Parameterization and Inference for Nonparametric Regression Problems, with Application to Dietary Intake Instruments

October 30, 2000

Wenxin Jiang[1]

Northwestern University

Victor Kipnis and Douglas Midthune

National Cancer Institute

Raymond J. Carroll

Texas A&M University

Summary

We consider local likelihood/local estimating equations, in which a multivariate function $\Theta(\cdot)$ is estimated but a derived function $\lambda(\cdot)$ of $\Theta(\cdot)$ is of interest. In many applications, when most naturally formulated the derived function is a nonlinear function of $\Theta(\cdot)$. In trying to understand whether the derived nonlinear function is constant/linear, a problem arises with this approach: when the function is actually constant/linear, the expectation of the function estimate need not be constant/linear, at least to second order. In such circumstances, the simplest standard methods in nonparametric regression for testing whether a function is constant/linear cannot be applied. We develop a simple general solution, applicable to nonparametric regression, varying coefficient models, nonparametric generalized linear models, etc. In fact, we show that in local linear kernel regression, inference about the derived function $\lambda(\cdot)$ is facilitated *without loss of power* by reparameterization so that $\lambda(\cdot)$ is itself a component of $\Theta(\cdot)$. Our approach is in contrast to the standard practice of choosing $\Theta(\cdot)$ for convenience and allowing $\lambda(\cdot)$ to be a nonlinear function of $\Theta(\cdot)$. The methods are applied to an important data set in nutritional epidemiology.

KEY WORDS: Asymptotic variance; Attenuation; Biomarkers; Breast cancer; Bias; Correlation; Generalized estimating equations; Goodness-of-fit test; Local estimating equations; Measurement error; Nonparametric regression; Reparameterization; Validation study; Varying coefficient models.

Short title: Local likelihood inference.

[1]Address for correspondence: Wenxin Jiang, Department of Statistics, Northwestern University, Evanston, IL 60208, USA. E–mail: wjiang@northwestern.edu

1 INTRODUCTION

We consider the problem of fitting a model involving several parameters, Θ, when the parameter of interest, λ, is a non-linear function of Θ in a standard parameterization. We compare and contrast this standard parameterization with a so-called 'direct parameterization', in which a reparameterization is used so that λ is an explicit component of Θ.

In parametric problems, reparameterization essentially does not affect the asymptotic inference. We point out in this paper that this is not the case for local estimation procedures, where the components in a given parameterization are to be modeled by local polynomials. We show that the bias that arises from the standard parameterization in local estimation procedures has a form that is more complex than is usually the case. In particular, the bias does not disappear under the assumption of constancy or linearity of the parameter of interest. We demonstrate that when the direct parameterization is applied the bias essentially can be ignored and tests of constancy or linearity are thus facilitated, without loss of statistical power. The methods are applied to an important data set in nutritional epidemiology. A small simulation is also presented.

2 MAIN RESULTS

The main technical point of our article is that inference about $\lambda(x) = \rho\{\Theta(x)\}$ is facilitated by explicitly parameterizing $\Theta(x)$ so that one of its components is $\lambda(x)$, i.e., $\lambda(x) = (1, 0, \ldots, 0)\Theta(x)$. We will denote the response by $\mathbf{Y}$, the scalar covariate by X, and the general q-dimensional function by $\Theta(x)$. Typically, a scalar function $\rho\{\Theta(x)\} = \lambda(x)$ is of primary interest. We consider inference about $\lambda(x)$, particularly whether it is constant over a fixed interval.

As in Carroll, Ruppert and Welsh (1998), the basic assumption is that $\Theta(x)$ can be identified by an unbiased estimating function $\psi\{\mathbf{Y}, \Theta(X)\}$, so that $E[\psi\{\mathbf{Y}, \Theta(X)\}|X] = 0$. Such methods include local likelihood, local GEE, varying coefficient models, etc., as special cases. The function $\Theta(\cdot)$ is to be estimated by an order p local polynomial estimating equation. Let $K(\cdot)$ be a symmetric density function, h a bandwidth, $K_h(v) = h^{-1}K(v/h)$, $G_p(v) = (1, v, \ldots, v^p)^T$, and $\otimes$ the Kronecker product. Let $\mathcal{B} = (\mathbf{b}_0^T, \mathbf{b}_1^T, \ldots, \mathbf{b}_p^T)^T$, and define $\widehat{\Theta}(x) = \widehat{\mathbf{b}}_0$, where $\widehat{\mathcal{B}}$ is the solution to

$$\mathbf{0} = \mathcal{S}(\mathcal{B}) = n^{-1} \sum_{i=1}^{n} K_h(X_i - x) G_p(X_i - x) \otimes \psi\{\mathbf{Y}_i,\ G_p^T(X_i - x)\mathcal{B}\}. \tag{1}$$

The *local estimating equation* (1) is equivalent to equation (2) of Carroll et al. (1998) and corresponds to modeling Θ as a polynomial near x, with the intercept of the polynomial $\mathbf{b}_0$ being the same as $\Theta(x)$. The parameter of interest, $\lambda(x)$, is estimated by $\widehat{\lambda}(x) = \rho\{\widehat{\Theta}(x)\}$. The basic analysis of and motivation for (1) are in Carroll et al. (1998). Let f_x be the probability density function (pdf) of X and define $\Theta^{(p+1)}(x)$ as the $(p+1)$th derivative of $\Theta(x)$, and $\lambda_\Theta(x) = \partial\lambda(x)/\partial\Theta(x) = \partial\rho\{\Theta(x)\}/\partial\Theta(x)$. The following result follows from Carroll et al. (1998).

Proposition 1 *(Bias and Variance). Let p be odd and $h \to 0$ such that $nh \to \infty$. There are asymptotically equivalent versions of $\widehat{\Theta}(\cdot)$ and $\widehat{\lambda}(\cdot)$ with the following properties:*

$$E\widehat{\lambda}(x) = \lambda(x) + \{h^{p+1}/(p+1)!\} a_1 \lambda_\Theta^T(x) \Theta^{(p+1)}(x)\{1 + o(1)\}; \tag{2}$$

$$var\{\widehat{\lambda}(x)\} = \{nhf_x(x)\}^{-1} a_2 \lambda_\Theta^T(x) \mathbf{B}(x,\Theta)^{-1} \mathbf{C}(x,\Theta)\{\mathbf{B}^{-1}(x,\Theta)\}^T \lambda_\Theta(x)\{1 + o(1)\}. \tag{3}$$

The symbols in the above formulas are defined in the following steps: Define $\mu(r) = \int z^r K(z) dz$, $\gamma(r) = \int z^r K(z)^2 dz$, $D_p(\mu)$ and $D_p(\gamma)$ to be $(p+1) \times (p+1)$ matrices with (j,k) entries $\mu(j+k-2)$ and $\gamma(j+k-2)$, respectively. Further define, for any L, $D_\mu(L) = \{\mu(L), \mu(L+1), \ldots, \mu(L+p)\}^T$, $\chi(\mathbf{y},\mathbf{v}) = (\frac{\partial}{\partial \mathbf{v}^T})\psi(\mathbf{y},\mathbf{v})$, $\mathbf{B}(x,\Theta) = E[\chi\{\mathbf{Y},\Theta(x)\}|X = x]$, and $\mathbf{C}(x,\Theta) = E[\psi\{\mathbf{Y},\Theta(x)\}\psi\{\mathbf{Y},\Theta(x)\}^T|X = x]$. Then a_1 is the first element of $D_p^{-1}(\mu) D_\mu(p+1)$ and a_2 is the first diagonal element of $D_p^{-1}(\mu) D_p(\gamma)\{D_p^{-1}(\mu)\}^T$.

We now show that inference about $\lambda(x) = \rho\{\Theta(x)\}$ is facilitated by parameterizing $\Theta(x)$ explicitly so that $\lambda(x)$ is one of its components, i.e., $\lambda(x) = (1, 0, \ldots, 0)\Theta(x)$.

First note that under the suggested parameterization, $\lambda_\Theta(x) = (1, 0, \ldots, 0)^T$, $\lambda_\Theta^T(x)\Theta^{(p+1)}(x) = \lambda^{(p+1)}(x)$, and hence the bias (2) is of order $O\{h^{p+1}\lambda^{(p+1)}(x)\}$. The variance from (3) is of order $O\{(nh)^{-1}\}$. We can then test whether $\lambda(\cdot)$ is constant or linear in x using standard methods. For specificity, we use local linear regression ($p = 1$) and apply the idea of Azzalini and Bowman (1993). Specifically, restrict the choices of the bandwidth to be in a grid of points of order $n^{-1/5}$. When

$\lambda(\cdot)$ is constant or linear in x, $\lambda^{(2)}(x)=0$ so that while the variance of $\widehat{\lambda}(\cdot)$ is of order $O(n^{-4/5})$, its squared bias is of order $o(n^{-4/5})$. Because the bias is asymptotically negligible, we can construct two tests as follows. First select a grid of values $\mathbf{x}=(x_1,\ldots,x_N)^T$. Let $\widehat{\mathbf{\Sigma}}_n$ be a consistent estimate of the covariance matrix of $(nh)^{1/2}\{\widehat{\lambda}(x_1)-\lambda(x_1),\ldots,\widehat{\lambda}(x_N)-\lambda(x_N)\}$. One way to test for constancy of the regression function is to fit a linear model to these estimates via generalized least squares based on $\widehat{\mathbf{\Sigma}}_n$, and then test whether the slope of the fitted regression is non–zero. A similar test for whether $\lambda(x)$ is linear fits a quadratic model to $\widehat{\lambda}(\cdot)$. These two tests assume specific alternatives. A test for a global alternative can be constructed as follows. Suppose that under the hypothesis, $\lambda(\mathbf{x})=\mathbf{t}(\mathbf{x})\boldsymbol{\beta}$ for a known function $\mathbf{t}(\mathbf{x})$ and a possibly unknown parameter $\boldsymbol{\beta}$. If $\lambda(\cdot)$ is constant, then $\mathbf{t}(\cdot)$ is a N-vector of ones, while if $\lambda(\cdot)$ is linear then $\mathbf{t}(\cdot)$ is an $N\times 2$ matrix with first column ones and second column $\mathbf{x}$. Let $\widehat{H}=\mathbf{t}(\mathbf{x})\{\mathbf{t}^T(\mathbf{x})\widehat{\mathbf{\Sigma}}_n^{-1}\mathbf{t}(\mathbf{x})\}^{-1}\mathbf{t}^T(\mathbf{x})\widehat{\mathbf{\Sigma}}_n^{-1}$, and $\widehat{Q}=(nh)\{\widehat{\lambda}(\mathbf{x})-\widehat{H}\widehat{\lambda}(\mathbf{x})\}^T\widehat{\mathbf{\Sigma}}_n^{-1}\{\widehat{\lambda}(\mathbf{x})-\widehat{H}\widehat{\lambda}(\mathbf{x})\}$. This is an adaptation of the idea of Azzalini and Bowman (1993) to the case of local estimating equations, although it is a little different from the original method of Azzalini and Bowman which uses the actual response vector in the context of nonparametric regression and an F-type statistic with a numerically approximated distribution. The basic idea is that $\widehat{H}$ is analagous to the hat matrix in ordinary least squares, so that $\widehat{H}\widehat{\lambda}(\mathbf{x})$ is the fit and $\widehat{\lambda}(\mathbf{x})-\widehat{H}\widehat{\lambda}(\mathbf{x})$ is the "residual" vector.

If $p=1$, under the hypothesis of constancy or linearity of $\lambda(\cdot)$, $\widehat{Q}$ has an asymptotic central chi-squared distribution with $N-\xi$ degrees of freedom, where $\xi=1$ or 2, respectively.

The key to this argument is that when $\lambda(\cdot)$ is constant/linear, $p=1$ and when $\lambda(\cdot)$ is explicitly a part of $\Theta(\cdot)$, then the squared bias in estimating $\lambda(\cdot)$ is of smaller order than its variance.

Now we contrast our approach with one in which $\lambda(\cdot)$ is a nonlinear function of $\Theta(\cdot)$. For specificity, consider the correlation function of a bivariate random variable $\mathbf{Y}=(Y,Z)$ as a function of X. The moments are $\Theta(X)=\{E(Y|X),E(Z|X),\mathrm{var}(Y|X),\mathrm{var}(Z|X),\mathrm{cov}(Y,Z|X)\}^T$ $=\{\theta_1(X),\ldots,\theta_5(X)\}^T$. The correlation function is $\lambda(X)=\theta_5(X)/\{\theta_3(X)\theta_4(X)\}^{1/2}$. Unbiased estimating equations for Θ are $\psi(\mathbf{Y},\Theta)=\{Y-\theta_1,Z-\theta_2,(Y-\theta_1)^2-\theta_3,(Z-\theta_2)^2-\theta_4,(Y-\theta_1)(Z-\theta_2)-\theta_5\}^T$. For this problem, consider the case that $E(Y|X)=E(Z|X)=0$, $\mathrm{var}(Y|X)=1$,

$\text{var}(Z|X) = X^2$, and $\text{cov}(Y, Z|X) = rX$, $0 < r < 1$. Then $\lambda(X) = \text{corr}(Y, Z|X) = r$ is a constant. In this parameterization, for local linear regression it is easily shown from (2) that the bias of $\widehat{\lambda}(x)$ is $-r(h^2/2)x^{-2} \int z^2 K(z)dz\{1 + o(1)\}$. In particular, this means that when $h \sim n^{-1/5}$, under the hypothesis the squared bias is no longer of smaller order than the variance. Application of the Azzalini and Bowman idea in this parameterization would lead to a test with elevated Type I error. Of course, some type of undersmoothing could be used to eliminate bias, but undersmoothing increases variance and any such device would therefore lead to a loss of power.

The unanswered question is whether parameterizing $\Theta(\cdot)$ explicitly so that $\lambda(\cdot)$ is one of its components inflates the variance of the estimate of $\lambda(\cdot)$. For example, in the bivariate problem, we suggest that the parameters should be $\Omega = g(\Theta) = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5/(\theta_3\theta_4)^{1/2}\}^T = (\omega_1, \ldots, \omega_5)^T$, or alternatively that $\Theta = g^{-1}(\Omega) = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5(\omega_3\omega_4)^{1/2}\}^T$. In fact, generally there is no effect on the variance for estimating the parameter of interest: the proof is immediate.

Proposition 2 *(Effects of Reparameterization). Let $\Omega = g(\Theta)$ be invertible and with g and g^{-1} both continuously differentiable, such that $\lambda = \rho(\Theta) = \eta(\Omega)$. Let $\widehat{\lambda}_1(\mathbf{x}) = \rho\{\widehat{\Theta}(\mathbf{x})\}$ and $\widehat{\lambda}_2(\mathbf{x}) = \eta\{\widehat{\Omega}(\mathbf{x})\}$, where $\widehat{\Theta}(\cdot)$ is the local estimator based on the estimating equation $\psi_1(\mathbf{Y}, \Theta)$ and $\widehat{\Omega}(\cdot)$ is the local estimator based on the estimating equation $\psi_2(\mathbf{Y}, \Omega)$. Suppose further that $\psi_2(\mathbf{Y}, \Omega) = M(\Omega)\psi_1\{\mathbf{Y}, g^{-1}(\Omega)\}$, where $M(\Omega)$ is a fixed non-singular matrix. Then for local polynomial regression with p odd, (i). The biases for $\widehat{\lambda}_1(x)$ and $\widehat{\lambda}_2(x)$ are the same to order $O(h^{p+1})$ if g is linear; (ii). If λ is linear in Ω and $\lambda(x)$ is linear in x, for $p = 1$ the bias of $\widehat{\lambda}_2(x)$ is of order $o(h^2)$. (iii). The variances for $\widehat{\lambda}_1(x)$ and $\widehat{\lambda}_2(x)$ are always the same to order $O\{(nh)^{-1}\}$.*

For implementation, consider a fixed grid of values $\mathbf{x}$. While asymptotically the elements of $\widehat{\lambda}(\mathbf{x})$ are independent, in finite samples they are not. The joint covariance matrix of $\widehat{\Theta}(\mathbf{x})$ can be estimated via the sandwich method (Carroll, et al., 1998) or via the bootstrap. In the standard parametrization, the joint covariance matrix for $\widehat{\lambda}(\mathbf{x})$ was computed using the delta method.

3 EXAMPLE: THE MRC DATA

The food frequency questionnaire (FFQ, Q) is the most common instrument for measuring nutrient intake in large epidemiological studies of nutrient-disease relationships. Despite its central role, the properties of FFQs and their relationship to usual nutrient intake (T, which is unobserved,) remain a matter of major dispute, see Kipnis, et al. (1999) and references therein.

While FFQs have biases and are subject to measurement error, the size of both is generally unknown. Two quantities are of particular concern: (a) the correlation between the FFQ and usual intake, which is a common measure of the quality of a particular FFQ; and (b) the slope of the regression of T on Q, known as the attenuation (Carroll, Ruppert & Stefanski, 1995). Write the attenuation as λ. If a regression, e.g., logistic, of disease on T has a regression coefficient ζ, then the regression of disease on Q generally has regression coefficient approximately $\lambda\zeta$.

The Medical Research Council data on protein intake (Plummer and Clayton, 1993) shed some light on this controversy. In addition to the FFQ (Q), 159 subjects provided protein intake as measured by weighed food records (F_j, $j = 1, 2, 3, 4$) four times over a year. In addition, a biomarker (M_j, $j = 1, 2, 3, 4$) for protein intake, urinary nitrogen, was also obtained at these four seasons in replicated pairs. We use a variation on the model of Kipnis, et al. (1999). For the ith subject, jth season and kth replicate, the model has three components: (a) $Q_i = \beta_0 + \beta_1 T_i + \epsilon_i$; (b) $F_{ij} = \alpha_0 + \alpha_1 T_i + s_i + U_{ij}$; and (c) $M_{ijk} = T_i + \xi_{ijk}$. The random variables $(\epsilon_i, s_i, U_{ij}, \xi_{ijk})$ are assumed to be independent of T_i, have mean zero and variances $(\sigma^2_\epsilon, \sigma^2_s, \sigma^2_u, \sigma^2_\xi)$. Nutritional insight and the fact that Q was obtained in season 3 makes it reasonable to assume that all random variables are independent except (ϵ_i, s_i), (ϵ_i, U_{i3}), (ϵ_i, ξ_{i3k}) $(k = 1, 2)$, $(U_{ij}, \xi_{ijk}), (j = 1, \ldots, 4,\ k = 1, 2)$. Denote by Θ the unique means and variances of the observed data arising from this model, i.e., $\theta_1 = E(M_{ijk})$, $\theta_2 = E(F_{ij})$, $\theta_3 = E(Q_i)$, $\theta_4 = \text{var}(Q_i)$, $\theta_5 = \text{cov}(Q_i, F_{ij} | j = 1, 2, 4)$, $\theta_6 = \text{cov}(Q_i, F_{i3})$, $\theta_7 = \text{cov}(Q_i, M_{ijk} | j = 1, 2, 4)$, $\theta_8 = \text{cov}(Q_i, M_{i3k})$, $\theta_9 = \text{var}(F_{ij})$, $\theta_{10} = \text{cov}(F_{ij}, F_{ik} | j \neq k)$, $\theta_{11} = \text{cov}(F_{ij}, M_{ijk})$, $\theta_{12} = \text{cov}(F_{ij}, M_{i\ell k} | j \neq \ell)$, $\theta_{13} = \text{var}(M_{ijk})$, $\theta_{14} = \text{cov}(M_{ijk}, M_{i\ell p} | j \neq \ell \text{ or } k \neq p)$. The notation indicates for example that $\theta_1 = E(M_{ijk})$ is independent of (i, j, k).

We estimated these moments using local likelihood for a Gaussian model, although normality is not necessary since only the moments are involved. While the FFQ was observed for all individuals, some individuals had missing food records and/or biomarkers in some seasons, and we adjusted the local likelihood accordingly.

In this context, the attenuation equals $\text{cov}(Q_i, M_{ijk}|j = 1,2,4)/\text{var}(Q_i)$, since for $j \neq 3$, $\text{cov}(Q_i, M_{ijk}) = \text{cov}(Q_i, T_i) + \text{cov}(Q_i, \xi_{ijk}) = \text{cov}(Q_i, T_i)$. Similarly, the correlation with usual intake can be shown to equal $\text{cov}(Q_i, M_{ijk}|j = 1,2,4)/[\{\text{var}(Q_i)\}^{1/2}\{\text{cov}(M_{ijk}, M_{i\ell p}|j \neq \ell)\}^{1/2}]$. Note that neither parameter involves the F's, but because the F's are related to the Q's and M's and because of the seemingly unrelated regression phenomenon, inclusion of the information from F increases the efficiency of the inference on the parameters of interest.

It is of considerable epidemiologic interest to know whether these parameters are independent of body mass index (BMI). We attack each parameter separately. One can in principle formulate the problem so that *both* the attenuation and the correlation functions are direct parameters in the problem. Our experience in this data example and in simulations related to it was that numerical problems can arise for the convergence of the scoring algorithm. Hence, in the analysis of these data, we ran one analysis for the correlation function, and another for the attenuation function. For both cases, the way we did the direct parameterization was to replace $\text{cov}(Q_i, M_{ijk}|j = 1,2,4)$ by the attenuation/correlation. In our analysis of the data, we used a transformation of BMI as the predictor, namely (rank of BMI)$/n$. This made the predictor, which we hereafter call BMI or B, equally spaced and improved the speed of convergence of the scoring method. The analysis of the raw BMI values was largely in line with that reported here.

We report here the analysis with $N = 5$ values of B equally spaced between 0.15 and 0.85: the results were only slightly sensitive to choices of N ranging from $N = 3$ to $N = 15$. The Epanechnikov kernel with a bandwidth of 0.25 was used: we varied the bandwidth somewhat and found that the results were fairly insensitive to the bandwidth chosen. The generally linear trends in both functions suggest that we use a trend test to test for constancy of the regression functions.

We used the sandwich method to estimate the covariance matrix of the function estimates. We then fit a simple linear model to the fitted functions regressed on the values of B, using generalized least squares, and tested for constancy of the regression function by testing whether the slope estimate was equal to zero. For the correlation of the FFQ and usual intake, the fitted slope

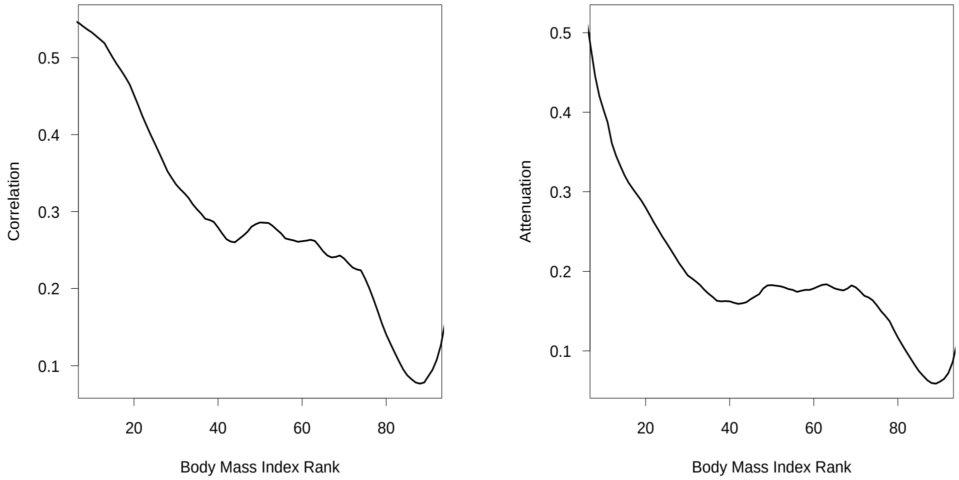

Figure 1: The estimated correlation between the FFQ and usual intake (left) and the estimated attenuation due to error and bias in the FFQ (right) as functions of Body Mass Rank scaled from 1 to 100, fit via a local linear smoother.

was -0.599 with significance level 0.035. The corresponding estimate and significance level for attenuation were -0.298 and 0.166, respectively. For attenuation, we conclude that there is little evidence in this example that this function depends on body mass index.

To understand the scale in Figure 1, for the values of $B = (.150, .325, .500, .675, .850)$, the estimates for the direct and standard parameterizations were as follows: the correlations were $(0.499, 0.321, 0.259, 0.241, 0.088)$ and $(0.496, 0.302, 0.282, 0.247, 0.093)$, and the standard errors were $(0.122, 0.158, 0.157, 0.119, 0.181)$ and $(0.122, 0.172, 0.163, 0.116, 0.169)$. The sandwich correlation matrix has elements r_{ij} for $i, j = 1, ..., 5$, with the standard parametrization in parentheses: $r_{12} = 0.525$ (0.446), $r_{13} = 0.066$ (0.069), $r_{23} = 0.764$ (0.811), $r_{24} = 0.137$ (0.155), $r_{34} = 0.390$ (0.412), $r_{35} = 0.025$ (0.029), $r_{45} = 0.508$ (0.520), and all other non–diagonal elements equal zero. The relatively high correlations explain why it is that while *pointwise* the function estimates ± 2 standard errors barely includes the constant function, the linear trend tests have significance level less that 0.05. The significance level for the standard parametrization was 0.042.

The fact that the results for the standard and direct parameterizations were very nearly the same can be explained as follows. Recall that the correlation of Q and T is $\text{cov}(Q_i, M_{ijk}|j = 1, 2, 4)/[\{\text{var}(Q_i)\}^{1/2}\{\text{cov}(M_{ijk}, M_{i\ell p}|j \neq \ell)\}^{1/2}]$. In the standard parameterization, the three functions making up this correlation are, as a function of BMI, either fairly constant ($\text{var}(Q)$) or fairly linear. In such cases, $\Theta^{(2)}(x)$ is nearly zero, so that Proposition 1 suggests that for this example, the standard parameterization will be nearly acceptable. The fact that in this instance the significance levels are nearly the same is of course predicted by Proposition 2.

4 SIMULATION

We performed a simulation in the special case of estimating a correlation function for two precisely observed variables Y and Z, as a function of a covariate X. We took the sample size to be $n = 100$, and ran $2,000$ simulated data sets. In all cases, X was uniformly distributed either on the interval $(1.0, 2.0)$ or on the interval $(0.2, 1.2)$, (Y, Z) given X was normally distributed with common mean equal to zero, Z had variance X^2, and Y had variance equal to 1.0. Local linear regression was used ($p = 1$). The kernel density function was the normal density with mean zero and variance one. The bandwidth chosen was $\widehat{\sigma}_x n^{-1/5}$, where $\widehat{\sigma}_x$ is the sample standard deviation of X.

Our major interest is in the null case that the correlation is constant as a function of X, with test cases having the common values 0.2 and 0.5. A nominal 5% level trend test was used, based on a grid of $N = 3$ values at 1.25, 1.50 and 1.75 when X was on the interval $(1.0, 2.0)$ and 0.45, 0.70 and 0.95 when X was on the interval $(0.2, 1.2)$.

Local estimating equations require an estimating equation. For speed of computations, in the simulation we used the method of moments, as follows. The basic parameters in the standard parameterization are $\theta_1(X) = E(Y|X)$, $\theta_2(X) = E(Z|X)$, $\theta_3(X) = \text{var}(Y|X)$, $\theta_4(X) = \text{var}(Z|X)$, $\theta_5(X) = \text{cov}(Y, Z|X)$. The estimating function in this standard parameterization $\psi(Y, Z, \Theta)$ has the five components $Y - \theta_1$, $Z - \theta_2$, $(Y - \theta_1)^2 - \theta_3$, $(Z - \theta_2)^2 - \theta_4$, $(Y - \theta_1)(Z - \theta_2) - \theta_5$. For the direct parameterization, the fifth parameter component is changed to be $\theta_5(X) = \text{corr}(Y, Z|X)$ and the last component of the estimating function becomes $(Y - \theta_1)(Z - \theta_2) - \theta_5(\theta_3\theta_4)^{1/2}$.

The joint covariance matrix of the functions at the grid values were computed in two ways: (a) via the sandwich method as in Carroll, et al. (1998); and (b) via the bootstrap (100 bootstrap data sets), although only the former is reported here. In the former case, for the direct parameterization to be able to compute the derivatives of the estimating function, we constrained the local linear fit to the two variances to be positive, setting the derivatives equal to zero if they were not. Also, both sandwich covariance matrix estimators were adjusted for degrees of freedom by multiplying them by $n/\{n - 5(p+1)\}$, where here $p = 1$ for local linear regression, so that there are $5(p+1)$ parameters estimated at each grid point.

The results of the simulation to check the level of the test are given in Table 1. When using the sandwich method, the direct parameterization clearly has level consistently closer to the nominal level, as suggested by our theory. Using the bootstrap instead of the sandwich method to estimate the joint covariance matrix of the function estimates did not appear to be advantageous.

		Test Level X Uniform on $(1.0, 2.0)$		**Test Level** X Uniform on $(0.2, 1.2)$	
Scale	Parameterization	$r = 0.2$	$r = 0.5$	$r = 0.2$	$r = 0.5$
Untransformed	Direct	0.077	0.055	0.057	0.048
	Standard	0.109	0.091	0.084	0.075
Transformed	Direct	0.066	0.056	0.050	0.048
	Standard	0.098	0.096	0.075	0.073

Table 1: The actual levels (Type I errors) of a nominal 5% level test. "Standard" refers to the standard parameterization, while "Direct" refers to the direct parameterization proposed in the paper. Also, "Untransformed" refers to inference based on the correlation function itself, while "Transformed" refers to the Fisher–transformation of the correlation: for the sandwich method the joint covariance matrix was derived using the delta–method.

In simulations not reported here, we also checked power when using the sandwich method, setting the correlation function to equal $c(X-1)$ for $c = 0.0, 0.2, 0.4, 0.6, 0.8$. The theory says that the standard and direct parameterizations should yield approximately the same power when the levels are equivalent. In order to make the levels of the tests equivalent, we first ran a simulation at the constant correlation case ($c = 0$) and then set the cut-off points of the tests so that the level exactly equaled 0.05. Having done this, the direct and standard parameterizations yielded almost the same power functions, a result predicted by our theory.

5 DISCUSSION

We have considered the problem in which a multivariate function $\Theta(x)$ is estimated, but a scalar function $\lambda(x) = \rho\{\Theta(x)\}$ is of major interest. Generally, $\Theta(\cdot)$ is chosen for convenience and $\lambda(\cdot)$ is a nonlinear function of $\Theta(\cdot)$: this is the standard parameterization. We have shown that inference for whether $\lambda(\cdot)$ is constant or linear is greatly facilitated without loss of power by the direct parameterization, namely reparameterizing so that $\lambda(\cdot)$ is an explicit component of $\Theta(\cdot)$. The methods were applied to a major study of protein intake in nutritional epidemiology.

It is important to note that our major result can be applied to testing in other contexts. There are many possible approaches (Hart, 1997). The essential feature of most these methods is that they require the bias term of the function estimate to be $o(h^2)$ under the hypothesis. For our suggested parameterization this will be the case, and at least in principle these alternative methods can be applied. In the standard parameterization, the bias does not disappear under the hypothesis, and application of these alternative methods will typically lead to tests with inflated levels.

ACKNOWLEDGMENTS

Carroll's research was supported by a grant from the National Cancer Institute and through the Texas A&M Center for Environment and Rural Health by a grant from the National Institute of Environmental Health Sciences. The authors are grateful to Laurence Freedman, Arthur Schatzkin, Sheila Bingham and Grant Izmirlian for help on this manuscript.

REFERENCES

Azzalini, A. and Bowman, A. (1993). On the use of nonparametric regression for checking linear relationships. *Journal of the Royal Statistical Society* B. 55, 549-557.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models.* New York: Chapman & Hall / CRC Press.

Carroll, R. J., Ruppert, D. and Welsh, A. H. (1998). Local estimating equations. *Journal of American Statistical Association*, 93, 214-227.

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* New York: Springer.

Kipnis, V., Carroll, R. J., Freedman, L. S. and Li, L. (1999). A new dietary measurement error model and its implications for the estimation of relative risk: application to four calibration studies. *American Journal of Epidemiology*, 150, 642-651.

Plummer, M. and Clayton, D. (1993). Measurement error in dietary assessment: an investigation using covariance structure models. *Statistics in Medicine*, 1993, 925–948.