1 Identification and Consistency

Let
\[ y_i = \theta_i g(X_i) \] (1)
be a production function for area \( i \) given a \( J \)-element vector of inputs \( X_i \) and a scalar random productivity parameter \( \theta_i \). Let
\[ C(y_i) = w_i'X_i \]
be a cost function. The goal of the area hospital/decision maker is to minimize cost subject to a quantity demand:
\[
\min_{X_i} C(y_i)
\text{st } y_i = \theta_i g(X_i).
\]
We can write the firm’s optimization problem as a Lagrangean equation as
\[
L = w_i'X_i + \lambda [y_i - \theta_i g(X_i)]
\]
with derivates,
\[
\frac{\partial L}{\partial X_i} = w_i - \lambda (y) \theta_i g_X(X_i) = 0 \tag{2}
\]
(\( g_X(X_i) = \begin{bmatrix} \frac{\partial g}{\partial X_{i1}} \\ \frac{\partial g}{\partial X_{i2}} \\ \vdots \\ \frac{\partial g}{\partial X_{iJ}} \end{bmatrix}_{J \times 1} \))
is the vector of “marginal products” before \( \theta_i \).
\[
\frac{\partial L}{\partial X} = y_i - \theta_i g(X_i).
\]
Equation (2) implies that

\[ X_i = g_X^{-1} \left( \frac{w}{\lambda(y) \theta_i} \right) \]  

(3)

if \( \theta_i \) is known at the time of the input decision and

\[ X_i = g_X^{-1} \left( \frac{w}{\lambda(y) E \theta_i} \right) \]  

(4)

if \( \theta_i \) is not known at the time of the input decision (where

\[ g_X^{-1}(z) = \begin{pmatrix} \left( \frac{\partial g}{\partial X_{i1}} \right)^{-1}(z_1) \\ \left( \frac{\partial g}{\partial X_{i2}} \right)^{-1}(z_2) \\ \vdots \\ \left( \frac{\partial g}{\partial X_{iJ}} \right)^{-1}(z_I) \end{pmatrix} \]

for

\[ z_j = \frac{w_j}{\lambda(y) \theta_i} \text{ or } \frac{w_j}{\lambda(y) E \theta_i}, \]

depending on whether \( \theta_i \) is known).

Let \( \hat{g}(X) \) be a nonparametric kernel estimator of \( g(\cdot) \). For example, we might set

\[ \hat{g}(X) = \frac{\sum y_i K(X_i - X)}{\sum K(X_i - X)}. \]  

(5)

Define

\[ g_j(X_i) = \frac{\partial g(X_i)}{\partial X_{ij}}. \]

**Theorem 1**

\[ \text{plim} \hat{g}_j(X) > g_j(X) \quad \forall j \]

if equation (3) is the appropriate input decision, and

\[ \text{plim} \hat{g}_j(X) = g_j(X) \quad \forall j \]

if equation (4) is the appropriate input decision.

**Proof.** Think of equation (5) as

\[ \hat{g}(X) = \sum y_i K^*(X_i - X) \]  

(6)

where

\[ K^*(X_i - X) = \frac{K(X_i - X)}{\sum K(X_i - X)}. \]
Substituting equation (1) into equation (6) leads to
\[ \hat{g}(X) = \sum \theta_i g(X_i) K^*(X_i - X). \]

A first order Taylor series approximation implies
\[ \hat{g}(X) = \sum \theta_i \left[ g(X_i (E\theta_i)) + g'_X(X_i (E\theta_i)) X_i \theta \right] K^*(X_i - X) \]
\[ = \sum \theta_i g(X_i (E\theta_i)) K^*(X_i - X) + \sum \theta_i g'_X(X_i (E\theta_i)) X_i \theta \theta - E\theta_i) K^*(X_i - X) \]
where
\[ g'_X(X_i (E\theta_i)) = \frac{\partial}{\partial X} g(X_i (E\theta_i)), \]
\[ X_i \theta = \frac{\partial X_i \theta}{\partial \theta}. \]

Taking plims leads to
\[ \text{plim} \hat{g}(X) = \text{plim} \sum \theta_i g(X_i (E\theta_i)) K^*(X_i - X) \]
\[ + \text{plim} \sum \theta_i g'_X(X_i (E\theta_i)) X_i \theta \theta - E\theta_i) K^*(X_i - X) \]
\[ = E \theta g(X (E\theta)) + g'_X(X (E\theta)) EX \theta \theta (\theta - E\theta). \]

If \( X \) is exogenous, then \( X \theta = 0 \), and
\[ \text{plim} \hat{g}_X(X) = E \theta g_X (X (E\theta)) = g_X(X). \]

If \( X \) is endogenous, then differentiation of equation (2) with respect to \( \theta \) implies that
\[ X \theta = -[\theta g_{XX} (X)]^{-1} g_X(X) > 0. \]

Plugging \( X \theta \) into equation (7) leads to
\[ \text{plim} \hat{g}_X(X) = E \theta g_X (X (E\theta)) + g'_X(X (E\theta)) EX \theta \theta (\theta - E\theta) \]
\[ = E \theta g_X (X (E\theta)) - g'_X(X (E\theta)) E [g_{XX} (X)]^{-1} g_X(X) (\theta - E\theta) \]
\[ = E \theta g_X (X (E\theta)) - E [g_X(X) (\theta - E\theta)] > E \theta g_X (X (E\theta)) \]

because \( X \theta > 0 \) and \( g_{XX}(X) < 0. \)

**Theorem 2** If either \( X \) is exogenous or \( g(\cdot) \) is homothetic in \( X \), then
\[ \text{plim} \frac{\partial g(X)}{g_k(X)} = \frac{g_j(X)}{g_k(X)} \forall j, k. \]
Proof. If \( X \) is exogenous, then \( \hat{g}_X(X) \) is consistent which implies the ratio is consistent. If \( g(\cdot) \) is homothetic in \( X \), then the asymptotic bias in \( \hat{g}_X(X) \) can be written as

\[
E[g_X(X)(\theta - E\theta)] = g_X(X(E\theta))E\left[\frac{g_X(X)}{g_X(X(E\theta))}(\theta - E\theta)\right]
\]

with

\[
\frac{g_X(X)}{g_X(X(E\theta))} = b(\theta)1
\]

where \( b(\theta) \) is a scalar and 1 is a vector of 1’s. Thus the bias is proportional to \( g_X(X(E\theta)) \) and cancels in a ratio. \( \blacksquare \)

2 Estimation Methodology

A significant problem in estimation is that, in general, \( g(X) \) has a large dimension. This will make nonparametric estimation too difficult. We might write a production function

\[
g(X) = \tilde{g}(Z, \alpha) \psi(A, N)
\] (8)

where \( Z \) is the vector of the amount of each input excluding anesthesiologists and CRNAs, \( A \) is the amount of anesthesiologists, and \( N \) is the amount of CRNAs. We could give \( \tilde{g}(\cdot, \cdot) \) a parsimonious flexible functional form (e.g., first or second order Taylor series approximation) depending on a small set of parameters \( \alpha \) and specify \( \psi(\cdot, \cdot) \) nonparametrically. We can take logs and then use a nonparametric method of moments (Ichimura and Lee, 1991) estimation procedure by solving

\[
\min_{\alpha} \sum_i \left[ \log y_i - \log \tilde{g}(Z_i, \alpha) - \log \hat{\psi}(A_i, N_i) \right]^2
\] (9)

where

\[
\log \hat{\psi}(A_i, N_i) = \frac{\sum_{k \neq i} [\log y_i - \log \tilde{g}(Z_i, \alpha)] K(A_i - A_k, N_i - N_k)}{\sum_{k \neq i} K(A_i - A_k, N_i - N_k)}.
\]

To construct the covariance matrix, define

\[
\bar{Z}_i = \frac{\partial \log \tilde{g}(Z_i, \alpha)}{\partial \alpha} + \frac{\partial \log \hat{\psi}(A_i, N_i)}{\partial \alpha}
\]

and

\[
\varphi_i(\alpha) = \log y_i - \log \tilde{g}(Z_i, \alpha) - \log \hat{\psi}(A_i, N_i).
\]
Consider a Taylor series approximation of the condition that $\tilde{Z}_i$ should be orthogonal to $\varphi_i(\alpha)$:

$$\frac{1}{n} \sum_i \left[ \tilde{Z}'_i \varphi_i (\tilde{\alpha}) - \tilde{Z}'_i \varphi_i (\tilde{\alpha}) (\tilde{\alpha} - \alpha) \right].$$

Then

$$\hat{\alpha} - \alpha = \left[ \frac{1}{n} \sum_i -\tilde{Z}'_i \varphi_i (\tilde{\alpha}) \right]^{-1} \left[ \frac{1}{n} \sum_i \tilde{Z}'_i \varphi_i (\tilde{\alpha}) \right],$$

and

$$\text{Cov} (\hat{\alpha}) = \left[ \frac{1}{n} \sum_i \tilde{Z}'_i \varphi_i (\tilde{\alpha}) \right]^{-1} \left[ \frac{1}{n} \sum_i \tilde{Z}'_i \varphi_i (\tilde{\alpha}) \varphi'_i (\tilde{\alpha}) \tilde{Z}_i \right] \left[ \frac{1}{n} \sum_i \varphi_i (\tilde{\alpha})' \tilde{Z}_i \right]^{-1}.$$ 

### 3 Allowing for Effects Across Nearby Counties

Let $d_{ik}$ be the distance between two counties $i$ and $k$. Let $\phi (d_{ik})$ be a function (somewhat like a kernel function) such that

$$\arg \max_d \phi (d) = 0,$$

$$\frac{\partial \phi (d)}{\partial d} \leq 0,$$

$$\phi (d) = 0 \quad \forall d : |d| \geq d_{\text{max}}.$$ 

Then we can generalize equation (5) to

$$\tilde{g}^* (X^*) = \frac{\sum y_i^* K (X_i^* - X^*)}{\sum K (X_i^* - X^*)}$$

where

$$y_i^* = \frac{\sum_k \phi (d_{ik}) y_k}{\sum_k \phi (d_{ik})};$$

$$X_i^* = \frac{\sum_k \phi (d_{ik}) X_k}{\sum_k \phi (d_{ik})}.$$ 

Note that $(y_i^*, X_i^*)$ satisfy a resource constraint (inputs used in one county can’t be used in another).

**Theorem 3** $\sum_i X_i^* = \sum_i X_i$. 

5
Proof.

\[ \sum_i X_i^* = \sum_i \sum_k \phi(d_{ik}) X_k = \sum_i \sum_k \phi(d_{ik}) \sum_k X_k \]
\[ = \sum_k \sum_i \phi(d_{ik}) X_k = \sum_i (\sum_k \phi(d_{ik})) (\sum_k X_k) \]
\[ = \sum_k X_k = \sum_i X_i. \]

Also, if we parameterize \( \phi(\cdot) \) subject to the conditions in equation (11), we can estimate the parameters by using an objective function, such as
\[ \min \sum_i [y_i^* - \tilde{g}^*(X_i^*)]^2 \]
as in Ichimura and Lee (1991). Also, it is straightforward how to generalize equation (9) to allow for cross county effects similarly: just replace \((y_i, Z_i, A_i, N_i)\) with corresponding \((y_i^*, Z_i^*, A_i^*, N_i^*)\).

4 Adjustment of the Covariance Matrix for Nearby County Correlations

If we start with equation (10) and square it, we get
\[ \text{Cov}(\hat{\alpha} - \alpha) = \text{plim} \left[ \frac{1}{n} \sum_i \tilde{Z}_i \varphi_{i\alpha}(\hat{\alpha}) \right]^{-1} \cdot \]
\[ \text{plim} \left[ \frac{1}{n} \sum_i \sum_j \tilde{Z}_i \varphi_i(\hat{\alpha}) \varphi_j(\hat{\alpha}) \tilde{Z}_j \right] \text{plim} \left[ \frac{1}{n} \sum_i \varphi_{i\alpha}(\hat{\alpha})' \tilde{Z}_i \right]^{-1} \]
\[ = \text{plim} \left[ \frac{1}{n} \sum_i \tilde{Z}_i \varphi_{i\alpha}(\hat{\alpha}) \right]^{-1} \cdot \]
\[ \text{plim} \left[ \frac{1}{n} \sum_i \sum_j \sigma_{ij} \tilde{Z}_i \tilde{Z}_j \right] \text{plim} \left[ \frac{1}{n} \sum_i \varphi_{i\alpha}(\hat{\alpha})' \tilde{Z}_i \right]^{-1} \]
where \( \sigma_{ij} = E \varphi_i(\hat{\alpha}) \varphi_j(\hat{\alpha}) \). In the earlier analysis, we assume that \( \sigma_{ij} = 0 \) if \( i \neq j \). But, \( \sigma_{ij} \) may be nonzero because a) our smoothing methodology causes correlation among geographically nearby counties and b) the errors in \( \varphi(\hat{\alpha}) \) may have been correlated naturally because of geographical proximity. The possibility of (b) suggests that a straightforward correction controlling for the correlation induced by smoothing would not be sufficient. An alternative,
following the lead of Newey and West (1987), is to estimate

\[ \hat{\sigma}_{ij} = \begin{cases} \varphi_i(\hat{\alpha}) \varphi_j(\hat{\alpha}) & \text{if } \phi(d_{ik}) > 0 \\
0 & \text{if } \phi(d_{ik}) = 0 \end{cases}. \]

While \( \hat{\sigma}_{ij} \) is not a consistent estimate of \( \sigma_{ij} \), as in Newey and West (1987),
\[ \frac{1}{n} \sum_i \sum_j \hat{\sigma}_{ij} Z_i^t Z_j^t \]
is a consistent estimator of \( \text{plim} \left[ \frac{1}{n} \sum_i \sum_j \sigma_{ij} Z_i^t Z_j^t \right] \) for the same reasons.

5 Covariance Matrix for \( \hat{\psi} \)

Let \( x = (A, N) \). Starting from Pagan and Ullah (1999), Theorem 3.5,

\[ (nh)^{1/2} \left[ \hat{\psi}(x) - E\hat{\psi}(x) \right] = \hat{f}^{-1}(0) (nh)^{-1/2} \sum_i K_i(x) u_i. \]

Therefore,

\[
\text{Cov} \left[ \hat{\psi}(x_1), \hat{\psi}(x_2) \right] = \hat{f}^{-2}(0) \text{plim} \left[ (nh)^{-1} \sum_i \sum_j K_i(x_1) K_j(x_2) u_iu_j \right] = \hat{f}^{-2}(0) \left[ (nh)^{-1} \sum_i \sum_j K_i(x_1) K_j(x_2) \sigma_{ij} \right].
\]

Note that, if \( \sigma_{ij} = 0 \) for all \( i \neq j \), then \( \text{Cov} \left[ \hat{\psi}(x_1), \hat{\psi}(x_2) \right] = 0 \) for all \( x_1 \neq x_2 \) because \( K_i(x_1) K_i(x_2) = 0 \).

Define

\[ \psi^*(x) = \begin{pmatrix} \hat{\psi}(x_1, x_2) \\ \hat{\psi}(x_1 + \delta_1, x_2) \\ \hat{\psi}(x_1, x_2 + \delta_2) \end{pmatrix} \]

Consider

\[ r(\psi^*(x), x) = \log \left[ \frac{\exp(\psi(x_1 + \delta_1, x_2)) - \exp(\psi(x_1, x_2))}{\exp(\psi(x_1, x_2 + \delta_1)) - \exp(\psi(x_1, x_2))} \right] \]

with estimator

\[ \hat{r}(\psi^*(x), x) = \log \left[ \frac{\exp(\hat{\psi}(x_1 + \delta_1, x_2)) - \exp(\hat{\psi}(x_1, x_2))}{\exp(\psi(x_1, x_2 + \delta_1)) - \exp(\psi(x_1, x_2))} \right]. \]
Then

\[ \text{Var} \left[ \hat{r} (\psi^* (x), x) \right] = r_{\psi} (x) \text{Cov} [\psi^* (x)] r_{\psi} (x) \]

where

\[
r_{\psi} (x) = \frac{\partial r (x)}{\partial \psi^* (x)} = \left( \begin{array}{c}
-1 - r(x) \\
\exp \left\{ \psi(x_1 + \delta_1, x_2) - \psi(x_1, x_2) \right\} - 1 \\
\exp \left\{ \psi(x_1, x_2) - \psi(x_1, x_2) \right\} - 1 \\
\exp \left\{ \psi(x_1, x_2 + \delta_2) - \psi(x_1, x_2) \right\} - 1
\end{array} \right)
\]

and \( \text{Cov} [\psi^* (x)] \) is given above.

References

