Econometric Theory, **21**, 2005, 1017–1025. Printed in the United States of America.
DOI: 10.1017/S0266466605050504

NOTES AND PROBLEMS

THE UNIQUENESS OF CROSS-VALIDATION SELECTED SMOOTHING PARAMETERS IN KERNEL ESTIMATION OF NONPARAMETRIC MODELS

QI LI
Texas A&M University
and
Tsinghua University

JIANXIN ZHOU
Texas A&M University

We investigate the issue of the uniqueness of the cross-validation selected smoothing parameters in kernel estimation of multivariate nonparametric regression or conditional probability functions. When the covariates are all continuous variables, we provide a necessary and sufficient condition, and when the covariates are a mixture of categorical and continuous variables, we provide a simple sufficient condition that guarantees asymptotically the uniqueness of the cross-validation selected smoothing parameters.

1. MOTIVATION AND RESULTS

The kernel method is the most popular technique used in the estimation of nonparametric/semiparametric models, and it is well known that the selection of smoothing parameters in nonparametric kernel estimation is of crucial importance. In the context of a regression model, Clarke (1975) proposes the leave-one-out least squares cross-validation method for selecting the smoothing parameters. The asymptotic optimality of this approach is studied by Härdle and Marron (1985) and Härdle, Hall, and Marron (1988) in the context of a univariate regression model, and Fan and Gijbels (1995) have studied band-

We thank a referee for the constructive comments. Address correspondence to Qi Li, Department of Economics, Texas A&M University, College Station, TX 77843-4228, USA; e-mail: qi@econmail.tamu.edu.

 0266-4666/05 $12.00

width selection in the context of local polynomial kernel regression. For a regression model with a single (univariate) continuous regressor, Härdle and Marron (1985) and Härdle et al. (1988) show that the cross-validation function has the following expression:

$$CV(h) \stackrel{def}{=} n^{-1} \sum_{i=1}^{n} [Y_i - \hat{g}_{-i}(X_i)]^2 w(X_i) = C_1 h^4 + \frac{C_2}{nh} + o_p(h^4 + (nh)^{-1}), \tag{1.1}$$

where $\hat{g}_{-i}(X_i) = \sum_{l \neq i}^{n} Y_l k((X_l - X_i)/h) / \sum_{l \neq i}^{n} k((X_l - X_i)/h)$ is the leave-one-out local constant kernel estimator of $g(X_i) \equiv E(Y_i|X_i)$, $k(.)$ is a second-order kernel function, h is the smoothing parameter, $w(.)$ is a weight function, $C_1 = \int \{\kappa_2/2[g''(x)f(x) + 2g'(x)f'(x)]\}^2 w(x) f(x)^{-1}\, dx$, $C_2 = \kappa \int \sigma^2(x) w(x)\, dx$, $\kappa_2 = \int k(v) v^2\, dv$, $\kappa = \int k(v)^2\, dv$, $g'(.)$ and $g''(.)$ denote first- and second-order derivative functions, and $\sigma^2(x) = \text{Var}(Y_i|X_i = x)$.

The terms of $C_1 h^4$ and $C_2/(nh)$ in (1.1) are the leading squared bias and variance of $CV(h)$, respectively. Let $\hat{h}$ denote the cross-validation selected smoothing parameter that minimizes $CV(h)$; then from (1.1) it is easy to show that the $\hat{h} = h_0 + o_p(h_0)$, where $h_0 = [C_2/(4C_1)]^{1/5} n^{-1/5}$. Note that C_1 is nonnegative and $C_2 > 0$. Therefore, a necessary and sufficient condition for the existence of the unique benchmark nonstochastic optimal smoothing parameter h_0 is that $C_1 > 0$. The assumption that $C_1 > 0$ puts some restrictions on $g(.)$; for example, $g(.)$ cannot be a constant function. A similar necessary and sufficient condition exists that guarantees an asymptotically uniquely defined cross-validation selected smoothing parameter in estimating a conditional probability density function (p.d.f.) with an univariate continuous conditional variable.

The cross-validation procedure can be easily extended to the multivariate (regression or p.d.f. estimation) settings for selecting the smoothing parameters. However, the conditions that ensure the uniqueness of cross-validation selected smoothing parameters become more complex. Recently, Hall, Racine, and Li (2004), Hall, Li, and Racine (2004), and Li and Racine (2003, 2004) have considered the problem of nonparametric estimation of conditional density and regression functions with mixed discrete and continuous data. They propose to use the data-driven cross-validation (CV) methods to select the smoothing parameters, and they have shown that the CV selected smoothing parameters are asymptotically equivalent to the nonstochastic optimal smoothing parameters that minimize the asymptotic weighted estimation mean square error. However, when discussing the existence of the asymptotically uniquely defined optimal smoothing parameters, Hall, Racine, and Li (2004) and Li and Racine (2004) impose overly strong conditions. In this note we provide substantially weaker sufficient conditions that guarantee the existence of the uniquely defined CV selected optimal smoothing parameters. We show that when all covariates are continuous random variables, the condition becomes necessary and sufficient for the existence of uniquely defined optimal smoothing parameters.

We consider a nonparametric regression model with mixed discrete and continuous covariates:

$$Y_i = g(X_i) + u_i, \tag{1.2}$$

where $g(\cdot)$ has an unknown functional form, $E(u_i|X_i) = 0$, $X_i = (X_i^c, X_i^d)$, X_i^d is a $q \times 1$ vector of regressors that assume discrete values, and $X_i^c \in R^p$ are the remaining continuous regressors. We use X_{ij}^d to denote the jth component of X_i^d, and we assume that X_{ij}^d takes $c_j \geq 2$ different values, that is, $X_{ij}^d \in \{0,1,\ldots,c_j - 1\}$ for $j = 1,\ldots,q$. We use $D = \prod_{j=1}^q \{0,1,\ldots,c_j - 1\}$ to denote the range assumed by x^d. We are interested in estimating $g(x) = E(Y_i|X_i = x)$ by the nonparametric kernel method. We use $f(x) = f(x^c, x^d)$ to denote the joint density function. For $x^c = (x_1^c,\ldots,x_p^c)$ we use the product kernel: $K^c(x^c, X_i^c) = \prod_{j=1}^p (1/h_j) k((x_j^c - X_{ij}^c)/h_j)$, where k is a symmetric, univariate density function and $0 < h_j < \infty$ is the smoothing parameter for x_j^c. For a discrete regressor we define, for $1 \leq j \leq q$,

$$l(X_{ij}^d, x_j^d, \lambda_j) = \begin{cases} 1 & \text{if } X_{ij}^d = x_j^d, \\ \lambda_j & \text{if } X_{ij}^d \neq x_j^d, \end{cases} \tag{1.3}$$

where $0 \leq \lambda_j \leq 1$ is the smoothing parameter for x_j^d. Therefore, the product kernel for $x^d = (x_1^d,\ldots,x_q^d)$ is given by $K^d(x^d, X_i^d) = \prod_{j=1}^q l(X_{ij}^d, x_j^d, \lambda_j)$. The kernel function for the mixed regressors $x = (x^c, x^d)$ is simply the product of K^c and K^d, that is, $\mathcal{K}(x, X_i) = K^c(x^c, X_i^c) K^d(x^d, X_i^d)$. The nonparametric estimate of $g(x)$ is given by $\hat{g}(x) = \sum_{i=1}^n Y_i \mathcal{K}(x, x_i) / \sum_{i=1}^n \mathcal{K}(x, x_i)$. We choose $(h, \lambda) = (h_1,\ldots,h_p,\lambda_1,\ldots,\lambda_q)$ by minimizing the following CV function:

$$CV_{LC}(h, \lambda) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{g}_{-i}(X_i))^2 w(X_i), \tag{1.4}$$

where $\hat{g}_{-i}(X_i) = \sum_{l \neq i}^n Y_l K(X_i, X_l) / \sum_{l \neq i}^n K(X_i, X_l)$ is the leave-one-out local-constant (LC) kernel estimator of $g(X_i)$ and $0 \leq w(.) \leq 1$ is a weight function that serves to avoid difficulties caused by dividing by zero, or by the slow convergence rate for when X_i is near the boundary of the support of X.

Define an indicator function $I_j(v^d, x^d) = I(v_j^d \neq x_j^d) \prod_{s \neq j, s=1}^q I(v_j^d = x_s^d)$. Note that $I_j(v^d, x^d) = 1$ if and only if v^d and x^d differ only in their jth component. Letting $m_j(x)$ and $m_{jj}(x)$ ($m = g$ or $m = f$) denote the first-order and second-order partial derivatives of $m(x^c, x^d)$ with respect to x_j^c, Hall, Li, and Racine (2004) have shown that ($\int dx = \sum_{x^d \in D} \int dx^c$, D is the support of X^d)

$$CV_{LC}(h,\lambda) = \int \left\{ \frac{\kappa_2}{2} \sum_{j=1}^{p} [g_{jj}(x)f(x) + 2g_j(x)f_j(x)]h_j^2 + \sum_{v^d \in D} I_j(v^d,x^d)[g(x^c,v^d) - g(x)]f(x^c,v^d)\lambda_j \right\}^2 \times w(x)f(x)^{-1}\,dx + \frac{\kappa^p}{nh_1\ldots h_p} \times \int \sigma^2(x)w(x)\,dx + o_p\left(\sum_{j=1}^{p} h_j^4 + \sum_{j=1}^{q} \lambda_j^2 + (nh_1\ldots h_p)^{-1} \right). \quad \textbf{(1.5)}$$

The preceding results are based on the LC kernel estimation result. Li and Racine (2004) have considered the local linear (LL) CV method. The CV objective function is the same as given in (1.4) but with $\hat{g}_{-i}(X_i)$ replaced by a leave-one-out LL kernel estimator. Li and Racine (2004) have shown that the resulting CV function has the same form as (1.5) with the term $2g_j(x)f_j(x)$ removed.

Define $z_j = n^{-2/(4+p)}h_j^2$ for $j = 1,\ldots,p$, and $z_{p+j} = n^{-2/(4+p)}\lambda_j$ for $j = 1,\ldots,q$; then both the leading terms of $CV_{LC}(h,\lambda)$ and $CV_{LL}(h,\lambda)$ can be written in the form of $c_0 n^{-4/(p+4)}\chi(z_1,\ldots,z_p,z_{p+1},\ldots,z_{p+q})$, where $c_0 = \kappa^p \int \sigma^2(x)w(x)\,dx > 0$ is a constant, and

$$\chi(z_1,\ldots,z_p,z_{p+1},\ldots,z_{p+q}) = \int \left\{ \sum_{j=1}^{p+q} B_j(x)z_j \right\}^2 dx + \frac{1}{(z_1\ldots z_p)^{1/2}} = z'Az + \frac{1}{(z_1\ldots z_p)^{1/2}}, \quad \textbf{(1.6)}$$

where $z = (z_1,\ldots,z_{p+q})'$ (the prime denotes transpose), A is a $(p+q) \times (p+q)$ symmetric positive semidefinite matrix with its (j,s)th element given by $A_{(j,s)} = \int B_j(x)B_s(x)\,dx$, where $B_j(x) = c_0^{-1/2}(\kappa_2/2)[g_{jj}(x)f(x) + 2g_j(x)f_j(x)]w(x)^{1/2}f(x)^{-1/2}$ (one removes $2g_j(x)f_j(x)$ if it is a local linear CV function) for $j = 1,\ldots,p$, and $B_{p+j}(x) = c_0^{-1/2}\sum_{v^d \in D} I_j(v^d,x^d)[g(x^c,v^d) - g(x)]f(x^c,v^d)w(x)^{1/2}f(x)^{-1/2}$ for $j = 1,\ldots,q$.

Hall, Racine, and Li (2004) have considered the CV selection of smoothing parameters in a conditional probability (density) estimation framework and show that their CV objective function also has a leading term of the form as given in (1.6) with of course a different definition of $B_j(x)$ for $j = 1,\ldots,p+q$. Therefore, the leading term of the CV objective function, in either a regression or a conditional probability model, has the expression as given by (1.6). The uniqueness of the CV selected optimal smoothing parameters replies on the uniqueness of a nonnegative vector $z \in \mathbb{R}_{++}^p \times \mathbb{R}_+^q$ that minimizes (1.6), where $\mathbb{R}_{++}^p = \{z \in \mathbb{R}^p, z_j > 0 \text{ for all } j = 1,\ldots,p\}$ and $\mathbb{R}_+^q = \{z \in \mathbb{R}^p, z_j \geq 0 \text{ for all } j = 1,\ldots,q\}$. Subsequently we will first focus on the simple case that all covariates are continuous.

When $q = 0$ (no discrete covariates), all covariates are continuous random variables, and (1.6) becomes

$$\chi_c(z_1,\dots,z_p) = \int \left\{ \sum_{j=1}^{p} B_j(x) z_j \right\}^2 dx + \frac{1}{(z_1 \dots z_p)^{1/2}}$$
$$= z'Az + \frac{1}{(z_1 \dots z_p)^{1/2}}, \tag{1.7}$$

with $z = (z_1,\dots,z_p)'$, and A is now of dimension $p \times p$. The uniqueness of the CV selected optimal smoothing parameters of $h_1,\dots,h_p$ hinges on the uniqueness of a vector $z \in \mathbb{R}^p_{++}$ that minimizes (1.7). Let z^* denote the vector of z that minimizes $\chi_c(z)$ over $\mathbb{R}^p_{++}$; we ask that

Each z_j^* ($j = 1,\dots,p$) is positive, finite, and uniquely determined. **(1.8)**

If (1.8) holds true, then the CV selected smoothing parameters are all well defined asymptotically. In fact, it follows from Hall, Li, and Racine (2004) and Hall, Racine, and Li (2004) that $n^{-1/(4+p)}\hat{h}_j \xrightarrow{p} \sqrt{z_j^*}$ for $j = 1,\dots,p$, or equivalently, $(\hat{h}_j - h_j^*)/h_j^* = \hat{h}_j/h_j^* - 1 \xrightarrow{p} 0$, where $h_j^* = \sqrt{z_j^*}n^{-1/(4+p)}$ is the benchmark nonstochastic optimal smoothing parameter ($j = 1,\dots,p$). The next theorem gives a simple necessary and sufficient condition for (1.8) to hold.

THEOREM 1.1. *Assume that $q = 0$ so that $z = (z_1,\dots,z_p)'$; define*

$$\mu = \inf_{z \in \mathbb{Z}, \|z\|=1} z'Az.$$

Then $\chi(z)$ has a unique minimizer $z^* \in \mathbb{R}^p_+$ with $0 < z_j^* < \infty$ for all $j = 1,\dots,p$ if and only if

$$\mu > 0. \tag{1.9}$$

Next, we discuss the general case with a mixture of continuous and discrete covariates. Now, $z = (z_1,\dots,z_{p+q})'$ and A is a $(p+q) \times (p+q)$ symmetric positive semidefinite matrix. Let $\mathbb{Z} = \mathbb{Z}_1 \times \mathbb{Z}_2$ where $\mathbb{Z}_1 = \mathbb{R}^p_{++}$ and $\mathbb{Z}_2 = \mathbb{R}^q_+$ and let $z^* \in \mathbb{Z}$ denote a minimizer of $\chi(z_1,\dots,z_{p+q})$. We seek conditions that ensure the following result:

For $j = 1,\dots,p$, each z_j^* is positive and finite, for $j = p+1,\dots,p+q$,

each z_j^* is nonnegative, and all z_j^*'s are uniquely determined. **(1.10)**

Condition (1.10) will lead to asymptotically uniquely defined CV selected smoothing parameters of $\hat{h}_1,\dots,\hat{h}_p,\hat{\lambda}_1,\dots,\hat{\lambda}_q$. We partition the A matrix as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A'_{12} & A_{22} \end{pmatrix}, \tag{1.11}$$

https://doi.org/10.1017/S0266466605050504 Published online by Cambridge University Press

where A_{11} is of dimension $p \times p$, and A_{22} is of dimension $q \times q$, and A_{12} has a comfortable dimension. The following theorem gives the existence and uniqueness of a minimizer for $\chi(z)$.

THEOREM 1.2. *Let*

$$\mu = \inf_{z \in \mathbb{Z}, \|z\|=1} z' A z.$$

If $\mu > 0$, then χ has a minimizer $z^* = (z^*_{(1)}, z^*_{(2)}) \in \mathbb{Z}$ with $\chi(z^*) < +\infty$, and a necessary and sufficient condition for a point $z = (z_{(1)}, z_{(2)}) \in \mathbb{Z}$ to be a minimizer of χ is that $z_{(1)} = z^*_{(1)}$ and $z_{(2)} = z^*_{(2)} + z^0_{(2)}$ for some $z^0_{(2)} \in \mathcal{N}(A_{22})$, the null space of A_{22}.[1] In particular, if $q = 0$ or A_{22} is positive definite, then the Hessian (second derivative) matrix $\mathcal{H}$ of χ is positive definite at every point $z \in \mathbb{Z}$ with $\chi(z) < +\infty$. Thus χ has a unique minimizer z^* satisfying (1.10).

2. Proofs and Discussions

Proof of Theorem 1.1. The "if" part of Theorem 1.1 is a special case of Theorem 1.2 with $q = 0$. Thus we only need to prove the "only if" part. Let $\mu = 0$ be attained at some $z^* = (z^*_1, \ldots, z^*_p) \in \mathbb{Z}$ with $\|z^*\| = 1$. If $z^*_i \neq 0$ for all $i = 1, \ldots, p$, then $\chi(tz^*) \to 0$ as $t \to +\infty$. This implies that χ has no minimizer. If $z^*_i = 0$ for some $1 \leq i \leq p$, without loss of generality, we assume that $z^*_1 = \cdots = z^*_r = 0$ for some $1 \leq r \leq p - 1$. Let $\varepsilon > 0$ be chosen such that $p(1 - \varepsilon) > r$. Let $z = (z_1, \ldots, z_p) \in \mathbb{Z}$ with $z_i = 1, 1 \leq i \leq r, z_i = 0, r + 1 \leq i \leq p$. Consider $z(t) = t^{\varepsilon - 1}(z^* + tz) \in \mathbb{Z}$ for all $t > 0$, because $\mathbb{Z}$ is a convex cone. We have

$$z(t)' A z(t) = t^{2\varepsilon - 2} z^{*\prime} A z^* + 2t^{\varepsilon} z^{*\prime} A z + t^{2\varepsilon} z' A z \to 0, \quad \text{as } t \to 0$$

and

$$(z(t)_1 \ldots z(t)_p) = t^{p(\varepsilon - 1) + r} \cdot (z^*_{r+1} \ldots z^*_p) \to +\infty, \quad \text{as } t \to 0$$

because $p(\varepsilon - 1) + r < 0$, which implies that $\chi(z(t)) \to 0$ as $t \to 0$. Therefore χ has no minimizer. ∎

Remark 2.1. From the proof of Theorem 1.1 we know that $\mu > 0$ is a necessary and sufficient condition for the existence of a minimizer z^* that minimizes $\chi_c(z)$; the uniqueness of the minimizer z^* comes from the fact that the Hessian matrix of $\chi_c(z)$ is positive definite.

Note that in Theorem 1.1 μ is defined as the infimum of $z'Az$, not of $\chi_c(z)$ as it does not contain the term of $1/\sqrt{z_1 \ldots z_p}$. Also note that the minimization is done over the unit sphere restricted to the first quadrant. Theorem 1.1 states that $\mu > 0$ is a necessary and sufficient condition for the existence of a unique minimizer z^* with each component z^*_j $(j = 1, \ldots, p)$ positive and finite. This condition is substantially weaker than the requirement that A be a positive def-

https://doi.org/10.1017/S0266466605050504 Published online by Cambridge University Press

inite matrix as assumed in Hall, Racine, and Li (2004) and Li and Racine (2004). It is obvious that when A is positive definite, then $\mu > 0$ because $z \neq 0$ when restricted to $\|z\| = 1$. However, consider the LL regression case with $p = 2$ and that $g(x_1, x_2) = x_1^2 + x_2^2$; then $g_{11}(x) = g_{22}(x) = 2$, and this leads to

$$A = c\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

where $c > 0$ is a constant. Thus, A is a singular matrix, and hence it is not positive definite. Nevertheless, it is easy to check that $\mu > 0$ because in this case $z'Az = c(z_1 + z_2)^2 > 0$ for any $z \in \mathbb{R}_+^2$ with $\|z\| = 1$. Therefore, by Theorem 1.1 we know that z^* is uniquely defined with $0 < z_j^* < \infty$ ($j = 1,2$); this implies that the CV selected smoothing parameters are well defined. In fact, $n^{-1/6}\hat{h}_j \xrightarrow{p} \sqrt{z_j^*}$ for $j = 1,2$. This result is quite intuitive; given that $g(x)$ is nonlinear in both x_1 and x_2, one would expect that the CV selected smoothing parameters should converge to zero with the rate of $O_p(n^{-1/(4+p)}) = O_p(n^{-1/6})$.

Proof of Theorem 1.2. It is clear that $\mathbb{Z}$ is a convex cone in $\mathbb{R}^{p+q}$. For each $z \in \mathbb{Z}$, we write $z = (z_{(1)}, z_{(2)})$ where $z_{(1)} \in \mathbb{Z}_1$ and $z_{(2)} \in \mathbb{Z}_2$. We have

$$\chi(z) < +\infty \Leftrightarrow z_{(1)} \in \text{int}(\mathbb{Z}_1).$$

By the definition (1.6), χ is a lower semicontinuous function from $\mathbb{Z}$ to $\mathbb{R} \cup \{+\infty\}$. For each $z \in \mathbb{Z}$ with $\|z\| = 1$ and $t > 0$, we have $tz \in \mathbb{Z}$ and $\chi(tz) > t^2\mu$. For $r > 0$, denote $B_r = \{z \in \mathbb{R}^{p+q+1} : \|z\| \leq r\}$ and $\mathbb{K}_r = \overline{\mathbb{Z}} \cap B_r$. Thus there exists $R > 0$ such that

$$\min_{z\in\mathbb{Z}} \chi(z) \Leftrightarrow \min_{z\in\mathbb{K}_R} \chi(z).$$

Because $\mathbb{K}_R = \overline{\mathbb{Z}} \cap B_R$ is a nonempty compact set, by the Weierstrass theorem, the lower semicontinuous function χ attains its minimum at $z^* = (z_{(1)}^*, z_{(2)}^*) \in \mathbb{K}_R$ with $\chi(z^*) < +\infty$.

To continue our proof of the theorem, let us examine the Hessian (the second-order derivative) matrix $\mathcal{H}$ of χ at each point $z \in \mathbb{Z}$ with $\chi(z) < +\infty$. A direct calculation shows that

$$\mathcal{H} = \frac{\partial^2\chi(.)}{\partial z \partial z'} = 2A + \begin{pmatrix} \frac{1}{4\sqrt{z_1 \ldots z_p}}[2G + J] & 0 \\ 0 & 0 \end{pmatrix}, \qquad \textbf{(2.1)}$$

where G is a $p \times p$ diagonal matrix with its jth diagonal element given by $1/z_j^2$ for $j = 1,\ldots,p$, and J is a $p \times p$ matrix with its (j,s)th element given by $1/(z_j z_s)$, $j,s = 1,\ldots,p$; that is, $J = (z_1^{-1},\ldots,z_p^{-1})'(z_1^{-1},\ldots,z_p^{-1})$ is positive semidefinite. Thus $2G + J$ is a symmetric positive definite matrix. Because A is symmetric positive semiefinite, $\mathcal{H}$ is always symmetric positive semidefinite. The case $q = 0$ implies that $\mathcal{H}$ is positive definite because $2G + J$ is positive definite;

https://doi.org/10.1017/S0266466605050504 Published online by Cambridge University Press

the case $q > 0$ and A_{22} being positive definite implies that the sum of the two matrices in the right-hand side of (2.1) is positive definite. That is, the Hessian matrix $\mathcal{H}$ is positive definite at any point $z \in \mathbb{Z}$ with $\chi(z) < +\infty$. Thus, $\chi(z)$ has a unique minimizer.

To prove the necessary and sufficient condition, let $z = (z_{(1)}, z_{(2)}) \in \mathbb{Z}$. If z is another minimizer of χ, then let $\chi(z^*) = \chi(z) = m$. Denote $z(\alpha) = \alpha z + (1-\alpha)z^* \in \mathbb{Z}$ for $0 \leq \alpha \leq 1$. Because χ is convex, we have

$$\chi(z(\alpha)) \leq \alpha\chi(z) + (1-\alpha)\chi(z^*) = m,$$

which implies $\chi(z(\alpha)) = m \ \forall 0 \leq \alpha \leq 1$. Because

$$0 = \chi(z(\alpha)) - \chi(z^*) = \alpha\nabla\chi(z^*)(z - z^*) + \frac{1}{2}\alpha^2(z - z^*)'\mathcal{H}(z^*)(z - z^*) + o(\|z - z^*\|^2),$$

where the last term $o(\|z - z^*\|^2)$ represents a higher order term, we must have $\nabla\chi(z^*)(z - z^*) = 0$ and $(z - z^*)\mathcal{H}(z^*)(z - z^*) = 0$. By (2.1), this can be true only if $z_{(1)} = z^*_{(1)}$. Then we have $z(\alpha)'Az(\alpha) = z^{*\prime}Az^* = C$. Denote $h(\alpha) = z(\alpha)'Az(\alpha) = (2\alpha^2 - 2\alpha + 1)C + (2\alpha - 2\alpha^2)z'Az^*$ for $0 \leq \alpha \leq 1$. For $0 < \alpha < 1$, we have $0 = h'(\alpha) = (4\alpha - 2)C + (2 - 4\alpha)z'Az^*$, which leads to $z'Az^* = C$, and then $(z - z^*)'A(z - z^*) = 0$. Because A is symmetric positive semidefinite, this implies $A(z - z^*) = 0$, and then $A_{22}(z_{(2)} - z^*_{(2)}) = 0$. Thus $z_{(2)} = z^*_{(2)} + z^0_{(2)}$ where $z^0_{(2)} = z_{(2)} - z^*_{(2)} \in \mathcal{N}(A_{22})$.

Conversely, if $z = (z_{(1)}, z_{(2)}) \in \mathbb{Z}$ with $z_{(1)} = z^*_{(1)}$ and $z_{(2)} = z^*_{(2)} + z^0_{(2)}$ for some $z^0_{(2)} \in \mathcal{N}(A_{22})$, to prove that z is a minimizer of χ, we only have to show that $z'Az = z^{*\prime}Az^*$. But this can be easily verified by substituting $z = z^* + (0, z^0_{(2)})$. This completes the proof of Theorem 1.2. ■

Let us apply Theorem 1.2 to show how to determine the existence and uniqueness of a minimizer for a simple case of $p = 1$ and $q = 2$ with

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}. \tag{2.2}$$

Then $z'Az = z_1^2 + (z_2 + z_3)^2$, and it is easy to see that $\mu > 0$ in this case. So by Theorem 1.2 we know there exists a minimizer z^*. However, $q = 2$ and A_{22} is not positive definite, so from the last part of Theorem 1.2 we cannot infer the uniqueness of z^*. Nevertheless, it is easy to check that in this case $\chi(z) = z_1^2 + (z_2 + z_3)^2 + 1/\sqrt{z_1}$ and that $(z_1^*, z_2^*, z_3^*) = ((\frac{1}{2})^{4/5}, 0, 0)$ is a minimizer of $\chi(z)$. Let $z = (z_{(1)}, z_{(2)}) \in \mathbb{Z}$ be another minimizer of χ. By the second part of Theorem 1.2, we have $z_{(1)} = z_1^* = (\frac{1}{2})^{4/5}$ and $z_{(2)} = z^*_{(2)} + z^0_{(2)} = z^0_{(2)}$ for some $z^0_{(2)} \in \mathcal{N}(A_{22})$ (because $z^*_{(2)} = (0,0)'$). However, $z_{(2)} = z^0_{(2)} \in \mathcal{N}(A_{22})$ implies that

https://doi.org/10.1017/S0266466605050504 Published online by Cambridge University Press

$z_3 = -z_2$; this together with $z_2, z_3 \in \mathbb{R}_+$ implies that $z_{(2)} = (0,0)'$. Hence, $z = z^*$, and z^* is the unique minimizer of $\chi(z)$.

NOTE

1. The null space of A_{22} is defined as $\mathcal{N}(A_{22}) = \{z_{(2)} \in \mathbb{R}^q : A_{22} z_{(2)} = 0\}$.

REFERENCES

Clarke, R.M. (1975) A calibration curve for radiocarbon dates. *Antiquity* 49, 251–256.

Fan, J. & I. Gijbels (1995) Data-driven bandwidth selection in local polynomial regression: Variable bandwidth selection and spatial adaptation. *Journal of the Royal Statistical Association, series B* 57, 371–394.

Härdle, W., P. Hall, & J.S. Marron (1988) How far are automatically chosen regression smoothing parameters from their optimum? *Journal of the American Statistical Association* 83, 86–99.

Härdle, W. & J.S. Marron (1985) Optimal bandwidth selection in nonparametric regression function estimation. *Annals of Statistics* 13, 1465–1481.

Hall, P., Q. Li, & J. Racine (2004) Estimation of Regression Function in the Presence of Irrelevant Variables. Working paper, Department of Economics, Texas A&M University.

Hall, P., J. Racine, & Q. Li (2004) Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association* 99, 1015–1026.

Li, Q. & J. Racine (2003) Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis* 86, 266–292.

Li, Q. & J. Racine (2004) Cross-validation on local linear estimators. *Statistica Sinica* 14, 485–512.