Robust Estimation of Additive Boundaries with Quantile Regression and Shape Constraints

Supplementary Material

Abstract. We consider the estimation of the boundary of a set when it is known to be sufficiently smooth, to satisfy certain shape constraints and to have an additive structure. Our proposed method is based on spline estimation of a conditional quantile regression and is resistant to outliers and/or extreme values in the data. This work is a desirable extension of Martins-Filho and Yao (2007) and Wang et al. (2020) and can also be viewed as an alternative to existing estimators that have been used in empirical analysis. The results of a Monte Carlo study show that the new method outperforms the existing methods when outliers or heterogeneity are present. Our theoretical analysis indicates that our proposed boundary estimator is uniformly consistent under a set of standard assumptions. We illustrate practical use of our method by estimating two production functions using real-world data sets.

Keywords: Polynomial spline; Robust estimation; Uniform consistency.

The supplementary document contains additional real data application of the proposed method. It also includes the relevant lemmas and detailed proofs of the theorems presented in the main paper.

U.S. high technology firm data

Data on U.S. high technology firms from Wharton Research Data Services are considered as another example. The main objective of this analysis is to evaluate the relative performance of high-tech firms in the United States. The data we consider are from 2017, and the high technology firms are defined as firms whose four-digit SIC code are in the following lists: biotechnology (2833-2836 and 8731-8734), computer (3570-3577 and 7371-7379), electronics (3600-3674), telecommunication (4810-4841) and Computer Programming, Data Process (7370) (Klobucnik and Sievers, 2013). We exclude all firms with sales greater than 5000 million dollars, which guarantees that the data are not too sparse. In addition, we only consider firms that reported sales in 2016. Eventually, this leaves us with 256 individual firms. For convenience, four input variables are considered: the number of employees (X_1), the research and development expenses (X_2), the total net property, plant and equipment (X_3) value, and operating expenses (X_4). The output variable (Y) is sales, which represents a firm's revenues. All the units are measured in millions. The relationships between output and each input variables are plotted in Figure 1. Obviously, firms' revenues increase as inputs increase. However, different from the previous example, there is no clear evidence to support concavity. Therefore, we consider an additive frontier with the monotone constraint to quantify the maximum revenue for U.S high tech firms.

In addition to MCRS and MCQS, both URS and UQS are used for comparison purposes to estimate the unknown frontier functions. The number of interior knots is set to be the integer part of $n^{1/(2p+3)}$ and the knot sequence is equally spaced in the quantile range of each input variable. Figures 2(a) and 2(b) show the estimates from the regression spline and quantile spline estimators, respectively. The circles are pseudo observations, the dashed lines correspond to the estimates with the unconstrained method, while the long-dashed line corresponds to the estimates with the monotone constrained method. It also plots the 95% point-wise confidence intervals (dotted lines) from 100 bootstrapped samples using the unconstrained method. All input variables except for the research and development expense have monotone effects on firms' revenues. Overall, the quantile method is more robust to outliers than the mean regression. Note also that as research and development expenses increase, the pseudo observations of firm revenue in the unconstrained model slightly decreases, which affects the results of a firm's efficiency estimation.

Furthermore, to assess the production efficiency of each firm, we estimate the frontier for U.S. high technology firms. In order to ensure the accuracy and robustness of our empirical analysis, we implement the outliers deletion in our frontier's estimates. Figure 3 plots both the estimated maximum revenue and the estimated kernel density of the relative efficiencies. It can be seen that the results are very similar for URS (solid), MCRS (dotted), UQS (dot-dashed) and MCQS (long-dashed). Most firms have high efficiencies with values above 0.95. As the revenue increases, the firm's production efficiency will increase correspondingly, which emphasizes the fact that the stronger the sales ability, the more efficient the asset utilization. It is important to note that the smaller efficiency observed from the MCQS method compared to the MCRS method could be due to the fact that the firm's revenue is slightly decreasing as the research and development expense increases.

In addition, we explore potential factors, or "environmental variables", that drive the firm's production efficiency. We use the long term debt ratio (DLTTR), market-to-book ratio (M-B), market value (MKVALT), age (AGE) and region (REG) as explanatory variables to examine the determinants of firm efficiency. Firm's efficiencies are obtained using the MCQS method. Table 1 reports the effects of firm other resources on the firm performance. Generally, all explanatory variables have a positive effect on firms' efficiencies. A firm with more long-term debt ratio is likely to be more efficient than its peers, which is consistent with the fact that firms with higher revenues are also predicted to have more long-term debt

(Albuquerque and Hopenhayn, 2004). We also consider the effect of with/without long term debt on firm's production efficiency, and find that firms with long term debt are more efficient in their production than the ones without long term debt (p value ≈ 0). In addition, as a firm is getting more experienced and mature, its production efficiency steadily improves (p value $= 0.6$). Neither market-to-book ratio nor market value are significantly associated with production efficiency. In the end, we use one-way ANOVA analysis to check whether there are regional (i.e., West, MidWest, South and East) differences in firm productive efficiency. However, the result shows that the difference is not significant with p-value of 0.154.

In this example, we use four variables as inputs to explain U.S. high tech firms revenue. Definitely, four inputs are not enough to explain a firm's capital structure and its productive efficiency. This may affect the results of our analysis. An extension to the case of more inputs should be accomplished for those who are interested in the study of firm efficiency.

Insert Figures 1, 2, 3 and Table 1 here.

Lemmas and Proofs

A.1 *Lemmas*

Lemma 1. Proof. Let $\|m\|_2^2 = E\left(m^2(\mathbf{X})\right)$ for any square integrable additive function $m(\mathbf{x})$. According to Theorem 1 in Horowitz and Lee (2005), $\sup_{\mathbf{x}\in[0,1]^d} |\widetilde{m}_\tau(\mathbf{x}) - m_\tau(\mathbf{x})| = O_p\left(N_n/\sqrt{n}\right)$ as $n \to \infty$, which implies that $\left\|\widetilde{m}_\tau\left(\mathbf{x}\right) - m_\tau\left(\mathbf{x}\right)\right\|_2 = O_p\left(N_n/\sqrt{n}\right)$. Furthermore, by Lemma 1 of Xue and Yang (2006), one has $\left\|\widetilde{m}_{l,\tau}(x_l) - m_{l,\tau}(x_l)\right\|_2 = O_p\left(N_n/\sqrt{n}\right)$ for $l = 1, \cdots, d$. Therefore, for each $l = 1, \cdots, d$, there exists $x_l^0 \in [0,1]$, such that $|\widetilde{m}_{l,\tau}\left(x_l^0\right) - m_{l,\tau}\left(x_l^0\right)| = O_p\left(N_n/\sqrt{n}\right)$. For any $x_l \in [0,1]$, let $\mathbf{x}^* = \left(x_1^0, \cdots, x_{l-1}^0, x_l, x_{l+1}^0, \cdots, x_d^0\right)^T \in [0,1]^d$. Then,

$$
\begin{aligned}
\sup_{x_l\in[0,1]} |\widetilde{m}_{l,\tau}(x_l) - m_{l,\tau}(x_l)| &= \sup_{x_l\in[0,1]} \left|\widetilde{m}_\tau(\mathbf{x}^*) - m_\tau(\mathbf{x}^*) - \sum_{l'\neq l}\left[\widetilde{m}_{l',\tau}(x_{l'}^0) - m_{l',\tau}(x_{l'}^0)\right]\right| \\
&\leq \sup_{\mathbf{x}^*\in[0,1]^d} \left|\widetilde{m}_\tau(\mathbf{x}^*) - m_\tau(\mathbf{x}^*)\right| + \sum_{l'\neq l}\left|\widetilde{m}_{l',\tau}(x_{l'}^0) - m_{l',\tau}\left(x_{l'}^0\right)\right| \\
&= O_p(N_n/\sqrt{n}) + O_p(N_n/\sqrt{n}) = O_p(N_n/\sqrt{n}).
\end{aligned}
$$

The proof of (12) in Lemma 1 follows similarly. ■

Lemma A.1 *For any function* $\alpha \in \mathcal{C}^{p+1}[0,1]$ *that is strictly monotone increasing with* $\alpha^{(1)}(x) \geq c > 0$*, for* $x \in [0,1]$*, then there exists a monotone increasing spline function* $g \in \mathcal{G}^p$ *such that* $\|\alpha - g\|_\infty \leq c \left\|\alpha^{(p+1)}\right\|_\infty / N_n^{p+1}$*, and* $\left\|\alpha^{(1)} - g^{(1)}\right\|_\infty \leq c \left\|\alpha^{(p+1)}\right\|_\infty / N_n^p$ *for some constant* $c > 0$ *and large enough* N_n*.*

Proof. Corollary 6.21 of Schumaker (2007) entails that there exists a spline function $g \in \mathcal{G}^p$ such that

$$\|\alpha(x) - g(x)\|_\infty \leq c\|\alpha^{(p+1)}\|_\infty / N_n^{p+1},$$

and

$$\left\|\alpha^{(1)}(x) - g^{(1)}(x)\right\|_\infty \leq c\|\alpha^{(p+1)}\|_\infty / N_n^p.$$

Therefore, $g^{(1)}(x) \geq c - c\|\alpha^{(p+1)}\|_\infty / N_n^p \geq c/2 > 0$ and g is monotone increasing when N_n is large enough. ■

Following the notation in He and Shi (1998), for any $1 \leq l \leq d$ and $1 \leq k \leq N_n + 1$, define interval $I_{nk}^l = (v_{l,k-1}, v_{l,k}]$ and denote $v_{l,k}^*$ as the midpoint of I_{nk}^l. Let $\delta_n^l = (v_{l,k} - v_{l,k-1})/2 \sim 1/2(N_n + 1)$. In addition, let $\check{m}_{l,\tau}$ be the one-step backfitted unconstrained estimator of $m_{l,\tau}$ and $\check{\boldsymbol{\beta}}_l$ be the corresponding regression coefficients. Note that for $x_l \in I_{nk}^l$, any B-spline function $\mathbf{B}_l^T(x_l)\boldsymbol{\beta}_l$ can be expressed as $\boldsymbol{\varphi}_{lk}^T(x_l)\boldsymbol{\alpha}_{lk}(\boldsymbol{\beta}_l)$ for some $\boldsymbol{\alpha}_{lk} \in \mathbb{R}^{p+1}$, where $\boldsymbol{\varphi}_{lk}(x_l) = \left(1, (x_l - v_{lk}^*)/\delta_n^l, \cdots, ((x_l - v_{lk}^*)/\delta_n^l)^p\right)^T$, and $\check{\boldsymbol{\alpha}}_{lk}$ is a function of $\check{\boldsymbol{\beta}}_l$ under this re-expression.

By Assumption (A5), there exists $\boldsymbol{\alpha}_{lk}^* \in \mathbb{R}^{p+1}$ such that for $x_l \in I_{nk}^l$, $\sup_{x_l \in I_{nk}^l} |r_n^l(x_l)| = O\left(N_n^{-(p+1)}\right)$, where $r_n^l(x_l) = m_{l,\tau}(x_l) - \boldsymbol{\varphi}_{lk}^T(x_l)\boldsymbol{\alpha}_{lk}^*$. In addition, one has the following error decomposition

$$m_{l,\tau}(x_l) - \check{m}_{l,\tau}(x_l) = \boldsymbol{\varphi}_{lk}^T(x_l)\left(\boldsymbol{\alpha}_{lk}^* - \check{\boldsymbol{\alpha}}_{lk}\right) + r_n^l(x_l). \quad \text{(A.1)}$$

From model (1) in the main paper, the conditional quantile function of order $\tau \in [0,1]$ of Y_i

given $\mathbf{X}_i$ can be written as

$$Y_i = g\left(\mathbf{X}_i\right)\mu_{R_\tau} + g\left(\mathbf{X}_i\right)\varepsilon_i = m_\tau\left(\mathbf{X}_i\right) + g\left(\mathbf{X}_i\right)\varepsilon_i,$$

where the error term $\varepsilon_i = R_i - \mu_{R_\tau}$ satisfies $Q_\tau\left(\varepsilon_i|\mathbf{X}_i\right) = 0$. The only difference between ε_i and R_i is that the locations of their distributions are different. Then when $x_{il} \in I_{nk}^l$, for pseudo response $Y_{i,-l} = Y_i - \widetilde{m}_{0,\tau} - \sum_{l'\neq l}\widetilde{m}_{l',\tau}\left(X_{il'}\right)$, one has the error decomposition

$$Y_{i,-l} - \boldsymbol{\varphi}_{lk}^T\left(X_{il}\right)\check{\boldsymbol{\alpha}}_{lk} = g(\mathbf{X}_i)\varepsilon_i - \boldsymbol{\varphi}_{lk}^T(X_{il})(\check{\boldsymbol{\alpha}}_{lk} - \boldsymbol{\alpha}_{lk}^*) + r_{ni}^l + e_{i,-l},$$

with $e_{i,-l} = m_{0,\tau} - \widetilde{m}_{0,\tau} + \sum_{l'\neq l}\left(m_{l',\tau}\left(X_{il'}\right) - \widetilde{m}_{l',\tau}\left(X_{il'}\right)\right)$, and $r_{ni}^l = r_n^l\left(X_{il}\right)$.

Assume there is a perturbation to the k-th component of $\check{\boldsymbol{\beta}}_l$, which causes changes to $p+1$ vectors of $\boldsymbol{\alpha}_{lk}$ (i.e., $\boldsymbol{\alpha}_{l,k-p}, \cdots, \boldsymbol{\alpha}_{l,k}$). Define $\boldsymbol{\theta}_{lk}\left(\boldsymbol{\beta}_l\right) = \left((\boldsymbol{\alpha}_{l,k-p}\left(\boldsymbol{\beta}_l\right) - \boldsymbol{\alpha}_{l,k-p}^*)^T, \cdots, (\boldsymbol{\alpha}_{l,k}(\boldsymbol{\beta}_l) - \boldsymbol{\alpha}_{l,k}^*)^T\right)^T$, $\boldsymbol{\phi}_{lk}(x_l) = \left(\boldsymbol{\varphi}_{l,k-p}^T(x_l), \cdots, \boldsymbol{\varphi}_{l,k}^T(x_l)\right)^T$ and $\check{\boldsymbol{\theta}}_{lk} = \boldsymbol{\theta}_{lk}\left(\check{\boldsymbol{\beta}}_l\right)$. Let $\mathcal{C}_k$ be the linear space of $\boldsymbol{\theta}_{lk}$, which is formed by perturbing the k-th component. Then $\check{\boldsymbol{\theta}}_{lk}$ is an inner point of $\mathcal{C}_k$. Let $S_{nk}^l = \{i : x_{il} \in (v_{l,k-p-1}, v_{l,k}]\}$, and M_{nk}^l be the number of data points in S_{nk}^l, with $M_{nk}^l = \#S_{nk}^l$. Define $M_n^l = \sup_k M_{nk}^l$. Under Assumptions (A1) and (A3), $M_n^l \sim n/N_n$. Furthermore, for any perturbed coefficients $\boldsymbol{\theta}_{lk} \in \mathcal{C}_k$, define

$$A_{nk}^l\left(\boldsymbol{\theta}_{lk}\right) = \sum_{i\in S_{nk}^l}\Psi\big(g\left(\mathbf{X}_i\right)\varepsilon_i - \mathbf{z}_{il}^T\boldsymbol{\theta}_{lk} + r_{ni}^l + e_{i,-l}\big)\mathbf{z}_{il},$$

with $\Psi(u) = 1/2 - I(u<0)$ and $\mathbf{z}_{il} = \boldsymbol{\phi}_{lk}\left(X_{il}\right)$. The term $A_{nk}^l\left(\boldsymbol{\theta}_l\right)$ quantifies the directional first order derivative of the objective function in the traditional polynomial spline method when the k-th component of $\check{\boldsymbol{\beta}}_l$ is perturbed, and we have the following result.

Lemma A.2 *For $l = 1, \cdots, d$, define $Q_{nk}^l = \left(f_\varepsilon(0)/M_{nk}^l\right)\sum_{i\in S_{nk}^l}\mathbf{z}_{il}\mathbf{z}_{il}^T/g(\mathbf{X}_i)$ to be a matrix of dimension $(p+1)^2\times(p+1)^2$, where f_ε is the density function of error term ε. Under Assumptions (A1)-(A5), for any constant $K>0$,*

$$E\big[A_{nk}^l\left(\boldsymbol{\theta}_{lk}\right)|\mathbf{X}\big] = -M_{nk}^l Q_{nk}^l\boldsymbol{\theta}_{lk} + O_p\big(nN_n^{-(p+2)} + \sqrt{n}\big), \tag{A.2}$$

uniformly for $\boldsymbol{\theta}_{lk} \in \boldsymbol{\Theta}_K^{lk} = \left\{\boldsymbol{\theta}_{lk} : \boldsymbol{\theta}_{lk} \in \mathbb{R}^{(p+1)^2}, |\boldsymbol{\theta}_{lk}| \leq K\sqrt{N_n\log n/M_{nk}^l}\right\}$ and $1 \leq k \leq$

N_n+1. *Here,* $\mathbf{X}=(\mathbf{X}_1^T,\ldots,\mathbf{X}_n^T)^T$.

Proof. By the definition of $A_{nk}^l(\boldsymbol{\theta}_{lk})$, for any $\boldsymbol{\theta}_{lk}\in\boldsymbol{\Theta}_K^{lk}$,

$$\begin{aligned}
&E\big[A_{nk}^l(\boldsymbol{\theta}_{lk})\,|\mathbf{X}\big]\\
=\;& E\Big[\sum_{i\in S_{nk}^l}\Psi\left(g(\mathbf{X}_i)\varepsilon_i-\mathbf{z}_{il}^T\boldsymbol{\theta}_{lk}+r_{ni}^l+e_{i,-l}\right)\mathbf{z}_{il}|\mathbf{X}\Big]\\
=\;& \sum_{i\in S_{nk}^l}\mathbf{z}_{il}E\Big[1/2-I\left(g(\mathbf{X}_i)\varepsilon_i<\mathbf{z}_{il}^T\boldsymbol{\theta}_{lk}-r_{ni}^l-e_{i,-l}\right)|\mathbf{X}\Big]\\
=\;& \sum_{i\in S_{nk}^l}\mathbf{z}_{il}\left[1/2-F_\varepsilon\left(\frac{\mathbf{z}_{il}^T\boldsymbol{\theta}_{lk}-r_{ni}^l-e_{i,-l}}{g(\mathbf{X}_i)}\right)\right]\\
=\;& \sum_{i\in S_{nk}^l}\mathbf{z}_{il}\left\{\frac{1}{2}-F_\varepsilon(0)-\frac{\mathbf{z}_{il}^T\boldsymbol{\theta}_{lk}-r_{ni}^l-e_{i,-l}}{g(\mathbf{X}_i)}f_\varepsilon(0)-\frac{1}{2}f_\varepsilon^{(1)}(\boldsymbol{\xi})\left(\frac{\mathbf{z}_{il}^T\boldsymbol{\theta}_{lk}-r_{ni}^l-e_{i,-l}}{g(\mathbf{X}_i)}\right)^2\right\}\\
=\;& \sum_{i\in S_{nk}^l}\left\{-\frac{f_\varepsilon(0)}{g(\mathbf{X}_i)}\mathbf{z}_{il}\mathbf{z}_{il}^T\boldsymbol{\theta}_{lk}+\frac{f_\varepsilon(0)\left(r_{ni}^l+e_{i,-l}\right)}{g(\mathbf{X}_i)}\mathbf{z}_{il}-\frac{1}{2}f_\varepsilon^{(1)}(\boldsymbol{\xi})\,\mathbf{z}_{il}\left(\frac{\mathbf{z}_{il}^T\boldsymbol{\theta}_{lk}-r_{ni}^l-e_{i,-l}}{g(\mathbf{X}_i)}\right)^2\right\}\\
=\;& -\left[\frac{f_\varepsilon(0)}{M_{nk}^l}\sum_{i\in S_{nk}^l}\frac{\mathbf{z}_{il}\mathbf{z}_{il}^T}{g(\mathbf{X}_i)}\right]M_{nk}^l\boldsymbol{\theta}_{lk}+\sum_{i\in S_{nk}^l}\frac{f_\varepsilon(0)\left(r_{ni}^l+e_{i,-l}\right)}{g(\mathbf{X}_i)}\mathbf{z}_{il}\\
&-\sum_{i\in S_{nk}^l}\frac{1}{2}f_\varepsilon'(\boldsymbol{\xi})\,\mathbf{z}_{il}\left(\frac{\mathbf{z}_{il}^T\boldsymbol{\theta}_{lk}-r_{ni}^l-e_{i,-l}}{g(\mathbf{X}_i)}\right)^2\\
=\;& -M_{nk}^lQ_{nv}^l\boldsymbol{\theta}_{lk}+\mathrm{I}_{1k}+\mathrm{II}_{2k}(\boldsymbol{\theta}_{lk}),
\end{aligned}$$

where $\boldsymbol{\xi}$ is a value between 0 and $\boldsymbol{\theta}_{lk}$, $\mathrm{I}_{1k}=\sum_{i\in S_{nk}^l}[f_\varepsilon(0)(r_{ni}^l+e_{i,-l})/g(\mathbf{X}_i)]\mathbf{z}_{il}$ and $\mathrm{II}_{2k}(\boldsymbol{\theta}_{lk})=-\sum_{i\in S_{nk}^l}\frac{1}{2}f_\varepsilon'(\boldsymbol{\xi})\,\mathbf{z}_{il}\left(\frac{\mathbf{z}_{il}^T\boldsymbol{\theta}_{lk}-r_{ni}^l-e_{i,-l}}{g(\mathbf{X}_i)}\right)^2$. For term I_{1k}, by the definition,

$$e_{i,-l}=m_{0,\tau}-\widetilde{m}_{0,\tau}+\sum_{l'\neq l}\left(m_{l',\tau}(X_{il'})-\widetilde{m}_{l',\tau}(X_{il'})\right).$$

According to Lemma 1, we have

$$\begin{aligned}
\sup_i|e_{i,-l}| &= \sup_i\left|m_{0,\tau}-\widetilde{m}_{0,\tau}+\sum_{l'\neq l}\left(m_{l',\tau}(X_{il'})-\widetilde{m}_{l',\tau}(X_{il'})\right)\right|\\
&\leq |m_{0,\tau}-\widetilde{m}_{0,\tau}|+\sum_{l'\neq l}\sup_{x_{l'}\in[0,1]}|m_{l',\tau}(x_{l'})-\widetilde{m}_{l',\tau}(x_{l'})|=O_p\left(N_n/\sqrt{n}\right).
\end{aligned}$$

Since both $f_\varepsilon(0)$ and $g(\mathbf{x}_i)$ are bounded, $\sup_i\left|r_{ni}^l\right|=O_p\left(N_n^{-(p+1)}\right)$ and $\sup_i|e_{i,-l}|=O_p\left(N_n/\sqrt{n}\right)$, one has

$$\sup_k |\mathrm{I}_{1k}| = \sup_k \left| \sum_{i \in S_{nk}^l} \frac{f(0)\left(r_{ni}^l + e_{i,-l}\right)}{g(\mathbf{X}_i)} \mathbf{z}_{il} \right| = O_p\left(M_n^l N_n^{-(p+1)} + M_n^l N_n/\sqrt{n}\right) = O_p\left(n N_n^{-(p+2)} + \sqrt{n}\right).$$

Similarly, we have

$$\sup_k \sup_{\boldsymbol{\theta}_{lk} \in \boldsymbol{\Theta}_K^{lk}} |\mathrm{II}_{2k}(\boldsymbol{\theta}_{lk})| = \sup_k \sup_{\boldsymbol{\theta}_{lk} \in \boldsymbol{\Theta}_K^{lk}} \left| - \sum_{i \in S_{nk}^l} \frac{1}{2} f_{\varepsilon}^{(1)}(\boldsymbol{\xi}) \mathbf{z}_{il} \left(\frac{\mathbf{z}_{il}^T \boldsymbol{\theta}_{lk} - r_{ni}^l - e_{i,-l}}{g(\mathbf{X}_i)} \right)^2 \right| = O_p\left(n N_n^{-(p+2)} + \sqrt{n}\right).$$

Since $M_n^l \sim n/N_n$ uniformly for $1 \le k \le N_n + 1$ and $\boldsymbol{\theta}_{lk} \in \boldsymbol{\Theta}_K^{lk}$, one has,

$$E\left[A_{nk}^l(\boldsymbol{\theta}_{lk}) \,|\mathbf{X}\right] = -M_{nk}^l Q_{nk}^l \boldsymbol{\theta}_{lk} + o_p\left(M_{nk}^l N_n^{-(p+1)} + M_{nk}^l N_n/\sqrt{n}\right) = -M_{nk}^l Q_{nk}^l \boldsymbol{\theta}_{lk} + o_p\left(n N_n^{-(p+2)} + \sqrt{n}\right).$$

■

A.2 *Proof of Theorem 1*

To prove Theorem 1, it is enough to show that $\sup_k \left|\check{\boldsymbol{\theta}}_{lk}\right| = O_p\left(N_n \sqrt{\log n/n}\right)$. By Lemma A.3 of He and Shi (1998), one has $\sup_k \sup_{\boldsymbol{\theta}_{lk} \in \boldsymbol{\Theta}_K^{lk}} \left|A_{nk}^l(\boldsymbol{\theta}_{lk}) - E\left[A_{nk}^l(\boldsymbol{\theta}_{lk})\,|\mathbf{X}\right]\right| = O_p\left(\sqrt{M_n^l N_n \log n}\right) = O_p(\sqrt{n \log n})$ for any fixed K. Together with Lemma A.2, for $\boldsymbol{\theta}_{lk} \in \Theta_K^{lk}$, one has

$$A_{nk}^l(\boldsymbol{\theta}_{lk}) = -M_{nk}^l Q_{nk}^l \boldsymbol{\theta}_l + O_p\left(n N_n^{-(p+2)} + \sqrt{n} + \sqrt{n \log n}\right). \tag{A.3}$$

We first define $G_{nk}^l(\boldsymbol{\theta}_{lk}) = -\boldsymbol{\theta}_{lk}^T A_{nk}^l(\boldsymbol{\theta}_{lk})$, which is a convex function in $\boldsymbol{\theta}_{lk}$. Similarly, following the proof of Theorem 1 in He and Shi (1998), we have

$$\begin{aligned}
& \inf_k \inf_{|\boldsymbol{\theta}_{lk}| > K\sqrt{N_n \log n / M_{nk}^l}} \left|A_{nk}^l(\boldsymbol{\theta}_{lk})\right| \ge \inf_k \inf_{|\boldsymbol{\theta}_{lk}| = K\sqrt{N_n \log n / M_{nk}^l}} G_{nk}^l(\boldsymbol{\theta}_{lk}) / |\boldsymbol{\theta}_{lk}| \\
= \ & \inf_k \inf_{|\boldsymbol{\theta}_{lk}| = K\sqrt{N_n \log n / M_{nk}^l}} \left[M_{nk}^l \boldsymbol{\theta}_{lk}^T Q_{nk}^l \boldsymbol{\theta}_{lk} / |\boldsymbol{\theta}_{lk}| + O_p\left(n N_n^{-(p+2)} + \sqrt{n} + \sqrt{n \log n}\right) \right] \\
\ge \ & \inf_k \left[K M_{nk}^l \lambda(Q_{nk}^l) \sqrt{N_n \log n / M_{nk}^l} \right] + O_p\left(n N_n^{-(p+2)} + \sqrt{n} + \sqrt{n \log n}\right) \\
\ge \ & c\sqrt{n \log n}\,(1 + o_p(1)),
\end{aligned} \tag{A.4}$$

where $c > 0$ is an arbitrary constant. The last step in (A.4) follows from the fact that $\inf_k M^l_{nk} = O_p(n/N_n)$, and there exists a constant c such that $\lambda\left(Q^l_{nk}\right)$, the smallest eigenvalue of Q^l_{nk}, satisfies that $\inf_k \lambda\left(Q^l_{nk}\right) > c$. Therefore when $|\boldsymbol{\theta}_{lk}| > K\sqrt{N_n \log n / M^l_{nk}}$, the term $\inf_k \left|A^l_{nk}(\boldsymbol{\theta}_{lk})\right|$ diverges to infinity with probability approaching one as sample size increases.

Furthermore, according to Lemma A.1 in He and Shi (1998), we can infer that $A^l_{nk}\left(\check{\boldsymbol{\theta}}_{lk}\right) = O_p(1)$. Therefore the inequality in (A.4) indicates that $\left|\check{\boldsymbol{\theta}}_{lk}\right| = O_p\big(\sqrt{N_n \log n/M^l_{nk}}\big) = O_p\big(N_n\sqrt{\log n/n}\big)$ uniformly for $k = 1, \cdots, N_n+1$. Thus $\sup_k |\check{\boldsymbol{\alpha}}_{lk} - \boldsymbol{\alpha}^*_{lk}| = O_p\big(N_n\sqrt{\log n/n}\big)$. Since $\check{m}_{l,\tau}(x_l) - m_{l,\tau}(x_l) = \boldsymbol{\varphi}^T_{lk}(x_l)\left(\check{\boldsymbol{\alpha}}_{lk} - \boldsymbol{\alpha}^*_{lk}\right) - r^l_n(x_l)$ and $\sup_{x_l \in I^l_{nk}} |\boldsymbol{\varphi}_{lk}(x_l)| \le \sqrt{p+1}$ uniformly in $x_l \in I^l_{nk}$ and $k = 1, \cdots, N_n + 1$, one has,

$$
\begin{aligned}
\sup_{x_l\in[0,1]} |\check{m}_{l,\tau}(x_l) - m_{l,\tau}(x_l)| &= \sup_{k=1,\cdots,N_n+1} \sup_{x_l\in I^l_{nk}} \left|\varphi^T_{lk}(x_l)\left(\check{\boldsymbol{\alpha}}_{lk} - \boldsymbol{\alpha}^*_{lk}\right) - r^l_n(x_l)\right| \\
&\le \sup_{k=1,\cdots,N_n+1} \sup_{x_l\in I^l_{nk}} \left|\varphi^T_{lk}(x_l)\right| \left|\check{\boldsymbol{\alpha}}_{lk} - \boldsymbol{\alpha}^*_{lk}\right| + \sup_{x_l\in[0,1]} \left|r^l_n(x_l)\right| \\
&\le \sup_{k=1,\cdots,N_n+1} \sqrt{p+1}\,|\check{\boldsymbol{\alpha}}_{lk} - \boldsymbol{\alpha}^*_{lk}| + O\left(N_n^{-(p+1)}\right) \\
&= O_p\left(N_n\sqrt{\log n/n} + N_n^{-(p+1)}\right).
\end{aligned}
$$

By Assumption (A5) and de Boor (2001) p.115, for $l = 1, \cdots, d$ and $x \in I^l_{nk}$, taking the first order derivative in Equation (A.1), one has

$$
\check{m}^{(1)}_{l,\tau}(x_l) - m^{(1)}_{l,\tau}(x_l) = \boldsymbol{\vartheta}^T_{lk}(x_l)\left(\check{\boldsymbol{\alpha}}_{lk} - \boldsymbol{\alpha}^*_{lk}\right) - \widetilde{r}^l_n(x_l),
$$

where $\boldsymbol{\vartheta}_{lk}(x_l) = \left(0, 1, 2\left(x_l - v^*_{lk}\right)/\delta^l_n \cdots, p\left(\left(x_l - v^*_{lk}\right)/\delta^l_n\right)^{p-1}\right)^T/\delta^l_n$, and $\sup_{x_l\in[0,1]}\left|\widetilde{r}^l_n(x_l)\right| = O\big((\delta^l_n)^p\big) = O(1/N^p_n)$. Then, we have

$$
\begin{aligned}
\sup_{x_l\in[0,1]} \left|\check{m}^{(1)}_{l,\tau}(x_l) - m^{(1)}_{l,\tau}(x_l)\right| &= \sup_{k=1,\cdots,N_n+1} \sup_{x_l\in I^l_{nk}} \left|\boldsymbol{\vartheta}^T_{lk}(x_l)\left(\check{\boldsymbol{\alpha}}_{lk} - \boldsymbol{\alpha}^*_{lk}\right) - \widetilde{r}^l_n(x_l)\right| \\
&\le \sup_{k=1,\cdots,N_n+1} \sup_{x_l\in I^l_{nk}} \left|\boldsymbol{\vartheta}^T_{lk}(x_l)\right| \left|\check{\boldsymbol{\alpha}}_{lk} - \boldsymbol{\alpha}^*_{lk}\right| + \sup_{x_l\in[0,1]} \left|\widetilde{r}^l_n(x_l)\right| \\
&= O_p\left(\frac{N_n}{\delta^l_n}\sqrt{\log n/n}\right) + O_p\left(N_n^{-p}\right) \\
&= O_p\left(N_n^2\sqrt{\log n/n} + N_n^{-p}\right).
\end{aligned}
$$

Consequently,

$$\sup_{x_l\in[0,1]}\left|\check{m}_{l,\tau}^{(1)}(x_l)-m_{l,\tau}^{(1)}(x_l)\right|=O_p\left(N_n^2\sqrt{\log n/n}+N_n^{-p}\right).$$

The rate of convergence for $\check{m}_{l,\tau}^{(2)}$ follows similarly by using

$$\boldsymbol{\vartheta}_{lk}^*(x_l)=\left(0,0,2,\cdots,p(p-1)\left(\left(x_l-v_{lk}^*\right)/\delta_n^l\right)^{p-2}\right)^T/\left(\delta_n^l\right)^2$$

in above arguments. ■

A.3 *Proof of Theorem 2*

When m_l is monotone increasing, condition (C6) and Theorem 1 imply that $\min_{x_l\in[0,1]}\check{m}_l^{(1)}(x_l)>c_3/2>0$ with probability approaching one. That is, the unconstrained spline estimator $\check{m}^{(1)}$ is monotone increasing asymptotically. In addition, when $p\leq 3$, Lemma A.1, together with Lemma 3 in Wang and Xue (2015) show that the linear constraints given in Section 3.1 are necessary and sufficient conditions for a spline function to be monotone. Therefore, the unconstrained and constrained estimators are asymptotically equivalent and enjoy the same asymptotic properties for $p\leq 3$. For m_l that is concave, results follow similarly from (C6*) and Theorem 1.

The rates convergence for $\hat{\mu}_{R_\tau}$ and $\hat{g}$ follow from similar arguments as the proof of Theorem 2 in Wang et al. (2020) and Theorem 2 in Martins-Filho and Yao (2007). In particular, note that $\left|\hat{\mu}_{R_\tau}-\mu_{R_\tau}\right|=\hat{\mu}_{R_\tau}\mu_{R_\tau}\left|\hat{\mu}_{R_\tau}^{-1}-\mu_{R_\tau}^{-1}\right|=\hat{\mu}_{R_\tau}\left|\max_i\left[m_\tau\left(X_i\right)R_i/\hat{m}_\tau\left(X_i\right)\right]-1\right|$, where $\hat{\mu}_{R_\tau}^{-1}=\left\{\max_i\left[Y_i/\tilde{m}\left(X_i\right)\right]\right\}^{-1}=O_p(1)$, and

$$\begin{aligned}\left|\max_i\frac{m_\tau\left(X_i\right)R_i}{\hat{m}_\tau\left(X_i\right)}-1\right| &\leq \left|\max_i\left[\frac{m_\tau\left(X_i\right)}{\hat{m}_\tau\left(X_i\right)}(R_i-1)\right]\right|+\left|\max_i\frac{m_\tau\left(X_i\right)}{\hat{m}_\tau\left(X_i\right)}-1\right|\\ &\leq \left|\max_i\frac{m_\tau\left(X_i\right)}{\hat{m}_\tau\left(X_i\right)}\right|\left|\max_i(R_i-1)\right|+\left|\max_i\frac{m_\tau\left(X_i\right)}{\hat{m}_\tau\left(X_i\right)}-1\right|,\end{aligned}$$

where Lemma 11 of Wang et al. (2020) entails that $\max_i(R_i-1)=O_p(n^{-1})$ and the first part of this theorem gives that $\left|\max_i\frac{m_\tau(X_i)}{\hat{m}_\tau(X_i)}\right|=O_p(1)$ and $\left|\max_i\frac{m_\tau(X_i)}{\hat{m}_\tau(X_i)}-1\right|=O_p(N_n\sqrt{\log(n)/n}+N_n^{-p-1})$. Therefore $\left|\hat{\mu}_{R_\tau}-\mu_{R_\tau}\right|=O_p(N_n\sqrt{\log(n)/n}+N_n^{-p-1})$. Finally the convergence of $\hat{g}$

follows from the fact that

$$
\begin{aligned}
\hat{g}(\mathbf{x}) - g(\mathbf{x}) &= \hat{m}_{\tau}(\mathbf{x}) / \hat{\mu}_{R_{\tau}} - m_{\tau}(\mathbf{x}) / \mu_{R_{\tau}} \\
&= \hat{m}_{\tau}(\mathbf{x}) \left(\hat{\mu}_{R_{\tau}}^{-1} - \mu_{R_{\tau}}^{-1} \right) + \mu_{R_{\tau}}^{-1} \left[\hat{m}_{\tau}(\mathbf{x}) - m_{\tau}(\mathbf{x}) \right]. \blacksquare
\end{aligned}
$$

A.4 *Proof of Theorem 3*

The proof is similar to the proof of Theorem 2.

References

Albuquerque, R., Hopenhayn, H., 2004. Optimal lending contracts and firm dynamics. The Review of Economic Studies 71, 285–315.

de Boor, C., 2001. A practical guide to splines. Springer, New York, NY.

He, X., Shi, P., 1998. Monotone B-spline smoothing. Journal of the American Statistical Association 93, 643–650.

Horowitz, J., Lee, S., 2005. Nonparametric estimation of an additive quantile regression model. Journal of the American Statistical Association 100, 1238–1249.

Klobucnik, J., Sievers, S., 2013. Valuing high technology growth firms. Journal of Business Economics 83, 947–984.

Martins-Filho, C., Yao, F., 2007. Nonparametric frontier estimation via local linear regression. Journal of Econometrics 141, 283–319.

Schumaker, L., 2007. Spline functions: basic theory. Cambridge University Press.

Wang, L., Xue, L., 2015. Constrained polynomial spline estimation of monotone additive models. Journal of Statistical Planning and Inference 167, 27–40.

Wang, L., Xue, L., Yang, L., 2020. Estimation of additive frontier with shape constraints. Journal of Nonparametric Statistics 32, 262–293.

Xue, L., Yang, L., 2006. Additive coefficient modeling via polynomial spline. Statistica Sinica 16, 1423–1446.

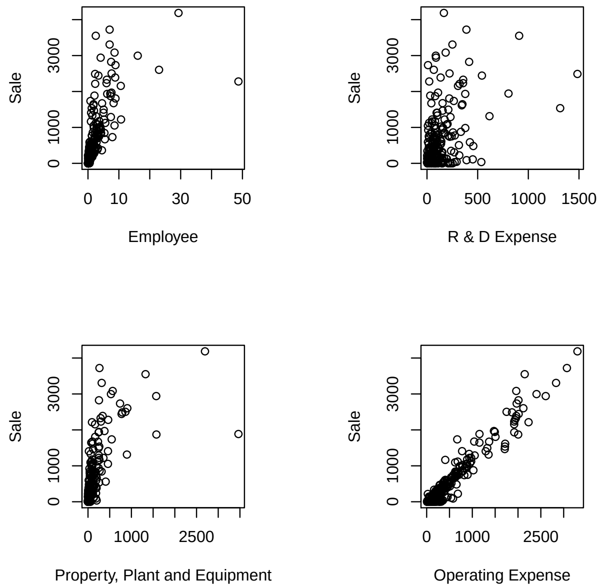
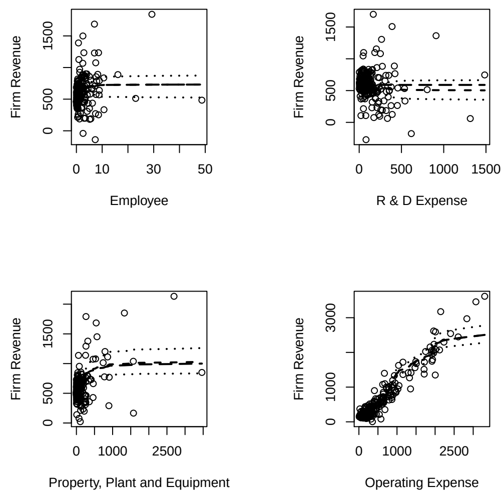


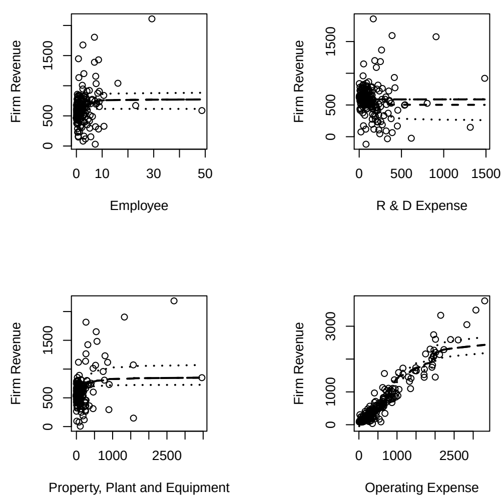
Figure 1: Scatter plot of the firm's revenue against each input variable for U.S. high tech firms.

Table 1: The estimated regression coefficients with standard errors in parentheses.

	DLTT	M-B	MKVALT	AGE
Coefficient	0.0240(0.0112)**	0.0000(0.0000)	0.0002(0.0006)	0.0002(0.0001)*

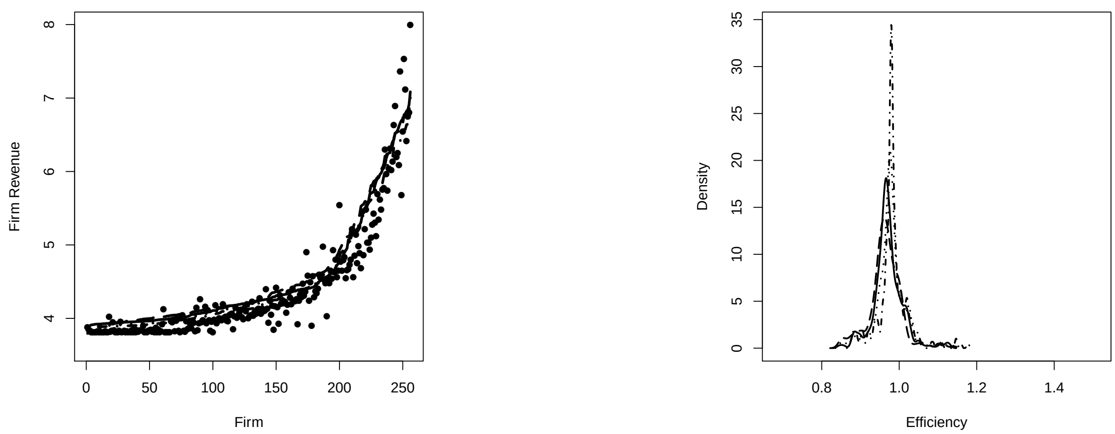


(a) Linear Regression



(b) Quantile Regression

Figure 2: Panel (a) plots the nonparametric estimate of the mean revenue, where the circles are pseudo observations with respect to each input, the dashed (- - -) and long-dashed (— —) lines denote the results from URS and MCRS, respectively, and the dotted (······) lines describes the 95% point-wise confidence interval from 100 bootstrap samples based on the URS method. While Panel (b) shows the nonparametric estimates of the median revenue using UQS and MCQS.



(a) Frontier

(b) Efficiency Density

Figure 3: Panel (a) plots the estimated maximum revenue for U.S. high tech firms, while Panel (b) gives the kernel densities of the relative efficiency estimates, where the solid (—–) , dotted (······) , dot-dashed (-·-·) and long-dashed (−−) lines represent the density estimated using URS, MCRS, UQS and MCQS, respectively. The solid circles in Fig. (a) are the true revenues.