Supplemental Materials: Additive Separability and Unobserved Heterogeneity: A Test for Hicks-neutral Productivity Shocks

Zhutong Gu*
Peking University, HSBC Business School

1 Asymptotic Proof

Notation. $C(x,v) = E(Y|X=x, V=v)$, $\Delta_i = \widehat{h}(V_i) - h(V_i)$, $f^*(x,v) = f(x)f(v)/f(x,v)$, where $f(\cdot)$ denotes marginal/joint densities. $t_i = \mathbf{1}\{X_i \in \mathcal{X}_0\}$. Remember that under $\mathbb{H}_0$, $Y = m_1(X) + h(V) + \epsilon$. Also let $\sum_{i,j}^N \equiv \sum_{i=1}^N \sum_{j=1}^N$, $\sum_{i,j,k}^N \equiv \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N$, $\sum_{j>i}^N \equiv \sum_{i=1}^{N-1} \sum_{j>i}^N$ and $\sum_{k>j>i}^N \equiv \sum_{i=1}^{N-2} \sum_{j>i}^{N-1} \sum_{k>j}^N$. Let $K_{i,j}^X = K_h(X_i - X_j) = k(X_i - X_j/h)/h$, $K_{i,\cdot}^X = K_h(X_i - x)$ and similar for other variables. I suppress superscript (or subscript) p for quantile(and trimming) indicators.

1.1 Main proofs

Proof of Theorem 4.1. Recall $\widehat{T}_N^p$ and $\widetilde{T}_N^p$ in the main text. $\widehat{T}_N^p$ is decomposed first into three components

$$\widehat{T}_N^p = I_1 + I_2 + I_3$$

As noted in the text, $I_3 = 0$ under $\mathbb{H}_0$. The following Lemma 1 aids to prove $I_2 = o_p(N^{-1/2})$. To analyze I_1, it suffices to study $\widetilde{T}_N^p$,

$$\widetilde{T}_N^p = D_N^g + D_N^a + D_N^h$$

as $\widehat{T}_N^p = \widetilde{T}_N^p + o_p(N^{-1/2})$, where D_N^g, D_N^a and D_N^h can be found in the main text, respectively. Lemma 2 handles D_N^a using second order U-statistic theorem and Lemma 3 deals with D_N^g using third order U-statistic theorem. Lemma 4 provides the intermediate result for Lemma 5

*Corresponding author email: zgu@phbs.pku.edu.cn. Tel: 86-18617094914. Address: Office 613, Peking University HSBC Business School, Xili University Town, Shenzhen, China

on D_N^h. Then it is shown that

$$\sqrt{N}\widehat{T}_N^p = N^{-1/2}\sum_{i=1}^{N}(\xi_{gi}^p + \xi_{ai}^p + \xi_{hi}^p) + o_p(1)$$

where

$$\begin{aligned}
\xi_{gi}^p &= t_i^p f^*(X_i, V_i)\epsilon_i + E(t^p)h(V_i) \\
\xi_{ai}^p &= -t_i^p \epsilon_i \\
\xi_{hi}^p &= E(t^p|V_i)f^*(X_i, V_i)\epsilon_i
\end{aligned}$$

Combine those three terms, $\xi_i^p = \xi_{gi}^p + \xi_{ai}^p + \xi_{hi}^p$.

$$\xi_i^p = [t_i^p + E(t^p|V_i)]f^*(X_i, V_i) - t_i^p]\epsilon_i + E(t^p)h(V_i)$$

By the CLT, this theorem establishes the limiting variance to be $\Omega_p = E(\xi_i^p \xi_i^{p\prime})$.

$\square$

Proof of Theorem 4.2 and Corollary 4.2.1. According to Theorem 4.1, it is true that $\widehat{T}_N$ follows a P-dimensional multivariate normal distribution.

$$\sqrt{N}\widehat{T}_N \xrightarrow{\text{D}} N(\mathbf{0}, \Omega)$$

So $W_N = NT_N'\Omega^{-1}T_N \xrightarrow{\text{D}} \chi_P^2$. By Slutsky's theorem, for any $\widehat{\Omega}_N \xrightarrow{\text{P}} \Omega$, then it holds that $\widehat{W}_N = NT_N'\widehat{\Omega}_N^{-1}T_N \xrightarrow{\text{D}} \chi_P^2$. $\square$

Lemma 1 I_2**.** *Suppose H_0 is true, under Assumption A.1-A.6, for each p*

$$\sqrt{N}I_2 \equiv \sqrt{N}\sum_{i=1}^{N}(\widehat{t}_i - t_i)(\widehat{D}(X_i) - D(X_i)) = o_p(1)$$

Proof. For any X_i,

$$|\widehat{D}(X_i) - D(X_i)| = \Big|[\widehat{g}(X_i) - g(X_i)] + [\widehat{a}(X_i) - a(X_i)]\Big| \le |\widehat{g}(X_i) - g(X_i)| + |\widehat{a}(X_i) - a(X_i)| = O_p((Nh)^{-1/2})$$

According to this, it is true that,

$$N^{-1}\sum_{i=1}^{N}|\widehat{D}(X_i) - D(X_i)|^2 = O((Nh)^{-1})$$

By Cauchy-Schwartz inequality,

$$\begin{aligned}
\sqrt{N}\sum_{i=1}(\widehat{t}_i - t_i)(\widehat{D}(X_i) - D(X_i))/N &\le \sqrt{N}\sqrt{\sum_{i=1}(\widehat{t}_i - t_i)^2/N}\sqrt{\sum_{i=1}[\widehat{D}(X_i) - D(X_i)]^2/N} \\
&= \sqrt{N}o_p(N^{-1/2})O_p(N^{-1/2}h^{-1/2}) = o_p(1)
\end{aligned}$$

$\square$

Lemma 2 D_N^a**.** *Suppose that Assumption A.1-A.6 hold, under $\mathbb{H}_0$, then D_N^a can be written as the following,*

$$D_N^a = -N^{-1}\sum_{i=1}^{N} t_i[Y_i - h(V_i) - a(X_i)] + o_p(N^{-1/2})$$

Proof. Let $Y_i^+ \equiv Y_i - h(V_i)$. Recall that

$$D_N^a = -N^{-1}\sum_{i=1}^{N} t_i[\widehat{E}(Y^+|X_i) - E(Y^+|X_i)]$$

Apply the intermediate Lemma A 2 in the supplemental material, it is true that $\widehat{E}(Y^+|X_i) - E(Y^+|X_i) = (N-1)^{-1}\sum_{j\neq i}^{N} K_{j,i}^X[Y_j^+ - E(Y^+|X_i)]/f(X_i) + o_p(N^{-1/2})$. Substitute this into D_N^a and note that $D_N^a = \widetilde{D}_N^a + o_p(N^{-1/2})$. From now on, it suffices to only work with $\widetilde{D}_a$ below,

$$\begin{aligned}
\widetilde{D}_N^a &= -\frac{1}{N(N-1)}\sum_{i,j}^{N} t_i f(X_i)^{-1} K_{j,i}^X[Y_j^+ - E(Y^+|X_i)] \\
&= -\binom{N}{2}^{-1}\sum_{j>i}^{N}(a_{ij} + a_{ji})/2
\end{aligned}$$

To apply the U-statistic theorem, we rewrite $\widetilde{D}_N^a$ as symmetric in i and j and where

$$\begin{aligned}
a_{ij} &= t_i f(X_i)^{-1} K_{j,i}^X[Y_j^+ - E(Y^+|X_i)] \\
a_{ji} &= t_j f(X_j)^{-1} K_{i,j}^X[Y_i^+ - E(Y^+|X_j)]
\end{aligned}$$

Moreover, by intermediate lemmas,

$$E(a_{ij}|X_i) = O(h^4); E(a_{ji}|X_i) = t_i[Y_i^+ - E(Y^+|X_i)] + O(h^4)$$

By Assumption A.1-A.3, it is true that $E|a_{ji}|^2 + E|a_{ij}|^2 = O(1) = o(N)$ as every multiplicative term is bounded. Therefore, the standard second order U-statistic applies,

$$\widetilde{D}_N^a = -N^{-1}\sum_{i=1}^{N} t_i[Y_i^+ - E(Y^+|X_i)] + O(h^4) + o_p(N^{-1/2})$$

Under $\mathbb{H}_0$ and Assumption A-5, $O(h^4) = o(N^{-1/2})$, it can be simplified to

$$D_N^a = -N^{-1}\sum_{i=1}^{N} t_i[Y_i - h(V_i) - a(X_i)] + o_p(N^{-1/2}) = -N^{-1}\sum_{i=1}^{N} t_i\epsilon_i + o_p(N^{-1/2})$$

$\square$

Lemma 3 D_N^g**.** *Suppose that Assumption A.1-A.6 hold and under* $\mathbb{H}_0$*, then* D_N^g *can be written as the following,*

$$D_N^g = N^{-1}\sum_{i=1}^{N}\{t_i f^*(X_i,V_i)\epsilon_i + E(t)h(V_i)\} + o_p(N^{-1/2})$$

Proof. Recall that D_N^g and by definition,

$$\begin{aligned} D_N^g &= N^{-1}\sum_{i=1}^{N} t_i[\widehat{g}(X_i) - g(X_i)] \\ &= \frac{1}{N(N-1)}\sum_{i,j}^{N} t_i\widehat{C}(X_i,V_j) - \frac{1}{N}\sum_{i=1}^{N}\int t_i C(X_i,v)dF(v) \\ &= \frac{1}{N(N-1)}\sum_{i,j}^{N} t_i\left[\widehat{C}(X_i,V_j) - C(X_i,V_j)\right] + \frac{1}{N(N-1)}\sum_{i,j}^{N} t_i\left[C(X_i,V_j) - \int C(X_i,v)dF(v)\right] \\ &\equiv D_N^{g1} + D_N^{g2} \end{aligned}$$

The third equality follows by adding and subtracting $\sum_{i,j}^{N} t_i C(X_i,V_j)/N(N-1)$.

Part a), for D_N^{g1}, by Lemma A 2, it is true that

$$\widehat{C}(X_i,V_j) - C(X_i,V_j) = \frac{1}{N-2}\sum_{k\neq i,j}^{N} K_{k,i}^X K_{K,j}^V[Y_k - C(X_i,V_j)]/f(X_i,V_j) + o_p(N^{-1/2})$$

Substitute this into D_N^{g1} and rewrite it as $D_N^{g1} = \widetilde{D}_N^{g1} + o_p(N^{-1/2})$, so from now on it suffices to work with $\widetilde{D}_N^{g1}$ defined below,

$$\begin{aligned} \widetilde{D}_N^{g1} &= \frac{1}{N(N-1)(N-2)}\sum_{i,j,k}^{N}\frac{t_i}{f(X_i,V_j)}K_{k,i}^X K_{k,j}^V\left[Y_k - C(X_i,V_j)\right] \\ &= \binom{N}{3}^{-1}\sum_{k>j>i}^{N}\sum_{l=1}^{6}\delta_{g1l}/6 \end{aligned}$$

To represent $\widetilde{D}_N^{g1}$ as a third-order U-statistic and where

$$\begin{aligned} \delta_{g11} &= \frac{t_i}{f(X_i,V_j)}K_{k,i}^X K_{k,j}^V\left[Y_k - C(X_i,V_j)\right];\, \delta_{g12} = \frac{t_i}{f(X_i,V_k)}K_{j,i}^X K_{j,k}^V\left[Y_j - C(X_i,V_k)\right] \\ \delta_{g13} &= \frac{t_j}{f(X_j,V_i)}K_{k,j}^X K_{k,i}^V\left[Y_k - C(X_j,V_i)\right];\, \delta_{g14} = \frac{t_k}{f(X_k,V_i)}K_{j,k}^X K_{j,i}^V\left[Y_j - C(X_k,V_i)\right] \\ \delta_{g15} &= \frac{t_k}{f(X_k,V_j)}K_{i,k}^X K_{i,j}^V\left[Y_i - C(X_k,V_j)\right];\, \delta_{g16} = \frac{t_j}{f(X_j,V_k)}K_{i,j}^X K_{i,k}^V\left[Y_i - C(X_j,V_k)\right] \end{aligned}$$

Moreover, by Lemma A 3,

$$\begin{gathered} E(\delta_{g11}|X_i) = O(h^4); E(\delta_{g12}|X_i) = O(h^4) \\ E(\delta_{g13}|V_i) = O(h^4); E(\delta_{g14}|V_i) = O(h^4) \\ E(\delta_{g15}|W_i) = t_i f^*(X_i, V_i)[Y_i - C(X_i, V_i)] + O(h^4) \\ E(\delta_{g16}|W_i) = t_i f^*(X_i, V_i)[Y_i - C(X_i, V_i)] + O(h^4) \end{gathered}$$

Proving the above results is nothing hard but tedious. To conserve space, I only show one item and the others follow the same line of reasoning. Now take $E(\delta_{g15}|W_i)$ as an example,

$$\begin{aligned} E(\delta_{g15}|W_i) &= \int \frac{t(x)}{f(x,v)} K(X_i - x) K(V_i - v) \left[Y_i - C(x, v) \right] f(x) f(v) dx dv \\ &= \int_{\{u_1 : t(X_i + u_1 h) = 1\}} f^*(X_i + u_1 h, V_i + u_2 h) \left[Y_i - C(X_i + u_1 h, V_i + u_2 h) \right] k(u_1) k(u_2) du_1 du_2 \\ &= t_i f^*(X_i, V_i)[Y_i - C(X_i, V_i)] + O(h^4) \end{aligned}$$

Note that the second equality holds by the transformation of variables, letting $x = X_i + u_1 h$ and $v = V_i + u_2 h$. The third equality follows from the Taylor expansion on h around $u_1 = u_2 = 0$.

By Assumption A.1-A.3, it is true that $\sum_{l=1}^{N} E|g_{1l}|^2 = O(1) = o(N^2)$ as every multiplicative term is bounded. Therefore, the standard second order U-statistic applies,

$$\begin{aligned} \widetilde{D}_N^{g1} &= N^{-1} \sum_{i=1}^{N} t_i f^*(X_i, V_i)[Y_i - C(X_i, V_i)] + O(h^4) + o_p(N^{-1/2}) \\ &= N^{-1} \sum_{i=1}^{N} t_i f^*(X_i, V_i) \epsilon_i + o_p(N^{-1/2}) \end{aligned}$$

Part b), for D_N^{g2}, one can also rewrite it as a second-order U-statistic,

$$\begin{aligned} D_N^{g2} &= \frac{1}{N(N-1)} \sum_{j,i}^{N} t_i \left[C(X_i, V_j) - \int C(X_i, v) dF(v) \right] \\ &= \binom{N}{2}^{-1} \sum_{j>i}^{N} (\delta_{g21} + \delta_{g22})/2 \end{aligned}$$

where in particular

$$\delta_{g21} = t_i \left[C(X_i, V_j) - \int C(X_i, v) dF(v) \right]; \delta_{g22} = t_j \left[C(X_j, V_i) - \int C(X_j, v) dF(v) \right]$$

It is obvious that $E(\delta_{g21}|W_i) = 0$. But for δ_{g22}, it can be shown that

$$\begin{aligned} E(\delta_{g22}|V_i) &= E_X[t(X)C(X,V_i)] - E\left[t(X)Yf^*(X,V)\right] \\ &= E(t)h(V_i) \end{aligned}$$

where E_X is the expectation taken with respect to only X. The second equality is true only under $\mathbb{H}_0$. Also by Assumption A.1-A.3, we have $E|g_{21}|^2 + E|g_{22}|^2 = O(1) = o(N)$, then the standard U-statistic theorem implies that

$$\widetilde{D}_N^{g2} = N^{-1}\sum_{i=1}^{N} E(t)h(V_i) + o_p(N^{-1/2})$$

Finally, combine $\widetilde{D}_N^{g1}$ and $\widetilde{D}_N^{g2}$, then Lemma 3 follows that

$$D_N^g = N^{-1}\sum_{i=1}^{N}\{t_i f^*(X_i,V_i)\epsilon_i + E(t)h(V_i)\} + o_p(N^{-1/2})$$

□

Lemma 4 and Lemma 5 use U-statistic theorem to analysis $N^{-1}\sum_{i=1}^{N}\widehat{E}(\Delta(V)|X_i)$.

Lemma 4 Δ_i**.** *Given* X_i *and* $V_i = v$*, let* $\Delta(v) = \widehat{h}(v) - h(v)$*, then it follows that*

$$\Delta(v) = \Delta_1(v) + \Delta_2(v) + \Delta_3 + o_p(N^{-1/2})$$

where

$$\begin{aligned} \Delta_1(v) &= \frac{1}{N}\sum_{i=1}^{N}\frac{f(X_i)K_{i,\cdot}^V}{f(X_i,v)}[Y_i - C(X_i,v)] \\ \Delta_2 &= \frac{1}{N}\sum_{i=1}^{N} m_1(X_i) - E[m_1(X)] \\ \Delta_3 &= E(Y) - \overline{Y} \end{aligned}$$

Proof. Given $V_i = v \in \mathcal{V}$, recall that $\Delta(v) = \widehat{h}(v) - h(v)$ with $\widehat{h}(\cdot)$ and $h(\cdot)$ defined in the main text. Following the similar argument in Lemma 2 $\Delta(v)$ can be decomposed into three components,

$$\Delta(v) = \underbrace{N^{-1}\sum_{i=1}^{N}\left[\widehat{C}(X_i,v) - C(X_i,v)\right]}_{\Delta_1(v)} + \underbrace{\left[\frac{1}{N}\sum_{i=1}^{N} C(X_i,v) - \int C(x,v)dF(x)\right]}_{\Delta_2(v)} + \underbrace{E(Y) - \overline{Y}}_{\Delta_3}$$

By Lemma A 2, $\Delta_1(v) = \widetilde{\Delta}_1(v) + o_p(N^{-1/2})$ where $\widetilde{\Delta}_1(v)$ is defined below,

$$\begin{aligned} \widetilde{\Delta}_1(v) &= \frac{1}{N(N-1)} \sum_{i,j}^{N} \frac{K^X_{j,i} K^V_{j,\cdot}}{f(X_i, v)} [Y_j - C(X_i, v)] \\ &= \binom{N}{2}^{-1} \sum_{j>i}^{N} (d_{ij} + d_{ji})/2 \end{aligned}$$

where in particular

$$d_{ij} = \frac{K^X_{j,i} K^V_{j,\cdot}}{f(X_i, v)} [Y_j - C(X_i, v)]; d_{ji} = \frac{K^X_{i,j} K^V_{i,\cdot}}{f(X_j, v)} [Y_i - C(X_j, v)]$$

It is straightforward to show that

$$E(d_{ij}|W_i) = O(h^4); E(d_{ji}|W_i) = \frac{f(X_i) K^V_{i,\cdot}}{f(X_i, v)} [Y_i - C(X_i, v)] + O(h^4)$$

The second moment condition $E|d_{ij}|^2 + E|d_{ji}|^2 = O(h^{-4})$ holds trivially according to Assumption A.1-A.3. So by Lemma A 1, we have

$$\Delta_1(v) = \widehat{\Delta}_1(v) + o_p(N^{-1/2})$$

Note that under $\mathbb{H}_0$, $\Delta_2 = N^{-1} \sum_i^N m_1(X_i) - E[m_1(X)]$. So it is true that $\Delta(v) = \widehat{\Delta}_1(v) + \Delta_2(v) + \Delta_3 + o_p(N^{-1/2})$. □

Lemma 5 D^h_N**.** *Suppose that Assumption A.1-A.6 hold and under $\mathbb{H}_0$, then D^h_N can be written as the following,*

$$D^h_N = N^{-1} \sum_{i=1}^{N} E(t|V_i) f^*(X_i, V_i) \epsilon_i + o_p(N^{-1/2})$$

Proof. To begin with, a result implied from Lemma 4 states that $E(\widehat{\Delta}|X_i) = o_p(N^{-1/2})$. As this can be seen from below,

$$E(\widehat{\Delta}|X_i) = E[\widehat{\Delta}_1(V)|X_i] + E(\Delta_2|X_i) + E(\Delta_3|X_i) + o_p(N^{-1/2})$$

Also, it is not hard to see the following from Lemma 4 that given X_i,

$$\begin{aligned} E[\widehat{\Delta}_1(V)|X_i] &= E\left[\frac{f(X_j) K^V_{i,j}}{f(X_j, V_i)} (Y_j - C(X_j, V_i)) \Big| X_i \right] = O(h^4) \\ E(\Delta_2|X_i) &= N^{-1}[m_1(X_j) - E(m_1(X))] = O_p(N^{-1}) \\ E(\Delta_3|X_i) &= -E[\overline{Y} - E(Y)|X_i] = -N^{-1}[E(Y|X_i) - E(Y)] = O_p(N^{-1}) \end{aligned}$$

Therefore, D_N^h can be further decomposed into four components like below,

$$D_N^h = \underbrace{N^{-1}\sum_{i=1}^N t_i[\widehat{E}(\widehat{\Delta}_1(V)|X_i) - E(\widehat{\Delta}_1(V)|X_i)]}_{D_N^{h1}} + \underbrace{N^{-1}\sum_{i=1}^N t_i[\widehat{E}(\Delta_2(V)|X_i) - E(\Delta_2(V)|X_i)]}_{D_N^{h2}}$$
$$+ \underbrace{N^{-1}\sum_{i=1}^N t_i[\widehat{E}(\Delta_3|X_i) - E(\Delta_3|X_i)]}_{D_N^{h3}} + o_p(N^{-1/2})$$

In what follows, only the first three components need to be analyzed separately.

For D_N^{h1}, we can represent it as a third-order U-statistic,

$$\begin{aligned}
D_N^{h1} &= \frac{1}{N(N-1)}\sum_{i,j}^N t_i f(X_i)^{-1} K_{i,j}^X[\widehat{\Delta}_{1j} - E(\widehat{\Delta}_1|X_i)] + o_p(N^{-1/2}) \\
&= \frac{1}{N(N-1)}\sum_{i,j}^N t_i f(X_i)^{-1} K_{i,j}^X \widehat{\Delta}_{1j} + o_p(N^{-1/2}) \\
&= \frac{1}{N(N-1)(N-2)}\sum_{i,j,k}^N \frac{t_i f(X_k) K_{i,j}^X K_{k,j}^V}{f(X_i) f(X_k, V_j)}[Y_k - C(X_k, V_j)] + o_p(N^{-1/2}) \\
&= \binom{N}{3}^{-1} \sum_{k>j>i}^N \sum_{l=1}^6 h_{1l}/6 + o_p(N^{-1/2})
\end{aligned}$$

The first equality holds as we can remove the random denominator of $\widehat{E}(\widehat{\Delta}_1|X_i)$ according to Lemma A 2. The second equality is because of the fact $E[\widehat{\Delta}_1|X_i] = O(h^4)$. Substitution of $\widehat{\Delta}_1$ gives the third equality. And where in particular,

$$\begin{aligned}
h_{11} &= \frac{t_i f(X_k) K_{i,j}^X K_{k,j}^V}{f(X_i) f(X_k, V_j)}[Y_k - C(X_k, V_j)]; h_{12} = \frac{t_i f(X_j) K_{i,k}^X K_{j,k}^V}{f(X_i) f(X_j, V_k)}[Y_j - C(X_j, V_k)] \\
h_{13} &= \frac{t_j f(X_k) K_{j,i}^X K_{k,i}^V}{f(X_j) f(X_k, V_i)}[Y_k - C(X_k, V_i)]; h_{14} = \frac{t_k f(X_j) K_{k,i}^X K_{j,i}^V}{f(X_k) f(X_j, V_i)}[Y_j - C(X_j, V_i)] \\
h_{15} &= \frac{t_k f(X_i) K_{k,j}^X K_{i,j}^V}{f(X_k) f(X_i, V_j)}[Y_i - C(X_i, V_j)]; h_{16} = \frac{t_j f(X_i) K_{j,k}^X K_{i,k}^V}{f(X_j) f(X_i, V_k)}[Y_i - C(X_i, V_k)]
\end{aligned}$$

It is easy to see that $E(h_{1l}|W_i) = O(h^4)$ for $l = \{1, 2, 3, 4\}$ and

$$E(h_{15}|W_i) = E(h_{16}|W_i) = E[t(X)|V_i] f^*(X_i, V_i)[Y_i - C(X_i, V_i)] + O(h^4)$$

It is also true that $\sum_{l=1}^6 E|h_{1l}|^2 = O(h^{-4}) = o(N^2)$ by Assumption A-5. Then standard

U-statistic theorem follows,

$$D_N^{h11} = N^{-1}\sum_{i=1}^{N} E[t(X)|V_i]f^*(X_i,V_i)\epsilon_i + o_p(N^{-1/2})$$

For D_N^{h2}, we can also represent it as a third-order U-statistic,

$$\begin{aligned}
D_N^{h2} &= N^{-1}\sum_{i=1}^{N} t_i[\widehat{E}(\Delta_2|X_i) - E(\Delta_2|X_i)] \\
&= N^{-1}\sum_{i,j}^{N} t_i f(X_i)^{-1} K_{i,j}^X[\Delta_{2j} - E(\Delta_2|X_i)] + o_p(N^{-1/2}) \\
&= N^{-1}\sum_{i,j}^{N} t_i f(X_i)^{-1} K_{i,j}^X[m_1(X_j) - m_1(X_i)]/N + o_p(N^{-1/2}) \\
&= o_p(N^{-1/2})
\end{aligned}$$

The second equality holds as we can remove the random denominator of $\widehat{E}(\Delta_2|X_i)$ according to Lemma A 2. Substitution of Δ_2 gives the third equality.

For D_N^{h13},

$$\begin{aligned}
D_N^{h13} &= -N^{-1}\sum_{i=1}^{N} t_i[\widehat{E}(\Delta_3|X_i) - E(\Delta_3|X_i)] \\
&= -N^{-1}\sum_{i,j}^{N} t_i f(X_i)^{-1} K_{i,j}^X[Y_j/N - E(Y|X_i)/N] + o_p(N^{-1/2}) \\
&= o_p(N^{-1/2})
\end{aligned}$$

So, combining all above, D_N^h follows.

□

1.2 Intermediate Lemmas

Lemma A 1 U-statistic. Serfling (2009)**.** *A "mth-order" U-statistic of the form*

$$U_N \equiv \binom{N}{m}^{-1} \sum_{i_1=1}^{N-m+1} \sum_{i_2>i_1}^{N-m+2} \cdots \sum_{i_m>i_{m-1}}^{N} p_N(W_{i_1},\ldots,W_{i_m})$$

where p_N is a symmetric in $W_{i_1}, \ldots, W_{i_m}$. Suppose that $\mathbb{E}||p_N(W_i, W_j)||^2 = o(N^{m-1})$. Also define

$$\begin{aligned} r_N(W_i) &\equiv \mathbb{E}[p_N(W_{i_1}, \ldots, W_{i_m})|W_{i_1}] \\ \theta_N &\equiv \mathbb{E}[r_N(W_i)] = \mathbb{E}[p_N(W_{i_1}, \ldots, W_{i_m})] \\ \widehat{U}_N &\equiv \theta_N + mN^{-1}\sum_{i=1}^{N}[r_N(W_{i_1}) - \theta_N] \end{aligned}$$

where θ_N is assumed to exist and $\widehat{U}_N$ is called the "projection" of U_N. Then

$$U_N - \widehat{U}_N = o_p(N^{-1/2})$$

Lemma A 2 Recursive Bias-reducing Conditional Mean Estimators**.** *Suppose $E|(Y|U)|^2 < \infty$, $\inf_{\mathcal{U}} f(u) > 0$ where $f(\cdot)$ is the density function of U that is everywhere positive, $d < 4$. For each $u \in \mathcal{U}$, the bias correction estimator $\widehat{C}(u) = \widehat{E}(Y|U = u)$, then it follows that*

$$\begin{aligned} &a). \quad |\widehat{C}(u) - C(u)| = f_U(u)^{-1}\left|N^{-1}\sum_{i=1}^{N} K^U_{i,\cdot}(\widetilde{Y}_i - C(u))\right| + o_p(N^{-1/2}) \\ &b). \quad N^{-1}\sum_{i=1}^{N} K^U_{i,\cdot}(\widetilde{Y}_i - C(u)) = N^{-1}\sum_{i=1}^{N} K^U_{i,\cdot}\epsilon_i + O(h^4) \end{aligned}$$

where $\widetilde{Y}_i = Y_i - \widehat{\delta}_i(u)$, $\widehat{\delta}_i(u) = \widehat{C}_0(U_i) - \widehat{C}_0(u)$ and $\epsilon_i = Y_i - C(U_i)$. $\widehat{C}_0(\cdot)$ is the preliminary conditional expectation estimator.

Proof. See Shen and Klein (2017) for detailed proof. For a). one must show that

$$DC = \left|\frac{N^{-1}\sum_{i=1}^{N} K^U_{i,\cdot}\widetilde{Y}_i}{\widehat{f}_U(u)} - C(u)\right|\left|\frac{\widehat{f}_U(u)}{f_U(u)} - 1\right| = o_p(N^{-1/2})$$

Under the well-known nonparametric rate, one would have

$$|\widehat{f}_U(u) - f_U(u)| = O_p(h^2 + N^{-1/2}h^{-d/2}); \left|N^{-1}\sum_{i=1}^{N} K^U_{i,\cdot}\widetilde{Y}_i - f_U(u)C(u)\right| = O_p(h^2 + N^{-1/2}h^{-d/2})$$

From this, let $c_f = \sup_u \widehat{f}(u)$,

$$\begin{aligned} \left|\frac{N^{-1}\sum_{i=1}^{N} K^U_{i,\cdot}\widetilde{Y}_i}{\widehat{f}_U(u)} - C(u)\right| &\leq c_f\left\{\left|N^{-1}\sum_{i=1}^{N} K^U_{i,\cdot}\widetilde{Y}_i - f_U(u)C(u)\right| + C(u)\left|\widehat{f}_U(u) - f_U(u)\right|\right\} \\ &= O_p(h^2 + N^{-1/2}h^{-d/2}) \end{aligned}$$

Then by Cauchy-Schwartz inequality, we have

$$DC \leq \sqrt{\left(\frac{N^{-1}\sum_{i=1}^N K_{i,\cdot}^U \widetilde{Y}_i}{\widehat{f}_U(u)} - C(u)\right)^2 \left(\frac{\widehat{f}_U(u)}{f_U(u)} - 1\right)^2} = O(h^4 + N^{-1}h^{-d} + N^{-1/2}h^{2-d/2})$$

Suppose $h = O(N^{-r})$, to make sure $\sqrt{N}DC = o_p(N^{-1/2})$ hold, $1/8 < r < 1/2d$ and generally requires $d < 4$.

For b). To begin with, first add and subtract $K_{i,\cdot}^U \delta_i(u)$ and use the fact $\delta_i(u) = C(U_i) - C(u)$,

$$\begin{aligned}
N^{-1}\sum_{i=1}^N K_{i,\cdot}^U(\widetilde{Y}_i - C(u)) &= N^{-1}\sum_{i=1}^N K_{i,\cdot}^U(Y_i - \delta_i(u) - C(u) - \widehat{\delta}_i(u) + \delta_i(u)) \\
&= N^{-1}\sum_{i=1}^N K_{i,\cdot}^U(Y_i - C(U_i) - \widehat{\delta}_i(u) + \delta_i(u)) \\
&= N^{-1}\sum_{i=1}^N K_{i,\cdot}^U \epsilon_i - N^{-1}\sum_{i=1}^N K_{i,\cdot}^U(\widehat{\delta}_i(u) - \delta_i(u))
\end{aligned}$$

And it is true that

$$\sqrt{N}^{-1/2}\sum_{i=1}^N K_{i,\cdot}^U(\widehat{\delta}_i(u) - \delta_i(u)) = o_p(N^{-1/2})$$

To see this, remember $\delta_i(u) - \delta_i(u) = O$

$$\begin{aligned}
E\left[N^{-1}\sum_{i=1}^N K_{i,\cdot}^U(\widehat{\delta}_i(u) - \delta_i(u))\right] &= E\left[K_{i,\cdot}^U(\widehat{\delta}_i(u) - \delta_i(u))\right] = O(h^2)O(h^2) = O(N^{-4r}) \\
\mathrm{Var}\left[N^{-1}\sum_{i=1}^N K_{i,\cdot}^U(\widehat{\delta}_i(u) - \delta_i(u))\right] &= N^{-1}\mathrm{Var}\left[K_{i,\cdot}^U(\widehat{\delta}_i(u) - \delta_i(u))\right] = O(N^{-2}h^{d+1})
\end{aligned}$$

As long as $1/8 < r < 1/(d+1)$, b). will hold. See Shen and Klein (2017) for higher order bias reduction. □

Lemma A 3. *Suppose that $R(\cdot)$ is any measurable function defined on $\mathbb{R}^d$ with continuous and bounded second derivatives. t_i is the quantile or trimming indicator and the density function $f_X(\cdot)$ satisfies Assumption A-2. For any $x_0 \in \mathcal{X}$, it is true that*

$$E\left[t_i^p R(X_i)K_h(X_i - x_0)\right] = t^p(x_0)R(x_0)f(x_0) + O(h^2)$$

where $t^p(x_0) \equiv \mathbf{1}\{x_0 \in \mathcal{X}_0\}$.

Proof. Define the upper bound $q_L \equiv q_X(\frac{p-1}{P})$ and lower bound $q_U \equiv q_X(\frac{p}{P})$,

$$
\begin{aligned}
E\left[t_i^p R(X_i) K_h(X_i - x_0)\right] &= \int_{q_U}^{q_L} k(\frac{x-x_0}{h}) R(x) f(x) dx \\
&= \int_{\frac{q_U - x_0}{h}}^{\frac{q_L - x_0}{h}} R(x_0 + uh) f(x_0 + uh) k(u) du \\
&= \int_{\frac{q_U - x_0}{h}}^{\frac{q_L - x_0}{h}} \left[R(x_0) f(x_0) + h(R(x_0) f(x_0))' u + h^2 (R(x) f(x))''|_{x^+} u^2/2\right] k(u) du \\
&= t^p(x_0) R(x_0) f(x_0) + \left[R(x_0) f(x_0)\right]' h \int_{\frac{q_U - x_0}{h}}^{\frac{q_L - x_0}{h}} u k(u) du + O(h^2)
\end{aligned}
$$

The second term is $o(h^2)$. □

2 An Example on Productivity Distribution

One of the most important aspects of estimating production functions is to extract measures of productivity. By studying the re-distribution of firm-specific productivity, one can examine the reallocation effects of particular trade or industrial policies, and identify the sources of productivity growth. Productivity is also interpreted as profitability or innovation potentials in many studies that focus on its relation with R&D expenditure. Previously, a common practice is to estimate or approximate the firm-specific productivity as the exponential of the "Solow" residuals from the log regression. However, the presence of the non-Hicks-neutral productivity shocks would not only pose difficulty in the interpretation of using a summarized productivity measure, but undermine its validity, as shown in the example below.

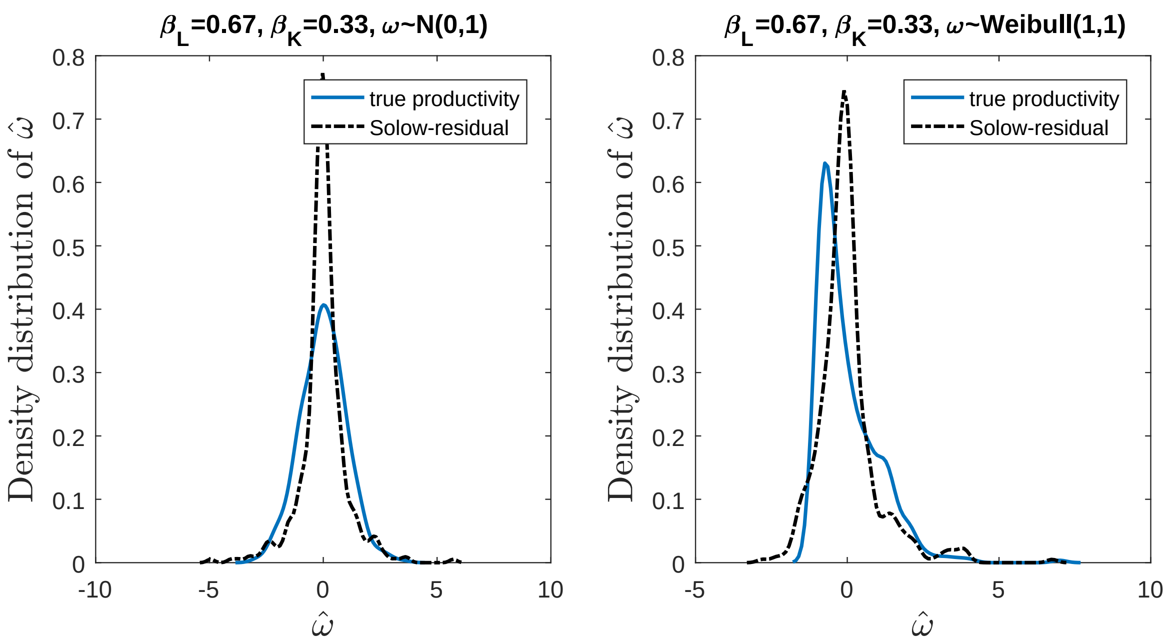
Example. Consider the random coefficient Cobb-Douglas production functions in log, a non-Hicks-neutral one. For simplicity, I focus on the cross-sectional production functions without endogenous inputs, i.e. both labor and capital are static and fixed inputs. For the purpose of demonstration, I also assume away the idiosyncratic shocks so that the only unobservable is the firm-specific productivity shock. Remember that adding more unobservables or shocks does nothing but make the interpretation hard, when only exogenous inputs exist.

$$y_t = \beta_L(\omega_t) l_t + \beta_K(\omega_t) k_t$$

Clearly, it is trivial to note that the distribution of productivity ω_t, if not degenerate, is different from the residuals that are uniformly 0 given any consistent estimator of random coefficients for each firm. A more interesting question is on how the estimated residual distribution looks like compared to the true distribution if estimated using OLS. Intuitively, it depends on the functional form of β_L and β_K as well as the distribution of firm heterogeneity and inputs. To get a first glimpse, I demonstrate it through the following simulation studies.

Suppose that a simple random coefficient CD production function with $\beta_L = 0.67 + \omega$, $\beta_K = 0.33 + \omega$, where ω is generated from either $N(0, 1)$ or Weibull$(1, 1)$. I use the same sample for labor and capital as the empirical application. For simplicity, attention is restricted to the sector of computer and electrics in 2011. The log output is generated according to $y = \beta_L l + \beta_K k + \varepsilon$, where $\varepsilon \sim N(0,1)$. In Figure 1, I plot the nonparametric kernel density estimates of true productivity versus estimated Solow-residuals. It is not hard to see that when the true productivity is normally distributed, Solow-residuals tends to display much higher than normal kurtosis. Turning to the right panel of Figure 1, the estimated Solow-residuals might have distorted skewness when the true is Weibull distributed. This example is used to illuminate one fact: severe distortions might occur to the distribution of residual productivity estimates in the presence of non-Hicks-neutral technological shocks.

Figure 1: Simulated Productivity Distributions



Note: 1. Total number of observations is 459. 2. Standardized values are plotted to ease comparison.

References

Robert J Serfling. *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons, 2009.

Chan Shen and Roger Klein. Market recursive differencing: Bias reduction with regular kernels. Working paper, 2017.