Kernel Estimation For Dyadic Data

Bryan Graham, Fengshi Niu, and
James Powell, UC Berkeley

Toulouse School of Economics
May 21, 2019

Topics

0. Estimation of mean using dyadic data.

1. Kernel density estimation using dyadic data.

2. Kernel regression estimation using dyadic data.

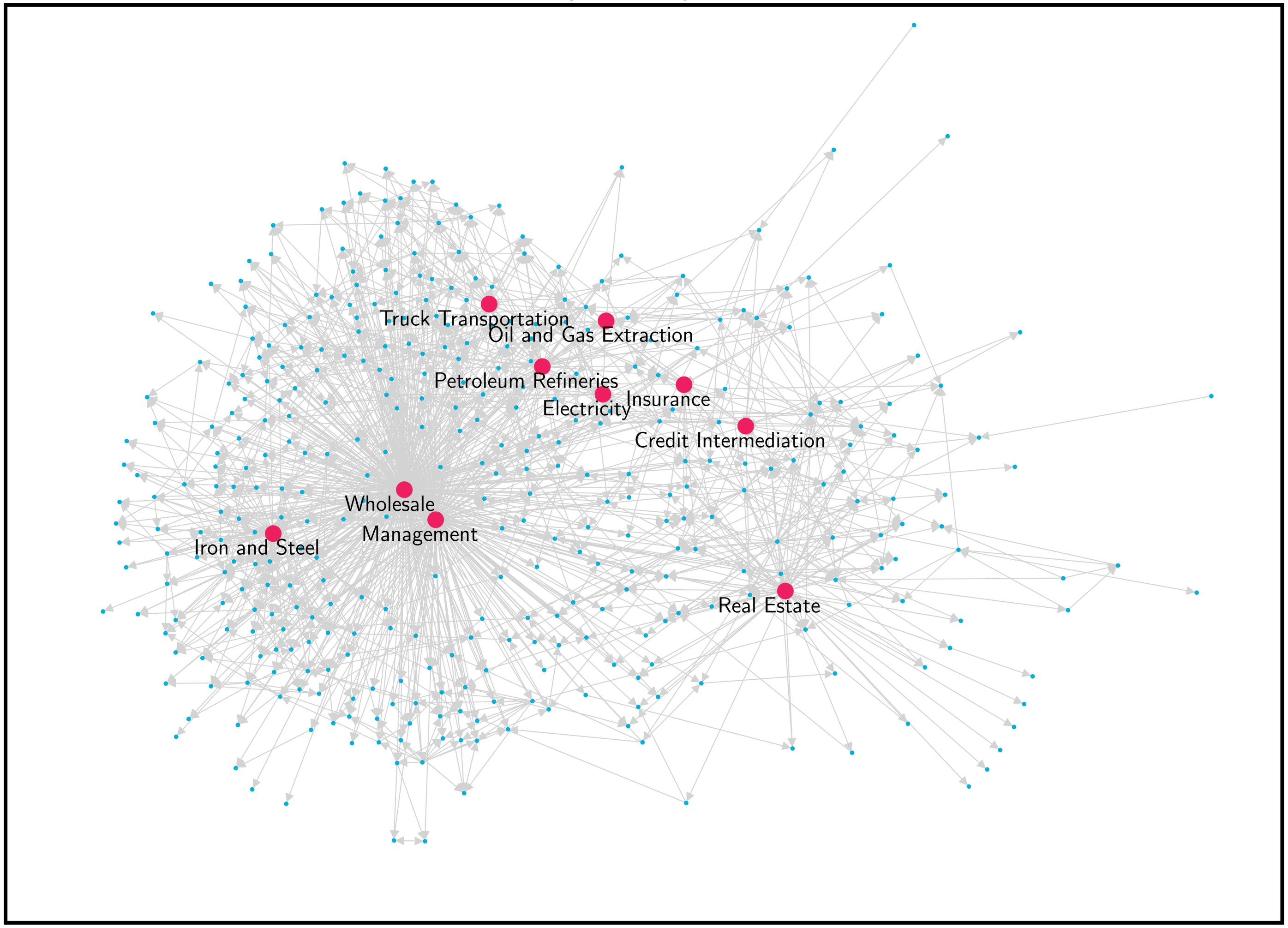
(3. Estimation of semiparametric network formation model.)

(All projects "in progress.")

Networks: Economic Examples

1. Trade/migration flows between countries
2. Buyer-supplier networks among firms
3. Research collaboration among scientists
4. Risk-sharing across households
5. Friendship/acquaintance networks
6. Phone call durations between customers

United States Input-Output Network, 2007



Source: Bureau of Economic Analysis (BEA) and author's calculations.

Raw data available at https://www.bea.gov/industry/input-output-accounts-data (Accessed September 2018)

Inter-Industry Linkages

- Input-output tables give dollar amounts of inputs Y_{ij} purchased from one industry to produce a dollar of output in another industry (cf., Miller and Blair 2009, Acemoglu *et al.* 2012).

- Graph based on U.S. Bureau of Economic Analysis 2007 Benchmark Industry-by-Industry table ($N = 388$ industries).

- Here Y_{ij} is "directed" variable (i.e., $Y_{ij} \neq Y_{ji}$ in general) indexed by "dyad" (i,j). May be interested in "undirected" dyadic variable, e.g., $\tilde{Y}_{ij} \equiv Y_{ij} + Y_{ji} = \tilde{Y}_{ji}$ or $D_{ij} = 1\{\tilde{Y}_{ij} > 0\}$.

Possible Parameters of Interest

- *Expected Value:* $\mu = E[Y_{ij}]$, assuming $E[|Y_{ij}|] < \infty$.
- *Density Function:* $f(y)$, assuming

$$\Pr\{Y_{ij} \leq y\} = \int_{-\infty}^{y} f(u)du.$$

- *Conditional Mean:* $g(w) = E[Y_{ij}|W_{ij} = w]$, some W_{ij}.
- *Other Parameters:* Moments, quantiles, etc.

Outline of Results

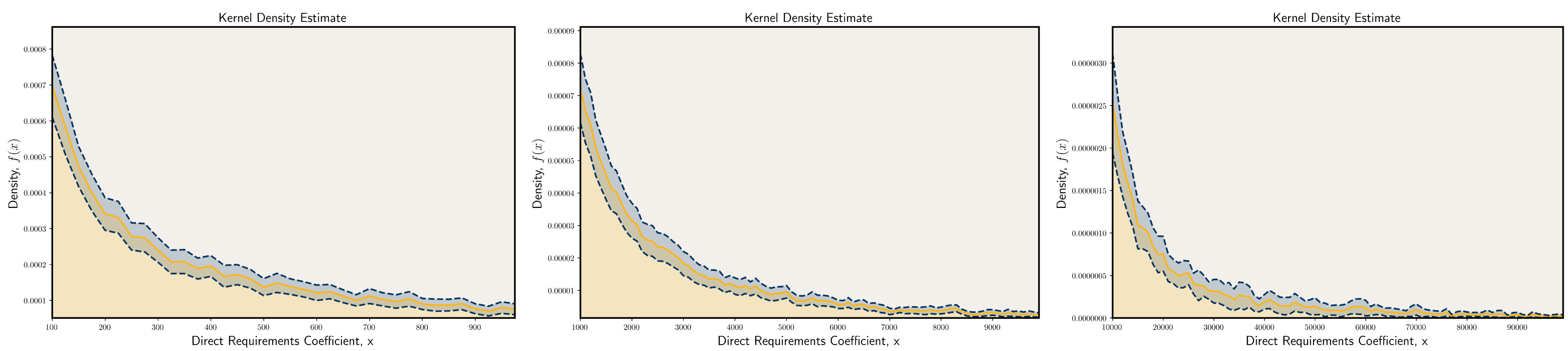
- Identification and estimation is straightforward via the "analogy principle." Get obvious expressions for kernel density and regression estimators.

- The variance formulae and estimators are also straightforward, but must account for dependence across observation with common indices (Fafchamps and Gubert 2007).

- Estimators should be asymptotically normal but with slow convergence rate. Relevant "sample size" is N, not $n \equiv N(N-1)/2 = O(N^2)$.

- Get same $\sqrt{N}$ rate for $\hat{\mu} = \bar{Y} = \hat{E}[Y_{ij}]$ and for $\hat{g}(w) = \hat{E}[Y_{ij}|W_{ij} = w]$ for continuous regressor W_{ij} (and for estimator $\hat{f}_W(w)$ of density of W_{ij}).

Direct Input Requirements in US$ per US$ One Million of Output



Model for Dyadic Outcome Variable

"**Undirected**" **Dyadic R**.**V**.: For $i,j \in \{1,\ldots,N\}$,

$$Y_{ij} = Y_{ji} = \xi(A_i, A_j, U_{ij}) = \xi(A_j, A_i, U_{ij}),$$

where it is assumed that:

1. Y_{ij} is scalar (convenience);
2. Both $\{A_i\}$ and $\{U_{ij}\}$ are i.i.d.;
3. $U_{ij} = U_{ji}$ is independent of A_i for all i and j;
4. $E[|Y_{ij}|^{2+\eta}] < \infty$, some $\eta > 0$.

Remarks:

- For "directed" case $Y_{ij} \neq Y_{ji}$, would assume

$$Y_{ij} = \xi(A_i, B_j, U_{ij}),$$

where $\{A_i\}$, $\{B_j\}$, $\{U_{ij}\}$ mutually independent and i.i.d.

- Assumptions imply that dependence of Y_{ij} is only "local,"

i.e., Y_{ij} is independent of Y_{kl} if $k \notin \{i,j\}$ and $l \notin \{i,j\}$, that is, Y_{ij} and Y_{kl} are independent unless they share an index (or two).

- Assumption that $\{A_i\}$ (and $\{B_j\}$) are i.i.d. is quite strong – rules out effects of other nodes in network (unless they affect all $\{A_i\}$ identically). Like notorious "IIA" condition for multinomial logit.

- In addition to the leading "nondegenerate" case, we will also consider the "degenerate" cases where $Y_{ij} = \xi(A_i, A_j)$ (e.g. $Y_{ij} = \|A_i - A_j\|^2$), and where $Y_{ij} = \xi(U_{ij})$ (so Y_{ij} i.i.d. over all i,j pairs).

Estimation of Mean of Outcome Variable

Sample Mean:

$$\bar{Y} = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} Y_{ij} \equiv \frac{1}{n} \sum_{i<j} Y_{ij},$$

where $n \equiv \binom{N}{2} = O(N^2)$.

Expected Value of $\bar{Y}$:

$$\bar{Y} = \frac{1}{n} \sum_{i<j} E[Y_{ij}] = \mu.$$

Variance of $\bar{Y}$:

$$
\begin{aligned}
Var[\bar{Y}] &= Var\left(\frac{1}{n} \sum_{i<j} Y_{ij} \right) \\
&= \left(\frac{1}{n}\right)^2 \sum_{i<j} \sum_{k<l} Cov(Y_{ij}, Y_{kl}) \\
&= \frac{1}{n}[Var(Y_{ij}) + 2(N-2) \bullet Cov(Y_{ij}, Y_{ik})] \\
&= \frac{4}{N} Cov(Y_{ij}, Y_{ik}) + \frac{1}{n}(Var(Y_{ij}) - 2Cov(Y_{ij}, Y_{ik})), \\
&= \frac{4}{N} \Sigma_1 + \frac{1}{n}(\Sigma_2 - 2\Sigma_1),
\end{aligned}
$$

where

$$\Sigma_2 \equiv Cov(Y_{ij}, Y_{ij}) = Var(Y_{ij})$$

$$= E[\xi(A_i, A_j, U_{ij})]^2 - \mu^2,$$

and

$$\Sigma_1 \equiv Cov(Y_{ij}, Y_{ik})$$

$$= Var(E[\xi(A_i, A_j, U_{ij})|A_i])$$

$$= E[(E[Y_{ij}|A_i])^2] - \mu^2.$$

Therefore,

$$
\begin{aligned}
Var(\bar{Y}) &= \frac{4}{N}\Sigma_1 + \frac{1}{n}(\Sigma_2 - 2\Sigma_1) \\
&= O\left(\frac{1}{N}\right) + O\left(\frac{1}{n}\right) \\
&= O\left(\frac{1}{N}\right).
\end{aligned}
$$

Asymptotic Normality: Write

$$\bar{Y} - \mu = \frac{1}{n}\sum_{i<j}(Y_{ij} - E[Y_{ij}|A_i, A_j])$$

$$+ \frac{1}{n}\sum_{i<j}(E[Y_{ij}|A_i, A_j] - \mu).$$

First term is

$$\frac{1}{n}\sum_{i<j}(Y_{ij} - E[Y_{ij}|A_i, A_j]) = O_p\left(\frac{1}{\sqrt{n}}\right) = o_p\left(\frac{1}{\sqrt{N}}\right).$$

By projection theorem for U-statistics, second term is

$$\frac{1}{n}\sum_{i<j}(E[Y_{ij}|A_i,A_j]-\mu) = \frac{2}{N}\sum_{i=1}^{N}(E[Y_{ij}|A_i]-\mu)+o_p\left(\frac{1}{\sqrt{N}}\right)$$

so

$$\sqrt{N}(\bar{Y}-\mu)\overset{d}{\to}N(0,4\Sigma_1),$$

where

$$\begin{aligned}\Sigma_1 &\equiv Cov(Y_{ij},Y_{ik})\\ &= Var(E[Y_{ij}|A_i]).\end{aligned}$$

Estimator of Variance (Fafchamps-Gubert 2007):

$$\widehat{Var}(\bar{Y}) = \left(\frac{1}{n}\right)^2 \sum_{i<j} \sum_{k<l} d_{ijkl}(Y_{ij} - \bar{Y})(Y_{kl} - \bar{Y}),$$

where

$d_{ijkl} = 1\{(i,j,k,l)$ has one or two elements in common$\}$.

Can show

$$N \cdot \widehat{Var}(\bar{Y}) \overset{p}{\to} 4\Sigma_1.$$

Simple Example:

$$Y_{ij} = \mu + A_i + A_j + U_{ij},$$

where A_i and U_{ij} are mutually independent and $i.i.d.$ Since

$$\bar{Y} = \mu + 2\bar{A} + \bar{U},$$

$$\sqrt{N}\,\bar{A} \overset{d}{\to} N(0, \sigma_A^2),$$

$$\sqrt{n}\,\bar{U} \overset{d}{\to} N(0, \sigma_U^2), \qquad \text{so}$$

$$\sqrt{N}\,(\bar{Y} - \mu) \overset{d}{\to} N(0, 4\sigma_A^2).$$

Get the same distribution for "degenerate error" case

$$Y_{ij} = \mu + A_i + A_j.$$

"Degenerate attribute" case has

$$Y_{ij} = \mu + U_{ij},$$

$$\bar{Y} = \mu + \bar{U},$$

so

$$\sqrt{n}\,(\bar{Y} - \mu) = \sqrt{n}\,\bar{U} \overset{d}{\to} N(0, \sigma_U^2).$$

In all cases, can estimate variance of $\bar{Y}$ by

$$\widehat{Var}(\bar{Y}) = \frac{4}{N^2}\sum_{i=1}^{N}(Y_{i\cdot} - \bar{Y})^2 + \frac{1}{n^2}\sum_{i<j}(Y_{ij} - \bar{Y})^2.$$

Estimation of Density

Structural Equation for W_{ij} ("**Regressor**"):

$$W_{ij} = W_{ji} = \omega(A_i, A_j, V_{ij}) = \omega(A_j, A_i, V_{ij}).$$

Nondegenerate Case: Assume

1. W_{ij} is scalar;
2. Both $\{A_i\}$ and $\{V_{ij}\}$ are i.i.d. sequences;
3. $V_{ij} = V_{ji}$ is independent of A_i for all i and j;
4. V_{ij} is continuous, smooth density $f_V(v)$ (estimation of f);
5. $\omega(a_1, a_2, v)$ is invertible & continuously differentiable in v.

Conditional Density of W_{ij} given $A_i = a_1,\ A_j = a_2$:

By change-of-variables formula,

$$f_{Y|AA}(w|a_1,a_2) = f_U(\omega^{-1}(a_1,a_2,w)) \bullet \left| \frac{\partial\omega(a_1,a_2,\omega^{-1}(a_1,a_2,w))}{\partial v} \right|^{-1}.$$

Conditional Density of W_{ij} given $A_i = a$:

$$f_{W|A}(y|a) \equiv E[f_{W|AA}(w|a,A_i)].$$

Marginal Density of Y_{ij}:

$$f_W(w) = E[f_{W|AA}(y|A_1,A_2)] = E[f_{W|A}(y|A)].$$

Kernel Estimator of Marginal Density of W_{ij}:

$$
\begin{aligned}
\hat{f}_W(w) &= \frac{1}{n}\sum_{i<j}\frac{1}{h}K\left(\frac{w - W_{ij}}{h}\right) \\
&\equiv \frac{1}{n}\sum_{i<j}K_{ij},
\end{aligned}
$$

where

$$
K_{ij} \equiv \frac{1}{h}K\left(\frac{w - W_{ij}}{h}\right).
$$

Kernel and Bandwidth: Assume

$$\int K(v)dv = 1,$$

$$\int v^j K(v)dv = 0 \qquad \text{if} \qquad j = 1, 2, \ldots, q-1,$$

$$\int v^q K(v)du \neq 0,$$

and

$$h = h_N \to 0 \qquad \text{as} \qquad N \to 0.$$

Also assume $f_W(w)$ has q continuous derivatives.

Expectation of Numerator:

$$
\begin{aligned}
E\left[\hat{f}_W(w)\right] &= E\left[\frac{1}{h}K\left(\frac{w - W_{ij}}{h}\right)\right] \\
&= E\left[\int \frac{1}{h}K\left(\frac{w-s}{h}\right)f_W(s)ds\right] \\
&= \int K(u)f_W(w-hu)du \qquad (u = (w-s)/h), \\
&= f_W(w) + E\left[\frac{(-h)^q}{q!}\frac{d^q f_W(w)}{(dw)^q}\int v^q K(v)dv\right] + o(h^q) \\
&\equiv f_W(w) + h^q B(w).
\end{aligned}
$$

As in "monadic" case, $Bias[\hat{f}_Y(y)] = o(1/\sqrt{N})$ if $Nh^{2q} \to 0$.

(If $q = 2$, need $h = o(N^{-1/4})$.)

Variance of Numerator:

$$
\begin{aligned}
Var[\hat{f}_W(w)] &= Var\left(\frac{1}{n}\sum_{i<j} K_{ij}\right) \\
&= \left(\frac{1}{n}\right)^2 \sum_{i<j}\sum_{k<l} Cov(K_{ij}, K_{kl}) \\
&= \left(\frac{1}{n}\right)^2 [n \cdot Cov(K_{ij}, K_{ij}) + 2n \cdot Cov(K_{ij}, K_{il})] \\
&= \frac{1}{n}[Var(K_{ij}) + 2(N-2) \cdot Cov(K_{ij}, K_{il})].
\end{aligned}
$$

Variance of K_{ij}:

$$
\begin{aligned}
Var(K_{ij}) &= E[(K_{ij})^2] - \left(E[\hat{f}_W(w)]\right)^2 \\
&= \frac{1}{h^2}\int\left[K\left(\frac{w-s}{h}\right)\right]^2 f_W(s)ds + O(1) \\
&= \frac{1}{h}\int[K(u)]^2 f_W(w-hu)du + O(1) \\
&= \frac{f_W(w)}{h}\cdot\int[K(u)]^2 du + O(1) \\
&\equiv \frac{1}{h}\Sigma_2(w) + O(1),
\end{aligned}
$$

also like monadic case.

Covariance of K_{ij} and K_{il}, $j \neq l$:

$$E[K_{ij} \bullet K_{il}] = E\left[\int\int \frac{1}{h^2}\left[K\left(\frac{w-s_1}{h}\right)\right] \cdot \left[K\left(\frac{w-s_2}{h}\right)\right] \bullet f_{W|AA}(s_1|A_i, A_j) f_{W|AA}(s_2|A_i, A_l) ds_1 ds_2\right]$$

$$= E\left[\int [K(u_1)] f_{W|A}(w - hu_1|A_i) du_1 \bullet \int [K(u_2)] f_{W|A}(w - hu_2|A_i) du_2\right],$$

$$= E[f_{W|A}(w|A_i)]^2 + o(1),$$

so

$$
\begin{aligned}
Cov(K_{ij}, K_{il}) &= E[K_{ij} \bullet K_{il}] - \left(E[\hat{f}(y)]\right)^2 \\
&= \left[E[f_{W|A}(w|A_i)]^2 - (f(w))^2\right] + O(h^q) \\
&= Var(f_{W|A}(w|A_i)) \ + o(1) \\
&\equiv \Sigma_1(w) + o(1).
\end{aligned}
$$

Variance of Estimator Redux:

$$Var[\hat{f}_W(w)] = \frac{1}{n}[2(N-2) \cdot Cov(K_{12}, K_{13}) + Var(K_{12})]$$

$$= \frac{4}{N}\Sigma_1(w) + \left(\frac{1}{nh}\Sigma_2(w) - \frac{2}{n}\Sigma_1(w)\right) + O\left(\frac{h}{N}\right),$$

Mean Squared Error:

$$MSE\left(\hat{f}_W(w); h\right) = \left[\left(E[\hat{f}_W(w)]\right)^2 - f(w)\right]^2 + Var[\hat{f}_W(w)]$$

$$= O(h^{2q}) + O\left(\frac{\Sigma_2(w)}{nh}\right) + O\left(\frac{\Sigma_1(w)}{N}\right).$$

Optimal Bandwidth Rate: *If* $\Sigma_1(w) \neq 0 \neq \Sigma_2(w)$,

$$h^* = O\left(\frac{1}{n}^{\frac{1}{1+2q}}\right) = O\left(\frac{1}{N}^{\frac{2}{1+2q}}\right),$$

$$(h^*)^{2q} = O\left(N^{\frac{4q}{2q+1}}\right) = o\left(\frac{1}{N}\right).$$

so

$$MSE\left(\hat{f}_W(w);h^*\right) = O((h^*)^{2q}) + O\left(\frac{\Sigma_2(w)}{nh^*}\right) + O\left(\frac{\Sigma_1(w)}{N}\right)$$
$$= o\left(\frac{1}{N}\right) + o\left(\frac{1}{N}\right) + O\left(\frac{1}{N}\right) = O\left(\frac{1}{N}\right).$$

Will hold whenever $Nh^{2q} \to 0$ and $Nh \to \infty$.

"**Degenerate Error" Case**: If $W_{ij} = \omega(A_i, A_j)$, get same MSE rates as nondegenerate case.

"**Degenerate Attribute" Case**: If $W_{ij} = \omega(U_{ij})$, then

$$f_{W|A}(w|a_1, a_2) = f(w)$$

and

$$\Sigma_1(w) = 0,$$

so

$$MSE\left(\hat{f}_W(w); h^*\right) = O((h^*)^{2q}) + O\left(\frac{\Sigma_2(w)}{nh^*}\right)$$
$$= O\left(\frac{1}{nh}\right) = o\left(\frac{1}{N}\right),$$

the nonparametric rate for sample size n.

Asymptotic Normality: Get usual decomposition

$$\begin{aligned}\hat{f}_W(w) - f_W(w) &= \frac{1}{n}\sum_{i<j}(K_{ij} - E[K_{ij}|A_i, A_j]) \\ &\quad + \frac{1}{n}\sum_{i<j}(E[K_{ij}|A_i, A_j] - E[K_{ij}]) \\ &\quad + E[K_{ij}] - f_W(w).\end{aligned}$$

First term is

$$\frac{1}{n}\sum_{i<j}(K_{ij} - E[K_{ij}|A_i, A_j]) = O_p\left(\sqrt{\frac{1}{nh}}\right) = o_p\left(\frac{1}{\sqrt{N}}\right),$$

second term is

$$\frac{1}{n}\sum_{i<j}(E[K_{ij}|A_i,A_j]-E[K_{ij}]) \;=\; \frac{2}{N}\sum_{i=1}^{N}(E[K_{ij}|A_i]-E[K_{ij}]) \;+O_p\left(\frac{1}{\sqrt{nh}}\right),$$

by U-statistic projection, and last (bias) term is

$$E[K_{ij}]-f(y) \;=\; O(h^q) \;=\; o\left(\frac{1}{\sqrt{N}}\right).$$

So

$$\sqrt{N}\,(\hat{f}_W(w)-f_W(w)) \;\overset{d}{\to}\; N(0,4\Sigma_1(w)) \qquad \text{if} \qquad Nh\to\infty,\ Nh^{2q}\to 0.$$

Remarks:

- For "directed" case $Y_{ij} \neq Y_{ji}$, kernel in "U-statistic" is

$$\tilde{K}_{ij} = \frac{1}{2}(K_{ij} + K_{ji});$$

get additional term in $\Sigma_2(w)$ from joint dependence of U_{ij} and U_{ji}.

- Consistent estimator of asymptotic covariance same as (Fafchamps-Gubert) estimator for sample mean, replacing "$Y_{ij} - \bar{Y}$" with "$K_{ij} - \hat{f}_W(w)$".

- If $Nh \to C \in (0,\infty)$ ($h = O(1/N)$), then $nh = O(N)$ and

$$\sqrt{N}\,(\hat{f}_W(w) - f_W(w)) \overset{d}{\to} N(0, 4\Sigma_1(w) + C^{-1}\Gamma(w))$$

for some $\Gamma(w)$.

- For p –dimensional density function,

$$\begin{aligned} MSE\left(\hat{f}_W(w); h\right) &= O\left(\frac{1}{N}\right) + O\left(\frac{1}{nh^p}\right) + O(h^{2q}) \\ &= O\left(\frac{1}{N}\right) \end{aligned}$$

if $Nh^p \to \infty,\ Nh^{2q} \to 0.$

- Estimator $\hat{f}_W(w)$ behaves like semiparametric "smoothed U-statistic", but with $N = O(\sqrt{n})$ observations.

Dyadic Kernel Regression

Regression Problem: Given Y_{ij} and W_{ij}, $i,j = 1,\ldots,N,$ estimate

$$g(w) \equiv E[Y_{ij}|W_{ij} = w]$$

Nondegenerate Case: Assume

1. Y_{ij} and W_{ij} are scalar;
2. $Y_{ij} = Y_{ji} = \xi(A_i, A_j, U_{ij}) = \xi(A_j, A_i, U_{ji}),$ observable;
3. $W_{ij} = W_{ji} = \omega(A_i, A_j, V_{ij}) = \omega(A_j, A_i, V_{ji}),$ observable;
4. Both $\{A_i\}$ and $\{(U_{ij}, V_{ij})\}$ are i.i.d., mutually independent;
5. W_{ij} given $A_i = a_1,\ A_j = a_2$ has conditional density $f_{W|AA}$;
6. $g(w)$ is smooth (lots of continuous derivatives).

Kernel Estimator of $g(w)$:

$$\hat{g}(w) \equiv \frac{\sum_{i<j} K\left(\frac{w-W_{ij}}{h}\right) Y_{ij}}{\sum_{i<j} K\left(\frac{w-W_{ij}}{h}\right)}$$

$$\equiv \frac{\frac{1}{n}\sum_{i<j} K_{ij} Y_{ij}}{\hat{f}_W(w)},$$

$$K_{ij} \equiv \frac{1}{h} K\left(\frac{w - W_{ij}}{h}\right)$$

Decomposition of $\hat{g}(w)$ (Nondegenerate Case):

$$\hat{g}(w) - g(w) = \frac{\frac{1}{n}\sum_{i<j} K_{ij}(Y_{ij} - g(w))}{\hat{f}_W(w)}$$

$$= \frac{1}{\hat{f}_W(w)}\left[\hat{T}_1 + \hat{T}_2 + \hat{T}_3\right]$$

where

$$T_1 = \frac{1}{n}\sum_{i<j}(K_{ij}Y_{ij} - E[K_{ij}Y_{ij}|A_i, A_j])$$

$$= O_p\left(\frac{1}{\sqrt{nh}}\right),$$

$$
\begin{aligned}
T_2 &= \frac{1}{n}\sum_{i<j}(E[K_{ij}Y_{ij}|A_i, A_j] - E[K_{ij}Y_{ij}]) \\
&= \frac{2}{N}\sum_{i}(E[Y_{ij}|A_i, W_{ij} = w] \cdot f_{W|A}(w|A_i) - g(w)f_W(w)) + o_p\left(\frac{1}{\sqrt{N}}\right) \\
&= O_p\left(\frac{1}{\sqrt{N}}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
T_3 &= E[K_{ij}Y_{ij}] - f_W(w)g(w) \\
&= O(h^q).
\end{aligned}
$$

Rates of Convergence: Under previous assumptions on $K,\ h$,

$$\hat{g}(w) = g(w) + \frac{1}{\hat{f}_W(w)}\left[\hat{T}_1 + \hat{T}_2 + \hat{T}_3\right]$$

$$= g(w) + O_p\left(\frac{1}{\sqrt{N}}\right).$$

Asymptotic Distribution of $\hat{g}(w)$:

$$\sqrt{N}\,(\hat{g}(w) - g(w)) \overset{d}{\to} N(0, 4\Gamma_1(w)),$$

for

$$\Gamma_1(w) \equiv Var\left(\frac{E[Y_{ij}|A_i, W_{ij} = w] \bullet f_{W|A}(w|A_i)}{f_W(w)} \right).$$

if $Nh \to \infty$, $Nh^q \to 0$ as $N \to \infty$.

Remarks:

- If $Nh \to C \in (0,\infty)$ and $Nh^q \to 0$ (so $h = O(1/N)$),

$$\sqrt{N}\,(\hat{g}(w) - g(w)) \overset{d}{\to} N(0, 4\Gamma_1(w) + C^{-1}\Gamma_2(w))$$

for some $\Gamma_2(w)$.

- Surprising (to me) fact:

$$E\left[E[Y_{ij}|A_i, W_{ij} = w] \cdot \frac{f_{W|A}(w|A_i)}{f_W(w)} \right] = E[Y_{ij}|W_{ij} = w] \equiv g(w).$$

- The consistent estimator of asymptotic covariance same as (Fafchamps-Gubert) variance estimator for sample mean, replacing "$Y_{ij} - \bar{Y}$" with "$(K_{ij} Y_{ij} / \hat{f}_W(w)) - \hat{g}(w)$".

- In the "degenerate attribute" case

$$W_{ij} = \omega(A_i, A_j)$$

we (currently) impose strong restrictions on the support of W_{ij} for $k \neq j$ but asymptotic distribution is similar.

- For general case of vector W_{ij} with p_1 jointly-continuous components, if (W_{ij}, W_{ik}) has p_2 distinct continuous components,

$$\hat{g}(w) = g(w) + O_p\left(\frac{1}{\sqrt{Nh^{2p_1 - p_2}}}\right).$$

In nondegenerate case $p_2 = 2p_1$, but possibly not in degenerate case.

- "Reduced form" nonparametric regression isn't structural:

supposing

$$Y_{ij} = \xi(W_{ij}, A_i, A_j, U_{ij}),$$

$$\begin{aligned}
g(w) &\equiv E[Y_{il} | W_{il} = w] \\
&= E[E[Y_{ij} | A_i, A_j, W_{ij}] | W_{ij} = w] \\
&= \int\int \xi(w, a_1, a_2, u) dF_{AAU|W}(a_1, a_2, u | W_{ij} = w) \\
&\neq \int\int \xi(w, a_1, a_2, u) dF_A(a_1) dF_A(a_2) dF_U(u) \\
&\equiv ASF(w).
\end{aligned}$$

So we would need to eliminate confounding effects of W_{ij} on A_i and A_j (and U_{ij}) to get direct effect of W_{il} on $Y_{il.}$

The End