Estimation of a Nonparametric Model of Profit Frontiers with an Application for the Swedish Paper Industry

by

Shan Zhou

B.A., Wuhan University, 2005

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Economics

2016

This thesis entitled:
Estimation of a Nonparametric Model of Profit Frontiers with an Application for the Swedish Paper Industry
written by Shan Zhou
has been approved for the Department of Economics

Professor Carlos Martins-Filho, Chair

Professor Xiaodong Liu

Date ____________________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Zhou, Shan (Ph.D., Economics)

Estimation of a Nonparametric Model of Profit Frontiers with an Application for the Swedish Paper Industry

Thesis directed by Professor Carlos Martins-Filho, Chair

In this thesis, I extend the conditional quantile frontier approach developed in Aragon et al. (2005) for production functions to profit functions. Instead of estimating a conventional profit function, that envelops all observed data and whose estimation is sensitive to outliers, I first define a class of profit frontiers based on conditional quantiles of order α associated with an appropriate joint distribution of profit, input and output prices. I show that these conditional quantiles are useful in defining and ranking production units in terms of profit efficiency. Then I propose a nonparametric conditional quantile estimator for the α-profit frontier by integrating a suitably defined estimator for the profit density. My estimator is inspired by that proposed in Martins-Filho and Yao (2008), but instead of adopting their traditional Rosenblatt-density estimator as a basis for the α-profit frontier, I use the class of density estimators introduced by Mynbaev and Martins-Filho (2010). I establish consistency and asymptotic normality of the α-profit frontier estimator. The estimator is more robust to the outliers since it does not envelope the data. Additionally, under some smoothness conditions on the distribution function, the bias of the proposed estimator converges to zero faster than that of an estimator constructed based on the Rosenblatt density estimator. A Monte-Carlo simulation study seems to support the asymptotic results and shows that the proposed estimator has better performance than its competitors in some scenarios. In the second chapter, I use the α-quantile frontier estimator proposed in the first chapter to study profit efficiency in the Swedish paper industry. I also study production efficiency by using a modification of our proposed estimator.

Dedication

To my family.

Acknowledgements

I am deeply indebted to my advisor Carlos Martins-Filho for his advise and support. I would like to thank my dissertation committee members: Professors Xiaodong Liu, Donald Waldman, Scott Savage and Vanja Dukic for the discussion and their helpful comments. I would also like to thank Professors Shawna Grosskopf, Rolf Färe and Tommy Lundgren for their help and suggestions. All errors are my own.

Contents

Chapter

Appendix

Tables

Table

Figures

Figure

Chapter 1

Estimation of a Nonparametric Model of Profit Frontiers: Theory

1.1 Introduction

Notions of efficiency permeate firm decision making in the classical microeconomic theory of the firm. For example, given a technology, firms are assumed to produce maximum output given a set of inputs; minimize cost and and maximize profit. However, there exists voluminous empirical evidence that suggests not all producers engage in successful optimizing behavior. (For instance, Berger and Humphrey (1993), Maudos et al. (2002), Färe et al. (2004), etc.) In various settings it is useful to measure and study the magnitude and nature of the inefficiency that pulls firms away from the relevant efficient frontier, be it a production, cost or profit function. Starting with Farrell (1957), a vast literature has emerged focusing on production or technical efficiency. These studies usually postulate a common production frontier for all firms, and measure the technical efficiency by a suitably defined distance between the production plan of each firm and the frontier. However, the ultimate objective of a firm is to maximize profit. Besides technical efficiency, a major component of a firm's profit-seeking behavior involves allocative efficiency, which captures the ability of choosing optimal proportions of inputs and outputs in the production process. In this paper, we provide a way of measuring and estimating a profit function and profit efficiency, which represents the combined effect of both technical and allocative efficiency. The basic idea, inspired by Aragon et al. (2005), is to describe profit functions at different efficiency levels as quantile functions of a suitably defined conditional distribution. We then estimate them by an improved nonparametric kernel method.

Empirical and theoretical models for measuring efficiency and estimating frontiers have fallen into two broad categories; stochastic and nonstochastic models of frontier. The basic principle in stochastic models is to describe the variable of interest (output, cost, profit) as being generated by a sum of the function of interest and a non-observed error term consisting of a noise and an inefficiency term. It was originally proposed by Aigner et al. (1977) and Meeusen and van den Broeck (1977) in the context of production functions. Greene (2008) and Kumbhakar and Lovell (2000) provide extensive reviews and applications of these models. Estimation of these models is normally conducted by maximum likelihood methods. Considering the possibility of mis-specification of parametric frontier models, Fan et al. (1996) investigate semiparametric estimation of a stochastic frontier model with nonparametric production function. Recent developments in the estimation of nonparametric stochastic frontier models include, among others, Kumbhakar et al. (2007) and Martins-Filho and Yao (2015).

A critical drawback of stochastic frontier models is that they generally require strong distributional assumptions regarding the inefficiency and noise terms, which are sometimes impractical for certain models. Nonstochastic frontier models assume that all observations lie inside the frontier and any deviation from the frontier is caused by inefficiency. The most popular nonparametric efficiency estimators are based on the idea of estimating the attainable set by the smallest set within some class that envelops the observed data. Data envelopment analysis (DEA) and free disposal hull (FDH) estimators are among the most popular and have been widely used in efficiency analysis since Charnes et al. (1978). The FDH estimator of the frontier is the free disposal hull of the observations and the DEA estimator is the convex cone of the FDH estimator. They rely on linear programming methods to search for the most efficient units, which are then connected to form a minimum enveloping frontier. DEA and FDH are very appealing to researchers because they rely on very few assumptions and are easy to implement; however, they suffer some critical drawbacks. Park et al. (2000, 2010) and Simar and Vanhems (2012) obtain general asymptotic properties and convergence rates of FDH and DEA estimators under certain assumptions. Convergence rates are also obtained in Korostelev et al. (1995); Kneip et al. (1998) and Gijbels et al. (1999) in special cases.

Similar to other nonparametric estimators, DEA and FDH estimators suffer from the "curse of dimensionality." The convergence speed of these estimators becomes much slower as the dimensionality of the problem increases. Another major drawback of these methods is that the estimation of the frontier is highly influenced by the most efficient firms, outliers or extreme values. Hence, these methods are not robust and highly sensitive to a small set of observations. Recently, Simar and Zelenyuk (2011) propose stochastic versions of the FDH and DEA estimators by allowing noise into the model, but the properties of these estimators remain unknown.

Considering these drawbacks, Cazals et al. (2002) introduce the concept of production frontier of order m and provide a robust envelopment estimator. Instead of the full production frontier, they consider the expected maximum output among m firms drawn from the population of firms using less than a given level of inputs. A new probabilistic interpretation of the frontiers and the efficiency scores is provided in the paper. Daouia and Simar (2005, 2007) and Daouia et al. (2010) further extend this idea and link frontier estimation to extreme value theory. Inspired by this idea, Aragon et al. (2005) introduces a quantile approach in production frontier analysis. They define a production function of continuous order α based on conditional quantiles of a distribution that describes the generation of inputs and outputs of a production process. Estimators for these conditional quantiles are by nature much more robust to outliers as they do not envelope all observations. Martins-Filho and Yao (2008) improves this method by introducing a smooth kernel estimator at the cost of introducing a bias term vanishing with the sample size. In this paper, inspired by Aragon et al. (2005), we define a profit frontier of continuous order α and propose an easy-to-implement nonparametric estimator for these profit frontiers. Our estimator is based on the kernel estimator proposed in Martins-Filho and Yao (2008), but we improve on their estimation method by using a new class of kernels proposed by Mynbaev and Martins-Filho (2010) that promise to reduce the order of the bias of the class of estimators under study.

Throughout the paper we consider competitive firms with technology represented by a production function $y = \rho(x)$ where $y \in \mathbb{R}_+^{d_1}$ is an output vector and $x \in \mathbb{R}_+^{d_2}$ is an input vector.[1] Given a vector of input prices $w = (w_1, \cdots, w_{d_2})' \in \mathbb{R}^{d_2}_{++}$ and a vector of output prices $p = (p_1, \cdots, p_{d_1})' \in \mathbb{R}^{d_1}_{++}$, profit is given by

$$\pi = p'\rho(x) - w'x.$$

We assume that for all input prices $w \in \mathbb{R}^{d_2}_{++}$ and output prices $p \in \mathbb{R}^{d_1}_{++}$, profit is bounded above as a function of x, i.e. there exists $0 < B_\pi < \infty$ such that $0 \leq \pi \leq B_\pi$ for all x. One can maximize profit with respect to x to get the input demand functions $x^* = x(p, w)$. If the maximum exists, then the maximum profit is given by

$$\pi(p, w) = p'\rho(x(p, w)) - w'x(p, w).$$

The value of maximum profit depends on p and w which are exogenous to the firm's decision making process. $\pi(p, w)$ is what we call the profit function throughout the paper. Given the existence of inefficiency, our objective is to estimate profit functions and assess firms' efficiency levels. There are a number of differences in estimating a profit function compared to estimating a production function. First, the derivation of the profit function relies on many assumptions on market structure and it is difficult to justify the assumption of a parametric form for the profit function. Second, the production function is monotonic nondecreasing with respect to its arguments (inputs), while the profit function is nondecreasing with respect to some of its arguments (output prices) and nonincreasing with other arguments (input prices). Perhaps, due to these difficulties there is a much smaller literature devoted to the analysis of profit efficiency. Existing empirical studies, such as Ali et al. (1994) and Maudos et al. (2002) are mostly based on parametric stochastic profit frontier models with very high probability of misspecification. We show in this paper that these problems can be solved by our nonparametric quantile approach.

The rest of the chapter is divided into four sections. Section 1.2 describes the model and its estimation in detail. Section 1.3 provides the main assumptions and theorems that establish the asymptotic behavior of our estimators. Section 1.4 contains a small Monte Carlo study that

[1] Throughout the paper we define $\mathbb{R}^d_+ = \times^d_{i=1}[0, \infty)$ as the d fold Cartesian product of $[0, \infty)$; and $\mathbb{R}^d_{++} = \times^d_{i=1}(0, \infty)$.

implements the estimator, investigates its finite sample properties and compares performances of smooth and nonsmooth estimators. Section 1.5 concludes. The proofs for all propositions and theorems are collected in the Appendix, where a set of auxiliary lemmas is also given.

1.2 Model and Estimation

1.2.1 Profit function of order α

Let $\{(\Pi_i, P_i, W_i)\}_{i=1}^n$ be a sequence of independent and identically distributed random vectors defined in the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and having the same distribution function as (Π, P, W), which is denoted by F with associated density function f. $\Pi_i \in \mathbb{R}_+$ denotes profit[2] , $P_i \in \mathbb{R}_+^{d_1}$ denotes a vector of output prices and $W_i \in \mathbb{R}_{++}^{d_2}$ denotes a vector of input prices associated with a firm or producing unit i. We denote the support of f by Ψ and focus on the set $\Psi^*(p, w) = \{(\Pi, P, W) \in \Psi : \mathcal{P}(P \leq p, W \geq w) > 0\}$ where $\mathcal{P}$ represent probability. Given $C_{p,w} = \{P \leq p, W \geq w\} \subset \mathcal{F}$ we let

$$F(\pi|C_{p,w}) = \mathcal{P}(\Pi \leq \pi|P \leq p, W \geq w) = \frac{\mathcal{P}(\Pi \leq \pi, P \leq p, W \geq w)}{\mathcal{P}(P \leq p, W \geq w)}. \tag{1.1}$$

and give the following probabilistic definition of a profit function

$$\pi(p, w) := \inf\{\pi \in [0, B_\pi] : F(\pi|C_{p,w}) = 1\}. \tag{1.2}$$

As defined, the value of the profit function at (p, w) is given by the "smallest" number that is larger than or equal to the highest attainable profit given input price larger than or equal to w and output prices less than or equal to p (vector inequalities are all taken element-wise). By definition, for any (Π_i, P_i, W_i) with $P_i \leq p$ and $W_i \geq w$, we must have $\Pi_i \leq \pi(p, w)$ with probability 1. That is, $\pi(p, w)$ envelopes all data points.

Similar to Aragon et al. (2005), in the context of production functions, our definition of profit function suggests the alternative concept of a profit function of continuous order $\alpha \in (0, 1]$, as the

[2] Of course, profit can be negative in the short run. Here we assume firms earning negative profit will exit the market eventually.

quantile function of order α of the conditional distribution of Π given $C_{p,w}$. Thus, we define

$$\pi_\alpha(p,w) := F^{-1}(\alpha|C_{p,w}) = \inf\{\pi \in [0, B_\pi] : F(\pi|C_{p,w}) \geq \alpha\} \tag{1.3}$$

where $F^{-1}(\cdot|C_{p,w})$ is the generalized inverse of $F(\cdot|C_{p,w})$. We call $\pi_\alpha(p,w)$ the profit function of order α. It is apparent that the profit function in (1.2) corresponds to that in (1.3) when $\alpha = 1$. By definition $F^{-1}(\alpha|C_{p,w})$ is the profit threshold exceeded by $100(1-\alpha)\%$ of firms that face input prices larger than or equal to w and output prices less than or equal to p. If the conditional distribution $F(\cdot|C_{p,w})$ is strictly increasing at $F^{-1}(\alpha|C_{p,w})$ for $\alpha \in (0,1]$, we have

Proposition 1. *Assume that for every (p,w) such that $\mathcal{P}(P \leq p, W \geq w) > 0$, the conditional distribution function $F(\cdot|C_{p,w})$ is strictly increasing at $F^{-1}(\alpha|C_{p,w})$ for $\alpha \in (0,1]$. Then, for any $(\pi,p,w) \in \Psi^*$, we have $\pi = \pi_\alpha(p,w)$ with $\alpha = F(\pi|C_{p,w})$.*

Proposition 1 shows that any vector $(\pi,p,w) \in \Psi^*$ belongs to some profit function of order α. That is, the quantile curves $\{(\pi_\alpha(p,w),p,w) : \mathcal{P}(P \leq p, W \geq w) > 0, \alpha \in (0,1]\}$ cover the entire set Ψ^* of attainable profits, input and output prices. Given a firm or production unit associated with $(\pi_\alpha(p,w),p,w)$, its profit is larger than $100\alpha\%$ of all units facing the same or less favorable prices (higher input prices and lower output prices) and less than $100(1-\alpha)\%$ all other firms or production units. Thus, the order of the conditional quantile curve to which (π,p,w) belongs, gives a measure of "profit efficiency" of the firm or production unit (π,p,w) relative to all other production units facing the same or less favorable prices.

It is clear that, for any fixed (p,w) such that $\mathcal{P}(P \leq p, W \geq w) > 0$, $\pi_\alpha(p,w)$ is a monotone nondecreasing function of α. The following proposition shows that as $\alpha \to 1$, $\{\pi_\alpha(p,w)\}_{0<\alpha\leq 1}$ converge to $\pi(p,w)$ pointwise, and under additional regularity condition, the convergence is uniform over a suitably defined set.

Proposition 2. *For any fixed (p,w) such that $\mathcal{P}(P \leq p, W \geq w) > 0$, $\lim_{\alpha\to 1}\pi_\alpha(p,w) = \pi(p,w)$. If, in addition, for every $\alpha \in (0,1]$, $\pi_\alpha(p,w)$ is continuous on the interior of the support of the* *marginal density of (P, W), denoted by S_0, then for any compact subset $\Phi \subset S_0$*

$$\sup_{(p,w)\in\Phi} |\pi_\alpha(p,w) - \pi(p,w)| \to 0 \text{ as } \alpha \to 1.$$

The most natural measure of profit efficiency of a firm or production unit i, compares its realized profit Π_i to the profits attained by all firms facing output prices $p \leq P_i$, the output prices faced by unit i, and input prices $w \geq W_i$, the input prices faced by unit i for $\alpha = 1$. However, in an attempt to decrease the sensitivity of our measurement of profit efficiency to outliers or extreme values, we introduce a new measure of efficiency that compares the profit of a production unit to a profit function of order α. Thus, we say that the firm or production unit i is α-profit efficient if its profit $\Pi_i \geq \pi_\alpha(P_i, W_i)$. Otherwise, such firm is labeled α-profit inefficient. Thus, we can define an α-efficiency score as $e_\alpha(\Pi_i, P_i, W_i) = \Pi_i/\pi_\alpha(P_i, W_i)$. Note, that different from efficiency scores that emerge from traditional frontiers that envelope all possible triples (Π_i, P_i, W_i), $e_\alpha(\Pi_i, P_i, W_i)$ may be greater than 1, since the profit function of order α does not provide an upper bound for the profits of all firms facing prices $p \leq P_i$ and $w \geq W_i$.

The concept of profit functions of order α can be easily extended to settings where additional constraints on profit and technology are appropriate. We give some examples below.

Example 1

Firms may face certain environmental variables z, and these constrain their choices of inputs x (or outputs y). For example, we consider a firm's production capacity is an exogenous variable $z \in \mathbb{R}^{d_1}_{++}$. Then its optimization problem becomes

$$\begin{aligned} \max_x \quad & p'\rho(x) - w'x, \\ \text{s.t.} \quad & \rho(x) \leq z. \end{aligned}$$

The profit function derived from above problem would be $\pi(z, p, w)$. In this case, we can adjust our definition of profit frontier by comparing units with the same or smaller production capacities. The definition of profit function and frontiers becomes

$$\pi(z, p, w) := \inf\{\pi \in [0, B_\pi] : F(\pi|C_{z,p,w}) = 1\}$$

and

$$\pi_\alpha(z,p,w) := F^{-1}(\alpha|C_{z,p,w}) = \inf\{\pi \in [0, B_\pi] : F(\pi|C_{z,p,w}) \geq \alpha\},$$

where random vector Z represents the production capacity of a firm. $C_{z,p,w}$ is a conditional set $\{Z \leq z, P \leq p, W \geq w\}$ and $F(\pi|C_{z,p,w})$ is a conditional distribution

$$F(\pi|C_{z,p,w}) = \mathcal{P}(\Pi \leq \pi | Z \leq z, P \leq p, W \geq w) = \frac{\mathcal{P}(\Pi \leq \pi, Z \leq z, P \leq p, W \geq w)}{\mathcal{P}(Z \leq z, P \leq p, W \geq w)}.$$

Therefore, $\pi(z,p,w)$ represents the smallest function that is larger than or equal to the highest attainable profit given input prices larger than or equal to w, output prices less than or equal to p, and with capacity less than or equal to z. $\pi_\alpha(z,p,w)$ is defined by comparing all units with same or smaller production capacities facing the same or less favorable input and output prices.

Example 2

Consider a firm that has monopoly power. Then, market demand affects output prices and p becomes endogenous. Assume the market price of output is determined by the inverse demand function $p(y)$, and the price elasticities of demand are represented by

$$\epsilon = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1d_1} \\ \epsilon_{21} & \epsilon_{22} & \cdots & \epsilon_{2d_1} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{d_11} & \epsilon_{d_12} & \cdots & \epsilon_{d_1d_1} \end{bmatrix},$$

where $\epsilon_{ij} = 1/(\frac{dp_j}{dy_i} \cdot \frac{y_i}{p_j})$, and p_j is the jth element of $p(y)$. Now the optimization problem becomes

$$\max_x p'(\rho(x))\rho(x) - w'x.$$

Take the first derivative for the objective function with respect to x, by matrix calculus we have

$$\frac{\partial \pi}{\partial x} = \frac{\partial \rho}{\partial x}(x) \cdot \frac{\partial p}{\partial \rho}(\rho(x)) \cdot \rho(x) + \frac{\partial \rho}{\partial x}(x) \cdot p(\rho(x)) - w,$$

where $\frac{\partial \rho}{\partial x}(x)$ is a $d_2 \times d_1$ matrix, $\frac{\partial p}{\partial \rho}(\rho(x))$ is a $d_1 \times d_1$ matrix. The first order condition implies

$$\frac{\partial \rho}{\partial x}(x^*) \cdot [\frac{\partial p}{\partial \rho}(\rho(x^*)) \cdot \rho(x^*) + p(\rho(x^*))] = w.$$

Assume price elasticity of demand ϵ does not depend on inputs x, By the definition of ϵ_{ij}, $(\frac{\partial p}{\partial \rho}(\rho(x^*)))_{ij} = (1/\epsilon_{ij})(\frac{p_j(\rho(x^*))}{\rho_i(x^*)})$, where $p_j(\rho(x^*))$ represents the jth element of vector $p(\rho(x^*))$ and $\rho_i(x^*)$ represents the ith element of vector $\rho(x^*)$. Hence,

$$\frac{\partial p}{\partial \rho}(\rho(x^*)) \cdot \rho(x^*) = \begin{bmatrix} (1/\epsilon_{11})p_1(\rho(x^*))+ & \dots & +(1/\epsilon_{1d_1})\,p_{d_1}(\rho(x^*)) \\ (1/\epsilon_{21})\,p_1(\rho(x^*))+ & \dots & +(1/\epsilon_{2d_1})\,p_{d_1}(\rho(x^*)) \\ \vdots & \ddots & \vdots \\ (1/\epsilon_{d_1 1})\,p_1(\rho(x^*))+ & \dots & +(1/\epsilon_{d_1 d_1})\,p_{d_1}(\rho(x^*)) \end{bmatrix}.$$

If there is only one input, the first order condition becomes

$$\frac{\partial \rho}{\partial x}(x^*) \cdot p(\rho(x^*))(1 + 1/\epsilon) = w.$$

Solving the first order condition gives the optimized inputs $x(\epsilon_{11}, \dots, \epsilon_{1d_1}, \dots, \epsilon_{d_1 d_1}, w)$ and thus the profit function in this setting is $\pi(\epsilon_{11}, \dots, \epsilon_{1d_1}, \dots, \epsilon_{d_1 d_1}, w)$. For a single output, we can define an α profit frontier as

$$\pi_\alpha(\epsilon, w) := \inf\{\pi \in [0, B_\pi] : F(\pi | C_{\epsilon,w}) \geq \alpha\},$$

where $F(\pi | C_{\epsilon,w}) = \mathcal{P}(\Pi \leq \pi | \Upsilon \geq \epsilon, W \geq w)$.

Example 3

More generally, we consider the alternative profit function in Humphrey and Pulley (1997). Assume firms (or banks, in their paper) have some market power but are not monopolists, and firm's decision is a mix of price-taking and price-setting behavior. They treat output as essentially exogenous at the time of decision, and focus on negotiating prices rather than output quantities to maximize profit. The optimization problem is

$$\max_{p,x} \quad p'y - w'x,$$
$$\text{s.t.} \quad g(y, x, p, w, z) = 0,$$

where $g(y, x, p, w, z)$ represents a firm's ability for transforming given values of y, x, w, and z into output prices. z includes variables capturing exogenous factors affecting firms' profitability. The profit function is derived by solving the constrained optimal choice for output prices $p = p(y, w, z)$

and input quantities $x = x(y, w, z)$. Therefore the alternative indirect profit function is given by $\pi(y, w, z) = p'(y, w, z)y - w'x(y, w, z)$; and a α frontier can be defined as:

$$\pi_\alpha(y, w, z) := \inf\{\pi \in [0, B_\pi] : F(\pi|C_{y,w,z}) \geq \alpha\},$$

where

$$F(\pi|C_{y,w,z}) = \begin{cases} \mathcal{P}(\Pi \leq \pi | Y \leq y, W \geq w, Z \geq z), & \text{if } \frac{d\pi(y,w,z)}{dz} \leq 0; \\ \mathcal{P}(\Pi \leq \pi | Y \leq y, W \geq w, Z \leq z), & \text{if } \frac{d\pi(y,w,z)}{dz} \geq 0. \end{cases}$$

The analysis and estimation procedures defined in the following subsection can easily be extended to these alternative settings with minor modifications.

1.2.2 Estimation

In order to estimate $\pi_\alpha(p, w)$, we first need an estimator for a conditional cumulative distribution function $F(\pi|C_{p,w})$. In a production function setting, Aragon et al. (2005) propose a simple estimator based on the empirical distribution function. Their empirical estimator is not smooth and as a result, it might be difficult to identify differences between firms that are similar in terms of profit efficiency. In the same setting Martins-Filho and Yao (2008) proposed a smooth kernel based estimator. The smoothness it provides might reduce the finite sample variance compared to the empirical estimator, but introduces a bias that does not vanish at the parametric rate. Here, we follow Martins-Filho and Yao (2008), but provide an alternative kernel that can produce biases of lower order. For convenience, we define the functions $P(\pi, p, w) = \mathcal{P}(\Pi \leq \pi, P \leq p, W \geq w)$ and $P_{PW}(p, w) = \mathcal{P}(P \leq p, W \geq w)$. We estimate $F(\pi|C_{p,w})$ by integrating a smooth kernel density estimator constructed using the observations $\{(\Pi_i, P_i, W_i)\}_{i \in \{i: P_i \leq p, W_i \geq w\}}$. Thus, we define

$$\hat{F}(\pi|C_{p,w}) = \begin{cases} 0 & \text{if } \pi \leq 0; \\ \frac{\hat{P}(\pi,p,w)}{\hat{P}_{PW}(p,w)} & \text{if } \pi > 0. \end{cases} \tag{1.4}$$

with

$$\hat{P}(\pi, p, w) = (nh_n)^{-1} \sum_{i=1}^{n} \left(\int_0^\pi M_k \left(\frac{\Pi_i - \gamma}{h_n} \right) d\gamma \right) I(P_i \leq p, W_i \geq w), \tag{1.5}$$

and

$$\hat{P}_{PW}(p,w) = n^{-1}\sum_{i=1}^{n} I(P_i \leq p, W_i \geq w). \tag{1.6}$$

h_n is a nonstochastic sequence of bandwidths such that $0 < h_n \to 0$ as $n \to \infty$, $I(A)$ is the indicator function for the set A and M_k for $k = 1, 2, \cdots$ is a class of kernels defined by Mynbaev and Martins-Filho (2010). The kernels M_k are defined as

$$M_k(x) = -\frac{1}{c_{k,0}}\sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K(\frac{x}{s}), \tag{1.7}$$

where $c_{k,s} = (-1)^{s+k} C_{2k}^{s+k}$, C_{2k}^{s+k} are the binomial coefficients and $K(\cdot)$ is a traditional (seed) kernel function, i.e., $K(\cdot)$ is a symmetric function such that $\int K(u)du = 1$. Lemma 1 shows that $M_k(x)$ is a kernel function for all k in that $\int M_k(x)dx = 1$. The main advantage of the definition of $M_k(x)$ is that it allows us to express the bias of our estimator in terms of higher order finite differences of the density function (see the proof in Lemma 2). It is clear that $\hat{F}(\pi|C_{p,w})$ depends on k through the dependence of $\hat{P}(\pi, p, w)$ on k. As a result, we are defining a class of estimators for $F(\pi|C_{p,w})$. The choice of k depends on the smoothness assumption on the distribution function, and will be discussed in the next section. Also, note that our estimator uses a smooth nonparametric estimator of the distribution function in the direction of profit π, but still uses an empirical distribution function in the direction of p and w. In the context of a production function, Martins-Filho and Yao (2008) showed that smooth kernel based estimator implemented in the output direction has a parametric ($\sqrt{n}$) rate of convergence. In the next section we will show that our estimator has the same convergence rate. Note that it is possible to smooth estimators in the directions of prices as well, but as a result the estimator would suffer from the well-known "curse of dimensionality."

Assuming that $\pi_\alpha(p,w)$ is the unique root of $F(\cdot|C_{p,w}) = \alpha$, we denote its estimator by $\pi_{\alpha,n}(p,w)$, the root of

$$\hat{F}(\pi_{\alpha,n}(p,w)|C_{p,w}) = \alpha \text{ for } \alpha \in (0,1],\ p \in \mathbb{R}_{++}^{d_1} \text{ and } w \in \mathbb{R}_{++}^{d_2}. \tag{1.8}$$

By the continuity of $F(\cdot|C_{p,w}))$, smoothness of the seed kernel, and the Mean Value Theorem, we write

$$\pi_{\alpha,n}(p,w) - \pi_\alpha(p,w) = \frac{F(\pi_\alpha(p,w)|C_{p,w}) - \hat{F}(\pi_\alpha(p,w)|C_{p,w})}{\hat{f}(\bar{\pi}_{\alpha,n}(p,w)|C_{p,w})},$$

where

$$\hat{f}(\pi|C_{p,w}) = \frac{\partial \hat{F}(\pi|C_{p,w})}{\partial \pi} = \begin{cases} 0, & \text{if } \pi = 0 \\ \frac{(nh_n)^{-1}\sum_{i=1}^n M_k(\frac{\Pi_i - \pi}{h_n}) I(P_i \le p, W_i \ge w)}{n^{-1}\sum_{i=1}^n I(P_i \le p, W_i \ge w)}, & \text{if } \pi > 0 \end{cases}$$

and $\bar{\pi}_{\alpha,n}(p,w) = \lambda \pi_{\alpha,n}(p,w) + (1-\lambda)\pi_\alpha(p,w)$ for some $\lambda \in (0,1)$. In the following section we provide some asymptotic characterizations for our estimator, including consistency and asymptotic normality.

1.3 Asymptotic Characterization of $\pi_{\alpha,n}$

In this section we provide theorems establishing asymptotic properties of our estimators. All proofs of the theorems and required lemmas can be found in Appendix. We begin by listing and discussing assumptions that are sufficient to establish our main theorems.

1.3.1 Assumptions

Assumption 1. *$\{(\Pi_i, P_i, W_i)\}_{i=1}^n$ is a sequence of independent random vectors taking values in a compact set $\Psi^* = [0, B_\pi] \times S_{PW}$ where S_{PW} is a compact set in $\mathbb{R}_{++}^{d_1} \times \mathbb{R}_{++}^{d_2}$. For any i, (Π_i, P_i, W_i) has the same joint distribution F and joint density function f as the vector (Π, P, W), f is defined on $\mathbb{R} \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ with support Ψ^*.*

Assumption 2. *(i) The seed kernel $K(\cdot)$ is a bounded symmetric density with compact support $[-B_K, B_K]$ and $\int_{-B_K}^{B_K} \gamma K(\gamma) d\gamma = 0$. (ii) $\int_{-B_K}^{B_K} \gamma^2 K(\gamma) d\gamma = \sigma_K^2$. (iii) For any $\gamma, \gamma' \in [-B_K, B_K]$, we have $|K(\gamma) - K(\gamma')| \le m_K |\gamma - \gamma'|$ for some $0 < m_K < \infty$. (iv) For all $\zeta, \zeta' \in [-B_K, \infty)$, we have $|\kappa(\zeta) - \kappa(\zeta')| \le m_\kappa |\zeta - \zeta'|$ for some $0 < m_\kappa < \infty$, where $\kappa(\zeta) = \int_{-B_K}^{\zeta} K(\gamma) d\gamma$. (v) For fixed k, $\int |K(t)| t^{2k} dt < \infty$.*

The first assumption is standard. Assumption 2 is similar to Martins-Filho and Yao (2008) except (v). We need Assumption 2 (v) in the proof of Lemma 2 for the purpose of bias restriction (see the similar assumption in Mynbaev and Martins-Filho (2010)). Note that Assumption 2 imposes some smoothness condition on the kernel, since a function satisfies Lipschitz condition if it has bounded first derivative. We can prove that for any $k \in \mathbb{N}$, if the seed kernel K satisfies Assumption 2, then it also holds for M_k by the definition (1.7).[3]

Assumption 3. *For all π and $\pi' \in G$, where G is a compact set, we have*

$$\left| \int_{\pi^{-1}([\pi,\pi'])} d(P,W) \right| \leq m_{\pi^{-1}} |\pi' - \pi|$$

for some $0 < m_{\pi^{-1}} < \infty$. Here, for any two sets $A \subseteq D_{p,w} := [0,p] \times [w,\infty)$ and $B \subseteq [0,\pi(p,w)]$, Define $\pi(A) = \{\pi(p,w) : (p,w) \in A)\}$ and $\pi^{-1}(B) = \{(p,w) \in D_{p,w} : \pi(p,w) \in B\}$.

Assumption 3 is similar to Assumption 4 in Martins-Filho and Yao (2008). It imposes a Lipschitz type condition on the inverse image π^{-1} of π.

Assumption 4. *(i) The joint density function f is continuous on Ψ^*, $0 < f(\pi,p,w) < B_f$ for all $(\pi,p,w) \in \Psi^*$. (ii) For all (π,p,w) and $(\pi',p,w) \in \Psi^*$, we have $|f(\pi',p,w) - f(\pi,p,w)| \leq m_f |\pi' - \pi|$ for some $0 < m_f < \infty$. (iii) For all (p,w) such that $\mathcal{P}(P \leq p, W \geq w) > 0$ and for all $\alpha \in (0,1]$, $f(\pi_\alpha(p,w)|C_{p,w}) > 0$, where $f(\cdot|C_{p,w})$ is the derivative of $F(\cdot|C_{p,w})$.*

Assumption 5. *Given p, w, for all $\pi \in (0, B_\pi)$,*

5A *Fix k, there exist functions $H_{2k}(\pi,p,w) > 0$ and $\varepsilon_{2k}(\pi,p,w) > 0$ such that*

$$|\Delta_h^{2k} F_f(\pi,p,w)| \leq H_{2k}(\pi,p,w) h^{2k}$$

for all $|h| \leq \varepsilon_{2k}(\pi,p,w)$. Here, $F_f(\pi,p,w) = \int_0^\pi f(\gamma,p,w)d\gamma$ and

$$\Delta_h^{2k} F_f(\pi,p,w) = \sum_{s=-k}^{k} c_{k,s} F_f(\pi + sh, p, w)$$

with $c_{k,s} = (-1)^{s+k} C_{2k}^{s+k}$.

5B *f is continuously differentiable with respect to π. $|f^{(1)}(\pi,p,w)| < \infty$, where $f^{(1)}(\pi,p,w)$ represent the first order derivative of f with respect to π.*

[3] See Lemma 1 in the Appendix.

Assumption 5A imposes an order $2k$ Lipschitz condition on $F_f(\pi, p, w)$ with respect to π. From the proof of Theorem 1 in Mynbaev and Martins-Filho (2010) we know that boundedness of $F_f^{(2k)}(\pi, p, w)$ implies a Lipshitz condition of order $2k$. As a result, Assumption 5B is a more strict condition than 5A in the special case $k = 1$. Given Assumption 5A, we can obtain the order of the bias for our estimator to be h^{2k}. Given Assumption 5B, we can obtain a specific structure for the asymptotic bias and variance by using a Taylor expansion.

1.3.2 Asymptotic Properties

We start by providing a proposition showing some asymptotic properties of $\hat{F}(\pi|C_{p,w})$ with Assumption 2.

Proposition 3. *Under Assumption 2, we have: (i) $\hat{F}(\pi|C_{p,w})$ is continuous; (ii) $\lim_{\pi\to 0} \hat{F}(\pi|C_{p,w}) = 0$; (iii) For any (p, w), there exists some $N(p, w)$ such that for all $n > N(p, w)$, $\lim_{\pi\to\infty} \hat{F}(\pi|C_{p,w}) = 1$.*

Note that since $M_k(.)$ is not necessarily positive, $\hat{F}(\pi|C_{p,w})$ is not necessarily monotonic. Except for that, Proposition 3 states that $\hat{F}(\pi|C_{p,w})$ has properties associated with a proper distribution function. The next main theorem establishes consistency of $\pi_{\alpha,n}$.

Theorem 1. *Let h_n be a nonstochastic sequence of bandwidths such that $0 < h_n \to 0$ as $n \to \infty$. Given $w \in \mathbb{R}^{d_2}_{++}$, $p \in \mathbb{R}^{d_1}_{++}$, suppose there exist $N(p, w)$, such that when $n > N(p, w)$ we have $\mathcal{P}\{\Pi < h_n B_M\} = 0$. Under Assumptions 1-4 along with Assumption 5A (or 5B), if $H_{2k}(\pi, p, w)$, $F_f(\pi, p, w)$ and $\varepsilon_{2k}(\pi, p, w)$ are bounded for all $(\pi, p, w) \in \Psi^*$, we have*

$$\pi_{\alpha,n}(p, w) - \pi_\alpha(p, w) = o_p(1). \tag{1.9}$$

The next main theorem shows that under suitable normalization and centering $\pi_{\alpha,n}(p, w)$ is asymptotically distributed as a standard normal.

Theorem 2. *Let h_n be a nonstochastic sequence of bandwidths such that $nh_n^2 \to \infty$ and $nh_n^4 = O(1)$ as $n \to \infty$. Given $w \in \mathbb{R}^{d_2}_{++}$, $p \in \mathbb{R}^{d_1}_{++}$, suppose there exist $N(p, w)$ such that when $n > N(p, w)$ we* *have* $\mathcal{P}\{\Pi < h_n B_M\} = 0$. *Then,*

(i) Under Assumption 1-4 and Assumption 5B, we have

$$v_n(p,w)^{-1}\sqrt{n}(\pi_{\alpha,n}(p,w) - \pi_\alpha(p,w) - B_n(p,w)) \xrightarrow{d} N(0,1)$$

where

$$B_n(p,w) = -\frac{1}{2}h_n^2\sigma_M^2 \frac{\int_{\pi^{-1}([\pi_\alpha(p,w),\pi(p,w)])} f^{(1)}(\pi_\alpha(p,w),P,W)d(P,W)}{P_{PW}(p,w)f(\pi_\alpha(p,w)|C_{p,w})} + o(h_n^2),$$

$$\begin{aligned} v_n^2(p,w) &= \frac{1}{(P_{PW}(p,w)f(\pi_\alpha(p,w)|C_{p,w}))^2}(F(\pi_\alpha(p,w),p,w) - \frac{F^2(\pi_\alpha(p,w),p,w)}{P_{PW}(p,w)} \\ &\quad -2h_n\sigma_\kappa \int_{\pi^{-1}([\pi_\alpha(p,w),\pi(p,w)])} f(\pi_\alpha(p,w),P,W)d(P,W)) + o(h_n), \end{aligned}$$

with $\sigma_\kappa = \int \gamma\kappa_M(\gamma)M_k(\gamma)d\gamma$, *and* $f^{(1)}(\pi,P,W)$ *denotes the first derivative of* f *with respect to* π.

(ii) Under Assumption 1-4 and Assumption 5A, we have

$$\begin{aligned} |B_n(p,w)| &\leq ch_n^{2k}[\int_{D_{p,w}} H_{2k}(\pi_\alpha(p,w),P,W)d(P,W) \\ &\quad + \int_{D_{p,w}} \sup_{\pi\in\mathbb{R}} |F_f(\pi,P,W)|\varepsilon_{2k}^{-2k}(\pi_\alpha(p,w),P,W)d(P,W)], \end{aligned}$$

where c *represent an arbitrary nonnegative constant.*

Part (i) of Theorem 2 shows the explicit structure for bias and variance when $k = 1$. Part (ii) shows that the bias decays to zero faster when we impose a stronger Lipschitz smoothness condition on the distribution function and increase the value of parameter k accordingly. From Theorem 2, we first observe that our estimator is $\sqrt{n}$ asymptotically normal although it is based on kernel smoothing. That is, the convergence speed of our estimator is independent of the dimensionality of the problem. Therefore, our estimator does not suffer the "curse of dimensionality". Second, note that the extra smoothness of our estimator provides a smaller variance compared to the empirical estimator at the cost of introducing a bias which vanishes asymptotically (see Aragon et al. (2005)). Finally, the order of the bias term is controlled by the smoothness assumptions on the density function. Note that under appropriate assumptions, the bias term is smaller than the order h_n^{2k}. Hence we can reduce the bias by increase the parameter k. For our estimator, the "smoother" the density function is, the faster the bias term would vanish.

1.4 Monte Carlo Study

1.4.1 Setup and Implementation

In this section, we design and conduct a small Monte-Carlo simulation to implement our estimator and investigate some of its finite sample properties. We also compare the performance of our estimator to that of a similar estimator based on the empirical distribution. The data generating process is given by

$$\begin{aligned} \Pi_i &= \pi(P_i, W_i) R_i \quad i = 1, .., n \\ R_i &= \exp(-Z_i), \quad Z_i \sim Exp(\beta) \end{aligned}$$

where Π_i represents profit, P_i and W_i represent output and input prices. In this simulation, we assume output is a scalar. $R_i = \exp(-Z_i)$ represents efficiency score for each unit i. Z_i are independently generated from an exponential distribution with parameter $\beta = 1/3$. As a result the density function of R_i is $f(r) = 3r^2$ with support $(0, 1]$ and a mean 0.75. $\pi(p, w)$ is the profit function. In this simulation we consider two profit functions $\pi(p, w) = p^{6/5} w^{-6/5}$ and $\pi(p, w_1, w_2) = \frac{1}{4} p^2 (w_1^{-1} + w_2^{-1})$. The first one considers only a single input; and the second one considers two inputs with their prices represented as w_1 and w_2. One can easily verify these functions satisfy all properties of a profit function: a) nondecreasing in p and nonincreasing in w; b) convex in both p and w; c) homogenous of degree one, and d) continuous. Prices are uniformly drawn from a meshgrid $[p_l, p_u] \times [w_l, w_u] = [1, 3] \times [1, 3]$ for the first profit function; and from a meshgrid $[p_l, p_u] \times [w_{1l}, w_{1u}] \times [w_{2l}, w_{2u}] = [1, 3] \times [1, 3] \times [1, 3]$ for the second profit function. Several experimental designs are considered: We estimate profit frontiers of order $\alpha = 0.25, 0.5, 0.75$ and 0.99 using M_k kernel functions with $k = 1, 2$ as well as an empirical distribution. In each experiment, We consider two sample sizes $n = 200$ and $n = 400$ and perform 2000 iterations to obtain the averaged absolute value of bias and root mean squared error of each estimator.

The empirical profit frontier of order α is estimated as follows: let $N_{p,w} = \sum_{i=1}^{n} I(P_i \leq p, W_i \geq w)$. For $j = 1, ..., N_{p,w}$, get the order statistic of the observation $\Pi_{(i_j)}$ such that $\Pi_{(i_1)} \leq$

$\Pi_{(i_2)} \leq \ldots \leq \Pi_{(i_{N_{p,w}})}$. The empirical conditional distribution $\hat{F}_e(\pi|C_{p,w})$ is

$$\begin{aligned}\hat{F}_e(\pi|C_{p,w}) &= \frac{\sum_{j=1}^{N_{p,w}} I(\Pi_{(i_j)} \leq \pi)}{N_{p,w}} \\ &= \begin{cases} 0 & \text{if } \pi < \Pi_{(i_1)}; \\ m/N_{p,w} & \text{if } \Pi_{(i_m)} \leq \pi < \Pi_{(i_{m+1})},\ 1 \leq m \leq N_{p,w}-1; \\ 1 & \text{if } \pi \geq \Pi_{(i_{N_{p,w}})}. \end{cases}\end{aligned}$$

Thus the empirical estimator for the conditional quantile $\pi_\alpha(p,w)$ can be computed as follows

$$\hat{\pi}_{e,\alpha}(p,w) = \begin{cases} \Pi_{(i_{\{\alpha N_{p,w}\}})} & \text{if } \alpha N_{p,w} \in \mathbb{N}; \\ \Pi_{(i_{\{[\alpha N_{p,w}]+1\}})} & \text{otherwise,} \end{cases}$$

where $[\alpha N_{p,w}]$ denotes the integer part of $\alpha N_{p,w}$.

The implementation of our estimator requires choices of kernel function as well as bandwidth. We use the Epanechnikov function $K(x) = \frac{3}{4}(1-x^2)I(|x| \leq 1)$ as the seed kernel. It is easy to show this kernel function satisfies Assumption 2. The bandwidth is chosen by minimizing the asymptotic approximation of our estimator's mean integrated squared error (AMISE). For $k = 1$, we get the global optimal bandwidth with respect to α as

$$h_n^* = \left(\frac{2\sigma_\kappa \int_0^1 \frac{I_2(p,w,\alpha)}{f^2(\pi_\alpha(p,w)|p,w)} d\alpha}{(\sigma_K^2)^2 \int_0^1 \frac{I_1^2(p,w,\alpha)}{f^2(\pi_\alpha(p,w)|p,w)} d\alpha} \right)^{1/3} n^{-1/3},$$

where

$$\begin{aligned} I_1(p,w,\alpha) &= \int_{\pi^{-1}([\pi_\alpha(p,w),\pi(p,w)])} f^{(1)}(\pi_\alpha(p,w),P,W)d(P,W), \text{ and} \\ I_2(p,w,\alpha) &= \int_{\pi^{-1}([\pi_\alpha(p,w),\pi(p,w)])} f(\pi_\alpha(p,w),P,W)d(P,W). \end{aligned}$$

In our simulations, since we know the true distribution, we can compute h_n^* directly. In practice, use of h_n^* requires the estimation of the unknown distribution. Applying a similar method described in Mynbaev and Martins-Filho (2010), we can estimate I_1, I_2 and f using a suitably defined Rosenblatt density estimator. The optimal bandwidths for the estimators with higher k are yet to be obtained. We use the same bandwidth as $k = 1$, when $k > 1$.

1.4.2 Results and Analysis

Table 1.1 gives the bias and root mean square error of our smoothed estimator with order of kernel $k = 1$ and $k = 2$ compared with the empirical estimator evaluated at prices $(p, w) = (2, 2)$ for the first profit function $\pi(p, w) = p^{6/5}w^{-6/5}$; as well as those at prices $(p, w_1, w_2) = (2, 2, 2)$ for the second profit function $\pi(p, w_1, w_2) = \frac{1}{4}p^2(w_1^{-1} + w_2^{-1})$. The simulations seem to confirm our

Table 1.1: Bias and RMSE under Each Experiment Design

$p^{6/5}w^{-6/5}$	|Bias|			RMSE		
n=200 α	Kernel k=1	Kernel k=2	Empirical	Kernel k=1	Kernel k=2	Empirical
0.25	.018	.019	.021	.024	.024	.027
0.50	.020	.021	.024	.033	.033	.037
0.75	.027	.027	.030	.031	.032	.037
0.99	.132	.261	.084	.175	.358	.095
n=400 α	Kernel k=1	Kernel k=2	Empirical	Kernel k=1	Kernel k=2	Empirical
0.25	.014	.013	.015	.017	.016	.019
0.50	.015	.012	.017	.018	.016	.019
0.75	.019	.016	.021	.023	.021	.028
0.99	.083	.098	.057	.102	.121	.068
$\frac{1}{4}p^2(w_1^{-1} + w_2^{-1})$	|Bias|			RMSE		
n=200 α	Kernel k=1	Kernel k=2	Empirical	Kernel k=1	Kernel k=2	Empirical
0.25	.023	.024	.027	.031	.032	.035
0.50	.035	.036	.039	.046	.046	.051
0.75	.026	.027	.031	.042	.040	.047
0.99	.169	.334	.107	.224	.460	.137
n=400 α	Kernel k=1	Kernel k=2	Empirical	Kernel k=1	Kernel k=2	Empirical
0.25	.018	.017	.020	.021	.021	.025
0.50	.025	.021	.022	.029	.027	.031
0.75	.023	.019	.027	.028	.022	.034
0.99	.108	.125	.074	.132	.154	.089

asymptotic results. In particular, the root mean squared error of all estimators decreases with the sample size. For both profit functions, the kernel estimator outperforms the empirical estimator in the cases with $\alpha = 0.25, 0.5$ and 0.75. Although we do not use the optimal bandwidth, the performance of the estimator with kernel order $k = 2$ is quite good. When the sample size is 200, the performance of estimators with $k = 1$ and $k = 2$ are very close. When the sample size grows from 200 to 400 we observe a larger improvement for the estimator with $k = 2$. For example, with $\alpha = 0.5$, the bias of the estimator with $k = 2$ decreases from .021 to .012, while the bias of the estimator with $k = 1$ just decreases from .020 to .015. We find the similar results for all α. This is consistent with the result in Theorem 2 which states the bias decays faster as k increases.

We also observe that as α increases, all estimators show larger bias and mean square error. This can be interpreted as resulting from the fact that there are less effective data available as α grows. As a result, when α is close to 1, profit functions of order α become more difficult to estimate. Note that the performance of our smoothed estimator is especially poor when $\alpha = 0.99$. This is most likely due to the fact that our distribution function has compact support, and it is not smooth near the boundary. Therefore, the smoothed estimator can generate large biases.

In summary, our simulation results indicate the proposed smooth estimator for the profit function of order α can outperform the empirical estimator in most cases as long as α is not very close to 1. Additionally, increases in the order k of the M_k kernel may increase the convergence speed of the bias. However, we do not suggest to use our method in approximating the full frontier where α is approaching to 1. Note that the full frontier is not required in estimating the efficiency in our method. According to the analysis in Section 1.2, any α frontier with $\alpha \in (0, 1)$ can be served as a standard in the efficiency analysis.

1.5 Conclusion and Discussion

In this paper we consider the construction and estimation of a profit function of continuous order $\alpha \in (0, 1]$. We define a class of such profit functions based on conditional quantiles of an appropriate distribution of profit, input and output prices. We show that they are useful in measuring and assessing profit efficiency. We show that our estimator is consistent and asymptotically normal with a parametric convergence speed of $\sqrt{n}$. Furthermore, the bias of our estimator decays to zero faster than the traditional kernel estimators. A Monte-Carlo simulation is performed to implement our estimator, investigate its finite sample performance and compare it to an estimator based on empirical distribution function. Simulation results seem to confirm the asymptotic results we have obtained and also seems to indicate that our proposed estimator can outperform its competitors in most cases. However, our estimator seems to possess large boundary bias. Finally, we conduct an efficiency analysis of the Swedish paper industry based on the α frontier estimation method developed in this paper. We find the industry is very profit inefficient compared to high average production efficiency score, and firms of different sizes may have different patterns in earning profit. Future research includes: (1) decrease the possible boundary bias for our estimator; (2) investigate asymptotic normality and the choice of optimal bandwidth for $k > 1$; (3) decomposition of technique efficiency and allocative efficiency for the profit efficiency estimator.

Chapter 2

Profit Efficiency in the Swedish Paper Industry: An Application

2.1 Introduction

Frontier estimation techniques for efficiency analysis are powerful tools of measuring firms' performances in many empirical studies. First, they allow individuals, such as firm managers with very little knowledge in economics to rank production units by assigning intuitive numerical scores representing their performances, and relate these results to their interests. Second, they help economists identify the sources of inefficiency and develop policies to improve the efficiency of individual firms and the entire industry. For example, if profit inefficiency is much larger than the production inefficiency, it suggests that there are profit improvements to be made among inefficient firms through better management of input and output, without actually investing in any technology improvement; it is simply about allocating resources more efficiently.

As mentioned before, both parametric and nonparametric methods have been widely employed in estimating production frontiers and efficiency scores. In contrast, despite the agreement that profitability is an important measure of performance, there has been very few empirical studies in the profit aspect due to the lack of estimation techniques as well as a generally accepted measure of profit efficiency (compared to the well-known concept of production or technical efficiency). So far, limited efforts in the profit efficiency analysis have been firstly and mostly conducted in the banking industry (for instance, Berger and Humphrey (1993); Berger and Mester (1997); Akhavein et al. (1997); Maudos et al. (2002) and Akhigbe and McNulty (2005)). Herr et al. (2010) is among very few empirical studies in other industries (hospitals). Most of these studies construct a profit frontier based on a parametric model, and measure the profit efficiency as the ratio of the actual profit of a bank and the potential or maximum level that could be obtained by the most efficient bank. Not surprisingly, these studies rely heavily on a number of strong assumptions, especially the parametric structure of the profit function, and are possibly exposed to the risk of mis-specification.

For nonparametric studies, Färe et al. (2004) adopt data envelopment analysis (DEA) and use the directional distance functions to compute profit inefficiency index for the USA banking sector. In their study, profit inefficiency is decomposed into technical and allocative inefficiency. They find that the allocative inefficiency is the major determinant of profit inefficiency for the USA banks. Maudos and Pastor (2003) also study an alternative profit efficiency of the Spanish banking based on DEA, and compared the profit efficiency to the cost efficiency. Recently, Färe et al. (2015) and Ruiz and Sirvent (2011) develop an alternative slack-based DEA method for the decomposition of profit efficiency. DEA is commonly criticized for its sensitivity to outliers, we hope the α-frontier estimator proposed in this paper could provide a good alternative for future profit efficiency analysis.

In this section, we try to apply the estimation method we developed in the previous sections to analyze performance of firms in the Swedish paper industry. Firms in the paper industry are probably different from the commonly studied banks in that: first, firm's ability of setting/influencing prices might be more restricted in this industry. Second, products (pulp and paper) are more homogenous. In other words, there is less output diversification. Therefore, we could expect a more competitive environment compared to the banking industry. The main objectives of this section are: (1) To demonstrate how to use the nonparametric method for estimating α-frontiers and the corresponding efficiency scores; (2) Based on efficiency scores, we try to discuss the source of profit inefficiency and investigate the relationship between production efficiency and profit efficiency.

2.2 Data and Estimation

Data and variables

The data set is a firm level panel taken from Sweden paper industry (pulp and paper sector) covering the years 1990 to 2008. 210 firms are included, and a few firms earning negative profit are excluded based on our positive profit assumption. Descriptive statistics for each of these variables are provided in the following table. Output y is constructed by sales divided by a sector level producer price index. We consider capital and labor as the two inputs. Capital k is derived from investment data and labor l is measured by the number of employees. r represents user cost of capital and w represents wage rate (salary/employees). Output price p is measured by the producer price index. Unit for monetary variables is thousands of SEK (Swedish Krona). Since the number of firms including in each year is relatively small (average of 60-70 firms), it would probably not be appropriate to construct a specified frontier for each year considering the sample size. To our knowledge, there is no significant structural change or technological advancement in the industry. Therefore we assume that the profit and the production frontier in this sector had not changed during the period. As a result, we treat each individual firm in each year as a single decision making unit (DMU), and assume at the moment that all DMU share a single production or profit frontier. The total number of valid observations is 1116.

Table 2.1: Descriptive Statistics of Variables

Variable	Mean	Median	Minimum	Maximum	Standard Deviation
Output y	947654994	254459957	5625046	14620107826	1796026213
Input 1 (Labor) l	437.021	147	4	9739	838.901
Input 2 (Capital) k	691164409	105882956	311140	8203495467	1337884052
Output Price p	117.901	120.9	89.9	138.6	14.396
Labor Price w	375286	364975	756.923	766010	106717
Capital Price r	0.144	0.138	0.121	0.186	0.019
Profit π	102145378	5690962	2	5724815000	370814498

Estimation

We denote an α profit frontier and an α production frontier as $\pi = \pi_\alpha(p, w, r)$ and $y = \rho_\alpha(l, k)$ respectively. Following the similar procedure developed in Section 2, the corresponding estimators for them are defined by $\pi_{\alpha,n}(p,w)$ and $\rho_{\alpha,n}(l,k)$, the root of

$$\begin{aligned}\hat{F}_\pi(\pi_{\alpha,n}(p,w,r)|C_{p,w,r}) &= \alpha, \quad \text{and} \\ \hat{F}_\rho(\rho_{\alpha,n}(l,k)|C_{l,k}) &= \alpha.\end{aligned}$$

where

$$\hat{F}_\pi(\pi|C_{p,w,r}) = \begin{cases} 0 & \text{if } \pi \le 0; \\ \frac{(nh_{\pi,n})^{-1}\sum_{i=1}^n \left(\int_0^\pi M_k\left(\frac{\Pi_i-\gamma}{h_{\pi,n}}\right)d\gamma\right) I(P_i\le p, W_i \ge w, R_i \ge r)}{n^{-1}\sum_{i=1}^n I(P_i \le p, W_i \ge w, R_i \ge r)} & \text{if } \pi > 0, \end{cases}$$

and

$$\hat{F}_\rho(\rho|C_{l,k}) = \begin{cases} 0 & \text{if } \rho \le 0; \\ \frac{(nh_{\rho,n})^{-1}\sum_{i=1}^n \left(\int_0^\rho M_k\left(\frac{Y_i-\gamma}{h_{\rho,n}}\right)d\gamma\right) I(L_i\le l, K_i \le k)}{n^{-1}\sum_{i=1}^n I(L_i \le l, K_i \le k)} & \text{if } \rho > 0. \end{cases}$$

Finally, the measurement of profit and production efficiency for DMU i are estimated respectively as $\hat{e}_{\pi,\alpha}(\Pi_i, P_i, W_i, R_i) = \Pi_i/\pi_{\alpha,n}(P_i, W_i, R_i)$ and $\hat{e}_{\rho,\alpha}(Y_i, L_i, K_i) = Y_i/\rho_{\alpha,n}(L_i, K_i)$.

Above nonparametric estimators for distribution functions are implemented using the Epanechnikov kernel. Bandwidths $h_{\rho,n}$ and $h_{\pi,n}$ are selected following by the similar plug-in method described in Martins-Filho and Yao (2008). We choose parameter $k = 1$ to keep things simple. To make profit and production efficiency to be comparable, α is chosen to be the same for both profit and production frontier estimators.

The remaining problem is the choice of parameter α. Theoretically speaking, a frontier with any order $\alpha \in (0, 1)$ can be used as a standard for estimating efficiency scores. Here we provide a data-driven selection method for α in practice. The basic motivation is that we want to choose an α to make the specified quantile function to be close to the true frontier; while remaining as robust as possible. Consider two cases: first, a small increase in α has little impact on the frontier estimator; second, a small increase in α leads to a significant change on the frontier estimator, making it envelope much more observations and consequently much closer to extreme values. Clearly, in the second case the estimator would be more affected by extreme values.

Figure 2.1 shows the relationship between α and the percentage of observations above the estimators of α frontiers. In the figure, we observe that for production frontiers, the percentage of 'over-efficient' observations decreases very slowly until $\alpha = 0.95$. It suggests that α frontiers of order $0 < \alpha < 0.95$ are very tightly distributed. That is, frontier estimators in this region are very close. The percentage of 'over-efficient' observations falls dramatically starting from $\alpha = 0.95$, which suggests that α frontiers of order $0.95 < \alpha < 1$ are very spaced and are spread out among 65% of observations. In this region, a small change in the α would lead to a large jump of the frontier estimator, and probably a big change in the shape of estimators. As a result, we tend to avoid picking an α in the interval $(0.95, 1)$ where frontier estimators are not very robust. For profit frontiers, we observe a very different pattern. There is no significant change for the percentage of 'over-efficient' observations for all $\alpha \in (0, 1)$. That is, estimators for α profit frontiers are basically evenly distributed for $\alpha \in (0, 1)$. From Figure 2.1, we have no strong preference in choosing α for profit frontier estimators. However, since we want to choose the same α for both profit and production frontier estimators for the purpose of comparison, we set $\alpha = 0.95$ based on the above discussion.

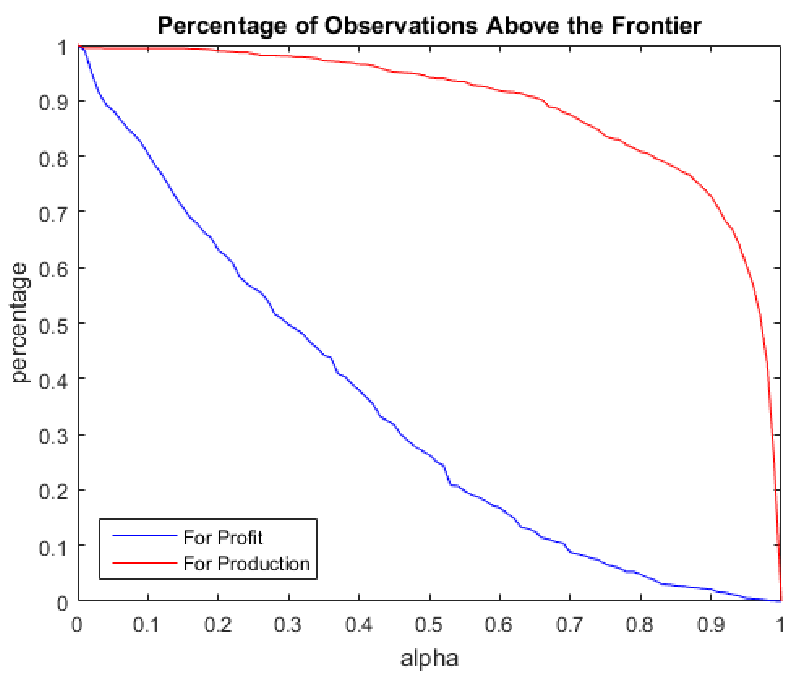

Figure 2.1: Percentage of Observations Above Frontiers

2.3 Empirical Results

Overall efficiency results

In this section, we obtain profit and production efficiency scores for each decision making unit based on the estimation method described above. We try to use these efficiency scores to explore the source of profit inefficiency. However, in the absence of a reliable theoretical model explaining the source of efficiency in this industry, we just analyze potential correlates of efficiency rather than explanatory factors. The major focus here is the relationship between production efficiency and profit efficiency.

Table 2.2 reports the average profit efficiency (PE) and production efficiency (TE) for all firms in the industry for each year. Average production efficiency is greater than 1 in each year, while average profit efficiency is less than 0.1 in each year, indicating that the industry is technically efficient, but significantly profit inefficient. Clearly, the main source of profit inefficiency in this industry is on the allocation side.

Table 2.2: Industry Average Profit and Production Efficiency

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
PE	0.099	0.043	0.076	0.078	0.070	0.152	0.075	0.088	0.093	0.039
	(0.253)	(0.145)	(0.219)	(0.196)	(0.161)	(0.375)	(0.196)	(0.205)	(0.217)	(0.140)
TE	1.314	1.364	1.340	1.267	1.408	1.222	1.101	1.424	1.432	1.242
	(0.933)	(1.697)	(1.434)	(0.693)	(1.188)	(1.049)	(0.632)	(0.816)	(0.832)	(0.814)
	2000	2001	2002	2003	2004	2005	2006	2007	2008	
PE	0.103	0.111	0.118	0.090	0.090	0.059	0.101	0.061	0.047	
	(0.323)	(0.373)	(0.298)	(0.207)	(0.237)	(0.165)	(0.330)	(0.161)	(0.156)	
TE	1.331	1.278	1.375	1.359	1.471	1.466	1.568	1.520	1.590	
	(0.952)	(0.816)	(0.738)	(0.781)	(0.744)	(0.785)	(0.918)	(0.871)	(0.981)	

The histogram in Figure 2.2 provides a clearer picture for the industry. For profit efficiency, we observe that over 90% of the observations lie inside the interval around 0.02, suggesting most firms can only earn less than 2% of potential profit compared to the 0.95 profit frontier. In contrast, for production efficiency, most firms have scores concentrated around 1. What is more, there are a number of observations with extremely high level of production efficiency at 3 to 5, explaining the high average production efficiency we observe in Table 3. The insignificant levels of profit efficiency and the significant levels of production efficiency suggest high levels of allocative inefficiency. That is, most firms in the industry might choose inadequately their input-output mix.

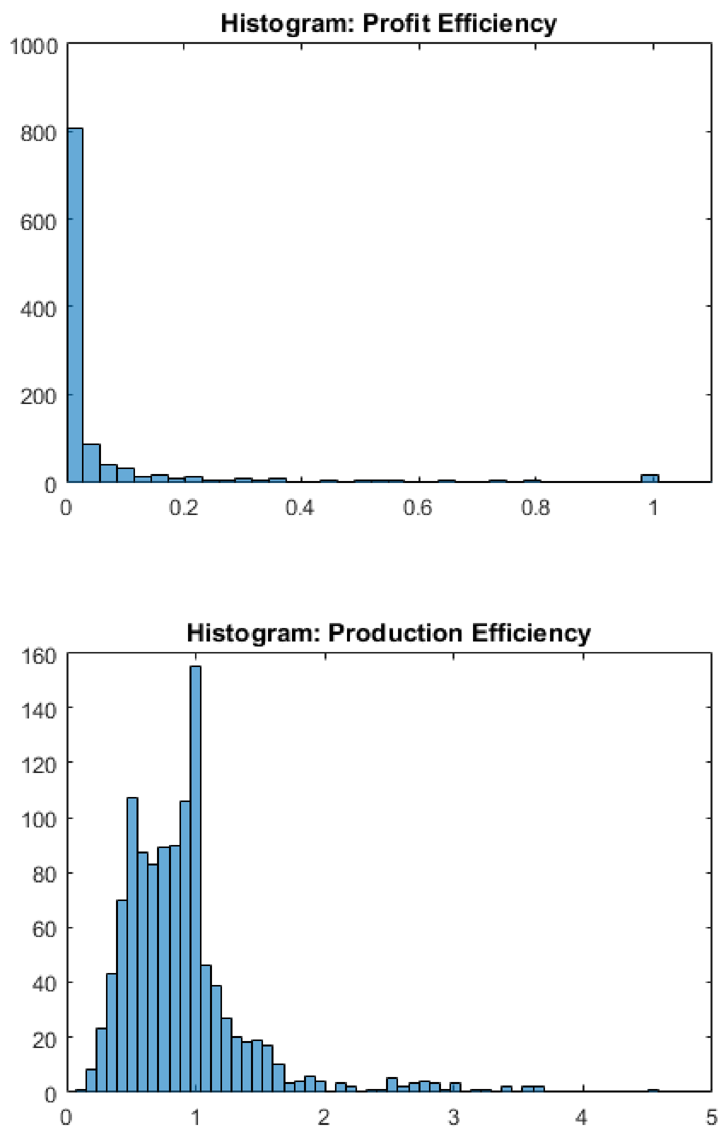

Figure 2.2: Histograms for Profit Efficiency and Production Efficiency

In order to compare performances among firms, we then calculate average efficiency scores across years for each individual firm, and rank them based on production and profit efficiency score respectively. We find the two rankings are not significantly correlated. In another word, firms with high rank of production efficiency may have low rank of profit efficiency; while firms with high rank of profit efficiency may have low rank of production efficiency. However, we do observe some patterns. For example, the lowest production efficiency rank for the top 10 profit efficient firms is 52; while the lowest profit efficiency rank for the top 10 production efficient firms is 96. Furthermore, 10 of the worst performing firms in production are all ranked in the last 15 in profit performance; while among 10 of the worst performing firms in profit, only 4 of them belong to the last 15 in production performance. Lastly, the top 25 firms in profit performance all have production efficiency score larger than 1. These observations suggest that profit efficient firms generally perform very well in the production aspect; production inefficient firms are generally profit inefficient. However, production efficient firms are not necessarily profit efficient; profit inefficient firms are not necessarily production inefficient.

Efficiency of firms with different sizes

The results above provide us a general idea about efficiency of the Swedish paper industry. We observe high technological efficiency with extreme low profit efficiency across firms. Naturally, we are concerned about the credibility of this result. Indeed, it is possible that many firms with low profit efficiency might be underestimated because they are compared to a frontier that is too high for them. That is, the assumption that all firms share a common frontier might be inappropriate. One possible reason is the impact of firm size.

Many papers studying profit efficiency treat firm size as an important factor that affects the variation in the profit efficiency across firms. Some may argue that high competitive pressures might induce more incentives for smaller banks to be efficient; others may support the statement that high profit efficiency usually is associated with larger banks because they have more market power and larger production capacities. Empirical results varies across papers and there is no consistent conclusion.

A common way of dealing size problem in previous studies is to divide the sample into several categories based on a measure of firm size, and then estimate efficiency scores separately for each category. The size of a firm can be measured in a number of ways: assets, sales, employees and value added are commonly used measures. Here we adopt 'value added' as our measure of firm size since it can capture more complexity of a firm compared to other measures (see Becker-Blease et al. (2010)). Based on value added, we separate the sample into three groups: we define small firms as those under 100 million in value added; medium-sized firms as those between 100 million and 1 billion, and large firms as those over 1 billion. We estimate different α frontiers for each sample, and calculate efficiency scores for firms in each group based on the group specified frontier respectively.

Table 2.3 reports our estimates of average profit and production efficiency in the industry by year and by size class. We make the following observations on the results. First, profit efficiency scores significantly increase for all three groups compared to the first model which ignores the size issue. The average profit efficiency across years increased from less than 0.1 in the first model to 0.37 for large firms; 0.17 for medium firms, and 0.27 for small firms. For most firms' profit efficiency now becomes more significant compared to the first model, the assumption of assigning different frontiers to different groups provides us more comparable efficiency indexes. Nevertheless, compared to the production efficiency, the industry is still very profit inefficient, suggesting most firms can probably improve their profit by adjusting their strategy of allocating inputs and output. This result is consistent with the first model. Finally, we observe that large firms possess highest average profit efficiency score, which is 10% higher than small firms and 20% higher than medium firms. It suggests that the link between profit efficiency and firm size in this industry may be more complicated than a simple linear relationship.

In order to analyze results in each group more clearly, Figure 2.3 shows plots of group average efficiency scores against years based on results in Table 2.3. We can see some different patterns for the time trends of production and profit efficiency among different groups. For large firms, we observe a sharp increase in the profit efficiency in the year 1992, which is the only year when the average profit efficiency exceeds the average production efficiency. This is probably because

Table 2.3: Average Efficiency for Different Groups

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
PE(L)	0.382	0.107	0.983	0.695	0.327	0.272	0.380	0.129	0.268	0.303
TE(L)	0.915	0.992	0.719	0.919	1.143	0.905	0.949	1.040	0.966	0.973
PE(M)	0.134	0.124	0.126	0.093	0.195	0.315	0.199	0.237	0.271	0.346
TE(M)	0.808	0.935	0.853	0.993	0.972	0.812	0.840	0.919	0.959	0.897
PE(S)	0.225	0.090	0.209	0.133	0.259	0.298	0.325	0.289	0.481	0.272
TE(S)	1.129	0.974	1.128	0.971	0.886	0.728	0.926	1.076	0.969	0.971
	2000	2001	2002	2003	2004	2005	2006	2007	2008	Mean
PE(L)	0.387	0.324	0.443	0.406	0.438	0.279	0.297	0.320	0.221	0.366
TE(L)	0.922	0.943	1.036	1.110	1.135	1.229	1.189	1.000	1.075	1.008
PE(M)	0.229	0.097	0.114	0.178	0.131	0.129	0.104	0.160	0.191	0.177
TE(M)	0.889	0.822	0.894	0.940	0.891	1.018	1.123	1.091	1.215	0.941
PE(S)	0.220	0.177	0.184	0.128	0.298	0.260	0.309	0.255	0.151	0.240
TE(S)	1.054	0.938	1.039	1.124	1.180	1.165	1.161	0.997	1.186	1.032

in 1992, the total number of large firms decreases suddenly from 5 to 1. The exit of competitors provides the only large firm in the industry greater market power, thus much more profit. In 1994, the number of large firms is back to 5, and the average profit efficiency fall back to normal. Another observation for large firms is that the trend of profit efficiency is basically not consistent with the trend of production efficiency. For example, in 1992 when the profit efficiency enjoyed a peak, the production efficiency actually decreases to the bottom at 0.7. The reason is probably the lack of motivation for improving production process given the high profitability gained from market power. This observation could be an indication that large firm's profitability does not necessarily rely on the production efficiency. For medium firms we don't see many dramatic changes compared to large firms. Two trends seem to be similar, and the movement of profit efficiency after the year 1996 is very consistent with the movement of production efficiency. It seems that production efficiency is the driving force for the profit efficiency change for medium firms, for there is little change in the allocative efficiency for these firms after 1996. Without improvement in the allocative efficiency, medium firms show lowest average profit efficiency score compared to the other two groups. For small firms, the co-movement of profit efficiency and production efficiency is more consistent than large firms but less consistent than medium firms, suggesting that production efficiency have some impact on the profit efficiency for small firms, but there are other factors. Here we observe more entering and exiting behaviors compared to other groups, and these behaviors may have considerable impact on the profit efficiency for small firms. For example, we see a peak for profit efficiency in 1998. This is followed by the exiting of a profit inefficient firm. Another increase in the profit efficiency in 2004 is followed by the entering of 4 new firms. The entering may introduce a lower efficiency score for the new firms, but brings more competition and consequently higher average profit efficiency in the group. To summarize, these observations may suggest that firms of different sizes may have different ways of achieving profitability.

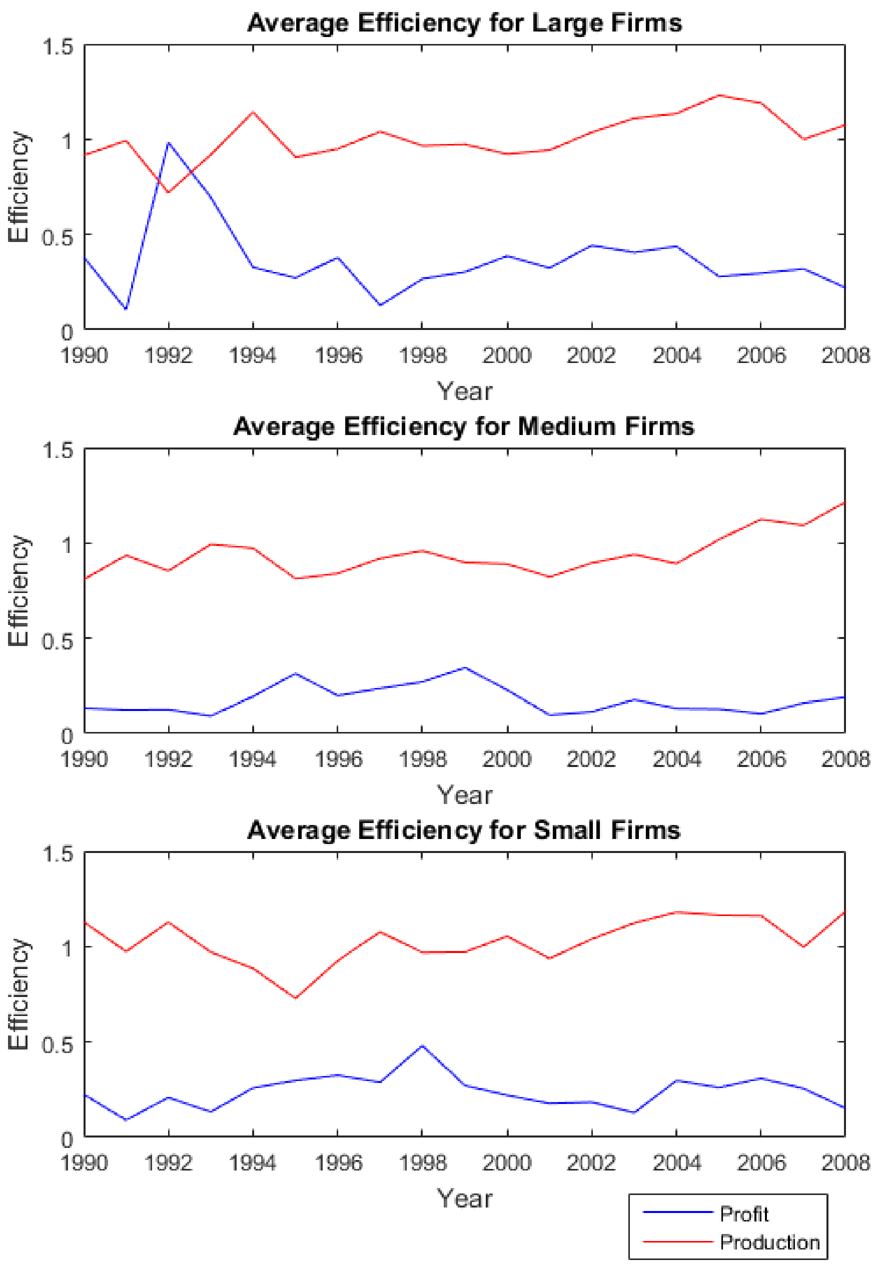

Figure 2.3: Efficiency Trends for Different Groups

Bibliography

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

Akhavein, J., Berger, A. and D.B., H. (1997). The effects of megamergers on efficiency and prices: evidence from a bank profit function, Review of Industrial Organization **12**: 95–139.

Akhigbe, A. and McNulty, J. (2005). Profit efficiency sources and differences among small and large u.s. commercial banks, Journal of Economics and Finance **29**: 289–299.

Ali, F., Parikh, A. and Shah, M. (1994). Measurement of profit efficiency using behavioural and stochastic frontier approaches, Applied Economics **26**: 181–188.

Aragon, Y., Daouia, A. and C., T.-A. (2005). Nonparametric frontier estimation: a conditional quantile-based approach, Econometric Theory **21**: 358–389.

Badin, L. and Simar, L. (2009). A bias-corrected nonparametric envelopment estimator of frontiers, Econometric Theory **25**: 1289–1318.

Becker-Blease, J. R., Kaen, F. R., Etebgari, A. and Baumann, H. (2010). Employees, firm size and profitability in u.s. manufacturing industries, Investment Management and Financial Innovations **7(2)**: 7–23.

Berger, A. N., D. H. and Humphrey, D. B. (1993). Bank efficiency derived from the profit function, Journal of Banking and Finance **17**: 317–347.

Berger, A. N. and Mester, L. J. (1997). Inside the black box: What explains differences in the efficiencies of financial institutions?, Journal of Banking and Finance **21**: 895–947.

Cazals, C., Florens, J. P. and Simar, L. (2002). Nonparametric frontier estimation: A robust appoach, Journal of Econometrics **106**: 1–25.

Charnes, A., Cooper, W. and Rhodes, E. (1978). Measuring the efficiency of decision making units, European Journal of Operational Research **2**: 429–444.

Daouia, A., Florens, J. P. and Simar, L. (2010). Frontier estimation and extreme value theory, Bernoulli **16**: 1039–1063.

Daouia, A. and Simar, L. (2005). Robust nonparametric estimators of monotone boundaries, Journal of Multivariate Analysis **96**: 311–331.

Daouia, A. and Simar, L. (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach, Journal of Econometrics **140**: 375–400.

Fan, Y., Li, Q. and Weersink, A. (1996). Semiparametric estimation of stochastic production frontier models, Journal of Business and Economic Statistics **14**: 460–468.

Färe, R., Fukuyama, H., Grosskopf, S. and Zelenyuk, V. (2015). Decomposing profit efficiency using a slack-based directional distance function, European Journal of Operational Research **247(1)**: 335–337.

Färe, R., Grosskopf, S. and Weber, W. L. (2004). The effect of risk-based capital requirements on profit efficiency in banking, Applied Economics **36:15**: 1731–1743.

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

Gijbels, I., Mammen, E., Park, B. and Simar, L. (1999). On estimation of monotone and concave frontier functions, Journal of American Statistical Association **94**: 220–228.

Greene, W., H. (2008). The econometric approach to efficiency analysis, Oxford University Press .

Herr, A., Schmitz, H. and Augurzky, B. (2010). Profit efficiency and ownership of german hospitals, Health Economics **20**: 660–674.

Humphrey, B. and Pulley, B. L. (1997). Banks' responses to deregulation: Profits, technology, and efficiency, Journal of Money, Credit and Banking **29(1)**: 73–93.

Jeong, O. S., Park, B. U. and Simar, L. (2010). Nonparametric conditional efficiency measures: asymptotic properties, Annals of Operational Research **173**: 105–122.

Jeong, O. S. and Simar, L. (2006). Linearly interpolated fdh efficiency score for nonconvex frontiers, Journal of Multivariate Analysis **97**: 2141–2161.

Kneip, A., Park, B. and Simar, L. (1998). A note on the convergence of nonparametric dea efficiency measures, Econometric Theory **14**: 783–793.

Knight, K. (2001). Limiting distributions of linear programming estimators, Extremes **4**: 87–103.

Korostelev, A., Simar, L. and Tsybakov, A. B. (1995). Efficient estimation of monotone boundaries, Annals of Statistics **23**: 476–89.

Kumbhakar, S. C., Park, B. U., Simar, L. and Tsionas, E. (2007). Nonparametric stochastic frontiers: A local maximum likelihood approach, Journal of Econometrics **137**: 1–27.

Kumbhakar, S. and Lovell, C. A. K. (2000). Stochastic frontier analysis, Cambridge University Press .

Martins-Filho, C. and Yao, F. (2008). A smooth nonparametric conditional quantile frontier estimator, Journal of Econometrics **143**: 317–333.

Martins-Filho, C. and Yao, F. (2015). Semiparametric stochastic frontier estimation via profile likelihood, Econometric Reviews **34**: 413–451.

Maudos, J. and Pastor, J. M. (2003). Cost and profit efficiency in the spanish banking sector (1985-1996): A non-parametric approach, Applied Financial Economics **13(1)**: 1–12.

Maudos, J., Pastor, J. M., Perez, F. and Quesada, J. (2002). Cost and profit efficiency in european banks, Journal of International Financial Markets, Institutions and Money **12**: 33–58.

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

Mynbaev, K. and Martins-Filho, C. (2010). Bias reduction in kernel density estimation via lipschitz condition, Journal of Nonparametric Statistics **22**: 219–235.

Park, B.and Simar, L. and Weiner, C. (2000). The fdh estimator for productivity efficient scores: asymptotic properties, Econometric Theory **16**: 855–877.

Park, U. B., Jeong, S.-O. and Simar, L. (2010). Asymptotic distribution of conical-hull estimators of directional edges, Annals of Statistics **38**: 1320–1340.

Park, U. B., Simar, L. and Weiner, C. (2000). The fdh estimator for productivity efficiency scores - asymptotic properties, Econometric Theory **16**: 855–877.

Ruiz, J, L. and Sirvent, I. (2011). A dea approach to derive individual lower and upper bounds for the technical and allocative components of the overall profit efficiency, The Journal of the Operational Research Society **62(11)**: 1907–1916.

Simar, L. and Vanhems, A. (2012). Probabilistic characterization of directional distances and their robust versions, Journal of Econometrics **166**: 342–354.

Simar, L. and Zelenyuk, V. (2011). Stochastic fdh/dea estimators for frontier analysis, Journal of Productivity Analysis **36**: 1–20.

Appendix A

Proofs and auxiliary lemmas

This appendix provides the proofs of lemmas, propositions, and the main theorems. Throughout the proofs, let $c > 0$ represent an arbitrary constant that may take different values in different contexts. Denote $D_{p,w} = (0, p_1] \times ...(0, p_{d_1}] \times [w_1, \infty) \times [w_{d_2}, \infty)$. By the definition of $\pi(p, w)$, for any point $(p^{'}, w^{'}) \in D_{p,w}$, we have $\pi(p^{'}, w^{'}) \leq \pi(p, w)$. Therefore, we can write $D_{p,w} = \pi^{-1}([0, \pi(p, w)])$. Denote $P(\pi, p, w) = \mathcal{P}(\Pi \leq \pi, P \leq p, W \geq w)$ and $\hat{P}(\pi, p, w)$ as defined in (1.5). $\sigma_\kappa = \int_{-B_M}^{B_M} M_k(\gamma)\gamma\kappa_M(\gamma)d\gamma$, assuming its existence. In order to prove Theorems 1 and 2, we need the following lemmas.

A.1 Lemmas

Lemma 1. *Let $M_k(x)$ be defined as in (1.7). Provided Assumption 2, we have: (i) $M_k(\cdot)$ is a symmetric bounded kernel function with compact support $[-B_M, B_M]$. $\int_{-B_M}^{B_M} \gamma M_k(\gamma)d\gamma = 0$; (ii) $\int_{-B_M}^{B_M} \gamma^2 M_k(\gamma)d\gamma := \sigma^2_{M_k} = 2\sigma^2_K \sum_{s=1}^{k} \lambda_{k,s}s^2$; (iii) For any $\gamma, \gamma' \in [-B_M, B_M]$, we have $|M_k(\gamma) - M_k(\gamma')| \leq m_{M_k}|\gamma - \gamma'|$ for some $0 < m_{M_k} < \infty$; (iv) For any $\zeta, \zeta' \in [-B_M, \infty)$, we have $|\kappa_M(\zeta) - \kappa_M(\zeta')| \leq m_{\kappa_k}|\zeta - \zeta'|$ for some $0 < m_{\kappa_k} < \infty$, where $\kappa_M(\zeta) = \int_{-B_M}^{\zeta} M_k(\gamma)d\gamma$.*

Proof. (i) For any fixed positive integer k, let $B_M = k \cdot B_K$. If $x \in (-\infty, -B_M) \cup (B_M, \infty)$, then $x/k \in (-\infty, -B_K) \cup (B_K, \infty)$. By Assumption 2, $K(x/k) = 0$. Therefore,

$$M_k(x) = -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K(\frac{x}{s}) = 0.$$

Furthermore, since $M_k(x)$ is just a finite linear combination of bounded functions $K(.)$, it is also bounded on the support $[-B_M, B_M]$. By construction, we can write

$$M_k(x) = \sum_{s=1}^{k} \frac{\lambda_{k,s}}{s}[K(\frac{x}{s}) + K(-\frac{x}{s})],$$

where $\lambda_{k,s} = -(c_{k,s}/c_{k,0}) = -(c_{k,-s}/c_{k,0}) = (-1)^{s+1}(k!)^2/(k+s)!(k-s)!,\ s = 1,...,k$. Therefore $M_k(\cdot)$ is symmetric: $M_k(x) = M_k(-x)$, and $\int_{-B_M}^{B_M} \gamma M_k(\gamma)d\gamma = 0$. Next, since $\sum_{|s|=0}^{k} c_{k,s} = (1-1)^{2k} = 0$, we have

$$-\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} = 1 \text{ or } \sum_{s=1}^{k} \lambda_{k,s} = \frac{1}{2}.$$

Therefore,

$$\begin{aligned}
\int_{-B_M}^{B_M} M_k(x)dx &= \sum_{s=1}^{k} \int_{-B_M}^{B_M} \frac{\lambda_{k,s}}{s}[K(\frac{x}{s}) + K(-\frac{x}{s})]dx \\
&= \sum_{s=1}^{k} \frac{\lambda_{k,s}}{s}[\int_{-kB_K}^{kB_K} K(\frac{x}{s})dx + \int_{-kB_K}^{kB_K} K(-\frac{x}{s})dx] \\
&= 2\sum_{s=1}^{k} \lambda_{k,s}[\int_{-\frac{k}{s}B_K}^{\frac{k}{s}B_K} K(\varphi)d\varphi] \\
&= 1.
\end{aligned}$$

The last equality comes from the fact that for $s = 1,...,k,\ \frac{k}{s}B_K \geq B_K$ and $\int_{-\frac{k}{s}B_K}^{\frac{k}{s}B_K} K(\varphi)d\varphi = \int_{-B_K}^{B_K} K(\varphi)d\varphi = 1$.

(ii)

$$\begin{aligned}
\int_{-B_M}^{B_M} \gamma^2 M_k(\gamma)d\gamma &= \sum_{s=1}^{k} \frac{\lambda_{k,s}}{s} \int_{-B_M}^{B_M} [\gamma^2 K(\frac{\gamma}{s}) + \gamma^2 K(-\frac{\gamma}{s})]d\gamma \\
&= \sum_{s=1}^{k} \frac{\lambda_{k,s}}{s}[\int_{-kB_K}^{kB_K} \gamma^2 K(\frac{\gamma}{s})d\gamma + \int_{-kB_K}^{kB_K} \gamma^2 K(-\frac{\gamma}{s})d\gamma].
\end{aligned}$$

For each $s = 1,...,k$,

$$\begin{aligned}
\int_{-kB_K}^{kB_K} \gamma^2 K(\frac{\gamma}{s})d\gamma &= \int_{-\frac{k}{s}B_K}^{\frac{k}{s}B_K} s(s\varphi)^2 K(\varphi)d\varphi \\
&= s^3 \int_{-B_K}^{B_K} \varphi^2 K(\varphi)d\varphi \\
&= s^3\sigma_K^2.
\end{aligned}$$

Therefore,

$$\int_{-B_M}^{B_M} \gamma^2 M_k(\gamma) d\gamma = 2\sigma_K^2 \sum_{s=1}^{k} \lambda_{k,s} s^2 = \sigma_{M_k}^2.$$

(iii) For any $\gamma, \gamma' \in [-B_M, B_M]$, we have $\gamma/s, \gamma'/s \in [-B_K, B_K]$ for any $s = 1, ..., k$. By Assumption 2,

$$\begin{aligned} |K(\frac{\gamma}{s}) - K(\frac{\gamma'}{s})| &\leq m_K |\frac{\gamma}{s} - \frac{\gamma'}{s}| \\ &= \frac{m_K}{s} |\gamma - \gamma'|. \end{aligned}$$

By triangle inequality,

$$\begin{aligned} & |M_k(\gamma) - M_k(\gamma')| \\ \leq & \sum_{s=1}^{k} \frac{|\lambda_{k,s}|}{s} [|K(\frac{\gamma}{s}) - K(\frac{\gamma'}{s})| + |K(-\frac{\gamma}{s}) - K(-\frac{\gamma'}{s})|] \\ \leq & \sum_{s=1}^{k} \frac{2|\lambda_{k,s}| m_K}{s^2} |\gamma - \gamma'| \\ = & \; m_{M_k} |\gamma - \gamma'|. \end{aligned}$$

(iv) Without loss of generality, assume $\zeta' < \zeta$, By the definition of $\kappa_M(.)$,

$$\begin{aligned} & |\kappa_M(\zeta) - \kappa_M(\zeta')| \\ = & \; |\int_{-B_M}^{\zeta} M_k(\gamma) d\gamma - \int_{-B_M}^{\zeta'} M_k(\gamma) d\gamma| \\ = & \; |\int_{\zeta'}^{\zeta} M_k(\gamma) d\gamma| \\ = & \; |\sum_{s=1}^{k} \frac{\lambda_{k,s}}{s} [\int_{\zeta'}^{\zeta} K(\frac{\gamma}{s}) d\gamma + \int_{\zeta'}^{\zeta} K(-\frac{\gamma}{s}) d\gamma]|. \end{aligned}$$

Since

$$\begin{aligned} |\int_{\zeta'}^{\zeta} K(\frac{\gamma}{s}) d\gamma| &= s |\int_{-\frac{\zeta'}{s}}^{\frac{\zeta}{s}} K(\varphi) d\varphi| \\ &= s |\int_{-B_K}^{\frac{\zeta}{s}} K(\varphi) d\varphi - \int_{-B_K}^{-\frac{\zeta'}{s}} K(\varphi) d\varphi| \\ &= s |\kappa(\zeta/s) - \kappa(\zeta'/s)| \\ &\leq s m_\kappa |\zeta/s - \zeta'/s| \\ &= m_\kappa |\zeta - \zeta'|, \end{aligned}$$

the result then follows by triangle inequality. ☐

Lemma 2. *Under Assumption 1-4 and Assumption 5A, we have:*

$$|E(\hat{P}(\pi,p,w)) - P(\pi,p,w)| \leq ch_n^{2k}\left[\int_{D_{p,w}} H_{2k}(\pi,P,W)d(P,W) + \int_{D_{p,w}} \sup_{\pi\in\mathbb{R}} |F_f(\pi,P,W)|\varepsilon_{2k}^{-2k}(\pi,P,W)d(P,W)\right],$$

for sufficiently small h_n. If we assume furthermore $H_{2k}(\pi,p,w)$, $F_f(\pi,p,w)$ and $\varepsilon_{2k}(\pi,p,w)$ are bounded for all $(\pi,p,w)\in\Psi^*$, we have $|E(\hat{P}(\pi,p,w)) - P(\pi,p,w)| = O(h_n^{2k})$

Proof.

$$\begin{aligned}
E(\hat{P}(\pi,p,w)) &= E[(nh_n)^{-1}\sum_{i=1}^{n}(\int_0^{\pi} M_k(\frac{\Pi_i-\gamma}{h_n})d\gamma)I(P_i\leq p, W_i\geq w)] \\
&= {h_n}^{-1}\int_{D_{p,w}}\int_{[0,\pi(P,W)]}\int_0^{\pi} M_k(\frac{\Pi-\gamma}{h_n})d\gamma f(\Pi,P,W)d\Pi d(P,W) \\
&= \int_{D_{p,w}}\int_{[0,\pi(P,W)]}\int_{-B_M}^{\frac{\pi-\Pi}{h_n}} M_k(\varphi)d\varphi f(\Pi,P,W)d\Pi d(P,W) \\
&= \int_{D_{p,w}}\int_{[0,\pi(P,W)]}\kappa_M(\frac{\pi-\Pi}{h_n})f(\Pi,P,W)d\Pi d(P,W).
\end{aligned}$$

The third equality follows from the fact that $M_k(.)$ is symmetric and $-\frac{\Pi}{h_n} < -B_M$ for sufficiently small h_n. We consider 3 cases: (1) $0<\pi<\pi(p,w)$; (2) $\pi>\pi(p,w)$; (3) $\pi=\pi(p,w)$. Here we only derive the result for case (1), for case (2) and (3) results are obtained in a similar manner.

For case (1),

$$\begin{aligned}
E(\hat{P}(\pi,p,w)) &= \int_{D_{p,w}}\int_{[0,\pi(P,W)]}\kappa_M(\frac{\pi-\Pi}{h_n})f(\Pi,P,W)d\Pi d(P,W) \\
&= \int_{\pi^{-1}([0,\pi(p,w)])}\int_{[0,\pi(P,W)]}\kappa_M(\frac{\pi-\Pi}{h_n})f(\Pi,P,W)d\Pi d(P,W) \\
&= \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])}\int_{[0,\pi(P,W)]}\kappa_M(\frac{\pi-\Pi}{h_n})f(\Pi,P,W)d\Pi d(P,W) \\
&\quad + \int_{\pi^{-1}(\{\pi\})}\int_{[0,\pi]}\kappa_M(\frac{\pi-\Pi}{h_n})f(\Pi,P,W)d\Pi d(P,W) \\
&= A_{1n} + A_{2n}.
\end{aligned}$$

Note that for the last term, for $(P,W)\in\pi^{-1}(\{\pi\})$, it must be $\Pi<\pi(P,W)=\pi$. Thus $\frac{\pi-\Pi}{h_n} > B_M$ for sufficient small h_n. Therefore $\kappa_M(\frac{\pi-\Pi}{h_n}) \to 1$ as $n\to\infty$. By Assumptions 2 and 4,

$|\kappa_M(\frac{\pi-\Pi}{h_n})f(\Pi,P,W)|<\infty$. By Lebesgue's dominated convergence theorem (LDC),

$$\begin{aligned}
A_{2n} &= \int_{\pi^{-1}(\{\pi\})}\int_{[0,\pi]}\kappa_M(\frac{\pi-\Pi}{h_n})f(\Pi,P,W)d\Pi d(P,W)\\
&= \int_{\pi^{-1}(\{\pi\})}\int_{[0,\pi)}\kappa_M(\frac{\pi-\Pi}{h_n})f(\Pi,P,W)d\Pi d(P,W)\\
&\to \int_{\pi^{-1}(\{\pi\})}\int_{[0,\pi]}f(\Pi,P,W)d\Pi d(P,W).
\end{aligned}$$

Now,

$$A_{1n}=\int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])}\int_{[0,\pi(P,W)]}\kappa_M(\frac{\pi-\Pi}{h_n})\frac{\partial F_f(\Pi,P,W)}{\partial\Pi}d\Pi d(P,W),$$

where $F_f(\Pi,P,W)=\int_0^\Pi f(\gamma,P,W)d\gamma$. Using integration by parts,

$$\begin{aligned}
&\int_{[0,\pi(P,W)]}\kappa_M(\frac{\pi-\Pi}{h_n})\frac{\partial F_f(\Pi,P,W)}{\partial\Pi}d\Pi\\
=&\int_{[0,\pi(P,W)]}\kappa_M(\frac{\pi-\Pi}{h_n})dF_f(\Pi,P,W)\\
=&\kappa_M(\frac{\pi-\Pi}{h_n})F_f(\Pi,P,W)|_{\Pi=0}^{\Pi=\pi(P,W)}-\int_{[0,\pi(P,W)]}F_f(\Pi,P,W)d\kappa_M(\frac{\pi-\Pi}{h_n})\\
=&\kappa_M(\frac{\pi-\pi(P,W)}{h_n})F_f(\pi(P,W),P,W)+\frac{1}{h_n}\int_{[0,\pi(P,W)]}F_f(\Pi,P,W)M_k(\frac{\pi-\Pi}{h_n})d\Pi\\
=&\kappa_M(\frac{\pi-\pi(P,W)}{h_n})F_f(\pi(P,W),P,W)+\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}}F_f(\pi-h_n\gamma,P,W)M_k(\gamma)d\gamma,
\end{aligned}$$

where the third equality follows from the fact that $F_f(0,P,W)=0$. Hence,

$$A_{1n}=E_{1n}+E_{2n},$$

where

$$\begin{aligned}
E_{1n} &= \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])}\kappa_M(\frac{\pi-\pi(P,W)}{h_n})F_f(\pi(P,W),P,W)d(P,W),\\
E_{2n} &= \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])}\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}}F_f(\pi-h_n\gamma,P,W)M_k(\gamma)d\gamma d(P,W).
\end{aligned}$$

Now,

$$\begin{aligned}
E_{1n} &= \int_{\pi^{-1}([0,\pi))}\kappa_M(\frac{\pi-\pi(P,W)}{h_n})F_f(\pi(P,W),P,W)d(P,W)\\
&\quad+\int_{\pi^{-1}((\pi,\pi(p,w)])}\kappa_M(\frac{\pi-\pi(P,W)}{h_n})F_f(\pi(P,W),P,W)d(P,W)\\
&= E_{11,n}+E_{12,n}.
\end{aligned}$$

For $E_{11,n}$, note that when $(P,W) \in \pi^{-1}([0,\pi))$, $\frac{\pi-\pi(P,W)}{h_n} \to \infty$ and $\kappa_M(\frac{\pi-\pi(P,W)}{h_n}) \to 1$ as $n \to \infty$. By Assumption 2 and Assumption 4, $|\kappa_M(\frac{\pi-\pi(P,W)}{h_n})F_f(\pi(P,W),P,W)| < \infty$. Thus by LDC we have

$$E_{11,n} \to \int_{\pi^{-1}([0,\pi))} F_f(\pi(P,W),P,W)d(P,W) = \int_{\pi^{-1}([0,\pi))} \int_{[0,\pi(P,W)]} f(\Pi,P,W)d\Pi d(P,W).$$

For $E_{12,n}$, note that when $(P,W) \in \pi^{-1}((\pi,\pi(p,w)])$, $\frac{\pi-\pi(P,W)}{h_n} \to -\infty$ and $\kappa_M(\frac{\pi-\pi(P,W)}{h_n}) \to 0$ as $n \to \infty$. Again by LDC $E_{12,n} \to 0$ as $n \to \infty$. As a result,

$$E_{1n} \to \int_{\pi^{-1}([0,\pi))} \int_{[0,\pi(P,W)]} f(\Pi,P,W)d\Pi d(P,W).$$

For E_{2n}, we consider

$$\begin{aligned}
& E_{2n} - \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} \int_{[0,\pi]} f(\Pi,P,W)d\Pi d(P,W) \\
= & \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} \int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} F_f(\pi - h_n\gamma,P,W)M_k(\gamma)d\gamma d(P,W) \\
& - \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} \int_{[0,\pi]} f(\Pi,P,W)d\Pi d(P,W) \\
= & \Bigg[\int_{\pi^{-1}([0,\pi))} \int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} F_f(\pi - h_n\gamma,P,W)M_k(\gamma)d\gamma d(P,W) \\
& - \int_{\pi^{-1}([0,\pi))} \int_{[0,\pi(P,W)]} f(\Pi,P,W)d\Pi d(P,W)\Bigg] \\
& + \Bigg[\int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} F_f(\pi - h_n\gamma,P,W)M_k(\gamma)d\gamma d(P,W) \\
& - \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{[0,\pi]} f(\Pi,P,W)d\Pi d(P,W)\Bigg].
\end{aligned}$$

When $(P,W) \in \pi^{-1}([0,\pi))$, $\frac{\pi-\pi(P,W)}{h_n} > B_M$, and $M_k(\gamma) = 0$ for $\gamma > \frac{\pi-\pi(P,W)}{h_n}$. By LDC $\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} F_f(\pi - h_n\gamma,P,W)M_k(\gamma)d\gamma \to 0$ as $n \to \infty$. Therefore the terms in the first square bracket converges to $\int_{\pi^{-1}([0,\pi))} \int_{[0,\pi(P,W)]} f(\Pi,P,W)d\Pi d(P,W)$ as $n \to \infty$. For the terms in the second square bracket, by the definition of $M_k(.)$ we have:

$$
\begin{aligned}
&\int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} F_f(\pi - h_n\gamma, P, W) M_k(\gamma) d\gamma d(P,W) \\
&- \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{[0,\pi]} f(\Pi, P, W) d\Pi d(P,W) \\
= \quad & -\frac{1}{c_{k,0}} \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} K(t) \sum_{|s|=1}^{k} c_{k,s} F_f(\pi - s h_n t, P, W) dt d(P,W) \\
&- \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_0^{\pi} f(\gamma, P, W) d\gamma d(P,W).
\end{aligned}
$$

When $(P,W) \in \pi^{-1}((\pi,\pi(p,w)])$, $\frac{\pi-\pi(P,W)}{h_n} < -B_K \leq -B_M$ (see Lemma 1 for the relationship between B_K and B_M), and $\frac{\pi}{h_n} > B_M \geq B_K$ for sufficient small h_n. As a result, as $n \to \infty$, by LDC we have

$$
\begin{aligned}
& -\frac{1}{c_{k,0}} \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} K(t) \sum_{|s|=1}^{k} c_{k,s} F_f(\pi - s h_n t, P, W) dt d(P,W) \\
\to \quad & \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{-B_K}^{B_K} K(t) \sum_{|s|=1}^{k} \frac{c_{k,s}}{c_{k,0}} F_f(\pi - s h_n t, P, W) dt d(P,W)
\end{aligned}
$$

Note that $\sum_{|s|=1}^{k} \frac{c_{k,s}}{c_{k,0}} = -1$, we can write

$$
\begin{aligned}
& -\int_{\pi^{-1}((\pi,\pi(p,w)])} \int_0^{\pi} f(\gamma, P, W) d\gamma d(P,W) \\
= \quad & \sum_{|s|=1}^{k} \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{-B_K}^{B_K} K(t) dt \frac{c_{k,s}}{c_{k,0}} F_f(\pi, P, W) d(P,W).
\end{aligned}
$$

As a result,

$$
\begin{aligned}
&\int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} F_f(\pi - h_n\gamma, P, W) M_k(\gamma) d\gamma d(P,W) \\
&- \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{[0,\pi]} f(\Pi, P, W) d\Pi d(P,W) \\
\to \quad & -\frac{1}{c_{k,0}} \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{-B_K}^{B_K} K(t) \Delta_{h_n t}^{2k} F_f(\pi, P, W) dt d(P,W).
\end{aligned}
$$

Hence,

$$\begin{aligned}
E_{2n} \quad \to \quad & \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} \int_{[0,\pi]} f(\Pi, P, W) d\Pi d(P, W) \\
& - \int_{\pi^{-1}([0,\pi))} \int_{[0,\pi(P,W)]} f(\Pi, P, W) d\Pi d(P, W) \\
& - \frac{1}{c_{k,0}} \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{-B_K}^{B_K} K(t) \Delta^{2k}_{h_n t} F_f(\pi, P, W) dt d(P, W),
\end{aligned}$$

and

$$\begin{aligned}
A_{1n} \quad = \quad & E_{1n} + E_{2n} \\
\to \quad & \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} \int_{[0,\pi]} f(\Pi, P, W) d\Pi d(P, W) \\
& - \frac{1}{c_{k,0}} \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{-B_K}^{B_K} K(t) \Delta^{2k}_{h_n t} F_f(\pi, P, W) dt d(P, W).
\end{aligned}$$

Combining above results,

$$\begin{aligned}
E(\hat{P}(\pi, p, w)) \quad = \quad & A_{1n} + A_{2n} \\
\to \quad & \int_{D_{p,w}} \int_{[0,\pi]} f(\Pi, P, W) d\Pi d(P, W) \\
& - \frac{1}{c_{k,0}} \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{-B_K}^{B_K} K(t) \Delta^{2k}_{h_n t} F_f(\pi, P, W) dt d(P, W).
\end{aligned}$$

By Assumption 5A, we have

$$\begin{aligned}
& |E(\hat{P}(\pi, p, w)) - P(\pi, p, w)| \\
\leq \; & c \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{-B_K}^{B_K} |K(t) \Delta^{2k}_{h_n t} F_f(\pi, P, W)| dt d(P, W) \\
\leq \; & c \int_{D_{p,w}} \Big(\int_{|h_n t| \leq \varepsilon_{2k}(\pi,P,W)} + \int_{|h_n t| > \varepsilon_{2k}(\pi,P,W)} \Big) |K(t) \Delta^{2k}_{h_n t} F_f(\pi, P, W)| dt d(P, W) \\
\leq \; & c \Big[\int_{D_{p,w}} \int_{|h_n t| \leq \varepsilon_{2k}(\pi,P,W)} |K(t)| (h_n t)^{2k} dt H_{2k}(\pi, P, W) d(P, W) \\
& + \int_{D_{p,w}} \sup_{\pi \in \mathbb{R}} |F_f(\pi, P, W)| \int_{|h_n t| > \varepsilon_{2k}(\pi,P,W)} |K(t)| dt d(P, W) \Big].
\end{aligned}$$

Since for any $N > 0$,

$$\int_{|t|>N} |K(t)| dt \leq \int_{|t|>N} |K(t)| |\frac{t}{N}|^{2k} dt \leq N^{-2k} \int_{-B_K}^{B_K} |K(t)| t^{2k} dt,$$

as a result, we have

$$
\begin{aligned}
& |E(\hat{P}(\pi,p,w)) - P(\pi,p,w)| \\
\leq \quad & ch_n^{2k}\left[\int_{D_{p,w}} H_{2k}(\pi,P,W)d(P,W) + \int_{D_{p,w}} \sup_{\pi\in\mathbb{R}} |F_f(\pi,P,W)|\varepsilon_{2k}^{-2k}(\pi,P,W)d(P,W)\right]
\end{aligned}
$$

$\square$

Lemma 2 gives the order of the bias as functions of k. Thus as we increase k, the speed of decay of bias increases. If we assume f has bounded first order derivative with respect to π, by applying Taylor's Theorem, the next lemma provides a more explicit structure for bias and variance when $k = 1$.

Lemma 3. *For $k = 1$, under Assumption 1-4 and Assumption 5B, we have: (a)*

$$
E(\hat{P}(\pi,p,w)) = \begin{cases} P(\pi,p,w) + \frac{1}{2}h_n^2\sigma_{M_k}^2 \int_{\pi^{-1}((\pi,\pi(p,w)])} f^{(1)}(\pi,P,W)d(P,W) + o(h_n^2) \\ \text{if } 0 < \pi < \pi(p,w); \\ \\ P(\pi,p,w) + o(h_n^2) \quad \text{if } \pi \geq \pi(p,w). \end{cases}
$$

(b)

$$
V(\hat{P}(\pi,p,w)) = \begin{cases} n^{-1}P(\pi,p,w)(1-P(\pi,p,w)) \\ -2n^{-1}h_n\sigma_\kappa \int_{\pi^{-1}((\pi,\pi(p,w)])} f(\pi,P,W)d(P,W) \ + o(h_n/n) \\ \text{if } 0 < \pi < \pi(p,w); \\ \\ n^{-1}P(\pi,p,w)(1-P(\pi,p,w)) + o(h_n/n) \\ \text{if } \pi \geq \pi(p,w). \end{cases}
$$

where $P(\pi,p,w) = \mathcal{P}(\Pi \leq \pi, P \leq p, W \geq w)$ and $\hat{P}(\pi,p,w)$ is defined in (1.5). $\sigma_\kappa = \int_{-B_M}^{B_M} M_k(\gamma)\gamma\kappa_M(\gamma)d\gamma$, assuming its existence.

Proof. (a) From the proof of Lemma 2,

$$
E(\hat{P}(\pi,p,w) = \int_{D_{p,w}} \int_{[0,\pi(P,W)]} \kappa_M(\frac{\pi - \Pi}{h_n}) f(\Pi,P,W)d\Pi d(P,W).
$$

We consider 3 cases: (1) $0 < \pi < \pi(p,w)$; (2) $\pi > \pi(p,w)$; (3) $\pi = \pi(p,w)$. For case (1),

$$\begin{aligned} E(\hat{P}(\pi,p,w)) &= \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} \int_{[0,\pi(P,W)]} \kappa_M(\frac{\pi-\Pi}{h_n}) f(\Pi,P,W) d\Pi d(P,W) \\ &\quad + \int_{\pi^{-1}(\{\pi\})} \int_{[0,\pi]} \kappa_M(\frac{\pi-\Pi}{h_n}) f(\Pi,P,W) d\Pi d(P,W) \\ &= A_{1n} + A_{2n}. \end{aligned}$$

Again, following the proof of Lemma 2,

$$A_{2n} \to \int_{\pi^{-1}(\{\pi\})} \int_{[0,\pi]} f(\Pi,P,W) d\Pi d(P,W),$$

and

$$A_{1n} = \kappa_M(\frac{\pi-\pi(P,W)}{h_n}) F_f(\pi(P,W),P,W) + \int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} F_f(\pi - h_n\gamma, P, W) M_k(\gamma) d\gamma.$$

By Taylor's theorem, $F_f(\pi - h_n\gamma, P, W) = F_f(\pi,P,W) - h_n\gamma f(\pi,P,W) + \frac{1}{2}h_n^2\gamma^2 f^{(1)}(\pi,P,W) + o(h_n^2)$. Hence, by LDC,

$$\begin{aligned} A_{1n} &= E_{1n} + \tilde{E}_{2n} + E_{3n} + E_{4n} + \int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} o(h_n^2) M_k(\gamma) d\gamma \\ &= E_{1n} + \tilde{E}_{2n} - E_{3n} + E_{4n} + o(h_n^2), \end{aligned}$$

where

$$\begin{aligned} E_{1n} &= \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} \kappa_M(\frac{\pi-\pi(P,W)}{h_n}) F_f(\pi(P,W),P,W) d(P,W); \\ \tilde{E}_{2n} &= \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} F_f(\pi,P,W) \int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma) d\gamma d(P,W); \\ E_{3n} &= h_n \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} f(\pi,P,W) \int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma d\gamma d(P,W); \\ E_{4n} &= \frac{1}{2}h_n^2 \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} f^{(1)}(\pi,P,W) \int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma^2 d\gamma d(P,W). \end{aligned}$$

Following the proof in Lemma 2,

$$E_{1n} \to \int_{\pi^{-1}([0,\pi))} \int_{[0,\pi(P,W)]} f(\Pi,P,W) d\Pi d(P,W)$$

For $\tilde{E}_{2n}$, we can write:

$$\begin{aligned}
\tilde{E}_{2n} &= \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} F_f(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)d\gamma d(P,W)\\
&= \int_{\pi^{-1}([0,\pi))} F_f(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)d\gamma d(P,W)\\
&\quad +\int_{\pi^{-1}((\pi,\pi(p,w)])} F_f(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)d\gamma d(P,W)\\
&= \tilde{E}_{21,n}+\tilde{E}_{22,n}.
\end{aligned}$$

For $\tilde{E}_{21,n}$, when $(P,W)\in\pi^{-1}([0,\pi))$, $\frac{\pi-\pi(P,W)}{h_n}\to\infty$ and $\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)d\gamma\to 0$ as $n\to\infty$ by LDC. For $\tilde{E}_{22,n}$, when $(P,W)\in\pi^{-1}((\pi,\pi(p,w)])$, $\frac{\pi-\pi(P,W)}{h_n}\to-\infty$ and $\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)d\gamma\to 1$ as $n\to\infty$. As a result, $\tilde{E}_{2n}\to\int_{\pi^{-1}((\pi,\pi(p,w)])}\int_{[0,\pi]} f(\Pi,P,W)d\Pi d(P,W)$.

$$\begin{aligned}
h_n^{-1}E_{3n} &= \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} f(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma d\gamma d(P,W)\\
&= \int_{\pi^{-1}([0,\pi))} f(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma d\gamma d(P,W)\\
&\quad +\int_{\pi^{-1}((\pi,\pi(p,w)])} f(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma d\gamma d(P,W)\\
&= E_{31,n}+E_{32,n}.
\end{aligned}$$

For $E_{31,n}$, when $(P,W)\in\pi^{-1}([0,\pi))$, $\frac{\pi-\pi(P,W)}{h_n}\to\infty$ and $\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma d\gamma\to 0$ as $n\to\infty$ by LDC. For $E_{32,n}$, when $(P,W)\in\pi^{-1}((\pi,\pi(p,w)])$, $\frac{\pi-\pi(P,W)}{h_n}\to-\infty$ and $\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma d\gamma\to 0$ as $n\to\infty$ by the symmetry of $M_k(.)$. As a result, $h_n^{-1}E_{3n}\to 0$.

$$\begin{aligned}
h_n^{-2}E_{4n} &= \frac{1}{2}\int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} f^{(1)}(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma^2 d\gamma d(P,W)\\
&= \frac{1}{2}\int_{\pi^{-1}([0,\pi])} f^{(1)}(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma^2 d\gamma d(P,W)\\
&\quad +\frac{1}{2}\int_{\pi^{-1}([\pi,\pi(p,w)])} f^{(1)}(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma^2 d\gamma d(P,W)\\
&= E_{41,n}+E_{42,n}.
\end{aligned}$$

Similarly, when $(P,W)\in\pi^{-1}([0,\pi))$, $\frac{\pi-\pi(P,W)}{h_n}\to+\infty$ and $E_{41,n}\to 0$ as $n\to\infty$. When $(P,W)\in\pi^{-1}((\pi,\pi(p,w)])$, $\frac{\pi-\pi(P,W)}{h_n}\to-\infty$ and $\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma^2 d\gamma\to\sigma^2_{M_k}$ as $n\to\infty$ by the symmetry of $M_k(.)$. As a result, $h_n^{-2}E_{4n} \to \frac{1}{2}\sigma^2_{M_k}\int_{\pi^{-1}((\pi,\pi(p,w)])} f^{(1)}(\pi,P,W)d(P,W)$. Therefore, if $0 < \pi < \pi(p,w)$,

$$\begin{aligned}
E(\hat{P}(\pi,p,w)) &= E_{1n} + \tilde{E}_{2n} + E_{3n} + E_{4n} + A_{2n} + o(h_n^2) \\
&= \int_{\pi^{-1}([0,\pi))}\int_{[0,\pi]} f(\Pi,P,W)d\Pi d(P,W) \\
&\quad + \int_{\pi^{-1}((\pi,\pi(p,w)])}\int_{[0,\pi]} f(\Pi,P,W)d\Pi d(P,W) \\
&\quad + \int_{\pi^{-1}(\{\pi\})}\int_{[0,\pi]} f(\Pi,P,W)d\Pi d(P,W) \\
&\quad + \frac{1}{2}h_n^2\sigma^2_{M_k}\int_{\pi^{-1}((\pi,\pi(p,w)])} f^{(1)}(\pi,P,W)d(P,W) + o(h_n^2) \\
&= P(\pi,p,w) + \frac{1}{2}h_n^2\sigma^2_{M_k}\int_{\pi^{-1}((\pi,\pi(p,w)])} f^{(1)}(\pi,P,W)d(P,W) + o(h_n^2).
\end{aligned}$$

For case (2), when $\pi > \pi(p,w)$, $\pi > \pi(P,W)$ for all $(P,W) \in D_{p,w}$, $\kappa_M(\frac{\pi-\pi(P,W)}{h_n}) \to 1$, $\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)d\gamma \to 0$, $\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma d\gamma \to 0$ and $\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma^2 d\gamma \to 0$. By LDC,

$$\begin{aligned}
\int_{D_{p,w}} \kappa_M(\frac{\pi-\pi(P,W)}{h_n})F_f(\pi(P,W),P,W)d(P,W) &\to P(\pi,p,w); \\
\int_{D_{p,w}} F_f(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)d\gamma d(P,W) &\to 0; \\
h_n\int_{D_{p,w}} f(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma d\gamma d(P,W) &\to 0; \\
\frac{1}{2}h_n^2\int_{D_{p,w}} f^{(1)}(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma^2 d\gamma d(P,W). &\to 0
\end{aligned}$$

Therefore, if $\pi > \pi(p,w)$, $E(\hat{P}(\pi,p,w)) = P(\pi,p,w) + o(h_n^2)$. For case (3), the proof is similar to case (1) with the set $(\pi,\pi(p,w)]$ replaced by the empty set.

(b) Note that $V(\hat{P}(\pi,p,w)) = \frac{1}{n}(V_{1n} - V_{2n})$, where

$$\begin{aligned}
V_{1n} &= E[{h_n}^{-2}(\int_0^\pi M_k(\frac{\Pi-\gamma}{h_n})d\gamma)^2 I(P_i \le p, W_i \ge w)]; \\
V_{2n} &= (E[h_n^{-1}\int_0^\pi M_k(\frac{\Pi-\gamma}{h_n})d\gamma I(P_i \le p, W_i \ge w)])^2.
\end{aligned}$$

From the proof in part (a), we know the limiting behavior of V_{2n}. Now for V_{1n}, since $h_n \to 0$ as $n \to \infty$, there exist $N(p,w) \in \mathbb{R}_+$ such that for all $n > N(p,w)$,

$$
\begin{aligned}
V_{1n} &= E[{h_n}^{-2}(\int_0^{\pi} M_k(\frac{\Pi-\gamma}{h_n})d\gamma)^2 I(P_i \le p, W_i \ge w)] \\
&= {h_n}^{-2}\int_{D_{p,w}}\int_{[0,\pi(P,W)]}(\int_0^{\pi} M_k(\frac{\Pi-\gamma}{h_n})d\gamma)^2 f(\Pi,P,W)d\Pi d(P,W) \\
&= \int_{D_{p,w}}\int_{[0,\pi(P,W)]}(\int_{-B_M}^{\frac{\pi-\Pi}{h_n}} M_k(\varphi)d\varphi)^2 f(\Pi,P,W)d\Pi d(P,W) \\
&= \int_{D_{p,w}}\int_{[0,\pi(P,W)]}(\kappa_M(\frac{\pi-\Pi}{h_n}))^2 f(\Pi,P,W)d\Pi d(P,W).
\end{aligned}
$$

Like part (a), we also consider 3 cases when (1) $0 < \pi < \pi(p,w)$; (2) $\pi > \pi(p,w)$; (3) $\pi = \pi(p,w)$.

For case (1),

$$
\begin{aligned}
V_{1n} &= \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])}\int_{[0,\pi(P,W)]}(\kappa_M(\frac{\pi-\Pi}{h_n}))^2 f(\Pi,P,W)d\Pi d(P,W) \\
&\quad + \int_{\pi^{-1}(\{\pi\})}\int_{[0,\pi]}(\kappa_M(\frac{\pi-\Pi}{h_n}))^2 f(\Pi,P,W)d\Pi d(P,W) \\
&= \tilde{A}_{1n} + \tilde{A}_{2n}.
\end{aligned}
$$

Note that for the last term, for $\Pi < \pi$, $\kappa_M(\frac{\pi-\Pi}{h_n}) \to 1$ as $n \to \infty$. By Assumptions 2 and 4, $|\kappa_M(\frac{\pi-\Pi}{h_n})^2 f(\Pi,P,W)| < \infty$. By Lebesgue's dominated convergence theorem,

$$
\tilde{A}_{2n} \to \int_{\pi^{-1}(\{\pi\})}\int_{[0,\pi]} f(\Pi,P,W)d\Pi d(P,W).
$$

Now,

$$
\tilde{A}_{1n} = \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])}\int_{[0,\pi(P,W)]}(\kappa_M(\frac{\pi-\Pi}{h_n}))^2 \frac{\partial F_f(\Pi,P,W)}{\partial \Pi} d\Pi d(P,W),
$$

where $F_f(\Pi,P,W) = \int_0^{\pi} f(\gamma,p,w)d\gamma$. Note that

$$
\int_{[0,\pi(P,W)]}(\kappa_M(\frac{\pi-\Pi}{h_n}))^2 \frac{\partial F_f(\Pi,P,W)}{\partial \Pi} d\Pi \quad = \quad \int_{[0,\pi(P,W)]}(\kappa_M(\frac{\pi-\Pi}{h_n}))^2 dF_f(\Pi,P,W)
$$

Using integral by parts,

$$
\begin{aligned}
&\int_{[0,\pi(P,W)]}(\kappa_M(\frac{\pi-\Pi}{h_n}))^2 dF_f(\Pi,P,W) \\
&= (\kappa_M(\frac{\pi-\Pi}{h_n}))^2 dF_f(\Pi,P,W)|_{\Pi=0}^{\Pi=\pi(P,W)} \\
&\quad - \int_{[0,\pi(P,W)]} F_f(\Pi,P,W)d(\kappa_M(\frac{\pi-\Pi}{h_n}))^2
\end{aligned}
$$

$$
\begin{aligned}
&= (\kappa_M(\frac{\pi-\pi(P,W)}{h_n}))^2 F_f(\pi(P,W),P,W) \\
&\quad + \frac{2}{h_n}\int_{[0,\pi(P,W)]} F_f(\Pi,P,W)\kappa_M(\frac{\pi-\Pi}{h_n})M_k(\frac{\pi-\Pi}{h_n})d\Pi \\
&= (\kappa_M(\frac{\pi-\pi(P,W)}{h_n}))^2 F_f(\pi(P,W),P,W) \\
&\quad + 2\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} F_f(\pi-h_n\gamma,P,W)\kappa_M(\gamma)M_k(\gamma)d\gamma.
\end{aligned}
$$

By Taylor's theorem, $F_f(\pi-h_n\gamma,P,W) = F_f(\pi,P,W) - h_n\gamma f(\pi,P,W) + o(h_n)$, Hence by LDC,

$$
\begin{aligned}
\tilde{A}_{1n} &= V_{11n} + V_{12n} + V_{13n} + \int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} o(h_n)\kappa_M(\gamma)M_k(\gamma)d\gamma \\
&= V_{11n} + V_{12n} + V_{13n} + o(h_n),
\end{aligned}
$$

where

$$
\begin{aligned}
V_{11n} &= \int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} (\kappa_M(\frac{\pi-\pi(P,W)}{h_n}))^2 F_f(\pi(P,W),P,W)d(P,W); \\
V_{12n} &= 2\int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} F_f(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\kappa_M(\gamma)d\gamma d(P,W); \\
V_{13n} &= -2h_n\int_{\pi^{-1}([0,\pi)\cup(\pi,\pi(p,w)])} f(\pi,P,W)\int_{\frac{\pi-\pi(P,W)}{h_n}}^{\frac{\pi}{h_n}} M_k(\gamma)\gamma\kappa_M(\gamma)d\gamma d(P,W).
\end{aligned}
$$

Using the same argument as in the proof of part (a),

$$
\begin{aligned}
V_{11n} &\to \int_{\pi^{-1}([0,\pi))}\int_{[0,\pi]} f(\Pi,P,W)d\Pi d(P,W); \\
V_{12n} &\to 2\int_{\pi^{-1}((\pi,\pi(p,w)])} F_f(\pi,P,W)\int_{-B_M}^{B_M} M_k(\gamma)\kappa_M(\gamma)d\gamma d(P,W).
\end{aligned}
$$

Note that,

$$
\begin{aligned}
\int_{-B_M}^{B_M} M_k(\gamma)\kappa_M(\gamma)d\gamma &= \int_{-B_M}^{B_M}\kappa_M(\gamma)d\kappa_M(\gamma) \\
&= \kappa_M(\gamma)^2|_{-B_M}^{B_M} - \int_{-B_M}^{B_M}\kappa_M(\gamma)d\kappa_M(\gamma) \\
&= 1 - \int_{-B_M}^{B_M}\kappa_M(\gamma)d\kappa_M(\gamma).
\end{aligned}
$$

As a result, $\int_{-B_M}^{B_M} M_k(\gamma)\kappa_M(\gamma)d\gamma = 1/2$. Therefore,

$$
V_{12n} \to \int_{\pi^{-1}((\pi,\pi(p,w)])}\int_{[0,\pi]} f(\Pi,P,W)d\Pi d(P,W).
$$

Similarly,

$$V_{13n} \to 2h_n\sigma_\kappa \int_{\pi^{-1}((\pi,\pi(p,w)])} f(\Pi,P,W)d\Pi d(P,W).$$

The result then follows by combining above results. Case (2) and (3) follow similarly. □

Lemma 4. *Let h_n be a sequence of nonstochastic bandwidths such that $0 < h_n \to 0$ as $n \to \infty$, Given $w \in \mathbb{R}^{d_2}_{++}$, $p \in \mathbb{R}^{d_1}_{++}$ and there exist $N(p,w)$ such that when $n > N(p,w)$, $\mathcal{P}\{\Pi < h_n B_M\} = 0$. Under Assumptions 1-4 along with Assumption 5B (or 5A) and if $H_{2k}(\pi,p,w)$, $F_f(\pi,p,w)$ and $\varepsilon_{2k}(\pi,p,w)$ are bounded for all $(\pi,p,w) \in \Psi^*$, then we have*

(a) $\sup_{\pi\in[0,\pi(p,w)]} |\hat{P}(\pi,p,w) - E(\hat{P}(\pi,p,w))| = o_p(1)$*; and*

(b) $\sup_{\pi\in[0,\pi(p,w)]} |E(\hat{P}(\pi,p,w)) - P(\pi,p,w))| = o(1)$.

Proof. (a): Since $[0,\pi(p,w)]$ is compact, there exist $\pi_0 \in [0,\pi(p,w)]$ and r_π such that $[0,\pi(p,w)] \subset B(\pi_0,r_\pi)$ where $B(\pi_0,r_\pi) = \{\pi \in \mathbb{R} : |\pi - \pi_0| < r_\pi\}$. Furthermore, for all $\pi \in [0,\pi(p,w)]$,

$$[0,\pi(p,w)] \subset \cup_{\{\pi:\pi\in[0,\pi(p,w)]\}} B(\pi, n^{-\frac{1}{2}})$$

By the Heine-Borel Theorem, there exists $\{B(\pi_l, n^{-\frac{1}{2}})\}_{l=1}^{L_n}$ such that

$$[0,\pi(p,w)] \subset \cup_{l=1}^{L_n} B(\pi_l, n^{-\frac{1}{2}})$$

with $L_n < r_\pi n^{\frac{1}{2}}$. Therefore, any $\pi \in [0,\pi(p,w)]$, there exists some $l \in \{1...L_n\}$, such that $\pi \in B(\pi_l, n^{-\frac{1}{2}})$. Then we have

$$\begin{aligned} & |\hat{P}(\pi,p,w) - E(\hat{P}(\pi,p,w))| \\ \le\ & |\hat{P}(\pi,p,w) - \hat{P}(\pi_l,p,w)| + |\hat{P}(\pi_l,p,w) - E(\hat{P}(\pi_l,p,w))| \\ & + |E(\hat{P}(\pi_l,p,w)) - E(\hat{P}(\pi,p,w))| \\ =\ & P_{1n} + P_{2n} + P_{3n}. \end{aligned}$$

For any $\varepsilon > 0$, note that:

$$
\begin{aligned}
& \mathcal{P}\{|\hat{P}(\pi,p,w)-\hat{P}(\pi_l,p,w)|>\varepsilon\} \\
= \; & \mathcal{P}\{|(nh_n)^{-1}\sum_{i=1}^{n}\left[\int_0^{\pi} M_k(\frac{\Pi_i-\gamma}{h_n})d\gamma-\int_0^{\pi_l} M_k(\frac{\Pi_i-\gamma}{h_n})d\gamma\right] I(P_i\le p, W_i\ge w)|>\varepsilon\} \\
= \; & \mathcal{P}\{n^{-1}|\sum_{i=1}^{n}\left[\int_{-\frac{\Pi_i}{h_n}}^{\frac{\pi-\Pi_i}{h_n}} M_k(\varphi)d\varphi-\int_{-\frac{\Pi_i}{h_n}}^{\frac{\pi_l-\Pi_i}{h_n}} M_k(\varphi)d\varphi\right] I(P_i\le p, W_i\ge w)|>\varepsilon\} \\
\le \; & \mathcal{P}\{n^{-1}\sum_{i=1}^{n}|\int_{-\frac{\Pi_i}{h_n}}^{\frac{\pi-\Pi_i}{h_n}} M_k(\varphi)d\varphi-\int_{-\frac{\Pi_i}{h_n}}^{\frac{\pi_l-\Pi_i}{h_n}} M_k(\varphi)d\varphi|>\varepsilon\}.
\end{aligned}
$$

For given $w\in\mathbb{R}_{++}^{d_2}$, $p\in\mathbb{R}_{++}^{d_1}$ there exist some $N(p,w)$ such that when $n>N(p,w)$, $\mathcal{P}\{\Pi<h_nB_M\}=0$. Since Π_i has the same distribution as Π, for any i, $\mathcal{P}\{\Pi_i<h_nB_M\}=0$. Therefore,

$$
\begin{aligned}
& \mathcal{P}\{n^{-1}\sum_{i=1}^{n}|\int_{-\frac{\Pi_i}{h_n}}^{\frac{\pi-\Pi_i}{h_n}} M_k(\varphi)d\varphi-\int_{-\frac{\Pi_i}{h_n}}^{\frac{\pi_l-\Pi_i}{h_n}} M_k(\varphi)d\varphi|>\varepsilon\} \\
= \; & \mathcal{P}\{n^{-1}\sum_{i=1}^{n}|\int_{-\frac{\Pi_i}{h_n}}^{\frac{\pi-\Pi_i}{h_n}} M_k(\varphi)d\varphi-\int_{-\frac{\Pi_i}{h_n}}^{\frac{\pi_l-\Pi_i}{h_n}} M_k(\varphi)d\varphi|>\varepsilon \mid \Pi_i\ge h_nB_M,\, i=1,...,n\} \\
= \; & \mathcal{P}\{n^{-1}\sum_{i=1}^{n}|\int_{-B_M}^{\frac{\pi-\Pi_i}{h_n}} M_k(\varphi)d\varphi-\int_{-B_M}^{\frac{\pi_l-\Pi_i}{h_n}} M_k(\varphi)d\varphi|>\varepsilon\} \\
= \; & \mathcal{P}\{n^{-1}\sum_{i=1}^{n}|\kappa_M(\frac{\pi-\Pi_i}{h_n})-\kappa_M(\frac{\pi_l-\Pi_i}{h_n})|>\varepsilon\} \\
\le \; & \mathcal{P}\{n^{-1}\sum_{i=1}^{n}m_\kappa|\pi-\pi_l|>\varepsilon\} \\
= \; & \mathcal{P}\{m_\kappa|\pi-\pi_l|>\varepsilon\},
\end{aligned}
$$

by Assumption 2 and Lemma 1. As a result, for $\pi\in B(\pi_l,n^{-\frac{1}{2}})$,

$$
\lim_{n\to\infty}\mathcal{P}\{|\hat{P}(\pi,p,w)-\hat{P}(\pi_l,p,w)|>\varepsilon\}\le\lim_{n\to\infty}\mathcal{P}\{m_\kappa n^{-\frac{1}{2}}>\varepsilon\}=0.
$$

Similarly, for P_{3n},

$$
\begin{aligned}
P_{3n} &= |E(\hat{P}(\pi_l,p,w))-E(\hat{P}(\pi,p,w))| \\
&= |\int_{D_{p,w}}\int_{[0,\pi(P,W)]}\kappa_M(\frac{\pi_l-\Pi}{h_n})f(\Pi,P,W)d\Pi d(P,W) \\
&\quad -\int_{D_{p,w}}\int_{[0,\pi(P,W)]}\kappa_M(\frac{\pi-\Pi}{h_n})f(\Pi,P,W)d\Pi d(P,W)|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \int_{D_{p,w}} \int_{[0,\pi(P,W)]} |\kappa_M(\frac{\pi_l - \Pi}{h_n}) - \kappa_M(\frac{\pi - \Pi}{h_n})| f(\Pi, P, W) d\Pi d(P, W) \\
&\leq m_\kappa n^{-\frac{1}{2}} \mathcal{P}\{P \leq p, W \geq w, \Pi \leq \pi(P, W)\} \\
&\leq m_\kappa n^{-\frac{1}{2}}.
\end{aligned}
$$

Given $n \to \infty$, we have $P_{1n} = o_p(1)$ and $P_{3n} = o(1)$. For any $l \in \{1, ...L_n\}$,

$$
\begin{aligned}
P_{2n} &= |\hat{P}(\pi_l, p, w) - E(\hat{P}(\pi_l, p, w))| \\
&\leq \max_{1 \leq l \leq L_n} |\hat{P}(\pi_l, p, w) - E(\hat{P}(\pi_l, p, w))|.
\end{aligned}
$$

We need to show that for any $\varepsilon > 0$, there exists some $\Delta_\varepsilon > 0$, such that

$$
\mathcal{P}\{(\frac{n}{\ln(n)})^{\frac{1}{2}} \max_{1 \leq l \leq L_n} |\hat{P}(\pi_l, p, w) - E(\hat{P}(\pi_l, p, w))| \geq \Delta_\varepsilon\} < \varepsilon.
$$

Note that

$$
\begin{aligned}
&\mathcal{P}\{(\frac{n}{\ln(n)})^{\frac{1}{2}} \max_{1 \leq l \leq L_n} |\hat{P}(\pi_l, p, w) - E(\hat{P}(\pi_l, p, w))| \geq \Delta_\varepsilon\} \\
\leq &\sum_{l=1}^{L_n} \mathcal{P}\{(\frac{n}{\ln(n)})^{\frac{1}{2}} |\hat{P}(\pi_l, p, w) - E(\hat{P}(\pi_l, p, w))| \geq \Delta_\varepsilon\}.
\end{aligned}
$$

Write $|\hat{P}(\pi_l, p, w) - E(\hat{P}(\pi_l, p, w))| = |\frac{1}{n}\sum_{i=1}^n W_{in}|$ where

$$
W_{in} = \kappa_M(\frac{\pi_l - \Pi_i}{h_n}) I(P_i \leq p, W_i \geq w) - E[\kappa_M(\frac{\pi_l - \Pi_i}{h_n}) I(P_i \leq p, W_i \geq w)].
$$

Obviously, $E(W_{in}) = 0$, $|W_{in}| \leq 2$ since both $I(.)$ and $\kappa_M(.)$ are less or equal to one. By Bernstein's inequality we have

$$
\mathcal{P}\{(\frac{n}{\ln(n)})^{\frac{1}{2}} |\hat{P}(\pi_l, p, w) - E(\hat{P}(\pi_l, p, w))| \geq \Delta_\varepsilon\} < 2\exp(-\frac{n\Delta_\varepsilon^2 \cdot (\frac{n}{\ln(n)})^{-1}}{2\bar{\sigma}_n^2 + \frac{4}{3}\Delta_\varepsilon \cdot (\frac{n}{\ln(n)})^{-\frac{1}{2}}}),
$$

with $\bar{\sigma}_n^2 = n^{-1}\sum_{i=1}^n V(W_{in}) \to P(\pi_l, p, w)(1 - P(\pi_l, p, w))$ by Lemma 3. Thus $2\bar{\sigma}_n^2 + \frac{4}{3}\Delta_\varepsilon \cdot (\frac{n}{\ln(n)})^{-\frac{1}{2}} \to 2P(\pi_l, p, w)(1 - P(\pi_l, p, w))$, Hence provided that $\Delta_\varepsilon^2 > 2P(\pi_l, p, w)(1 - P(\pi_l, p, w))$,

$$
\begin{aligned}
&L_n \mathcal{P}\{(\frac{n}{\ln(n)})^{\frac{1}{2}} |\hat{P}(\pi_l, p, w) - E(\hat{P}(\pi_l, p, w))| \geq \Delta_\varepsilon\} \\
< \; &r_\pi n^{\frac{1}{2}} \cdot 2\exp(-\ln(n)) \\
= \; &2r_\pi n^{-\frac{1}{2}}.
\end{aligned}
$$

Therefore, $P_{2n} = o_p(1)$ and as a result, $\sup_{\pi\in[0,\pi(p,w)]} |\hat{P}(\pi,p,w) - E(\hat{P}(\pi,p,w))| = o_p(1)$.

(b) Note that for $\pi \in [0, \pi(p,w)]$,

$$\begin{aligned}
E(\hat{P}(\pi,p,w)) &= \int_{D_{p,w}} \int_{[0,\pi(P,W)]} \kappa_M(\frac{\pi-\Pi}{h_n}) f(\Pi,P,W) d\Pi d(P,W) \\
&= \int_{\pi^{-1}([0,\pi))} \int_{[0,\pi(P,W)]} \kappa_M(\frac{\pi-\Pi}{h_n}) f(\Pi,P,W) d\Pi d(P,W) \\
&\quad + \int_{\pi^{-1}(\{\pi\})} \int_{[0,\pi(P,W)]} \kappa_M(\frac{\pi-\Pi}{h_n}) f(\Pi,P,W) d\Pi d(P,W) \\
&\quad + \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{[0,\pi]} \kappa_M(\frac{\pi-\Pi}{h_n}) f(\Pi,P,W) d\Pi d(P,W) \\
&\quad + \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{[\pi,\pi(P,W)]} \kappa_M(\frac{\pi-\Pi}{h_n}) f(\Pi,P,W) d\Pi d(P,W).
\end{aligned}$$

Therefore, by triangular inequality, we have

$$\begin{aligned}
&\sup_{\pi\in[0,\pi(p,w)]} |E(\hat{P}(\pi,p,w) - P(\pi,p,w))| \\
= &\sup_{\pi\in[0,\pi(p,w)]} G_{1n} + \sup_{\pi\in[0,\pi(p,w)]} G_{2n} + \sup_{\pi\in[0,\pi(p,w)]} G_{3n} + \sup_{\pi\in[0,\pi(p,w)]} G_{4n},
\end{aligned}$$

where

$$\begin{aligned}
G_{1n} &= |\int_{\pi^{-1}([0,\pi))} \int_{[0,\pi(P,W)]} \kappa_M(\frac{\pi-\Pi}{h_n}) f(\Pi,P,W) d\Pi d(P,W) \\
&\quad - \int_{\pi^{-1}([0,\pi))} \int_{[0,\pi(P,W)]} f(\Pi,P,W) d\Pi d(P,W)|;
\end{aligned}$$

$$\begin{aligned}
G_{2n} &= |\int_{\pi^{-1}(\{\pi\})} \int_{[0,\pi(P,W)]} \kappa_M(\frac{\pi-\Pi}{h_n}) f(\Pi,P,W) d\Pi d(P,W) \\
&\quad - \int_{\pi^{-1}(\{\pi\})} \int_{[0,\pi(P,W)]} f(\Pi,P,W) d\Pi d(P,W)|; \\
G_{3n} &= |\int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{[0,\pi]} \kappa_M(\frac{\pi-\Pi}{h_n}) f(\Pi,P,W) d\Pi d(P,W) \\
&\quad - \int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{[0,\pi]} f(\Pi,P,W) d\Pi d(P,W)|; \\
G_{4n} &= |\int_{\pi^{-1}((\pi,\pi(p,w)])} \int_{[\pi,\pi(P,W)]} \kappa_M(\frac{\pi-\Pi}{h_n}) f(\Pi,P,W) d\Pi d(P,W)|.
\end{aligned}$$

For the first term, when $(P,W) \in \pi^{-1}([0,\pi))$, $\Pi \leq \pi(P,W) < \pi$. This implies $\kappa_M(\frac{\pi-\Pi}{h_n}) \to 1$ as $n \to \infty$. First, by LDC,

$$\int_{\pi^{-1}([0,\pi))} \int_{[0,\pi(P,W)]} \kappa_M(\frac{\pi-\Pi}{h_n}) f(\Pi,P,W) d\Pi d(P,W) \to \int_{\pi^{-1}([0,\pi))} \int_{[0,\pi(P,W)]} f(\Pi,P,W) d\Pi d(P,W).$$

Second, $\int_{\pi^{-1}([0,\pi))}\int_{[0,\pi(P,W)]}\kappa_M(\frac{\pi-\Pi}{h_n})f(\Pi,P,W)d\Pi d(P,W)$ is increasing with n. Furthermore, By the Lipschitz condition imposed on $\kappa_M(.)$,

$\int_{\pi^{-1}([0,\pi))}\int_{[0,\pi(P,W)]}\kappa_M(\frac{\pi-\Pi}{h_n})f(\Pi,P,W)d\Pi d(P,W)$ is a continuous function in π. As a result, by Dini's Theorem,

$$\int_{\pi^{-1}([0,\pi))}\int_{[0,\pi(P,W)]}\kappa_M(\frac{\pi-\Pi}{h_n})f(\Pi,P,W)d\Pi d(P,W)\to\int_{\pi^{-1}([0,\pi))}\int_{[0,\pi(P,W)]}f(\Pi,P,W)d\Pi d(P,W)$$

uniformly. Thus, $\sup_{\pi\in[0,\pi(p,w)]}G_{1n}=o(1)$. Similarly, we can prove that $\sup_{\pi\in[0,\pi(p,w)]}G_{2n}=o(1)$ and $\sup_{\pi\in[0,\pi(p,w)]}G_{3n}=o(1)$. For the last term, note when $\Pi\in[\pi,\pi(P,W)]$, $\kappa_M(\frac{\pi-\Pi}{h_n})\to 0$ Similarly, by LDC and Dini's theorem, $\sup_{\pi\in[0,\pi(p,w)]}G_{4n}=o(1)$. □

A.2 Proof of Propositions

Proposition 1 *Proof.* For any $(\pi,p,w)\in\Psi^*$, if $\pi<\pi_\alpha(p,w)=\inf\{\pi\in[0,B_\pi]:F(\pi|C_{p,w})\geq\alpha\}$, then $\pi\notin\{\pi\in[0,B_\pi]:F(\pi|C_{p,w})\geq\alpha\}$. That is, $F(\pi|C_{p,w})<\alpha$. If $\pi>\pi_\alpha(p,w)$, there exist some $\delta>0$ such that $\pi>\pi_\alpha(p,w)+\delta$. By the definition of $\pi_\alpha(p,w)$, for any $\delta>0$, there exist some $\pi_0\in\{\pi\in[0,B_\pi]:F(\pi|C_{p,w})\geq\alpha\}$ such that $\pi_0<\pi_\alpha(p,w)+\delta$. By the strict monotonicity of $F(\cdot|C_{p,w})$, $F(\pi|C_{p,w})>F(\pi_\alpha(p,w)+\delta|C_{p,w})>F(\pi_0|C_{p,w})\geq\alpha$. The result then follows. □

Proposition 2 *Proof.* From its definition, $\{\pi_\alpha(p,w)\}_{0\leq\alpha\leq1}$ is monotonically non-decreasing in α. The first result follows immediately by the fact $\sup_{0\leq\alpha\leq1}\{\pi_\alpha(p,w)\}=\pi(p,w)$. Let Φ be a compact set interior to the support of marginal distribution of (P,W). Define $\phi_n(p,w)=\pi_{1-\frac{1}{n}}(p,w)$. Since $\{\pi_\alpha(p,w)\}_{0\leq\alpha\leq1}$ is monotone nondecreasing in α, for any $n\in\mathbb{N}$, $\phi_n(p,w)\leq\phi_{n+1}(p,w)$, with $\lim_{n\to\infty}\phi_n(p,w)=\pi(p,w)$ pointwise. By Dini's Theorem, $\sup_{(p,w)\in\Phi}|\phi_n(p,w)-\pi(p,w)|\to 0$. Thus, for any $\varepsilon>0$, there exist some N such that when $n>N$, $\sup_{(p,w)\in\Phi}|\phi_n(p,w)-\pi(p,w)|<\varepsilon$. That is, there exist $\delta=1-\frac{1}{N}$ such that when $|\alpha-1|<\delta$, $\sup_{(p,w)\in\Phi}|\pi_\alpha(p,w)-\pi(p,w)|<\varepsilon$. □

Proposition 3 *Proof.* (i) For any $\pi_0\in[0,B_\pi]$, let $|\pi-\pi_0|<\delta$ for some $\delta>0$. Then,

$$\begin{aligned}
& |\hat{P}(\pi,p,w)-\hat{P}(\pi_0,p,w)| \\
= \ & |(nh_n)^{-1}\sum_{i=1}^{n}(\int_0^{\pi} M_k(\frac{\Pi_i-\gamma}{h_n})d\gamma-\int_0^{\pi_0} M_k(\frac{\Pi_i-\gamma}{h_n})d\gamma)I(P_i\le p, W_i\ge w)| \\
\le \ & (nh_n)^{-1}\sum_{i=1}^{n}\int_{\pi_0}^{\pi}|M_k(\frac{\Pi_i-\gamma}{h_n})|d\gamma \\
\le \ & (nh_n)^{-1}\sup_{\varphi\in[-B_M,B_M]} M_k(\varphi)\sum_{i=1}^{n}|\pi-\pi_0| \\
\le \ & {h_n}^{-1}\delta\cdot\sup_{\varphi\in[-B_M,B_M]} M_k(\varphi) \\
< \ & \varepsilon,
\end{aligned}$$

for any $\varepsilon > 0$ with a sufficiently small δ. (ii) follows directly from (i). For (iii) we need only prove that for any (p, w), there exists some $N(p, w)$ such that for all $n > N(p, w)$,

$$ {h_n}^{-1}\lim_{\pi\to\infty}\int_0^{\pi} M_k(\frac{\Pi_i-\gamma}{h_n})d\gamma = 1.$$

Now, note that

$$ {h_n}^{-1}\lim_{\pi\to\infty}\int_0^{\pi} M_k(\frac{\Pi_i-\gamma}{h_n})d\gamma = \lim_{\pi\to\infty}\int_{\frac{\Pi_i-\pi}{h_n}}^{\frac{\Pi_i}{h_n}} M_k(\varphi)d\varphi.$$

Since $h_n \to 0$ as $n \to \infty$, there exists $N(p, w)$ for any (p, w), such that for all $n > N(p, w)$, $\frac{\Pi_i}{h_n} > B_M$ and $\frac{\Pi_i-\pi}{h_n} < -B_M$. The result follows from the fact that M_k integrates to 1. □

A.3 Proof of Theorems

Theorem 1 *Proof.* First we consider the event $A = \{\omega : |\pi_{\alpha,n}(p, w) - \pi_\alpha(p, w)| > \varepsilon\}$. Given (p, w), provided that $\pi_\alpha(p, w)$ is unique, for any $\varepsilon > 0$, we have $F(\pi_\alpha(p, w)+\varepsilon|C_{p,w}) > F(\pi_\alpha(p, w)|C_{p,w}) > F(\pi_\alpha(p, w) - \varepsilon|C_{p,w})$. For $\omega \in A = \{\omega : |\pi_{\alpha,n}(p, w) - \pi_\alpha(p, w)| > \varepsilon\}$, $\pi_{\alpha,n}(p, w) > \pi_\alpha(p, w) + \varepsilon$ or $\pi_{\alpha,n}(p, w) < \pi_\alpha(p, w) - \varepsilon$. By the strict monotonicity of $F(.|C_{p,w})$, we have $F(\pi_{\alpha,n}(p, w)|C_{p,w}) > F(\pi_\alpha(p, w) + \varepsilon|C_{p,w})$ or $F(\pi_{\alpha,n}(p, w)|C_{p,w}) < F(\pi_\alpha(p, w) - \varepsilon|C_{p,w})$. Let

$$\begin{aligned}
\delta(\varepsilon,p,w) \ &= \ \min\{F(\pi_\alpha(p,w)+\varepsilon|C_{p,w})-F(\pi_\alpha(p,w)|C_{p,w}), F(\pi_\alpha(p,w)|C_{p,w}) \\
& \quad -F(\pi_\alpha(p,w)-\varepsilon|C_{p,w})\} \\
&> \ 0.
\end{aligned}$$

For all $\omega \in A$,

(1) when $F(\pi_{\alpha,n}(p,w)|C_{p,w}) - F(\pi_\alpha(p,w)|C_{p,w}) > 0$, we have $\pi_{\alpha,n}(p,w) > \pi_\alpha(p,w) + \varepsilon$. By monotonicity,

$$F(\pi_{\alpha,n}(p,w)|C_{p,w}) - F(\pi_\alpha(p,w)|C_{p,w}) > F(\pi_\alpha(p,w)+\varepsilon|C_{p,w}) - F(\pi_\alpha(p,w)|C_{p,w}) \geq \delta(\varepsilon,p,w).$$

(2) Similarly, when $F(\pi_{\alpha,n}(p,w)|C_{p,w}) - F(\pi_\alpha(p,w)|C_{p,w}) < 0$, we have

$$F(\pi_{\alpha,n}(p,w)|C_{p,w}) - F(\pi_\alpha(p,w)|C_{p,w}) < F(\pi_\alpha(p,w)-\varepsilon|C_{p,w}) - F(\pi_\alpha(p,w)|C_{p,w}) \leq -\delta(\varepsilon,p,w).$$

As a result, For $\omega \in A$, $|F(\pi_{\alpha,n}(p,w)|C_{p,w}) - F(\pi_\alpha(p,w)|C_{p,w})| > \delta(\varepsilon,p,w)$. i.e., $A \subseteq B = \{\omega : |F(\pi_{\alpha,n}(p,w)|C_{p,w}) - F(\pi_\alpha(p,w)|C_{p,w})| > \delta(\varepsilon,p,w)\}$. Thus, $\mathcal{P}(A) \leq \mathcal{P}(B)$. Therefore, we just need to prove $|F(\pi_{\alpha,n}(p,w)|C_{p,w}) - F(\pi_\alpha(p,w)|C_{p,w})| = o_p(1)$.

$$\begin{aligned}
& |F(\pi_{\alpha,n}(p,w)|C_{p,w}) - F(\pi_\alpha(p,w)|C_{p,w})| \\
= \;& |F(\pi_{\alpha,n}(p,w)|C_{p,w}) - \hat{F}(\pi_{\alpha,n}(p,w)|C_{p,w})| \\
\leq \;& \sup_{\pi\in\mathbb{R}_+} |F(\pi|C_{p,w}) - \hat{F}(\pi|C_{p,w})| \\
\leq \;& \sup_{\pi\in\mathbb{R}_+} \left|\frac{P(\pi,p,w)}{P_{PW}(p,w)} - \frac{\hat{P}(\pi,p,w)}{\hat{P}_{PW}(p,w)}\right| \\
\leq \;& \sup_{\pi\in\mathbb{R}_+} \left|\frac{P(\pi,p,w)}{P_{PW}(p,w)} - \frac{P(\pi,p,w)}{\hat{P}_{PW}(p,w)}\right| + \sup_{\pi\in\mathbb{R}_+} \left|\frac{P(\pi,p,w)}{\hat{P}_{PW}(p,w)} - \frac{\hat{P}(\pi,p,w)}{\hat{P}_{PW}(p,w)}\right| \\
\leq \;& \sup_{\pi\in\mathbb{R}_+} P(\pi,p,w)\left|\frac{1}{P_{PW}(p,w)} - \frac{1}{\hat{P}_{PW}(p,w)}\right| + \left|\frac{1}{\hat{P}_{PW}(p,w)}\right| \sup_{\pi\in\mathbb{R}_+} |P(\pi,p,w) - \hat{P}(\pi,p,w)| \\
\leq \;& P_{PW}(p,w)\left|\frac{1}{P_{PW}(p,w)} - \frac{1}{\hat{P}_{PW}(p,w)}\right| + \left|\frac{1}{\hat{P}_{PW}(p,w)}\right| \sup_{\pi\in\mathbb{R}_+} |P(\pi,p,w) - \hat{P}(\pi,p,w)|.
\end{aligned}$$

Note that $\hat{P}_{PW}(p,w) - P_{PW}(p,w) = o_p(1)$ by the properties of indicator function. By Slutsky theorem we have $\frac{1}{P_{PW}(p,w)} - \frac{1}{\hat{P}_{PW}(p,w)} = o_p(1)$. Since $\hat{P}_{PW}(p,w) = O_p(1)$, we just need to prove $\sup_{\pi\in\mathbb{R}_+} |P(\pi,p,w) - \hat{P}(\pi,p,w)| = o_p(1)$.

$$\begin{aligned}
& \sup_{\pi\in\mathbb{R}_+} |P(\pi,p,w) - \hat{P}(\pi,p,w)| \\
\leq \;& \sup_{\pi\in[0,\pi(p,w)]} |P(\pi,p,w) - \hat{P}(\pi,p,w)| + \sup_{\pi\in(\pi(p,w),\infty)} |P(\pi,p,w) - \hat{P}(\pi,p,w)|.
\end{aligned}$$

From Lemma 4, $\sup_{\pi\in[0,\pi(p,w)]}|P(\pi,p,w)-\hat{P}(\pi,p,w)| = o_p(1)$. For all $\pi \in (\pi(p,w),\infty)$,

$$\begin{aligned}
P(\pi,p,w) &= \mathcal{P}(\Pi\leq\pi, P\leq p, W\geq w)\\
&= \mathcal{P}(\Pi\leq\pi(p,w), P\leq p, W\geq w)\\
&= \mathcal{P}(P\leq p, W\geq w)\\
&= P_{PW}(p,w).
\end{aligned}$$

Given $\min_{\{i:P_i\leq p,W_i\geq w\}}\Pi_i \geq h_nB_M$, and for any i, $\Pi_i\leq\pi(p,w)<\pi$. There exist $N(p,w)$ such that for all $n>N(p,w)$,

$$\begin{aligned}
\hat{P}(\pi,p,w) &= (nh_n)^{-1}\sum_{i=1}^{n}(\int_0^{\pi}M_k(\frac{\Pi_i-\gamma}{h_n})d\gamma)I(P_i\leq p,W_i\geq w)\\
&= n^{-1}\sum_{i=1}^{n}\int_{-\frac{\Pi_i}{h_n}}^{\frac{\pi-\Pi_i}{h_n}}M_k(\varphi)d\varphi I(P_i\leq p,W_i\geq w)\\
&= n^{-1}\sum_{i=1}^{n}\int_{-B_M}^{B_M}M_k(\varphi)d\varphi I(P_i\leq p,W_i\geq w)\\
&= n^{-1}\sum_{i=1}^{n}I(P_i\leq p,W_i\geq w)\\
&= \hat{P}_{PW}(p,w).
\end{aligned}$$

As a result, as $\hat{P}_{PW}(p,w)\to P_{PW}(p,w)$ as $n\to\infty$. By triangular inequality,

$$\begin{aligned}
&\sup_{\pi\in(\pi(p,w),\infty)}|P(\pi,p,w)-\hat{P}(\pi,p,w)|\\
\leq\ &\sup_{\pi\in(\pi(p,w),\infty)}|P(\pi,p,w)-P_{PW}(p,w)|+\sup_{\pi\in(\pi(p,w),\infty)}|\hat{P}(\pi,p,w)-P_{PW}(p,w)|\\
=\ &o_p(1)
\end{aligned}$$

The result then follows. □

Theorem 2 *Proof.* (i) By Mean Value Theorem,

$$\begin{aligned}
\pi_{\alpha,n}(p,w)-\pi_\alpha(p,w) &= \frac{\hat{F}(\pi_{\alpha,n}(p,w)|C_{p,w})-\hat{F}(\pi_\alpha(p,w)|C_{p,w})}{\hat{f}(\bar{\pi}_{\alpha,n}(p,w)|C_{p,w})}\\
&= \frac{F(\pi_\alpha(p,w)|C_{p,w})-\hat{F}(\pi_\alpha(p,w)|C_{p,w})}{\hat{f}(\bar{\pi}_{\alpha,n}(p,w)|C_{p,w})},
\end{aligned}$$

where $\hat{f}(\pi|C_{p,w}) = \frac{\partial \hat{F}(\pi|C_{p,w})}{\partial \pi}$ and $\bar{\pi}_{\alpha,n}(p,w) = \lambda\pi_{\alpha,n}(p,w) + (1-\lambda)\pi_\alpha(p,w)$ for some $\lambda \in (0,1)$. Write

$$\pi_{\alpha,n}(p,w) - \pi_\alpha(p,w) = (A_n + C_n)(\frac{1}{f(\pi_\alpha(p,w)|C_{p,w})} + \beta_n),$$

where

$$\begin{aligned} A_n &= F(\pi_\alpha(p,w)|C_{p,w}) - \frac{E(\hat{P}(\pi_\alpha(p,w),p,w))}{E(\hat{P}_{PW}(p,w))}; \\ C_n &= \frac{E(\hat{P}(\pi_\alpha(p,w),p,w))}{E(\hat{P}_{PW}(p,w))} - \hat{F}(\pi_\alpha(p,w)|C_{p,w}); \\ \beta_n &= \frac{1}{\hat{f}(\bar{\pi}_{\alpha,n}(p,w)|C_{p,w})} - \frac{1}{f(\pi_\alpha(p,w)|C_{p,w})}. \end{aligned}$$

The theorem follows if (a) $\beta_n = o_p(1)$; (b) $A_n = -\frac{1}{2}h_n^2\sigma_{M_k}^2 \frac{\int_{\pi^{-1}([\pi_\alpha(p,w),\pi(p,w)])} f^{(1)}(\pi_\alpha(p,w),P,W)d(P,W)}{P_{PW}(p,w)} + o(h_n^2)$; (c) $(\frac{s_n(p,w)}{\hat{P}_{PW}(p,w)})^{-1}\sqrt{n}C_n \to N(0,1)$ where

$$\begin{aligned} s_n^2(p,w) &= P(\pi_\alpha(p,w),p,w) - \frac{(P(\pi_\alpha(p,w),p,w))^2}{P_{PW}(p,w)} \\ &\quad -2h_n\sigma_\kappa \int_{\pi^{-1}([\pi_\alpha(p,w),\pi(p,w)])} f(\pi_\alpha(p,w),P,W)d(P,W) + o(h_n). \end{aligned}$$

(a) By Slutsky theorem, it is suffice to prove $\hat{f}(\bar{\pi}_{\alpha,n}(p,w)|C_{p,w}) - f(\pi_\alpha(p,w)|C_{p,w}) = o_p(1)$. Since $\pi_{\alpha,n}(p,w) - \pi_\alpha(p,w) = o_p(1)$ by theorem 1, also, $\bar{\pi}_{\alpha,n}(p,w) - \pi_\alpha(p,w) = o_p(1)$.

$$\begin{aligned} & |\hat{f}(\bar{\pi}_{\alpha,n}(p,w)|C_{p,w}) - f(\pi_\alpha(p,w)|C_{p,w})| \\ \leq\ & |\hat{f}(\bar{\pi}_{\alpha,n}(p,w)|C_{p,w}) - f(\bar{\pi}_\alpha(p,w)|C_{p,w})| + |f(\bar{\pi}_{\alpha,n}(p,w)|C_{p,w}) - f(\pi_\alpha(p,w)|C_{p,w})| \\ \leq\ & |\hat{f}(\bar{\pi}_{\alpha,n}(p,w)|C_{p,w}) - f(\bar{\pi}_\alpha(p,w)|C_{p,w})| + o_p(1) \end{aligned}$$

by continuity of f. Therefore it is suffice to prove $\sup_{\pi\in G} |\hat{f}(\pi|C_{p,w}) - f(\pi|C_{p,w})| = o_p(1)$. where G is a compact set and $G \subset (0, \pi(p,w))$.

When $(P,W) \in \pi^{-1}([0,\pi])$, $\Pi \leq \pi(P,W) \leq \pi$. $F(\pi|C_{p,w}) = 1$ and $\frac{\partial F(\pi|C_{p,w})}{\partial \pi} = 0$, Therefore,

$$f(\pi|C_{p,w}) = \frac{\int_{\pi^{-1}((\pi,\pi(p,w)])} f(\pi,P,W)d(P,W)}{P_{PW}(p,w)}.$$

$$
\begin{aligned}
& \sup_{\pi\in G}|\hat{f}(\pi|C_{p,w})-f(\pi|C_{p,w})| \\
= \ & \sup_{\pi\in G}|\frac{(nh_n)^{-1}\sum_{i=1}^n M_k(\frac{\Pi_i-\pi}{h_n})I(P_i\leq p,W_i\geq w)}{\hat{P}_{PW}(p,w)}-\frac{\int_{\pi^{-1}((\pi,\pi(p,w)])}f(\pi,P,W)d(P,W)}{P_{PW}(p,w)}| \\
\leq \ & \frac{1}{\hat{P}_{PW}(p,w)}\sup_{\pi\in G}|(nh_n)^{-1}\sum_{i=1}^n M_k(\frac{\Pi_i-\pi}{h_n})I(P_i\leq p,W_i\geq w) \\
- \ & \int_{\pi^{-1}((\pi,\pi(p,w)])}f(\pi,P,W)d(P,W)| \\
& +|\frac{1}{P_{PW}(p,w)}-\frac{1}{\hat{P}_{PW}(p,w)}|\sup_{\pi\in G}\int_{\pi^{-1}((\pi,\pi(p,w)])}f(\pi,P,W)d(P,W).
\end{aligned}
$$

Since $\frac{1}{P_{PW}(p,w)}-\frac{1}{\hat{P}_{PW}(p,w)}=o_p(1)$ by Slutsky theorem,

$$
\sup_{\pi\in G}\int_{\pi^{-1}((\pi,\pi(p,w)])}f(\pi,P,W)d(P,W)\leq B_f\int_{\pi^{-1}((\pi,\pi(p,w)])}d(P,W)=O(1)
$$

by Assumptions 3 and 4.

Denote $Q_n(p,w)=(nh_n)^{-1}\sum_{i=1}^n M_k(\frac{\Pi_i-\pi}{h_n})I(P_i\leq p,W_i\geq w)$, Thus,

$$
\begin{aligned}
& \sup_{\pi\in G}|Q_n(p,w)-\int_{\pi^{-1}((\pi,\pi(p,w)])}f(\pi,P,W)d(P,W)| \\
\leq \ & \sup_{\pi\in G}|Q_n(p,w)-E(Q_n(p,w))| \\
& +\sup_{\pi\in G}|E(Q_n(p,w))-\int_{D_{p,w}}\kappa_M(\frac{\pi(P,W)-\pi}{h_n})f(\pi,P,W)d(P,W)| \\
& +\sup_{\pi\in G}|\int_{\pi^{-1}((\pi,\pi(p,w)])}\kappa_M(\frac{\pi(P,W)-\pi}{h_n})f(\pi,P,W)d(P,W) \\
- \ & \int_{\pi^{-1}((\pi,\pi(p,w)])}f(\pi,P,W)d(P,W)| \\
& +\sup_{\pi\in G}|\int_{\pi^{-1}([0,\pi])}\kappa_M(\frac{\pi(P,W)-\pi}{h_n})f(\pi,P,W)d(P,W)| \\
= \ & Q_{1n}+Q_{2n}+Q_{3n}+Q_{4n}.
\end{aligned}
$$

Follow the similar proof process as in Lemma 4 (a), we can prove that $Q_{1n}=O_p((\frac{\ln n}{nh_n})^{\frac{1}{2}})$ if $nh_n^2\to\infty$. For any (p,w), there exist some $N(p,w)$ such that when $n>N(p,w)$

$$
\begin{aligned}
E(Q_n(p,w)) &= {h_n}^{-1}\int_{D_{p,w}}\int_{[0,\pi(P,W)]}M_k(\frac{\Pi-\pi}{h_n})f(\Pi,P,W)d\Pi d(P,W) \\
&= \int_{D_{p,w}}\int_{-\frac{\pi}{h_n}}^{\frac{\pi(P,W)-\pi}{h_n}}M_k(\varphi)f(\pi+h_n\varphi,P,W)d\varphi d(P,W) \\
&= \int_{D_{p,w}}\int_{-B_M}^{\frac{\pi(P,W)-\pi}{h_n}}M_k(\varphi)f(\pi+h_n\varphi,P,W)d\varphi d(P,W).
\end{aligned}
$$

By Taylor's theorem, for any $\pi \in G$,

$$\begin{aligned}
& |\int_{D_{p,w}} \int_{-B_M}^{\frac{\pi(P,W)-\pi}{h_n}} M_k(\varphi)(f(\pi + h_n\varphi, P, W) - f(\pi, P, W))d\varphi d(P,W)| \\
\leq & \int_{D_{p,w}} \int_{-B_M}^{\frac{\pi(P,W)-\pi}{h_n}} M_k(\varphi)|(f(\pi + h_n\varphi, P, W) - f(\pi, P, W))|d\varphi d(P,W) \\
\leq & \; m_f h_n \int_{D_{p,w}} \int_{-B_M}^{B_M} M_k(\varphi)|\varphi| d\varphi d(P,W) + o(h_n) \\
= & \; O(h_n).
\end{aligned}$$

Therefore, $Q_{2n} = o(1)$. Since when $(P,W) \in \pi^{-1}([0,\pi])$, $\kappa_M(\frac{\pi(P,W)-\pi}{h_n}) \to 0$ and when $(P,W) \in \pi^{-1}((\pi, \pi(p,w)])$, $\kappa_M(\frac{\pi(P,W)-\pi}{h_n}) \to 1$. By LDC, for any $\pi \in G$,

$$\int_{\pi^{-1}((\pi,\pi(p,w)])} \kappa_M(\frac{\pi(P,W)-\pi}{h_n}) f(\pi, P, W)d(P,W) \to \int_{\pi^{-1}((\pi,\pi(p,w)])} f(\pi, P, W)d(P,W)|,$$

and

$$\int_{\pi^{-1}([0,\pi])} \kappa_M(\frac{\pi(P,W)-\pi}{h_n}) f(\pi, P, W)d(P,W) \to 0.$$

Therefore, $Q_{3n} = o(1)$ and $Q_{4n} = o(1)$. In sum, Noting that $\frac{1}{\hat{P}_{PW}(p,w)} = O_p(1)$, we have

$$\sup_{\pi \in G} |\hat{f}(\pi|C_{p,w}) - f(\pi|C_{p,w})| = o_p(1).$$

As a result, $\beta_n = o_p(1)$.

(b):

$$\begin{aligned}
A_n &= F(\pi_\alpha(p,w)|C_{p,w}) - \frac{E(\hat{P}(\pi_\alpha(p,w),p,w))}{E(\hat{P}_{PW}(p,w))} \\
&= \frac{E(\hat{P}_{PW}(p,w))F(\pi_\alpha(p,w)|C_{p,w})}{E(\hat{P}_{PW}(p,w))} - \frac{P(\pi_\alpha(p,w),p,w)}{E(\hat{P}_{PW}(p,w))} \\
&\quad + \frac{P(\pi_\alpha(p,w),p,w)}{E(\hat{P}_{PW}(p,w))} - \frac{E(\hat{P}(\pi_\alpha(p,w),p,w))}{E(\hat{P}_{PW}(p,w))} \\
&= \frac{1}{E(\hat{P}_{PW}(p,w))}[(E(\hat{P}_{PW}(p,w))F(\pi_\alpha(p,w)|C_{p,w}) - P(\pi_\alpha(p,w),p,w)) \\
&\quad + (P(\pi_\alpha(p,w),p,w) - E(\hat{P}(\pi_\alpha(p,w),p,w)))] \\
&= \frac{1}{E(\hat{P}_{PW}(p,w))}(A_{1n} + A_{2n}).
\end{aligned}$$

We know $E(\hat{P}_{PW}(p,w)) = P_{PW}(p,w)$. Clearly $A_{1n} = 0$. Since given $\alpha \in (0,1)$, $\pi_\alpha(p,w) \in (0, \pi(p,w))$, by Lemma 3,

$$A_{2n} = -\frac{1}{2} h_n^2 \sigma_M^2 \int_{\pi^{-1}((\pi_\alpha(p,w),\pi(p,w)])} f^{(1)}(\pi_\alpha(p,w), P, W) d(P,W) + o(h_n^2).$$

The result then follows.

(c):

$$\begin{aligned}
\sqrt{n} C_n &= \sqrt{n}\Big(\frac{E(\hat{P}(\pi_\alpha(p,w),p,w))}{E(\hat{P}_{PW}(p,w))} - \hat{F}(\pi_\alpha(p,w)|C_{p,w})\Big) \\
&= \sqrt{n}\Big(\frac{E(\hat{P}(\pi_\alpha(p,w),p,w))\hat{P}_{PW}(p,w)}{E(\hat{P}_{PW}(p,w))\hat{P}_{PW}(p,w)} - \frac{\hat{P}(\pi_\alpha(p,w),p,w)}{\hat{P}_{PW}(p,w)}\Big) \\
&= \frac{1}{\hat{P}_{PW}(p,w)} \sum_{i=1}^n Z_{in},
\end{aligned}$$

where

$$Z_{in} = \frac{1}{\sqrt{n}}\Big(\frac{E(\hat{P}(\pi_\alpha(p,w),p,w))}{P_{PW}(p,w)} I(P_i \le p, W_i \ge w) - \frac{1}{h_n}\int_0^{\pi_\alpha(p,w)} M_k\Big(\frac{\Pi_i - \gamma}{h_n}\Big) d\gamma I(P_i \le p, W_i \ge w)\Big).$$

Here,

$$\begin{aligned}
E(Z_{in}) &= \frac{1}{\sqrt{n}}(E(\hat{P}(\pi_\alpha(p,w),p,w)) - E(\hat{P}(\pi_\alpha(p,w),p,w))) \\
&= 0,
\end{aligned}$$

$$\sum_{i=1}^n E(Z_{in}^2) = s_n^2(p,w) = s_{1n} + s_{2n} + s_{3n},$$

where

$$\begin{aligned}
s_{1n} &= \frac{\{E(\hat{P}(\pi_\alpha(p,w),p,w))\}^2}{P_{PW}(p,w)^2} E(I(P_i \le p, W_i \ge w)) = \frac{\{E(\hat{P}(\pi_\alpha(p,w),p,w))\}^2}{P_{PW}(p,w)}; \\
s_{2n} &= E[(\frac{1}{h_n}\int_0^{\pi_\alpha(p,w)} M_k(\frac{\Pi_i - \gamma}{h_n}) d\gamma I(P_i \le p, W_i \ge w)))^2] = E[(\hat{P}(\pi_\alpha(p,w),p,w)^2]; \\
s_{3n} &= -2\frac{E(\hat{P}(\pi_\alpha(p,w),p,w)}{P_{PW}(p,w)} E(\frac{1}{h_n}\int_0^{\pi_\alpha(p,w)} M_k(\frac{\Pi_i - \gamma}{h_n}) d\gamma I(P_i \le p, W_i \ge w)) \\
&= -2 s_{1n}.
\end{aligned}$$

By Lemma 3,

$$
\begin{aligned}
E(\hat{P}(\pi_\alpha(p,w),p,w)) &= P(\pi_\alpha(p,w),p,w) + \frac{1}{2}h_n^2\sigma_M^2 \int_{\pi^{-1}((\pi,\pi(p,w)])} f^{(1)}(\pi,P,W)d(P,W) + o(h_n^2);\\
E[(\hat{P}(\pi_\alpha(p,w),p,w)^2] &= P(\pi_\alpha(p,w),p,w) - 2h_n\sigma_\kappa \int_{\pi^{-1}((\pi_\alpha(p,w),\pi(p,w)])} f(\pi_\alpha(p,w),P,W)d(P,W)\\
&+ o(h_n).
\end{aligned}
$$

As a result,

$$
\begin{aligned}
s_{1n} &= \frac{1}{P_{PW}(p,w)}(P(\pi_\alpha(p,w),p,w) + \frac{1}{2}h_n^2\sigma_M^2 \int_{\pi^{-1}((\pi,\pi(p,w)])} f^{(1)}(\pi,P,W)d(P,W) + o(h_n^2))^2\\
&= \frac{(P(\pi_\alpha(p,w),p,w))^2}{P_{PW}(p,w)} + o(h_n);\\
s_{2n} &= P(\pi_\alpha(p,w),p,w) - 2h_n\sigma_\kappa \int_{\pi^{-1}((\pi_\alpha(p,w),\pi(p,w)])} f(\pi_\alpha(p,w),P,W)d(P,W) + o(h_n);\\
s_{3n} &= -2s_{1n} = -2\frac{(P(\pi_\alpha(p,w),p,w))^2}{P_{PW}(p,w)} + o(h_n),
\end{aligned}
$$

$$
\begin{aligned}
\sum_{i=1}^n E(Z_{in}^2) &= s_{1n} + s_{2n} + s_{3n}\\
&= P(\pi_\alpha(p,w),p,w) - \frac{(P(\pi_\alpha(p,w),p,w))^2}{P_{PW}(p,w)}\\
&\quad -2h_n\sigma_\kappa \int_{\pi^{-1}((\pi_\alpha(p,w),\pi(p,w)])} f(\pi_\alpha(p,w),P,W)d(P,W) + o(h_n).
\end{aligned}
$$

By Liapounov's CLT, $\sum_{i=1}^n \frac{Z_{in}}{s_n(p,w)} \xrightarrow{d} N(0,1)$ if $\lim_{n\to\infty}\sum_{i=1}^n E(|\frac{Z_{in}}{s_n(p,w)}|^{2+\delta}) = 0$ for some $\delta > 0$.

$$
\sum_{i=1}^n E(|\frac{Z_{in}}{s_n(p,w)}|^{2+\delta}) \le \sum_{i=1}^n E(|Z_{in}|^{2+\delta}|\frac{1}{s_n(p,w)}|^{2+\delta}).
$$

Since $s_n(p,w) = O(1)$, we just need to prove $\lim_{n\to\infty}\sum_{i=1}^n E(|Z_{in}|^{2+\delta}) = 0$. By C_r Inequality,

$$
\begin{aligned}
\sum_{i=1}^n E(|Z_{in}|^{2+\delta}) &\le 2^{1+\delta}(n^{-2/\delta}E(|\frac{E(\hat{P}(\pi_\alpha(p,w),p,w))}{P_{PW}(p,w)}I(P_i \le p, W_i \ge w)|^{2+\delta})\\
&\quad +n^{-2/\delta}E(|\frac{1}{h_n}\int_0^{\pi_\alpha(p,w)} M_k(\frac{\Pi_i - \gamma}{h_n})d\gamma I(P_i \le p, W_i \ge w)|^{2+\delta}))\\
&= 2^{1+\delta}(n^{-2/\delta}E|\frac{E(\hat{P}(\pi_\alpha(p,w),p,w))}{P_{PW}(p,w)}|^{2+\delta}E(I(P_i \le p, W_i \ge w)))\\
&\quad +n^{-2/\delta}\int_{D_{p,w}}\int_{[0,\pi(P,W)]} \kappa_M(\frac{\pi_\alpha(p,w) - \Pi}{h_n})f(\Pi,P,W)d\Pi d(P,W).
\end{aligned}
$$

Since $E(I(P_i \leq p, W_i \geq w) = O(1)$,

$$\begin{aligned} n^{-2/\delta} E|\frac{E(\hat{P}(\pi_\alpha(p,w),p,w))}{P_{PW}(p,w)}|^{2+\delta} &= n^{-2/\delta} \frac{|E(\hat{P}(\pi_\alpha(p,w),p,w))|^{2+\delta}}{P_{PW}(p,w)^{2+\delta}} \\ &= O(n^{-2/\delta}) \end{aligned}$$

Since $\kappa_M(.) \leq 1$, $f < B_f$ and $\pi \leq B_\pi$,

$$\begin{aligned} & n^{-2/\delta} \int_{D_{p,w}} \int_{[0,\pi(P,W)]} \kappa_M(\frac{\pi_\alpha(p,w) - \Pi}{h_n}) f(\Pi, P, W) d\Pi d(P,W) \\ \leq \; & n^{-2/\delta} B_f \int_{D_{p,w}} \int_{[0,\pi(P,W)]} d\Pi d(P,W) \\ \leq \; & n^{-2/\delta} B_f \int_{\pi^{-1}[0,B_\pi]} \int_{[0,B_\pi]} d\Pi d(P,W) \\ = \; & O(n^{-2/\delta}). \end{aligned}$$

The result then follows.

(ii) Note that in the proof of part (i), $A_n = \frac{1}{E(\hat{P}_{PW}(p,w))}(A_{1n} + A_{2n})$ is the bias term and $A_{1n} = 0$. by Lemma 2,

$$\begin{aligned} |A_{2n}| &= |P(\pi_\alpha(p,w),p,w) - E(\hat{P}(\pi_\alpha(p,w),p,w))| \\ &\leq c h_n^{2k} [\int_{D_{p,w}} H_{2k}(\pi_\alpha(p,w), P, W) d(P,W) \\ &\quad + \int_{D_{p,w}} \sup_{\pi \in \mathbb{R}} |F_f(\pi, P, W)| \varepsilon_{2k}^{-2k}(\pi_\alpha(p,w), P, W) d(P,W)]. \end{aligned}$$

The result then follows. $\square$