ	θ^{True}	$\hat{\theta}^{\text{ML}}$	θ	θ^{True}	$\hat{\theta}^{\text{ML}}$	θ	θ^{True}	$\hat{\theta}^{\text{ML}}$
β_0	10.0000	10.0213	β_1^w	0.6000	0.5984	σ_v	0.3162	0.3164
		(0.0971)			(0.0076)			(0.0249)
β_1^y	0.8250	0.8288	β_{11}^{ww}	0.0500	0.0491	σ_u	-	0.1924
		(0.0112)			(0.0098)			(0.1122)
β_{11}^{yy}	0.0500	0.0496	β_{11}^{yw}	0.0100	0.0099	σ_{e_2}	-	0.0604
		(0.0019)			(0.0010)			(0.0004)

Note: Standard deviations are in parentheses.

Table 2.4.6: Result of Simulation II ($J = 2,\ M = 2,\ \theta_{12} = 0.4$)

θ	θ^{True}	$\hat{\theta}^{\text{ML}}$	θ	θ^{True}	$\hat{\theta}^{\text{ML}}$	θ	θ^{True}	$\hat{\theta}^{\text{ML}}$
β_0	10.0000	10.0096	β_{22}^{yy}	0.0400	0.0395	σ_v	0.3162	0.3143
		(0.1045)			(0.0034)			(0.0300)
β_1^y	0.4000	0.4004	β_1^w	0.6000	0.5987	σ_u	-	0.2030
		(0.0147)			(0.0050)			(0.1161)
β_2^y	0.3000	0.3048	β_{11}^{ww}	0.0500	0.0494	σ_{e_2}	-	0.0624
		(0.0156)			(0.0026)			(0.0012)
β_{11}^{yy}	0.0500	0.0496	β_{11}^{yw}	0.0100	0.0099			
		(0.0039)			(0.0016)			
β_{12}^{yy}	-0.0100	-0.0099	β_{21}^{yw}	-0.0150	-0.0148			
		(0.0034)			(0.0008)			

Note: Standard deviations are in parentheses.

Table 2.4.7: Result of Simulation II ($J = 3,\ M = 1,\ \theta_{12} = \theta_{13} = 0.2$)

θ	θ^{True}	$\hat{\theta}^{\text{ML}}$	θ	θ^{True}	$\hat{\theta}^{\text{ML}}$	θ	θ^{True}	$\hat{\theta}^{\text{ML}}$
β_0	10.0000	10.0755	β_{11}^{ww}	0.0350	0.0335	σ_v	0.3162	0.3342
		(0.0543)			(0.0015)			(0.0114)
β_1^y	0.8250	0.8286	β_{12}^{ww}	-0.0150	-0.0148	σ_u	-	0.1338
		(0.0113)			(0.0016)			(0.0588)
β_{11}^{yy}	0.0500	0.0499	β_{22}^{ww}	0.0250	0.0238	σ_{e_2}	-	0.0931
		(0.0019)			(0.0032)			(0.0004)
β_1^w	0.2500	0.2425	β_{11}^{yw}	0.0100	0.0101	σ_{e_3}	-	0.0914
		(0.0029)			(0.0004)			(0.0004)
β_2^w	0.4000	0.3965	β_{12}^{yw}	-0.0050	-0.0038	ρ_e	-	-0.5559
		(0.0043)			(0.0009)			(0.0014)

Note: Standard deviations are in parentheses.

Table 2.4.8: Result of Simulation II ($J = 3,\ M = 2,\ \theta_{12} = \theta_{13} = 0.2$)

θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$
β_0	10.0000	10.0712	β_2^w	0.4000	0.4008	β_{22}^{yw}	-0.0050	-0.0041
		(0.0545)			(0.0044)			(0.0007)
β_1^y	0.4000	0.3963	β_{11}^{ww}	0.0350	0.0330	σ_v	0.3162	0.3328
		(0.0140)			(0.0016)			(0.0111)
β_2^y	0.3000	0.3070	β_{12}^{ww}	-0.0150	-0.0149	σ_u	-	0.1369
		(0.0130)			(0.0018)			(0.0551)
β_{11}^{yy}	0.0500	0.0495	β_{22}^{ww}	0.0250	0.0224	σ_{e_2}	-	0.0842
		(0.0038)			(0.0035)			(0.0004)
β_{12}^{yy}	-0.0100	-0.0097	β_{11}^{yw}	0.0100	0.0097	σ_{e_2}	-	0.0796
		(0.0033)			(0.0008)			(0.0004)
β_{22}^{yy}	0.0400	0.0391	β_{12}^{yw}	-0.0050	-0.0064	ρ_e	-	-0.5208
		(0.0032)			(0.0013)			(0.0023)
β_1^w	0.2500	0.2418	β_{21}^{yw}	0.0150	0.0150			
		(0.0035)			(0.0004)			

Note: Standard deviations are in parentheses.

2.5 Conclusion

The estimation strategy proposed in Ryu (2021) involves multidimensional integral and two-step transformations. Therefore, it would be necessary to conduct a set of Monte Carlo simulations to confirm their validity. Like Amsler et al. (2021), the simulation results suggest that the parameters of the model in which APS copulas are employed can be reliably estimated in complex settings. In addition, I conduct another set of simulations to check the plausibility of assumptions in Greene (1980). Simulation results imply that it would lead biased estimates of inefficiency to ignore the relationship between technical and allocative inefficiency when they are indeed dependent.

APPENDICES

APPENDIX A

Simulating from the APS-2-A Copula

This section is written based on the supplemental material for Amsler et al. (2021). Let $c_{12}(\omega_1, \omega_2)$ be a copula density, where $(\omega_1, \omega_2) \in [0,1]^2$ are copula arguments that are uncorrelated, but ω_1 is correlated with $|\omega_2 - 0.5|$.

For computational ease, ω_2 and ω_1 are generated sequentially, which requires $C_{1|2}(\omega_1|\omega_2)$ is used for the Rosenblatt transformation. In this section, $C_{1|2}(\omega_1|\omega_2)$ is derived, then methods to obtain ω_2 and ω_1 are presented.

2.A.1 Derivation of the Conditional Distribution

Given that $C_{1|2}(\omega_1|\omega_2) = \int c_{1|2}(t_1|\omega_2)dt_1$, it is required to obtain the conditional density $c_{1|2}(\omega_1|\omega_2)$, which is

$$c_{1|2}(\omega_1|\omega_2) = \frac{c_{12}(\omega_1, \omega_2)}{f_2(\omega_2)} = c_{12}(\omega_1, \omega_2).$$

The second equality holds because $\omega_2 \sim U[0,1]$.

Hence, the conditional copula $C_{1|2}(\omega_1|\omega_2)$ is

$$\begin{aligned}
C_{1|2}(\omega_1|\omega_2) &= \int_0^{\omega_1} c_{1|2}(t_1|\omega_2)dt_1 \\
&= \int_0^{\omega_1} [1 + \theta_{12}(1 - 2t_1)\{1 - 12(\omega_2 - 0.5)^2)\}]dt_1 \\
&= [t_1 + \theta_{12}(t_1 - t_1^2)\{1 - 12(\omega_2 - 0.5)^2)\}]\Big|_0^{\omega_1} \\
&= \omega_1 + \theta_{12}\omega_1(1 - \omega_1)\{1 - 12(\omega_2 - 0.5)^2)\} \\
&= \omega_1 + g\omega_1(1 - \omega_1),
\end{aligned}$$

where $g = \theta_{12}\{1 - 12(\omega_2 - 0.5)^2\}$.

2.A.2 Obtain Copula Arguments

2.A.2.1 Obtain ω_2

As $\omega_2 \sim U[0,1]$, $\zeta_2 = F_2(\omega_2) = \omega_2$. Therefore, draw ζ_2 from $U[0,1]$ and define $\omega_2 = \zeta_2$.

2.A.2.2 Obtain ω_1

First, draw ζ_1 from $U[0,1]$. Then, one can obtain ω_1 by solving the equation $C_{1|2}(\omega_1|\omega_2) - \zeta_1 = 0$. It yields

$$\begin{aligned}
&\omega_1 + g\omega_1(1-\omega_1) - \zeta_1 = 0 \\
\Rightarrow\ & g\omega_1^2 - (1+g)\omega_1 + \zeta_1 = 0 \\
\Rightarrow\ & \omega_1 = \frac{(1+g) \pm \sqrt{(1+g)^2 - 4g\zeta_1}}{2g}.
\end{aligned}$$

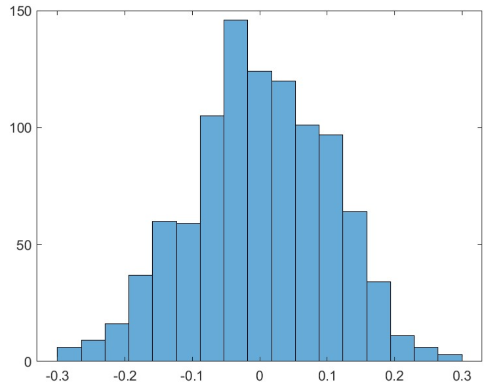
Due to the upper and lower bounds of ω_1, the solution $\omega_1 = \frac{(1+g)+\sqrt{(1+g)^2-4g\zeta_1}}{2g}$ is ruled out (Amsler et al., 2021). Let $A = 1+g$. Then,

$$\begin{aligned}
\omega_1 &= \frac{A - \sqrt{A^2 - 4(A-1)\zeta_1}}{2(A-1)} \\
&= \frac{(A - \sqrt{A^2 - 4(A-1)\zeta_1})(A + \sqrt{A^2 - 4(A-1)\zeta_1})}{2(A-1)(A + \sqrt{A^2 - 4(A-1)\zeta_1})} \\
&= \frac{A^2 - A^2 + 4(A-1)\zeta_1}{2(A-1)(A + \sqrt{A^2 - 4(A-1)\zeta_1})} \\
&= \frac{2\zeta_1}{A + \sqrt{A^2 - 4(A-1)\zeta_1}}
\end{aligned}$$

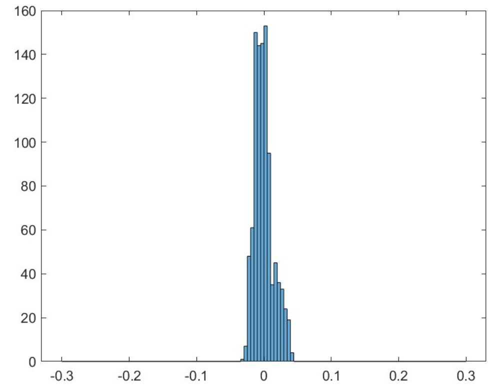
APPENDIX B

Tables and Figures

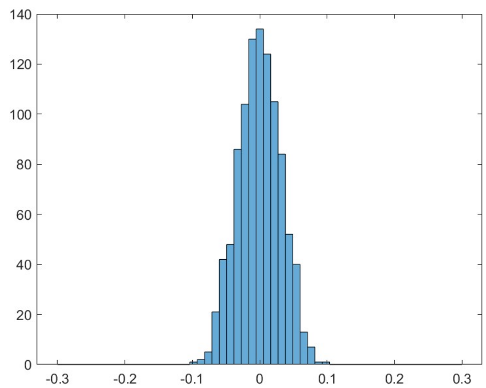
Figure 2.B.1: Sample Correlations between ζ_1 and ζ_2



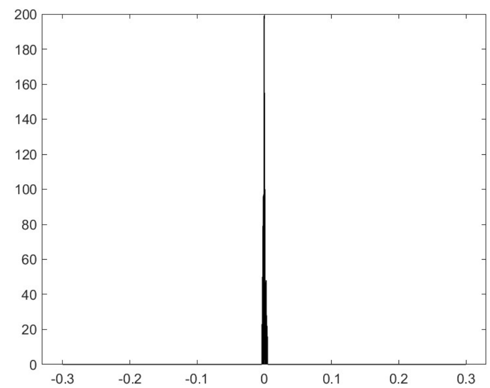
(a) PRNG - `rand` of MATLAB
($N = 100$)



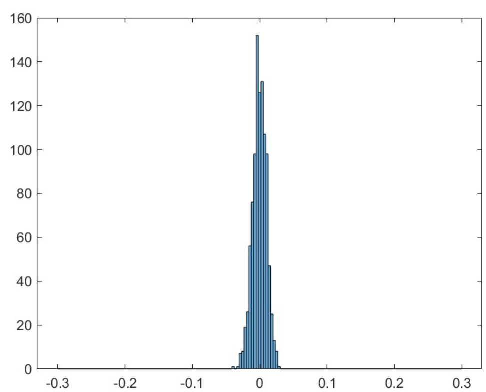
(b) QRNG - `haltonset` of MATLAB
($N = 100$)



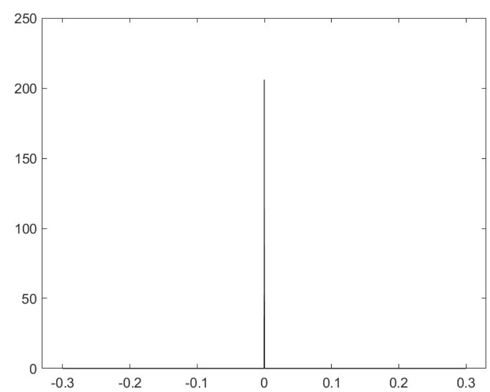
(c) PRNG - `rand` of MATLAB
($N = 1000$)



(d) QRNG - `haltonset` of MATLAB
($N = 1000$)



(e) PRNG - `rand` of MATLAB
($N = 10000$)



(f) QRNG - `haltonset` of MATLAB
($N = 10000$)

Table 2.B.1: Result of Simulation I ($J = 2,\ M = 1,\ \theta_{12} = 0$)

θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$
β_0	10.0000	9.9993	β_1^w	0.6000	0.6002	σ_v	0.3162	0.3153
		(0.0305)			(0.0130)			(0.0071)
β_1^y	0.8250	0.8250	β_{11}^{ww}	0.0500	0.0505	σ_{ν_2}	0.0100	0.0074
		(0.0111)			(0.0133)			(0.0109)
β_{11}^{yy}	0.0500	0.0500	β_{11}^{yw}	0.0100	0.0101	θ_{12}	0.0000	0.0052
		(0.0018)			(0.0010)			(0.1679)
						σ_T	0.2236	0.2243
								(0.0170)
						σ_{ξ_2}	0.3162	0.3146
								(0.0329)

Note: Standard deviations are in parentheses.

Table 2.B.2: Result of Simulation I ($J = 2,\ M = 1,\ \theta_{12} = 0.2$)

θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$
β_0	10.0000	9.9996	β_1^w	0.6000	0.6008	σ_v	0.3162	0.3153
		(0.0304)			(0.0127)			(0.0071)
β_1^y	0.8250	0.8250	β_{11}^{ww}	0.0500	0.0510	σ_{ν_2}	0.0100	0.0074
		(0.0110)			(0.0129)			(0.0103)
β_{11}^{yy}	0.0500	0.0500	β_{11}^{yw}	0.0100	0.0101	θ_{12}	0.2000	0.2056
		(0.0018)			(0.0010)			(0.1596)
						σ_T	0.2236	0.2243
								(0.0170)
						σ_{ξ_2}	0.3162	0.3167
								(0.0294)

Note: Standard deviations are in parentheses.

Table 2.B.3: Result of Simulation I ($J = 2,\ M = 2,\ \theta_{12} = 0$)

θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$
β_0	10.0000	9.9989	β_{22}^{yy}	0.0400	0.0401	σ_v	0.3162	0.3145
		(0.0352)			(0.0031)			(0.0072)
β_1^y	0.4000	0.4002	β_1^w	0.6000	0.5998	σ_{ν_2}	0.0100	0.0160
		(0.0140)			(0.0056)			(0.0164)
β_2^y	0.3000	0.2996	β_{11}^{ww}	0.0500	0.0499	θ_{12}	0.0000	0.0610
		(0.0130)			(0.0030)			(0.2049)
β_{11}^{yy}	0.0500	0.0501	β_{11}^{yw}	0.0100	0.0100	σ_T	0.2236	0.2263
		(0.0036)			(0.0017)			(0.0173)
β_{12}^{yy}	-0.0100	-0.0101	β_{21}^{yw}	-0.0150	-0.0150	σ_{ξ_2}	0.3162	0.2934
		(0.0032)			(0.0008)			(0.0475)

Note: Standard deviations are in parentheses.

Table 2.B.4: Result of Simulation I ($J = 2,\ M = 2,\ \theta_{12} = 0.2$)

θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$
β_0	10.0000	9.9983	β_{22}^{yy}	0.0400	0.0402	σ_v	0.3162	0.3144
		(0.0352)			(0.0031)			(0.0072)
β_1^y	0.4000	0.4002	β_1^w	0.6000	0.5999	σ_{ν_2}	0.0100	0.0132
		(0.0140)			(0.0057)			(0.0138)
β_2^y	0.3000	0.2996	β_{11}^{ww}	0.0500	0.05000	θ_{12}	0.2000	0.2525
		(0.0130)			(0.0031)			(0.1654)
β_{11}^{yy}	0.0500	0.0501	β_{11}^{yw}	0.0100	0.0100	σ_T	0.2236	0.2266
		(0.0036)			(0.0017)			(0.0174)
β_{12}^{yy}	-0.0100	-0.0101	β_{21}^{yw}	-0.0150	-0.0150	σ_{ξ_2}	0.3162	0.3144
		(0.0032)			(0.0008)			(0.0072)

Note: Standard deviations are in parentheses.

Table 2.B.5: Result of Simulation I ($J=3,\ M=1,\ \theta_{12}=\theta_{13}=0$)

θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$
β_0	10.0000	9.9990	β_{22}^{ww}	0.0250	0.0251	θ_{12}	0.0000	-0.0049
		(0.0310)			(0.0044)			(0.1237)
β_1^y	0.8250	0.8247	β_{11}^{yw}	0.0100	0.0100	θ_{13}	0.0000	-0.0002
		(0.0113)			(0.0004)			(0.1186)
β_{11}^{yy}	0.0500	0.0500	β_{12}^{yw}	-0.0050	-0.0050	ρ	-0.5000	-0.4971
		(0.0019)			(0.0010)			(0.0339)
β_1^w	0.2500	0.2495	σ_v	0.3162	0.3154	σ_T	0.2236	0.2252
		(0.0029)			(0.0072)			(0.0171)
β_2^w	0.4000	0.4005	σ_{ν_2}	0.0100	0.0094	σ_{ξ_2}	0.3162	0.3177
		(0.0060)			(0.0027)			(0.0199)
β_{11}^{ww}	0.0350	0.0347	σ_{ν_3}	0.0100	0.0100	σ_{ξ_3}	0.3162	0.3166
		(0.0017)			(0.0009)			(0.0056)
β_{12}^{ww}	-0.0150	-0.0149	ρ_ν	0.0000	-0.0509			
		(0.0023)			(0.1088)			

Note: Standard deviations are in parentheses.

Table 2.B.6: Result of Simulation I ($J=3,\ M=2,\ \theta_{12}=\theta_{13}=0$)

θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$	θ	θ^{True}	$\hat{\theta}^{\text{MSL}}$
β_0	10.0000	9.9999	β_{11}^{ww}	0.0350	0.0348	σ_{ν_2}	0.0100	0.0095
		(0.0348)			(0.0015)			(0.0024)
β_1^y	0.4000	0.3994	β_{12}^{ww}	-0.0150	-0.0149	σ_{ν_3}	0.0100	0.0100
		(0.0139)			(0.0019)			(0.0005)
β_2^y	0.3000	0.3001	β_{22}^{ww}	0.0250	0.0250	ρ_ν	0.0000	-0.0445
		(0.0129)			(0.0032)			(0.0992)
β_{11}^{yy}	0.0500	0.0500	β_{11}^{yw}	0.0100	0.0100	θ_{12}	0.0000	0.0005
		(0.0038)			(0.0007)			(0.1171)
β_{12}^{yy}	-0.0100	-0.0100	β_{12}^{yw}	-0.0050	-0.0050	θ_{13}	0.0000	-0.0009
		(0.0033)			(0.0011)			(0.1200)
β_{22}^{yy}	0.0400	0.0400	β_{21}^{yw}	0.0150	0.0150	ρ	-0.5000	-0.4996
		(0.0032)			(0.0004)			(0.0116)
β_1^w	0.2500	0.2496	β_{22}^{yw}	-0.0050	-0.0050	σ_T	0.2236	0.2243
		(0.0026)			(0.0007)			(0.0172)
β_2^w	0.4000	0.4003	σ_v	0.3162	0.3150	σ_{ξ_2}	0.3162	0.3163
		(0.0034)			(0.0071)			(0.0073)
						σ_{ξ_3}	0.3162	0.3165
								(0.0040)

Note: Standard deviations are in parentheses.

CHAPTER 3

DEMAND ESTIMATION OF DEPOSITS: A CASE OF THE KOREAN FINANCIAL INDUSTRY

3.1 Introduction

During the last several decades, tools in structural economic modeling have developed remarkably, especially in the field of industrial organization. These techniques and tools were recently applied in finance to some extent, presenting many promising directions.[1] For example, Hortaçsu et al. (2018) estimate a structural model of the uniform price auctions of U.S. Treasury bills and notes in order to analyze market power across the three different bidder groups: primary dealers, direct bidders, and indirect bidders. Bonaldi et al. (2015) propose a framework for estimating spillover effects between individual banks' short-term funding costs and measure systemic risk using data from the main refinancing operations of the European Central Bank.

The other intersecting field of finance and industrial organization is estimating a demand system for financial assets, which are viewed as differentiated products. There are two main directions with respect to this field: the first one is based on a product-space demand model like the one proposed by Deaton and Muellbauer (1980) that approximates the demand function by a flexible functional form; and the second direction is based on a characteristics space demand model such as Berry et al. (1995, hereafter BLP), where consumer choices are based on products' characteristics rather than the products themselves.

The aim of this paper is to estimate a structural demand model for the financial instruments of Korea in order to measure the effect of deregulation in the payment and settlement systems. From 2009, securities companies were given access to participate in retail payment systems, which were previously restricted to banks only. Consequently, cash management ac-

[1]Please refer to Kastl (2017) for more details.

counts (hereafter CMAs) provided by securities companies, which were similar to traditional deposits of banks but had limitations in transferring funds, became the close substitutes for deposits in terms of services.

CMAs, which were introduced in Korea in the 1980s, have similarities to the checking accounts of banks that consumers can deposit and withdraw funds from without limitations. In addition, as securities companies generally invest funds from CMAs in government or public corporations bonds with repurchase agreements[2], they offer interest rates of CMAs around the policy interest rates, whereas the checking accounts usually provide almost zero interest. That is, CMAs share the features of the checking accounts and the time deposits of banks. However, securities companies are regarded as less safe than banks due to the differences in the business model and the size of institutions, as well as regulatory gaps between banks and securities companies. Also, most CMAs are not protected by deposit insurance.[3] Furthermore, as the retail payment systems were only accessible by banks, CMAs were not used as a means of exchange.

Ever since the Capital Market and Financial Investment Business Act, enacted in August 2007, was enforced in February 2009, securities companies were allowed to participate in the retail payment systems operated by the Korea Financial Telecommunications and Clearings Institute (KFTC).[4] Therefore, from the depositor's perspective, traditional deposits and CMAs became indistinguishable in terms of services they provide. For instance, consumers who have CMAs are able to transfer funds to bank accounts via internet or mobile banking services and vice versa. Also, CMA holders can pay off their credit card balances by deducting from their CMAs. Reflecting these changes, CMAs were included in M2 from July 2009.[5]

[2]As of Q4 2016, CMAs with RP agreement count for 59.0% of total CMAs, while those investing in MMF and fiduciarily managed by Korea Securities Finance Corporation count for 6.1% and 30.3%, respectively.

[3]CMAs of a securities company that also has the merchant bank license are protected by deposit insurance. However, its share of CMAs is only 4.6% as of Q4 2016 since only two securities companies hold the merchant bank license.

[4]Securities companies began to join the retail payment systems from July 2009.

[5]Although CMAs are transferable like the checking accounts included in M1, depositors in CMAs have to sacrifice interests if they use balance in CMAs for transaction. Therefore, CMAs are classified into M2. Please refer to International Monetary Fund (2016) for more details on the definitions of money aggregates.

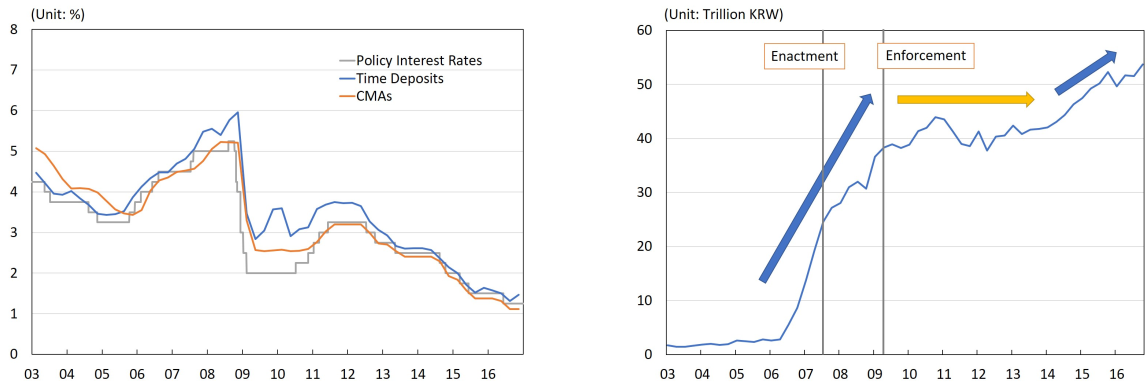
The enactment of the Capital Market and Financial Investment Business Act sparked a fierce debate on securities companies' participation in the retail payment systems. People who supported the deregulation claimed that consumer welfare would be improved by promoting competition among financial institutions, and it would be necessary to promote the financial investment businesses that were less developed than the banking industry. On the other hand, people who were against the measure contended that it would be harmful to the financial system as it would cause an increase in payment and settlement risks. Also, receiving deposits is considered the banks' own business[6] and only a few countries allowed securities companies to participate in the retail payment systems.

As a result, the Capital Market and Financial Investment Business Act permitted securities companies to participate in the retail payment systems. However, although CMAs have the advantages of interest rates and services compared to bank deposits, its total amount has stabilized after a sharp increase between mid-2006 and mid-2008. In particular, its balance had remained around forty trillion KRW for five years since the global financial crisis. This suggests that depositors' choice may depend on the financial stability situation that would affect their risk attitudes, referred to as the market discipline in banking. Based on this phenomenon, this paper evaluates whether consumer welfare has significantly increased with the enforcement of the act when considering consumer's risk attitudes.

In order to measure the effect of deregulation, I develop a structural demand model following the characteristic space approach. As in Petrin (2002), the researcher is able to evaluate welfare gains for consumers from the introduction of new products by constructing a structural model. Furthermore, as Nevo (2000) notes, the econometrician can reduce the number of parameters that need to be estimated. To estimate the model, I apply the random coefficient discrete choice approach. This approach can estimate the model using only market-level price and quantity data, deal with the price endogeneity, and allow for a

[6] For example, the U.S. Bank Holding Companies Act defines banks as an institution which both (i) accepts demand deposits or deposits that the depositor may withdraw by check or similar means for payment to third parties or others; and (ii) is engaged in the business of making commercial loans.

Figure 3.1.1: Interest Rates and Total Amount of CMAs



(a) Interest Rates (b) Total Amount

Source: Bank of Korea, Korea Financial Investment Association

more realistic substitution pattern reflecting the heterogeneity in consumer tastes.

The characteristics space approach model in the finance literature relates to asset pricing and portfolio choice. Markowitz (1952), the classical reference in finance, views a portfolio as bundles consisting of mean-variance characteristics. In addition to the mean and the variance of returns, other relevant characteristics of financial instruments may include maturity, probability of default, asset covariance with the market return, etc. However, as Kastl (2017) points out, although those might be the relevant characteristics that capture important parts of variation in demand for portfolio, it might be hard to succinctly capture other important ones.

Other than analyzing portfolio choice that considers whole financial markets, another way of defining the relevant characteristics is by restricting the scope of the financial instruments, such as deposits. This paper focuses on deposits instruments, which include checking, savings, and time deposit accounts generally provided by commercial banks, and CMAs.[7] The reasons are as follows: (i) the banking sector holds more than 50% of the financial assets among Korean financial institutions[8]; (ii) deposits are the major source of Korean

[7] One can consider to separately construct models by products. However, as deposits cannot be disaggregated at the bank level as well as CMAs hold characteristics of both checking and time deposits, I focus on whole deposit services.

[8] As of Q4 2016, the banking sector holds 50.8% of the financial assets among Korean financial institutions, while insurance sector and securities sector hold 15.9% and 5.8%, respectively.

banks' funding[9], and (iii) CMAs, the interest of this paper, became the close substitute for traditional deposits by the deregulation in payment and settlement systems.

Recently, some papers have applied a discrete choice model to estimate the demand for deposits. For example, Dick (2008) estimates a structural demand model for commercial bank deposit services in order to measure the effects on consumers, given changes in bank services owing to the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994 that allowed for nationwide branching. Following the discrete choice literature, it assumes that consumer decisions are based on prices and bank characteristics, such as deposit rates, account fees, the age, size, and geographic diversification.

Based on the demand estimation for deposits, some papers extended a structural model of the banking sector to analyze the financial fragility. For example, Egan et al. (2017) develop a structural empirical model of the U.S. banking sector that considers both demand and supply sides. After estimating the demand and supply for deposit, the researchers evaluate several proposed bank regulations. The results, for instance, suggest a capital requirement below eighteen percent could lead to significant instability in the U.S. banking system.

However, those papers do not explicitly take market discipline in banking into consideration. Market discipline in banking, in its broad terms, is defined as the mechanism via which market participants monitor and discipline excessive risk-taking behavior by banks (Stephanou, 2010). It is often described as a situation where depositors face costs that are positively related to bank risk and react on the basis of these costs (Berger, 1991). For instance, given that the bank's fragility increases, depositors respond by withdrawing their funds or by demanding higher interest rates on their deposits. Since it is known that market discipline would lower the probability of individual bank's failures and the incidence of banking crises by reducing the problems of moral hazard and asymmetric information in banking, policymakers have increasingly recognized its role and have incorporated it in their regulatory frameworks. One example of its codification is Pillar 3 in the Basel III, which is

[9]As of Q4 2016, deposits consists of 83.6% of banks' funding.

the global supervisory framework for internationally active banks.

Much work has been done on the existence of market discipline.[10] Previous studies provide evidence of market discipline in both developed countries and developing countries. Most of these studies examine the existence of market discipline by analyzing either how yields on uninsured deposits or the level or growth of uninsured deposits respond to measures of bank risk. However, a number of papers have found that the typical test for existence of market discipline might fail in some developing economies in non-crisis periods, as traditional indicators of bank soundness tend to become less significant and explain a smaller fraction of the total variance of deposits and interest rates during financial turmoil than during stable periods. The results imply that depositors behave differently by the financial stability situation.

The remainder of the chapter is organized as follows. Section 3.2 outlines the model specifications and estimation strategies. Section 3.3 describes the data and instruments. Section 3.4 reports the estimation results. Section 3.5 concludes the chapter.

3.2 Empirical Framework

3.2.1 Assumptions

I assume that, following Dick (2008), consumers[11] cluster their deposits within one primary bank for acquiring banking services together. Based on this assumption, one can apply the discrete choice model. It might be possible for consumers to demand multiple banking services. However, if banks were to provide benefits to depositors who use the bank as the primary one, which is common in Korea[12], consumers would then have incentives to

[10]Please refer to Flannery (1998), Arena (2003), and Levy-Yeyati et al. (2010) for more details.

[11]Due to the limitation of data that it does not divide depositors into households and corporates by financial institutions, I assume that two groups of depositors choose a depository institution in a similar manner. Dick (2008) also assumes that their behavior is similar based on the consumer and business survey.

[12]For instance, banks offer higher deposit interest rates and lower fees on transactions to depositors depending on their class, which is decided by the amount of deposit, the records of direct deposit of salary, the number of accounts, etc.

consolidate their deposits in a single financial institution. In addition, according to the Survey of Household Finances and Living Conditions[13], the median amount of deposits per household is thirty three million KRW as of the end of March 2017, which is lower than the amount of deposits protected by deposit insurance (fifty million KRW). These suggest that it is reasonable to assume that consumers choose a single bank for deposits. Given that CMAs have become the close substitutes since the deregulation, securities companies providing them are assumed to be treated as banks in the deposits market, albeit it seems to be a strong assumption.

I define market share based on the amount of deposits, and outside goods as deposits in financial institutions other than banks[14] and securities companies; these include merchant banking corporations, mutual savings banks, credit cooperatives, and postal savings. This implies, along with the first assumption, that depositors can have a number of accounts as long as they cluster deposits into one bank. The definition of market share using the amount of deposits, not the number of accounts, makes up for the shortcomings of the first assumption, which enables to apply a discrete choice model within a multinomial choice setting. For instance, even though consumers hold accounts in multiple banks, the problems that stem from the first assumption could be mitigated as long as the amount of deposits in banks other than the primary one is negligible. In addition, given that transferring funds is easier to do than opening and closing accounts, it will reduce the fixed cost to change one's primary bank if consumers have accounts in multiple banks. The definition of outside goods has limitations as it might not capture the true market share since some people may choose to invest funds in financial instruments other than deposits. However, the results of the Survey of Household Finances and Living Conditions, which shows that households' preference for financial instruments have remained stable, suggest that this study's definition of outside goods would therefore be reasonable.

[13]The survey is annually conducted of twenty thousand households by the Statistics Korea, the Financial Supervisory Service of Korea, and the Bank of Korea since 2012.

[14]I exclude KDB and KEXIM from the category of banks due to their heterogeneous business model.

Table 3.2.1: Households' Preferences for Financial Instruments

	Deposits	Pension	Stock	Etc.	Total
2012	89.9	1.7	5.9	2.5	100.0
2013	90.7	1.8	4.7	2.8	100.0
2014	91.6	2.2	3.4	2.8	100.0
2015	90.6	2.3	4.7	2.4	100.0
2016	91.6	1.9	4.0	2.5	100.0
2017	91.8	1.8	4.1	2.3	100.0

Source: Statistics Korea, Financial Supervisory Service of Korea, and Bank of Korea, *"Survey of Household Finances and Living Conditions"*

3.2.2 Models

In the characteristics space demand model, the price of a product can be correlated with an omitted product attribute, which is relevant but not observed by the econometrician. If an omitted product attribute is positively correlated with the price, estimates of the price sensitivity term will be biased toward zero and those of the price elasticities will be biased as well.[15] To deal with the potential price endogeneity problem, one can use instrumental variables and/or apply a random coefficient discrete choice model.

Thus, I construct the following models to estimate the demand for deposits: (i) the simple conditional logit model that does not include an omitted product attribute (hereafter Conditional Logit); (ii) the Berry (1994) type logit model that includes an omitted product attribute (hereafter IV Logit); (iii) the simple random coefficients logit model that does not include an omitted product attribute (hereafter RC Logit); and (iv) the BLP (1995) type random coefficient logit model (hereafter BLP (1995) RC Logit).

3.2.2.1 Conditional Logit and IV Logit Models

Similar to most discrete choice models following the Random Utility Maximization (RUM) hypothesis, I assume that individual agents $i = 1, \ldots, I$ $(= \infty)$ at $t = 1, \ldots, T$ markets make choices between $j = 1, ..., J$ alternatives in order to maximize their indirect utility,

[15] Kim and Petrin (2015) provide a literature review about this problem.

u_{ijt}, specified as

$$u_{ijt} = x'_{jt}\beta + \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt}$$
$$= \delta_{jt} + \epsilon_{ijt},$$

where $x_{jt} = (x_{jt,1}, \ldots, x_{jt,K})'$ is a $K \times 1$ vector of observed characteristics for deposit product j at the market t, p_{jt} is the spread or interest rates paid by banks on j at t, ξ_{jt} is an unobserved characteristic for j at t, and ϵ_{ijt} is the error term. As Conditional Logit model does not take account for unobserved heterogeneity, $\xi_{jt} = 0$ for all j and t. $\delta_{jt} = x'_{jt}\beta + \alpha p_{jt} + \xi_{jt}$ is referred to as the mean utility, which is common to all agents. The $K+1$ dimensional vector $\theta = (\beta, \alpha)$ represents the taste parameters.

Now, assume that ϵ_{ijt} are identically and independently distributed according to the Type I extreme-value distribution. Then, by integrating over ϵ_{ijt}, the predicted market share for j at t is derived such that

$$s_{jt}(x, \beta, \alpha, \xi) = \frac{\exp(x'_{jt}\beta + \alpha p_{jt} + \xi_{jt})}{\sum_{r=1}^{J} \exp(x'_{rt}\beta + \alpha p_{rt} + \xi_{rt})}. \tag{3.2.1}$$

Berry (1994) assumes that at the true parameter values, β_0 and α_0, the following equality must hold

$$s_{jt}(x, \beta_0, \alpha_0, \xi) = S_{jt},$$

where S_{jt} is the true market share from the aggregated data. In other words, conditioning on the true values of δ_0, the model should exactly fit the data.

Berry (1994) uses the following transformation of equation (3.2.1) such that

$$\log(s_{jt}(x, \beta, \alpha, \xi)) = e_t + x'_{jt}\beta + \alpha p_{jt} + \xi_{jt},$$

where $e_t = -\log(\sum_{r=1}^{J} exp(x'_{rt}\beta + \alpha p_{rt} + \xi_{rt}))$. By normalizing the mean utility of the outside good, denoted as $j = 0$, to zero that implies

$$\log(s_{0t}(x, \beta, \alpha, \xi)) = e_t,$$

equation (3.2.2) is obtained such that

$$\begin{aligned}\log(S_{jt}) - \log(S_{0t}) &= \delta_{jt} \\ &= x'_{jt}\beta + \alpha p_{jt} + \xi_{jt},\end{aligned} \tag{3.2.2}$$

where S_{0t} is the share of the outside good at t.

Given equation (3.2.2), one can estimate the Conditional Logit model with ordinary least squares by regressing $\log(S_{jt}) - \log(S_{0t})$ on (x'_{jt}, p_{jt}), as well as IV Logit model with instrumental variables estimation given the assumption $\mathbb{E}[\xi_{jt}|Z_{jt}] = 0$.

3.2.2.2 RC Logit and BLP (1995) RC Logit Models

For RC Logit and BLP (1995) RC Logit models, I specify the indirect utility similar to Nevo (2000) that allows the price coefficient to be random without taking the natural log. Therefore, the indirect utility of an agent i from consuming j at the market t is specified as

$$u_{ijt} = x_{jt}\beta_i + \alpha_i p_{jt} + \xi_{jt} + \epsilon_{ijt},$$

where $\beta_{i,k} = \beta_k + \sigma_k \eta_{i,k}$, $\alpha_i = \alpha + \sigma_p \eta_{i,p}$, $\eta_{i,k}, \eta_{i,p} \sim N(0,1)$, ξ_{jt} is an unobserved characteristic for j at t, and ϵ_{ijt} is the error term.

Now, I decompose indirect utility by two parts: the mean utility, δ_{jt}, and the heteroskedastic error terms, ν_{ijt}, that captures the effect of random tastes parameters such that

$$u_{ijt} = \delta_{jt} + \nu_{ijt},$$

where $\delta_{jt} = x_{jt}\beta + \alpha p_{jt} + \xi_{jt}$ represents a mean level of utility and $\nu_{ijt} = [\sum_k x_{jt,k}\sigma_k \eta_{i,k}] + \sigma_p \eta_{i,p} p_{jt} + \epsilon_{ijt}$ represents a heteroskedastic error terms that captures the effect of random tastes parameters.

In order to estimate the model, I define the set of values of error terms, A_{jt}, that make j maximizing utility at t given the J dimensional vector $\delta_t = (\delta_{1t}, \ldots, \delta_{Jt})$, such that

$$A_{jt}(\delta_t) = \{\nu_{it} = (\nu_{ijt}) \mid \delta_{jt} + \nu_{ijt} > \delta_{j't} + \nu_{ij't}, \;\; \forall\, j' \neq j\}.$$

Then, the market share for j at t is written as

$$s_{jt}(\delta_t(x,p,\xi),x,\beta,\alpha,\sigma) = \int_{A_{jt}(\delta_t)} f(\nu)d\nu.$$

In order to estimate the models, I take the following steps. First, I compute the market shares given δ_t and σ such that

$$s_{jt}(\delta_t,\sigma) = \int \frac{\exp(\delta_{jt} + \sum_k x_{jt,k}\sigma_k\eta_{i,k} + \sigma_p\eta_{i,p}p_{jt})}{1+\sum_{r=1}^{J}\exp(\delta_{rt} + \sum_k x_{rt,k}\sigma_k\eta_{i,k} + \sigma_p\eta_{i,p}p_{rt})} df(\eta_i).$$

Second, given σ, I find δ_{jt} by contraction mapping. Third, given δ_{jt}, β, and α, obtain ξ_{jt}. Last, choose β, α and σ to minimize the sample criterion function. For example, I use the moment condition of $\mathbb{E}[\xi_{jt}(\beta_0,\alpha_0,\sigma_0)|Z_{jt}] = 0$ to estimate BLP (1995) RC Logit by GMM.

3.3 Data and Instruments

3.3.1 Data

The data mainly come from two sources: financial institution-level data from the Financial Statistics Information System (FISIS) of the Financial Supervisory Service of Korea (http://fisis.fss.or.kr), and country-level aggregate data from the Economic Statistics System (ECOS) of the Bank of Korea (http://ecos.bok.or.kr). The data on each financial institution's deposits and its attributes are obtained from the balance sheet, the income statement, and other reporting forms uploaded on FISIS. The data on the total amount of deposits from the Flow of Funds and the policy interest rates are taken from ECOS. The amount of CMAs is obtained from the Korea Financial Investment Association Portal (FreeSIS, http://freesis.kofia.or.kr). The sample covers the period from Q1 2003 to Q2 2015 considering the completion of the restructuring Korean financial industry after the Asian financial crisis (Q4 2002), the enforcement of Capital Market and Financial Investment Business Act (Q2 2009), and the merger of Hana and KEB banks (Q3 2015).

An observation is defined as a financial institution[16]- quarter combination in the estimation exercises. I choose the attributes of financial institutions from available data, which are important and easily observable by depositors. Table 3.B.4 shows summary statistics of data.

I use spread, which is the difference between interest rates paid on deposits and the policy interest rates, as the price variable. This is because deposit rates are decided in line with the policy interest rates and the interest rates regime shifts before and after the global financial crisis. The deposit rates are driven by dividing interest expense on deposits by the amount of deposits from each institution's quarterly income statement and annualized.

In addition to the price variable, four categories of observed characteristics are chosen: (i) size, (ii) quality of service, (iii) quantity of service, and (iv) financial soundness. Similar to Dick (2008), I classify financial institutions into five groups, considering their asset sizes and other characteristics[17], and use them to control for size rather than using the asset size itself. The reason is that the asset size itself should increase as the financial institution receives more deposits by the law of accounting.[18] In addition, it would capture features associated with the size of financial institutions, including larger infrastructures, product diversity, and know-how. The quality and quantity of service are proxied by the number of employees per branches[19] and the number of branch[20], respectively.

I include the financial soundness indicator and a dummy variable for the period of financial turmoil in order to test the existence of market discipline in the deposit market. Egan et al. (2017) use the implied probability of default of banks from credit default swap (CDS) spreads when estimating the demand for deposits. However, it is not easily available

[16] As the amount of CMAs of each securities firm is not available in public, I assume them as a single entity.

[17] For more details, please refer to Table 3.B.5.

[18] For the sample period, the correlation between dependent variable and asset size is 0.91.

[19] Dick (2008) argues that it can capture consumers' waiting time, the types of services specific to bank, and the value of human interaction to consumers who are not able to use the online service.

[20] One can consider the number of ATMs as a proxy for the quantity of service. However, the data on the banks' number of ATMs does not cover the whole sample period as well as these on securities companies are not provided. Therefore, I do not include the number of branch although it seems to be relevant.

to depositors, and its value might highly depend on the model and assumptions. Therefore, I use the risk-based capital ratios, which are representative, well-known, and publicly disclosed indicators: the BIS ratio for banks and the net operating capital ratio for securities companies. I assume the period from Q3 2008 to Q2 2013 as a time of financial instability[21] reflecting major financial events and financial stability indices.

3.3.2 Instruments

I use three categories of instrumental variables: (i) financial institutions' characteristics themselves; (ii) mark-up shifters; and (iii) cost shifters. The set of mark-up shifters includes BLP instruments, which are the sum of characteristics of other products in the market, following the convention of the literature on discrete choice models. This is based on the intuition from models of oligopoly that suggest the more isolated the firm is in the product space, the more likely it is to have a higher price relative to the cost[22].

The set of cost shifters includes variables related to marginal costs, funding costs and labor costs. I use the policy interest rates as a proxy for funding costs, as the interest rates of funding sources other than deposits, such as bank debenture and call money, are also decided based on it. Labor costs come from the average wage data of the financial business from Statistics Korea. These two variables are chosen, although the income statement provides data for each financial institution, because the financial institution's technology and quality are already controlled through other covariates. For example, if a financial institution hires more skilled workers whose wages tend to be higher than those of low-skilled workers, the actual salary data may contain the hidden quality components, therefore leading it to violating the independent assumption.

[21] I exclude the credit card debacle in 2003, since the problem stemming from credit card companies might not affect depositors' risk attitude as well as the debacle was recovered in the short time.

[22] Please refer to BLP (1995) for more details.

3.4 Results

3.4.1 Model Estimation

Table 3.B.7 presents the estimation results, where column (1) corresponds to the Conditional Logit model, columns (2) and (3) correspond to the IV Logit model, column (4) corresponds to the RC Logit model, and columns (5) and (6) correspond to the BLP (1995) RC Logit model. In columns (2) and (5), financial institution's characteristics themselves and mark-up shifters are used as instrumental variables, whereas cost shifters are included as well in columns (3) and (6). Coefficients of RC Logit and BLP (1995) RC Logit are the mean values of random coefficients (β, α).

In order to test whether the existence of market discipline depends on the financial stability situation, the financial turmoil dummy variable interacts with both the spread and the risk-based capital ratios. In addition, since the risk-based capital ratios between banks and securities companies are different, an additional dummy variable that represents securities companies interacts with them.

The results from the Conditional Logit model show that the coefficients on spread in both the stable period and the financially distressed period are significantly negative, implying that an unobserved attribute is correlated with spread; thus, it is biased toward zero. Although the random coefficient model is known to deal with the price endogeneity problem, the result in column (4) shows that the coefficient on spread is statistically insignificant in the stable period while its sign is reversed to positive. However, the results from the IV Logit and the BLP (1995) RC Logit models for which the price variable is instrumented show that the spread coefficients in the stable period have the expected sign and are statistically significant. Furthermore, the magnitude of coefficient substantially increases in the BLP (1995) RC Logit model compared to that in the IV Logit model. This is in line with the finding from related studies (e.g., BLP (1995), Petrin (2002)). Table 3.B.6 shows the distribution of own-price elasticities for the tranquil times obtained from the IV Logit model and the BLP

(1995) RC Logit model.

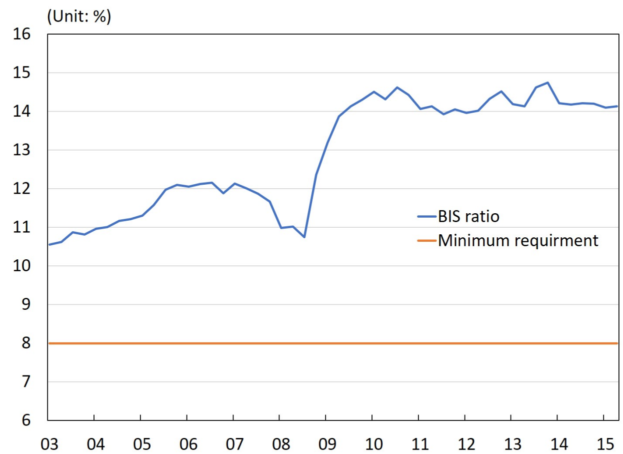
The coefficients on spread as well as the risk-based capital ratio in financial turmoil suggest that the market discipline in Korean banking sector has appeared differently depending on the financial stability situation. In both the IV Logit model and the BLP (1995) RC Logit model, the coefficients on spread are not significantly different from zero during the period of financial instability. Instead, the coefficients on the banks' risk-based capital ratio[23] become significantly positive, whereas those in the stable period are significantly negative. That is, regardless of the deposit rates, consumers prefer to deposit in a safer depository institution when the financial system is unstable. This phenomenon is similar to the flight to quality in the bond and equity markets occurred during the financial crisis.

It is counter-intuitive that the coefficients on the risk-based capital ratio are negative in tranquil times. However, for instance, the BIS ratios of Korean banks have been maintained over the minimum requirement during the sample period due to the experience of the Asian financial crisis. Figure 3.4.1 represents the unweighted BIS ratios of Korean banks and the minimum requirement. Therefore, it would be possible that depositors might regard them as safe regardless of the level of the BIS ratio. In addition, if consumers with low credit scores can use other services provided by a bank, such as loans, by depositing, it would lead to a lower risk-based capital ratio of the bank.

The signs, magnitudes, and significance of other coefficients are in accord with expectations. Depositors respond favorably to the size, the branch staffing, and the number of branches of depository institutions. The result that the coefficients on Group 2 financial institutions are significantly positive in the BLP (1995) RC Logit model reflects the characteristics of banks in the group: one is specialized in the transaction of foreign exchange, and the other is established in order to support financing of small to medium enterprises. The reasons for the negative coefficients on Group 5 financial institutions, securities companies,

[23]The coefficients on the securities companies' risk-based capital ratio are statistically insignificant in financial turmoil. However, considering that those in the stable period are negative and depositors in CMAs would have different risk attitude from depositors in banks, one can interpret this that depositors, even who are less risk-averse, become more risk-averse in the times of financial instability.

Figure 3.4.1: BIS ratios of Korean Banks



Source: Financial Supervisory Service of Korea

seems to be (i) differences in institutional framework from banks, including deposit insurance and regulation; and (ii) stigma effects from the collapse of Dongyang Securities whose CMAs market share was one of the highest[24] before the bankruptcy.

3.4.2 Consumer Welfare

In order to measure the effect of deregulation on consumer welfare, I calculate the equivalent variation (EV) following Small and Rosen (1981) in the context of the discrete choice model. According to Dick (2008), the equivalent variation (EV) can be calculated as

$$EV = S_t(p, x; \theta) - S_{t-1}(p, x; \theta), \tag{3.4.1}$$

where $S(p, x; \theta) = \ln[\Sigma_j \exp(\delta_j(p_j, x_j; \theta))]/\alpha$, and $\delta_j = x_j'\beta + \alpha p_j + \xi_j$, $\theta = (\beta, \alpha)$.[25]

However, the estimation results show that depositors do not respond to spread in the times of financial turmoil. This implies that even though CMAs offer higher interests than banks[26] that might induce banks to increase the deposit rates they offer, the deregulation

[24]Although CMAs balance for each securities company is not disclosed, it was known that CMAs balance of Dongyang Securities was around 10 trillion KRW in the peak.

[25]In this formula, α, the coefficient on spread, represents the marginal utility of income.

[26]During the sample period, the average spread of securities firms is 6.4 bps, whereas that of banks is -41.9 bps.

may not affect the consumer welfare at all in terms of a monetary unit. Furthermore, one cannot exclude the possibility that it might have a negative effect on consumer welfare due to the weakness of CMAs or securities companies illustrated in Section 3.1: (i) most CMAs are not included in the scope of the financial instruments protected by the deposit insurance and (ii) there exists a regulatory gap between banks and securities companies.

Therefore, I compare changes in welfare focusing on the stable period following equation (3.4.1). Depositors experience a gain in welfare due to deregulation between tranquil times, with a mean of KRW 0.0005-0.005 per consumer per year. This implies, for example, with a welfare gain of KRW 0.002, a depositor carrying a median balance (33 million KRW as of the end of March 2017) can gain 66 thousand KRW per year. However, it should be noted that the welfare gain has been diluted due to the prolonged financial stress since the global financial crisis.

3.5 Conclusion

This paper sought to apply structural econometric modeling in the field of industrial organization to finance. The results suggest that unlike other products (e.g., automobiles, Petrin (2002)), a new financial instrument does not necessarily improve consumer welfare even if it seems competitive in terms of price; this finding may be due to the existence of market discipline within financial markets. This implies that in order to achieve the goal of deregulation in the payment and settlement systems, it is necessary to devise an institutional framework that can reduce the difference in risk between products and financial institutions, which would foster a level playing field for financial institutions.

The model of the paper relies on simplifying assumptions. For instance, given that the services provided by securities companies are different from banks, the assumption of treating securities companies that provide CMAs as banks might not reflect the reality. Also, some consumers might split a significant amount of deposits in multiple banks. To manage the problem, one can consider applying multiple-discrete choice model (e.g., Hendel (1999)), which may need micro-level data. It is important to note that the approach in this paper uses only market-level data.

Having taken basic but important steps in estimating a demand system, this model has the potential to lead to future research with improvements. For example, the model can be used to measure the effect of changes in prudential regulation. Also, given that two internet-only banks were newly established in Korea in 2017, the demand model for deposits taking account of both price and service competition can be extended to measure the effect of introducing internet-only banks.

APPENDICES

APPENDIX A

An Overview on the Korean Financial System

3.A.1 Financial Industry

The Korean financial system has been developed as bank-based, in that banks play a leading role in mobilizing savings, allocating capital, overseeing the investment decisions of corporate managers, and providing risk management vehicles.[27] Table 3.A.1 shows the total assets of the major financial institutions in Korea and their shares. Although banks' asset shares in the financial system have decreased after the Asian financial crisis, they still account for the largest portion with more than 50%.

Table 3.A.1: Total Assets of Major Financial Institutions in Korea

(Unit: Trillion KRW, %)

	1990	1995	2000	2005	2010	2015
Banks	249.7 (63.3)	595.8 (62.9)	982.2 (63.4)	1,213.5 (57.8)	1,884.1 (54.8)	2,440.7 (57.3)
Merchant Banking Corporations[1)]	23.7 (6.0)	45.9 (4.8)	21.3 (1.4)	13.2 (0.6)	24.2 (0.7)	11.1 (0.3)
Mutual Savings Banks	11.5 (2.9)	32.6 (3.4)	24.2 (1.6)	44.9 (2.1)	91.3 (2.7)	43.9 (1.0)
Credit Cooperatives	24.6 (6.2)	75.9 (8.0)	145.1 (9.4)	220.2 (10.5)	360.9 (10.5)	533.5 (12.5)
Postal Savings	3.4 (0.9)	7.0 (0.7)	24.5 (1.6)	37.8 (1.8)	55.4 (1.6)	65.6 (1.5)
Insurance Companies	34.6 (8.8)	86.6 (9.1)	163.6 (10.6)	308.6 (14.7)	507.5 (14.8)	816.0 (19.2)
Securities Companies	16.6 (4.2)	27.8 (2.9)	42.0 (2.7)	62.7 (3.0)	189.4 (5.5)	344.5 (8.1)
Collective Investment Business Entities	30.6 (7.8)	76.0 (8.0)	146.7 (9.5)	198.4 (9.4)	325.3 (9.5)	4.8 (0.1)
Total	394.8 (100.0)	947.6 (100.0)	1,549.5 (100.0)	2,099.2 (100.0)	3,438.0 (100.0)	4,260.1 (100.0)

Note: 1) Including consolidated financial accounts of banks and securities companies.
Source: Bank of Korea, Financial Supervisory Service of Korea

Since the 1980s, the Korean government had eased regulations on the financial market entry in order to foster competition among financial institutions. As a result, the number

[27] Please refer to Demirgüç-Kunt and Levine (1999) for more details on bank-based and market-based financial systems.

of banks increased to thirty-three before the Asian financial crisis in 1997. However, as the soundness of banks deteriorated during the crisis, insolvent financial institutions were resolved through liquidation or mergers and acquisitions based upon judgements as to their survivability. Therefore, as of Q4 2002 when restructuring due to the crisis was finalized, the number of banks decreased to nineteen. As of Q4 2016[28], there were seventeen banks, including six nationwide banks, six local banks, and five specialized banks.[29] Table 3.A.2 shows the list of banks in Korea.

Table 3.A.2: List of Banks in Korea (as of Q4 2016)

Nationalwide Banks	Local Banks	Specialized Banks
Kookmin	Kyongnam	Nonghyup
Shinhan	Kwangju	Suhyup
Woori	Daegu	IBK
KEB Hana	Busan	KDB
SC Korea	Jeonbuk	KEXIM
Citibank Korea	Jeju	

Source: Financial Supervisory Service of Korea

Deposits are major funding sources for Korean Banks. To be specific, as of Q4 2016, deposits consisted of 83.6% of banks' funding. Therefore, banks have the largest portion, 68.4%, in the deposit market, and it is made up of deposits in depository institutions, such as banks, merchant banking corporations, mutual savings banks, credit cooperatives, postal savings, and CMAs.

3.A.2 Payment and Settlement Systems

The payment and settlement systems in Korea consist of a large-value payment system, retail payment systems, securities settlement systems, and foreign exchange settlement systems. While a large-value payment system is used for transactions between financial institutions,

[28]Shinhan and Chohung were merged in Q2 2006, and Hana and KEB were merged in Q3 2015.

[29]Specialized banks are established with specific purposes of bolstering financing in areas encountering funding difficulties due to shortages of finance, profitability and expertise. However, except Korea Development Bank (KDB) and Export-Import Bank of Korea (KEXIM), their business model, such as the funding structure, is similar to commercial banks.

retail payment systems are used for those among individuals or corporations. By the enforcement of the Capital Market and Financial Investment Business Act, as of Q4 2016, twenty-five securities companies[30] are participating in six retail payment systems operated by KFTC[31]. Thus, CMA holders became able to use them as a means of exchange.

However, retail payment systems are processed by net settlements that net obligations arising from transactions in the retail payment systems are transferred between the current accounts of the financial institutions involved at a designated time. Therefore, unlike the real-time gross settlement system, financial institutions are exposed to settlement risks such as credit risk when the counterpart fails to transfer fund.

3.A.3 Financial Stability Situation

In order to test the hypothesis that the existence of market discipline depends on the financial stability condition, it is essential to identify a period of financial distress. However, as Aspachs-Bracons et al. (2012) point out, it is hard to measure financial fragility, whereas inflation can be measured by a relatively simple and intuitive variable, the consumer price index. After the global financial crisis, there has been a growth in literature concerning the field of devising the financial stability index.[32]

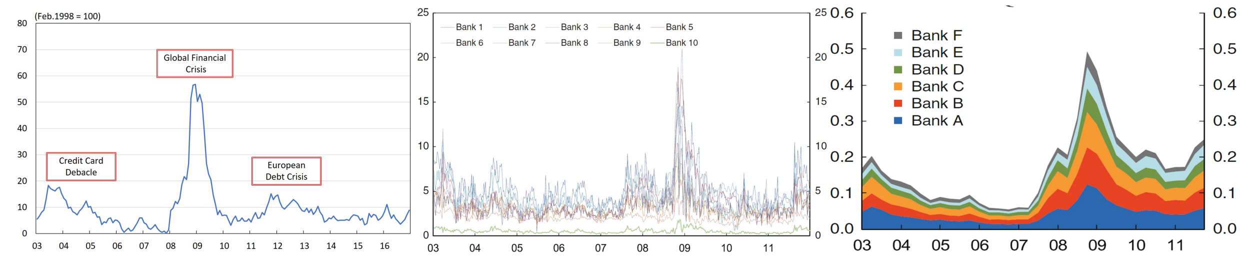
Figure 3.A.1 shows the financial stability indices of Korea, where (a) represents a composite financial stability index published in the Financial Stability Report of the Bank of Korea, (b) represents CoVaR based on Adrian and Brunnermeier (2016), and (c) represents Marginal Expected Shortfall (MES) based on Tarashev et al. (2010). It shows that these indices all have similar trends reflecting the major events in the Korean financial system. Table 3.A.3 shows major financial events in Korea since 2000.

[30] Sixteen major securities companies joined the retail payment systems in Q3 2009.

[31] These include Electronic Banking System, Cash Management Service Network, Interbank Remittance System, Giro System, CD Network, and Payment Gateway System. As of Q4 2016, those systems counts for 89.2% of total transaction volume.

[32] Please refer to Silva et al. (2017) for more details.

Figure 3.A.1: Financial Stability Indices of Korea



(a) Composite Index (b) CoVaR (c) MES

Source: Bank of Korea *Financial Stability Report*, Lee et al. (2013)

Table 3.A.3: Major Financial Events in Korea

Period	Events
2003	Credit Card Debacle
2008-	Global Financial Crisis
2010-	European Debt Crisis
2011	Bankruptcy of Mutual Savings Banks
2013	Bankruptcy of Dongyang Securities

APPENDIX B

Tables

Table 3.B.4: Summary Statistics

	Mean	Std. Dev.	Max	Min
Market Share	0.0412	0.0394	0.1604	0.0014
Spread	-0.0041	0.0073	0.0238	-0.0218
- Stable Period	-0.0076	0.0049	0.0042	-0.0218
- Financial Turmoil Period	0.0010	0.0072	0.0238	-0.0186
Deposit Interest Rates	0.0284	0.0070	0.0534	0.0107
Policy Interest Rates	0.0325	0.0096	0.0515	0.0170
Group 2 Financial Institutions	0.1214	0.3267	1	0
Group 3 Financial Institutions	0.1214	0.3267	1	0
Group 4 Financial Institutions	0.4248	0.4946	1	0
Group 5 Financial Institutions	0.0291	0.1683	1	0
Emplyees per Branch	12.6582	3.1767	28.0752	7.2697
Number of Branch	475.5334	411.8725	1,789	31
BIS Ratio	0.1288	0.0183	0.1825	0.0855
- Stable Period	0.1221	0.0172	0.1825	0.0855
- Financial Turmoil Period	0.1389	0.0151	0.1771	0.0940
Net Operating Capital Ratio	0.1654	0.0274	0.2221	0.1217
- Stable Period	0.1379	0.0113	0.1557	0.1217
- Financial Turmoil Period	0.1792	0.0220	0.2221	0.1469
Observations		824		

Table 3.B.5: Classification of Financial Institutions

Group	Description
Group 1	Banks with assets over 100 Trillion KRW as of Q2 2015 and belonging to a holding company
Group 2	Banks with assets over 100 Trillion KRW as of Q2 2015 (excluding Group 1 banks)
Group 3	Banks with assets less than 100 Trillion KRW as of Q2 2015 and foreign owned
Group 4	Banks with assets less than 100 Trillion KRW as of Q2 2015 (excluding Group 3 banks)
Group 5	Securities Companies

Table 3.B.6: Distribution of Own Price Elasticities

	10%	25%	Median	75%	90%
IV Logit Model	0.0512	0.0992	0.1435	0.1950	0.2581
BLP (1995) RC Model	0.2175	0.3067	0.4539	0.6034	0.7599

Table 3.B.7: Estimation Results

	Conditional Logit	IV Logit		RC Logit	BLP (1995) RC Logit	
	(1)	(2)	(3)	(4)	(5)	(6)
Spread						
- Stable Period	-11.9073***	18.8223**	17.4081**	3.5915	49.4622***	48.1135***
	(4.5920)	(7.8120)	(6.9278)	(4.6821)	(8.5157)	(7.5949)
- Financial Turmoil Period	-14.7139***	8.9737	4.6535	-19.2045***	6.0866	-1.3014
	(3.4651)	(6.6238)	(5.6131)	(3.6118)	(7.5065)	(6.0896)
Group 2 Financial Institutions	-0.1335**	0.0946	0.0682	-0.0725	0.2203***	0.1748**
	(0.0640)	(0.0773)	(0.0709)	(0.0876)	(0.0848)	(0.0758)
Group 3 Financial Institutions	-0.2875***	-0.1615**	-0.1792**	-0.3011***	-0.1469*	-0.1786**
	(0.0765)	(0.0814)	(0.0789)	(0.0971)	(0.0884)	(0.0847)
Group 4 Financial Institutions	-0.9126***	-0.7756***	-0.7925***	-1.1262***	-0.9514***	-0.9753***
	(0.0996)	(0.1088)	(0.1053)	(0.1124)	(0.1159)	(0.1110)
Group 5 Financial Institutions	-2.9003***	-2.6662***	-2.6643***	-5.4742***	-5.1946***	-5.2284***
	(0.3346)	(0.3246)	(0.3199)	(0.9226)	(0.3374)	(0.3316)
Employees per Branch	0.0706***	0.0650***	0.0657***	0.0652***	0.0568***	0.0590***
	(0.0077)	(0.0076)	(0.0075)	(0.0089)	(0.0084)	(0.0082)
Number of Branch	0.0019***	0.0021***	0.0020***	0.0019***	0.0022***	0.0021***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Risk-based Capital Ratio						
- Banks	-4.2529***	-6.5529***	-6.5063***	-4.2852***	-7.5974***	-7.6293***
× Stable Period	(1.2916)	(1.2830)	(1.2536)	(1.3371)	(1.4098)	(1.3793)
- Banks	3.4807**	2.7265**	3.2595**	5.9185***	5.0045***	5.7892***
× Financial Turmoil Period	(1.4694)	(1.3635)	(1.3145)	(1.7733)	(1.4116)	(1.3614)
- Securities Companies	-2.9404	-8.3174***	-8.2239***	-7.0095	-14.1804***	-13.9986***
× Stable Period	(2.2782)	(2.1462)	(2.0458)	(6.6984)	(2.1963)	(2.0748)
- Securities Companies	1.0803	-1.4192	-0.9217	2.8133	-0.1593	0.7458
× Financial Turmoil Period	(1.3535)	(1.3286)	(1.2767)	(4.9401)	(1.4447)	(1.3664)
Financial Turmoil	-0.9872***	-1.4363***	-1.4870***	-1.3268***	-1.9962***	-2.0884***
	(0.2588)	(0.2442)	(0.2313)	(0.2971)	(0.2594)	(0.2473)
Constant	-3.4206***	-3.0279***	-3.0279***	-2.3475***	-1.7301***	-1.7213***
	(0.2171)	(0.2082)	(0.2041)	(0.2147)	(0.2290)	(0.2238)

Note: *** Significant at 1%, ** Significant at 5%, * Significant at 10%. Standard errors are in the parentheses.

BIBLIOGRAPHY

BIBLIOGRAPHY

Adrian, T. and Brunnermeier, M. K. (2016). Covar. *American Economic Review*, 106(7):1705–1741.

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

Altunbas, Y., Carbo, S., Gardener, E. P., and Molyneux, P. (2007). Examining the relationships between capital, risk and efficiency in European banking. *European Financial Management*, 13(1):49–70.

Amsler, C., Prokhorov, A. B., and Schmidt, P. (2021). A new family of copulas, with application to estimation of a production frontier system. *Journal of Productivity Analysis*, 55(1):1–14.

Arena, M. (2003). Bank fundamentals, bank failures, and market discipline: An empirical analysis for Latin America and Asia during the nineties. Working Paper.

Aspachs-Bracons, O., Goodhart, C. A., Segoviano, M., Tsomocos, D. P., and Zicchino, L. (2012). Searching for a metric for financial stability. In Goodhart, C. A. and Tsomocos, D. P., editors, *Financial Stability in Practice*, chapter 12, pages 251–276. Edward Elgar Publishing, Northampton, MA.

Battese, G. E. and Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, 38(3):387–399.

Bauer, P. W. (1985). *An Analysis of Multiproduct Technology and Efficiency Using the Joint Cost Function and Panel Data: An application to U.S. Airline Industry.* PhD thesis, University of North Carolina, Chapel Hill, NC.

Bauer, P. W. (1990). Recent developments in the econometric estimation of frontiers. *Journal of Econometrics*, 46(1-2):39–56.

Berger, A. (1991). Market discipline in banking. Proceedings 328, Federal Reserve Bank of Chicago.

Berger, A. N. and Humphrey, D. B. (1992). Measurement and efficiency issues in commercial banking. In Griliches, Z., editor, *Output Measurement in the Service Sectors*, chapter 7, pages 245–300. University of Chicago Press, Chicago, IL.

Berger, A. N. and Humphrey, D. B. (1997). Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research*, 98(2):175–212.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2):242–262.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Bhatia, V., Basu, S., Mitra, S. K., and Dash, P. (2018). A review of bank efficiency and productivity. *OPSEARCH*, 55(3-4):557–600.

Bonaldi, P., Hortaçsu, A., and Kastl, J. (2015). An empirical analysis of funding costs spillovers in the EURO-zone with application to systemic risk. NBER Working Paper 21462, National Bureau of Economic Research.

Brown, B. W. and Walker, M. B. (1995). Stochastic specification in random production models of cost-minimizing firms. *Journal of Econometrics*, 66(1-2):175–205.

Caflisch, R. E. (1998). Monte carlo and quasi-monte carlo methods. *Acta Numerica*, 7:1–49.

Cambou, M., Hofert, M., and Lemieux, C. (2017). Quasi-random numbers for copula models. *Statistics and Computing*, 27(5):1307–1329.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186.

Chavas, J.-P. and Segerson, K. (1987). Stochastic specification and estimation of share equation systems. *Journal of Econometrics*, 35(2-3):337–358.

Christensen, L. R. and Greene, W. H. (1976). Economies of scale in U.S. electric power generation. *Journal of Political Economy*, 84(4):655–676.

Christensen, L. R., Jorgenson, D. W., and Lau, L. J. (1971). Transcendental logarithmic production frontiers. *Econometrica*, 39(4):255–256.

Deaton, A. and Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review*, 70(3):312–326.

Demirgüç-Kunt, A. and Levine, R. (1999). Bank-based and market-based financial systems: Cross-country comparisons. Policy, Research Working Paper 2143, The World Bank.

Dick, A. A. (2008). Demand estimation and consumer welfare in the banking industry. *Journal of Banking & Finance*, 32(8):1661–1676.

Ding, D. and Sickles, R. C. (2019). Capital regulation, efficiency, and risk taking: A spatial panel analysis of US banks. In Tsionas, M., editor, *Panel Data Econometrics*, chapter 13, pages 405–466. Academic Press, Cambridge, MA.

Egan, M., Hortaçsu, A., and Matvos, G. (2017). Deposit competition and financial fragility: Evidence from the US banking sector. *American Economic Review*, 107(1):169–216.

Embrechts, P., Lindskog, F., and Mcneil, A. (2003). Modelling dependence with copulas and applications to risk management. In Rachev, S. T., editor, *Handbook of Heavy Tailed Distributions in Finance*, volume 1 of *Handbooks in Finance*, chapter 8, pages 329–384. North-Holland, Amsterdam.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19(3):1257–1272.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3):253–290.

Feng, G. and Serletis, A. (2009). Efficiency and productivity of the US banking industry, 1998–2005: Evidence from the fourier cost function satisfying global regularity conditions. *Journal of Applied Econometrics*, 24(1):105–138.

Flannery, M. J. (1998). Using market information in prudential bank supervision: A review of the U.S. empirical evidence. *Journal of Money, Credit and Banking*, 30(3):273–305.

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

Guarda, P., Rouabah, A., and Vardanyan, M. (2013). Identifying bank outputs and inputs with a directional technology distance function. *Journal of Productivity Analysis*, 40(2):185–195.

Hendel, I. (1999). Estimating multiple-discrete choice models: An application to computerization returns. *Review of Economic Studies*, 66(2):423–446.

Hortaçsu, A., Kastl, J., and Zhang, A. (2018). Bid shading and bidder surplus in the US treasury auction system. *American Economic Review*, 108(1):147–169.

International Monetary Fund (2016). Monetary and financial statistics manual and compilation guide.

Joe, H. (2014). *Dependence Modeling with Copulas*. Monographs on Statistics and Applied Probability 134. CRC Press, Boca Raton, FL.

Jondrow, J., Lovell, C. A. K., Materov, I. S., and Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19(2-3):233–238.

Judd, K. L. (1998). *Numerical Methods in Economics*. The MIT Press, Cambridge, MA.

Kastl, J. (2017). Recent advances in empirical analysis of financial markets: Industrial organization meets finance. In Honoré, B., Pakes, A., Piazzesi, M., and Samuelson, L., editors, *Advances in Economics and Econometrics: Eleventh World Congress*, volume 2 of *Econometric Society Monographs*, pages 231–270. Cambridge University Press, Cambridge, UK.

Kim, K. and Petrin, A. (2015). Tests for price endogeneity in differentiated product models. *Journal of Econometric Methods*, 4(1):47–69.

Kumbhakar, S. C. (1989). Modelling technical and allocative inefficiency in a translog production function. *Economics Letters*, 31(2):119–123.

Kumbhakar, S. C. (1991). The measurement and decomposition of cost-inefficiency: The translog cost system. *Oxford Economic Papers*, 43(4):667–683.

Kumbhakar, S. C. (1997). Modeling allocative inefficiency in a translog cost function and cost share equations: An exact relationship. *Journal of Econometrics*, 76(1-2):351–356.

Kumbhakar, S. C. and Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge, UK.

Kumbhakar, S. C. and Tsionas, E. G. (2005a). The joint measurement of technical and allocative inefficiencies: An application of bayesian inference in nonlinear random-effects models. *Journal of the American Statistical Association*, 100(471):736–747.

Kumbhakar, S. C. and Tsionas, E. G. (2005b). Measuring technical and allocative inefficiency in the translog cost system: a bayesian approach. *Journal of Econometrics*, 126(2):355–384.

Kumbhakar, S. C. and Tsionas, E. G. (2011). Stochastic error specification in primal and dual production systems. *Journal of Applied Econometrics*, 26(2):270–297.

Lee, J. H., Ryu, J., and Tsomocos, D. P. (2013). Measures of systemic risk and financial fragility in Korea. *Annals of Finance*, 9(4):757 – 786.

Levy-Yeyati, E., Pería, M. S. M., and Schmukler, S. L. (2010). Depositor behavior under macroeconomic risk: Evidence from bank runs in emerging economies. *Journal of Money, Credit and Banking*, 42(4):585–614.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

McElroy, M. B. (1987). Additive general error models for production, cost, and derived demand or share systems. *Journal of Political Economy*, 95(4):737–757.

Meeusen, W. and van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18(2):435–444.

Melchers, R. E. and Beck, A. T. (2018). *Structural Reliability Analysis and Prediction.* JohnWiley & Sons Ltd., Hoboken, NJ, Third edition.

Melfi, C. A. (1984). *Estimation and Decomposition of Productive Efficiency in a Panel Data Model: An Application to Electric Utilities.* PhD thesis, University of North Carolina, Chapel Hill, NC.

Morokoff, W. J. and Caflisch, R. E. (1995). Quasi-monte carlo integration. *Journal of Computational Physics*, 122(2):218–230.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy*, 9(4):513–548.

Parmeter, C. F. and Kumbhakar, S. C. (2014). Efficiency analysis: A primer on recent advances. *Foundations and Trends in Econometrics*, 7(3-4):191–385.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, 110(4):705–729.

Reiss, P. C. and Wolak, F. A. (2007). Structural econometric modeling: Rationales and examples from industrial organization. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6A, chapter 64, pages 4277–4415. Elsevier B.V. and North-Holland, Amsterdam and Oxford.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472.

Ryu, J. (2021). Measurement and decomposition of cost inefficiency using copulas: An application to the U.S. banking industry. Working Paper.

Schmidt, P. (1984). An error structure for systems of translog cost and share equations. Econometrics Workshop Paper 8309, Department of Economics, Michigan State University, East Lansing, MI.

Schmidt, P. and Lovell, C. A. K. (1979). Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers. *Journal of Econometrics*, 9(3):343–366.

Schmidt, P. and Lovell, C. A. K. (1980). Estimating stochastic production and cost frontiers when technical and allocative inefficiency are correlated. *Journal of Econometrics*, 13(1):83–100.

Sealey, C. W. and Lindley, J. T. (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions. *The Journal of Finance*, 32(4):1251–1266.

Sickles, R. C. and Zelenyuk, V. (2019). *Measurement of Productivity and Efficiency: Theory and Practice*. Cambridge University Press, Cambridge, UK.

Silva, W., Kimura, H., and Sobreiro, V. A. (2017). An analysis of the literature on systemic financial risk: A survey. *Journal of Financial Stability*, 28:91–114.

Small, K. A. and Rosen, H. S. (1981). Applied welfare economics with discrete choice models. *Econometrica*, 49(1):105–130.

Stefanski, L. A. and Carroll, R. J. (1990). Deconvolving kernel density estimators. *Statistics*, 21(2):169–184.

Stephanou, C. (2010). Rethinking market discipline in banking : Lessons from the financial crisis. World Bank Policy Research Working Paper 5227, The World Bank.

Tarashev, N., Borio, C., and Tsatsaronis, K. (2010). Attributing systemic risk to individual institutions. Working Paper 308, Bank for International Settlements.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, NY, Second edition.

Wang, X.-F. and Ye, D. (2015). Conditional density estimation in measurement error problems. *Journal of Multivariate Analysis*, 133:38–50.