The Annals of Statistics
1995, Vol. 23, No. 3, 946–967

ON A SEMIPARAMETRIC VARIANCE FUNCTION MODEL AND A TEST FOR HETEROSCEDASTICITY[1]

BY HANS-GEORG MÜLLER AND PENG-LIANG ZHAO

University of California, Davis and Merck Research Laboratories

We propose a general semiparametric variance function model in a fixed design regression setting. In this model, the regression function is assumed to be smooth and is modelled nonparametrically, whereas the relation between the variance and the mean regression function is assumed to follow a generalized linear model. Almost all variance function models that were considered in the literature emerge as special cases. Least-squares-type estimates for the parameters of this model and the simultaneous estimation of the unknown regression and variance functions by means of nonparametric kernel estimates are combined to infer the parametric and nonparametric components of the proposed model. The asymptotic distribution of the parameter estimates is derived and is shown to follow usual parametric rates in spite of the presence of the nonparametric component in the model. This result is applied to obtain a data-based test for heteroscedasticity under minimal assumptions on the shape of the regression function.

1. Introduction. Regression models with nonconstant error variance are common in practice. Consider the fixed design regression model

$$y_{i,n} = g(t_{i,n}) + \delta_{i,n}, \tag{1.1}$$

where $y_{i,n}$ are measurements of the regression function g at $t_{i,n}$, contaminated with independent errors $\delta_{i,n}$ such that $E\delta_{i,n} = 0$ and $E\delta_{i,n}^2 = \sigma^2(t_{i,n})$. It is thus assumed that $\operatorname{var}(y_{i,n}) = \sigma^2(t_{i,n})$, where $\sigma^2(\cdot)$ is the variance function.

We address here the problem of inference for heteroscedasticity in (1.1) whenever we do not want to make parametric assumptions on the regression function g. The idea is to find an intrinsic measure of heteroscedasticity, no matter what (linear, nonlinear or other) shape the regression function has; we will show that, even under nonparametric assumptions on the regression function, $\sqrt{n}$-consistent inference is possible (Theorem 4.3) and that this extends to a data-adaptive version (Theorem 5.1).

To recognize heteroscedasticity, that is, a nonconstant variance function $\sigma^2(\cdot)$ in (1.1), is an important task for the data analyst: Efficient inference for the regression function itself requires that the heteroscedasticity be taken

Received February 1992; revised August 1994.

[1] Research supported by NSF Grants DMS-90-02423, DMS-93-05484 and Air Force Grant AFOSR-89-0386.

AMS 1991 subject classifications. 62G07, 62G10, 62J12.

Key words and phrases. Constant coefficient of variation model, exponential variance model, generalized linear model, nonparametric regression, polynomial variance model, power of the mean model, rate of convergence, smoothing transformation.

into account; this may lead to transformations of the data [Box and Hill (1974)], weighted least squares procedures [Fuller and Rao (1978)], variable bandwidth nonparametric regression smoothing [see Müller and Stadtmüller (1987), Section 4] or modified likelihood procedures [Davidian and Carroll (1987)]. Furthermore, establishing heteroscedasticity and estimating variance functions which relate the error variance to the predictors is sometimes of interest in its own right [see Carroll and Ruppert (1988) or Müller and Stadtmüller (1987) for examples].

Diagnostic plots to assess heteroscedasticity are a time-honored tool of exploratory data analysis [cf., e.g., Cook and Weisberg (1982) for a discussion]. Most of these procedures are based on residuals which are obtained after fitting a parametric linear or nonlinear model to the data, usually by least squares or likelihood methods. The same applies to formal tests of heteroscedasticity; a rather complete discussion of such tests is given in Carroll and Ruppert [(1988), Section 3.4]. Without exception, these tests are based on completely parametrically specified regression and variance functions, g and $\sigma^2(\cdot)$, and usually are variants of tests which make use of least squares [Harrison and McCabe (1979), Jobson and Fuller (1980)] or likelihood methods, such as score tests [Breusch and Pagan (1979), Cook and Weisberg (1983)], quasilikelihood ratio tests and pseudolikelihood tests [Davidian and Carroll (1987), Carroll and Ruppert (1988)]. Koenker and Bassett (1981) discuss tests for heteroscedasticity in econometric contexts.

While some of these authors extend the classical framework somewhat to include "robust" tests [Carroll and Ruppert (1981)], the case where no parametric form is assumed for the regression function has not been investigated so far, to the knowledge of the authors. We develop here a model where it is only assumed that the true regression function is known to be "smooth" (nonparametric part of the model) and that the relation between variance function and regression function follows a generalized linear model (parametric part). In our approach, the role of the residuals is replaced by nonparametric estimates of the variance function. No distributional assumptions are made on the distribution of the errors, except for some moment properties, essentially amounting to existence of the eighth moment.

The relation between regression function g and variance function $\sigma^2(\cdot)$ is assumed to follow the generalized linear model

$$G(\sigma^2(t)) = \theta_0 + \sum_{j=1}^{p-1} \theta_j H_j(g(t)), \tag{1.2}$$

with known link functions G, H_i, $1 \le i \le p-1$, and unknown parameters θ_i, $0 \le i \le p-1$. As special cases of our rather general semiparametric approach, one obtains almost all completely parametrized models that were considered previously in the literature, including the so-called Poisson, Gamma and lognormal models [McCullagh and Nelder (1989)], as well as the "power-of-the-mean model" and the "exponential variance model" which are described in detail in Carroll and Ruppert (1988).

We give now three example scenarios motivating the usefulness of the proposed model by specifications of (1.2).

(A) Consider an experiment where one has reason to assume that the responses follow a binomial distribution, possibly with overdispersion. However, the relation between mean response and covariate t is not known well enough to be parametrically (linearly or nonlinearly) specified. We are interested in the nature of the dependency of the mean on the covariate and in an assessment of overdispersion.

The proposed approach models the mean–covariate relationship $g(t)$ nonparametrically and the variance function by $\sigma^2(t) = \sigma_0^2 g(t)(1 - g(t))$; this is a special case of the polynomial variance model, a submodel of (1.2) which is discussed further in Section 2. Inference for $g(\cdot)$ and σ_0^2 as developed below is then of interest. The test introduced in Section 5 could be applied in this context to test $\sigma_0^2 = 1$ versus $\sigma_0^2 > 1$, that is, no overdispersion versus overdispersion.

(B) Consider an experiment where continuous measurements are recorded and it is suggested that the responses have approximately constant coefficient of variation. One would like to know whether this is indeed the case. At the same time, the dependency of the mean on the covariate cannot be parametrically specified. We model $g(t)$ nonparametrically and embed the variance function in the power-of-the-mean model (see Section 2), $\sigma^2(t) = \sigma_0^2 g(t)^{2\theta_1}$.

Inference for $g(\cdot)$, σ_0^2 and θ_1 is of interest. In particular, the suggested constant coefficient variation model corresponds to $\theta_1 = 1$. An estimate for the functional relation between variance and mean is obtained by estimating θ_1. This requires, of course, that the power-of-the-mean model is not grossly inappropriate.

(C) Assume the responses are count data. "Quasi-Poisson," "overdispersed Poisson" and "quasi–negative binomial" models are considered as possibilities for the relation between variance and mean, whereas the regression function g, relating the mean to the covariate, cannot be specified. One would like to obtain information on g and at the same time differentiate between the possible variance functions. Then incorporate $g(t)$ nonparametrically and adopt the polynomial variance model in the following form:

$$\sigma^2(t) = \theta_1 g(t) + \theta_2 g^2(t).$$

Estimation of g and of (θ_1, θ_2) is of interest. Tests can be performed to compare the fit of models with $\theta_1 = 1$, $\theta_2 = 0$ (Poisson), with $\theta_1 > 1$, $\theta_2 = 0$ (overdispersed Poisson) and with $\theta_1 = 1$, $\theta_2 > 0$ [negative binomial, see McCullagh and Nelder (1989), page 373].

Our semiparametric variance function model is described in full detail in Section 2. Least-squares-type estimators for the parametric and kernel estimators for the nonparametric components of our model as well as iterative simultaneous estimation of these components by a Gauss–Seidel scheme will be discussed in Section 3. We establish uniform convergence with rates for estimates of the regression and variance function in Theorems 4.1 and 4.2. One of our central results is Theorem 4.3, which provides the joint limiting distribution for the estimates of the parameters relating variance and regression functions under parametric rates of convergence.

Since the asymptotic covariance matrix of the parameter estimates contains unknowns, consistent estimators are developed in Lemma 5.1. This result is then applied to arrive at a data-based test for heteroscedasticity as given in Theorem 5.1. The null hypothesis of homoscedasticity corresponds to $\theta_1 = \cdots = \theta_{p-1} = 0$ in (1.2). The power of this test for a family of contiguous alternatives is investigated. Auxiliary results and proofs are collected in Sections 6 and 7.

To conclude this section, a word of caution. The nonparametric procedures which are a part of our semiparametric approach typically require moderate to large sample sizes and therefore may not be fully feasible for small sample size situations. Note that the main results providing justification for our approach are asymptotic in nature.

2. A semiparametric variance function model. We consider the regression model

$$\text{(M1)} \qquad y_{i,n} = g(t_{i,n}) + \delta_{i,n},$$

where $g(\cdot)$ is an unknown regression function and $\sigma^2(\cdot) = \text{var}(\delta(\cdot))$ is an unknown variance function, both defined on [0, 1]. The points $t_{i,n}$ at which measurements $y_{i,n}$ are taken are assumed to be fixed but not necessarily equidistant for each n. They could also have arisen from a random design regression experiment, in which case our analysis would be conditional on the observed realization of the design. In order to have available strict bounds on the spacings of the $t_{i,n}$, we assume that they are generated as

$$\text{(M2)} \qquad t_{i,n} = F^{-1}\left(\frac{i-1}{n-1}\right), \qquad 1 \le i \le n,$$

where F is a distribution function on [0, 1] which has a Lipschitz continuous density $f = F'$ satisfying

$$\text{(M3)} \qquad 0 < \inf_{[0,1]} f(\cdot) \le \sup_{[0,1]} f(\cdot) < \infty.$$

Note that the common equidistant case corresponds to $f \equiv 1$ on [0, 1].

For the errors $\delta_{i,n} = \delta(t_{i,n})$, it is assumed that they are independent for each n, with

$$\text{(M4)} \qquad \begin{aligned} E\delta_{i,n} &= 0, & E\delta_{i,n}^2 &= \sigma^2(t_{i,n}), \\ E\delta_{i,n}^3 &= \mu_3(t_{i,n}), & E\delta_{i,n}^4 &= \mu_4(t_{i,n}), \end{aligned}$$

for moment functions $\mu_3(\cdot)$ and $\mu_4(\cdot)$, which are continuous on [0, 1], and that

(M5) there exists an $s > 8$ such that $E|\delta_{i,n}|^s < c < \infty$.

Assumptions (M4) and (M5) are rather general, allowing for all kinds of error distributions, including skewed distributions.

The nonparametric part of our model consists of smoothness assumptions on regression and variance functions g and $\sigma^2(\cdot)$. Let $k \geq 2$ be an integer. We require

(M6) $$g,\ \sigma^2(\cdot) \in \mathscr{C}^k([0,1]),$$

where $\mathscr{C}^k([0,1])$ denotes the space of k times continuously differentiable functions on [0, 1]. The parametric part of our semiparametric model relates regression function g and variance function $\sigma^2(\cdot)$ as follows. Assuming that, for some integer $p \geq 1$, p link functions $G, H_1, \dots, H_{p-1}: \mathbb{R} \to \mathbb{R}$ are given, let

(M7) $$G(\sigma^2(\cdot)) = \theta_0 + \sum_{j=1}^{p-1} \theta_j H_j(g(\cdot)) = \sum_{j=0}^{p-1} \theta_j H_j(g(\cdot)),$$

defining $H_0(\cdot) \equiv 1$. The link functions G and H_j are supposed to satisfy some regularity conditions:

(M8) $$G, H_j \in \mathscr{C}^2(\mathbb{R}), \qquad 0 \leq j \leq p-1;$$

and

(M9) $$G(\cdot) \neq \text{const and, for any constant vector } a = (a_0, \dots, a_{p-1})^T \neq 0,$$
$$\int_0^1 f(u) \left[\sum_{j=0}^{p-1} a_j H_j(g(u)) \right]^2 du > 0,$$
where a^T denotes the transpose of a vector a.

Our first aim is to estimate the functions $g(\cdot)$ and $\sigma^2(\cdot)$ and the parameters $\theta_0, \dots, \theta_{p-1}$. Of particular interest is the question whether rates of convergence for the parameter estimates are the usual parametric ones in spite of the nonparametric smoothness assumptions (M6). Theorem 4.3 answers this in the affirmative, while Theorems 4.1 and 4.2 guarantee the usual nonparametric rates for function estimates $\hat{g}$ and $\hat{\sigma}^2(\cdot)$.

Assumptions (M8) and (M9) impose weak technical restrictions on the link functions G and H_j. Model (M7) is thus very general. Consider, for instance, the following models, which generalize other models that have been previously considered in the literature:

1. *power-of-the-mean model*, $\sigma^2(t) = \sigma_0^2 g(t)^{2\theta}$ [Carroll and Ruppert (1988), McCullagh and Nelder (1989)];
2. *exponential variance model*, $\sigma^2(t) = \sigma_0^2 \exp(2\theta g(t))$ [Carroll and Ruppert (1988)];
3. *polynomial variance model*, $\sigma^2(t) = \beta_0 + \beta_1(g(t))^{\alpha_1} + \cdots + \beta_{p-1}(g(t))^{\alpha_{p-1}}$, with known powers $\alpha_i > 0$, all different, where it is assumed that always $\sigma^2(\cdot) \geq 0$.

Model 1 comprises Poisson and Gamma regression, and model 3 contains the "constant coefficient of variation model" $\sigma^2(t) = \sigma_0^2 g(t)^2$, among many other special fully parametric models discussed in McCullagh and Nelder.

It is easy to see how models 1–3 can be expressed in the general form (M7). For model 1 take (denoting the identity function by id) $G \equiv \log$, $H_1 \equiv \log$, whence $\theta_0 = 2 \log \sigma_0$ and $\theta_1 = 2\theta$. For model 2, the corresponding choices are $G \equiv \log$ and $H_1 \equiv \text{id}$, whence $\theta_0 = 2 \log \sigma_0$ and $\theta_1 = 2\theta$, and for model 3, $G \equiv \text{id}$, $H_i \equiv x^{\alpha_i}$, $\alpha_0 = 0$, $\theta_i = \beta_i$, $i = 0, \dots, p-1$.

We note that, for instance, in the power-of-the-mean model, a test for heteroscedasticity corresponds to testing H_0: $\theta = 0$. Beyond developing this test as a special case of a general procedure, our approach will also provide an estimate of θ, which then helps to determine the relation between mean and variances within this particular model.

One can ask what happens in case the assumed specification of the variation–mean relationship (M7) is inaccurate. Of course the specific models can be chosen as fairly large parametric families, as in Models 1–3 above, and this flexibility may render such misspecifications less likely. If nevertheless a misspecification occurs, the estimates for the nonparametric parts $g(\cdot)$ and $\sigma^2(\cdot)$ of the model (Section 3.1) will still be consistent with the same rates of convergence as in Theorems 4.1 and 4.2. They may, however, suffer from a loss in efficiency. The estimate for the parametric part β of the model, on the other hand, possibly would cease to be consistent. The estimates then correspond to "projections" on the misspecified family. Often one has some basic idea about the nature of the response data which allows specification of a family of parametric models (M7), whereas knowledge often is lacking regarding the nature of the dependency of the means on the covariates. Our approach targets this situation.

3. The estimators.

3.1. *Estimators for the nonparametric components.* The nonparametric components of our semiparametric model (M1)–(M9) consist of the regression function g and the variance function $\sigma^2(\cdot)$. These functions, which are smooth according to (M6), could in principle be estimated by various available nonparametric regression procedures [see Eubank (1988) or Müller (1988)]. We consider here a class of kernel estimates [see Gasser and Müller (1984)], but the following results can be easily extended to cover other nonparametric regression methods as well, like those proposed by Priestley and Chao (1972), local polynomial fits or smoothing splines.

Define

$$W_i(t) = W_{in}(t) = \frac{1}{b}\int_{s_{i-1}}^{s_i} K\left(\frac{t-u}{b}\right) du, \qquad 1 \le i \le n, \tag{3.1}$$

where $s_0 = 0$, $s_n = 1$ and $s_i = \frac{1}{2}(t_i + t_{i+1})$, $1 \le i \le n-1$. Then the kernel estimators for g and $\sigma^2(\cdot)$ are given by

$$\hat{g}(t) = \sum_{i=1}^{n} W_i(t) y_i, \tag{3.2}$$

$$\hat{\sigma}^2(t) = \sum_{i=1}^{n} W_i(t) y_i^2 - \Big(\sum_{i=1}^{n} W_i(t) y_i\Big)^2 = \sum_{i=1}^{n} W_i(t) y_i^2 - (\hat{g}(t))^2, \tag{3.3}$$

where the measurements y_i come from model (M1).

The second estimator is motivated by the formula $\mathrm{var}(Y|X = t) = E(Y^2|X = t) - \{E(Y|X = t)\}^2$, which applies to the random design case. Alternative estimators were considered previously but are less suited for our purposes here [Carroll (1982), Müller and Stadtmüller (1987)].

Observe that $b = b(n)$ is a sequence of bandwidth parameters and K is a kernel function. The kernel function K is supposed to have compact support and to satisfy

(K1)
$$\mathrm{supp}(K) = [-1, 1], \qquad \sup|K| \le c < \infty, \qquad \int K(u)\, du = 1,$$
$$\int K(u) u^j\, du = 0, \qquad 0 < j < k, \qquad \int K(u) u^k\, du \ne 0$$
$$\text{and} \quad K(-x) = K(x).$$

Basic requirements for the sequence of bandwidths are

(K2)
$$b \to 0, \qquad nb \to \infty.$$

Furthermore, for uniform convergence of $\hat{g}(\cdot)$ and $\hat{\sigma}^2(\cdot)$, we need the following condition:

(K3) For some ζ, $0 < \zeta \le 1$, it holds that, for all sequences α_n, $\alpha_n \to 0$, as $n \to \infty$, $\int |K(v + \alpha_n) - K(v)|\, dv = O(\alpha_n^{\zeta})$.

Condition (K3) is for instance satisfied if K is Lipschitz continuous of order ζ on $\mathbb{R}$ except at a finite number of points where K could have a discontinuity.

For uniform considerations, we will need a further assumption on the bandwidth sequence $b = b(n)$. Let s be a constant as in (M5), and let r be another constant such that $2 < r < s$, with

(K4)
$$\liminf_{n \to \infty} \left(\frac{nb}{\log n} \right)^{1/2} n^{-2/r} > 0.$$

The following assumptions are occasionally needed:

(K5) $nb^2(\log n)^2 \to \infty$, $\quad n^{1/2} b^k \to d^*$ for a constant d^* with $0 \le d^* < \infty$;

(K6) $nb^{2k} \to 0$ as $n \to \infty$.

The kernel estimators (3.2) and (3.3) will be subject to boundary effects when estimating near the endpoints 0 or 1 [see, e.g., Müller (1991)]. We will consider uniform convergence on an interval $I \subset [0, 1]$ and assume that boundary effects can be ignored on I, either by reducing I sufficiently or by properly modifying estimates near the boundaries; for some smoothing methods such as locally weighted least squares, such modifications will be automatic.

3.2. *Estimators for the parametric components.* It is natural to estimate the parameters $\theta_0, \dots, \theta_{p-1}$ in model (M7) by a weighted least squares method,

$$(3.4)\qquad \left(\hat{\theta}_0, \dots, \hat{\theta}_{p-1}\right) = \underset{(\theta_1, \dots, \theta_{p-1})}{\arg\min} \sum_{i=1}^{n} q(t_i)\left[G(\hat{\sigma}^2(t_i)) - \sum_{l=0}^{p-1} \theta_l H_l(\hat{g}(t_i))\right]^2,$$

where $H_0(\cdot) \equiv 1$ as in (M7) and $q(\cdot)$ is a weight function, which is assumed to be Lipschitz continuous. Let

$$\beta = (\theta_0, \dots, \theta_{p-1})^T, \qquad \hat{\beta} = \left(\hat{\theta}_0, \dots, \hat{\theta}_{p-1}\right)^T,$$

$$X = (x_{ij})_{1 \le i \le n, 1 \le j \le p}, \qquad \hat{X} = (\hat{x}_{ij})_{1 \le i \le n, 1 \le j \le p} \quad \text{with } x_{ij} = H_{j-1}(g(t_i)),$$

$$\hat{x}_{ij} = H_{j-1}(\hat{g}(t_i)) \quad (\text{note that } H_0 \equiv 1);$$

furthermore, let

$$Z = \left(G(\sigma^2(t_1)), \dots, G(\sigma^2(t_n))\right)^T,$$

$$\hat{Z} = \left(G(\hat{\sigma}^2(t_1)), \dots, G(\hat{\sigma}^2(t_n))\right)^T,$$

$$Q^{-1} = \operatorname{diag}(q(t_1), \dots, q(t_n)) \quad (\text{denoting the corresponding } n \times n \text{ diagonal matrix}).$$

Here, estimates $\hat{g}(\cdot)$ and $\hat{\sigma}^2(\cdot)$ are as given by (3.2) and (3.3).

Then model (M7) implies

$$(3.5)\qquad Z = X\beta$$

and the estimator for the parameters (3.4) becomes the classical least squares estimator,

$$(3.6)\qquad \hat{\beta} = \left(\hat{X}^T Q^{-1} \hat{X}\right)^{-1} \hat{X}^T Q^{-1} \hat{Z},$$

provided the r.h.s. is well defined.

3.3. *Simultaneous estimation of parametric and nonparametric components.* We sketch here an iterative procedure for the simultaneous estimation of β, g and $\sigma^2(\cdot)$ by taking advantage of model (M7) and some additional assumptions. Once estimates

$$\hat{\beta} = \left(\hat{\theta}_0, \dots, \hat{\theta}_{p-1}\right)$$

have been obtained, they can be used to improve estimation of g and $\sigma^2(\cdot)$ in the following way: Assume that G is strictly monotone. Then, according to (M7), defining a mapping $\psi_\sigma(\beta, g) = G^{-1}(\sum_{j=0}^{p-1}\theta_j H_j(g(\cdot)))$, one has $\sigma^2(\cdot) = \psi_\sigma(\beta, g(\cdot))$. Assume furthermore that the specific form of (M7) allows a representation

$$g(\cdot) = \psi_g(\beta, \sigma^2(\cdot)), \tag{3.7}$$

for a suitable mapping ψ_g. The parametric estimator (3.6) can be rewritten as

$$\hat{\beta} = \psi_\beta(\hat{g}(\cdot), \hat{\sigma}^2(\cdot)) \tag{3.8}$$

with a mapping ψ_β. These observations motivate the following Gauss–Seidel type of iteration procedure. Set $\hat{g}_{(0)}(\cdot) = \hat{g}(\cdot)$ [(3.2)] and $\hat{\sigma}^2_{(0)}(\cdot) = \hat{\sigma}^2(\cdot)$ [(3.3)], with undersmoothing bandwidths (3.10) given below and $\hat{\beta}_{(0)} = \hat{\beta}$ [(3.6)]. Then iterate

$$\hat{\beta}_{(i+1)} = \psi_\beta\big(\hat{g}_{(i)}, \hat{\sigma}^2_{(i)}(\cdot)\big),$$
$$\hat{g}_{(i+1)}(\cdot) = \psi_g\big(\hat{\beta}_{(i+1)}, \hat{\sigma}^2_{(i)}(\cdot)\big),$$
$$\hat{\sigma}^2_{(i+1)}(\cdot) = \psi_\sigma\big(\hat{\beta}_{(i+1)}, \hat{g}_{(i+1)}(\cdot)\big)$$

until convergence.

It is not difficult to see that the resulting estimates $\hat{g}(\cdot)$ and $\hat{\sigma}^2(\cdot)$ at each iteration satisfy (4.1) and (4.2), so that the resulting $\hat{\beta}$ will still satisfy (4.3). Moreover, relation (M7) will be satisfied for estimates $\hat{\beta}$, $\hat{\sigma}^2(\cdot)$ and $\hat{g}(\cdot)$ after each iteration.

In case the mapping ψ_g (3.7) does not exist, the updating step for $\hat{g}$ in the Gauss–Seidel algorithm is omitted. Then one would update g via locally adaptive bandwidth choice. The first iteration step estimate for $\hat{\beta}$, when starting with undersmoothing bandwidths for $\hat{g}(\cdot)$ and $\hat{\sigma}^2(\cdot)$ will be asymptotically unbiased (see Theorem 4.3) and can therefore be used for the test in Theorem 5.1. Estimates for $\hat{g}(\cdot)$ and $\hat{\sigma}^2(\cdot)$ can then be obtained by further iteration or by applying local optimal bandwidths (3.9).

For kernel estimates, locally optimal bandwidths [in the sense of minimizing the asymptotic mean square error of $\hat{g}(t)$] are given by

$$b^*(t) = \left\{\frac{V}{2kB^2}\frac{\sigma^2(t)}{f(t)g^{(k)2}(t)}\frac{1}{n}\right\}^{1/(2k+1)} \quad \text{for } g^{(k)}(t) \neq 0, \tag{3.9}$$

where $V = \int K^2(x)\,dx$ and $B = ((-1)^k/k!)\int K(x)x^k dx$. Global analogs b^* are obtained by replacing $\sigma^2(t)$ and $f(t)g^{(k)^2}(t)$ by $\int\sigma^2$ and $\int fg^{(k)^2}$.

Efficient estimates for g are achieved by replacing the unknowns $\sigma^2(\cdot)$ and $g^{(k)}(\cdot)$ by consistent estimates [cf. Müller and Stadtmüller (1987) for details].

Undersmoothing bandwidths which satisfy (K4) and (K5) can be obtained from (3.9) by

$$\bar{b}(t) = b^*(t)n^{-1/k(2k+1)}, \tag{3.10}$$

and analogously for global bandwidths.

4. Asymptotic properties and rates of convergence. Consider first uniform convergence of the estimates $\hat{g}(\cdot)$ and $\hat{\sigma}^2(\cdot)$ [(3.2) and (3.3)] of the nonparametric components of our model. Lemmas 6.1–6.3 lead to the following uniform rates; analogous results for $\hat{g}$ under homoscedastic errors were discussed in Cheng and Lin (1981) and Müller and Stadtmüller (1987). For more details see Section 6.

THEOREM 4.1. *Under* $(M1)$–$(M6)$ *and* $(K1)$–$(K4)$,

$$\sup_{t\in I}|\hat{g}(t) - g(t)| = O_p\left(\left[\frac{\log n}{nb}\right]^{1/2} + b^k\right) \quad \text{where } I \subset (0,1). \tag{4.1}$$

THEOREM 4.2. *Under* $(M1)$–$(M6)$ *and* $(K1)$–$(K4)$,

$$\sup_{t\in I}|\hat{\sigma}^2(t) - \sigma^2(t)| = O_p\left(\left[\frac{\log n}{nb}\right]^{1/2} + b^k\right) \quad \text{where } I \subset (0,1). \tag{4.2}$$

This provides consistency of the nonparametric parts, with the usual nonparametric rates of convergence. Under somewhat stricter moment conditions, (4.1) and (4.2) can be modified to yield almost sure convergence results with the same rate.

Turning now to the parameter estimates $\hat{\beta}$ [(3.4)], the following asymptotic normality result establishes parametric rates of convergence. It provides the basic tool for assessing heteroscedasticity in our semiparametric variance function model. The derivation which is given in Section 7 requires the uniform convergence results (4.1) and (4.2).

THEOREM 4.3. *Under* $(M1)$–$(M9)$ *and* $(K1)$–$(K5)$,

$$\sqrt{n}\,(\hat{\beta} - \beta) \to_{\mathscr{L}} \mathscr{N}(\mu, \Sigma), \tag{4.3}$$

where $\mu = \Sigma_0^{-1}\eta$, $\Sigma = \Sigma_0^{-1}\Sigma_1\Sigma_0^{-1}$ *and* $\eta = (\eta_0, \dots, \eta_{p-1})^T$, $\Sigma_0 = (\rho_{\kappa,\lambda})_{0\le\kappa,\lambda\le p-1}$ *and* $\Sigma_1 = (\tau_{\kappa,\lambda})_{0\le\kappa,\lambda\le p-1}$ *are* $(p\times 1)$, $(p\times p)$ *and* $(p\times p)$ *matrices, respectively, with*

$$\begin{aligned}\eta_l = d^*B\Bigg\{&\int_0^1 f(u)q(u)H_l(g(u))G'(\sigma^2(u)) \\ &\times\Big[\big(g^2(u)+\sigma^2(u)\big)^{(k)} - 2g(u)g^{(k)}(u)\Big]\,du \\ &-\int_0^1 f(u)q(u)H_l(g(u))\left[\sum_{j=0}^{p-1}\theta_j H_j'(g(u))\right]g^{(k)}(u)\,du\Bigg\}, \\ &\qquad 0\le l\le p-1,\end{aligned} \tag{4.4}$$

d^* being defined in (K5) *and B as defined after* (3.9);

$$\rho_{\kappa,\lambda} = \int_0^1 f(u) q(u) H_\kappa(g(u)) H_\lambda(g(u))\, du; \tag{4.5}$$

and

$$\begin{aligned} \tau_{\kappa,\lambda} = \int_0^1 & f(u) q^2(u) H_\kappa(g(u)) H_\lambda(g(u)) \\ & \times \Bigg[\{G'(\sigma^2(u))\}^2 \{\mu_4(u) - \sigma^4(u)\} \\ & + 2G'(\sigma^2(u)) \left\{ \sum_{j=0}^{p-1} \theta_j H_j'(g(u)) \right\} \mu_3(u) \\ & + \left\{ \sum_{j=0}^{p-1} \theta_j H_j'(g(u)) \right\}^2 \sigma^2(u) \Bigg] du, \qquad 0 \le \kappa, \lambda \le p-1. \end{aligned} \tag{4.6}$$

Note that it follows from (4.3) and (4.4) that $\hat{\beta}$ will be asymptotically unbiased whenever $d^* = 0$ in (K5), which can be achieved by choosing bandwidths (3.10).

Note also that, by (4.4)–(4.6), the mean μ and the asymptotic covariance matrix Σ depend on the unknown parameter $\beta = (\theta_0, \dots, \theta_{p-1})^T$; therefore this result is not immediately applicable for the construction of confidence intervals or testing. Condition (M9) guarantees that both matrices Σ_0 and Σ_1 are symmetric and positive definite when $q \equiv 1$, which is a natural initial choice of the weight function $q(\cdot)$. Optimal weights $q(t_i)$ would minimize $|\det \Sigma|$ and depend in a complex way on the unknowns $g(\cdot)$, $\sigma^2(\cdot)$ and β.

As a final note, we observe that the asymptotic covariance matrix Σ depends on third and fourth moment functions $\mu_3(\cdot)$ and $\mu_4(\cdot)$ through Σ_1 [(4.6)]. These functions are usually unknown. Their estimation is discussed in the next section. In much of the classical parametric literature it is assumed for the sake of simplicity that the error distribution is symmetric and thus $\mu_3(\cdot) \equiv 0$, which simplifies the expression for Σ, but these classical results can be extended as well to allow for general third moment functions $\mu_3(\cdot)$. In our context it is necessary to consider this more general situation, as we wish to cover cases involving quasi-Poisson or quasi-Gamma errors which necessarily have $\mu_3(\cdot) \neq 0$.

5. A test for heteroscedasticity. In this section, we develop a data-dependent test statistic whose distribution under the null hypothesis of homoscedasticity is derived. Besides this test, we obtain asymptotic confidence regions for the unknown parameter vector $\beta = (\theta_0, \dots, \theta_{p-1})^T$ as defined in (M7).

Observe that assumption (K5) as needed for Theorem 4.3 amounts to some undersmoothing of the nonparametric curve estimates even in the case $d^* > 0$. The mean squared error optimal bandwidth sequence is seen to be

$b^* \sim n^{-1/(2k+1)}c^*$ for some constant c^*, as given in (3.9) and the following remark. We will always have $n^{1/2}b^{*k} \to \infty$, so that bandwidth sequences satisfying (K5) must be undersmoothing. If $d^* > 0$ in assumption (K5), the mean μ in the limiting distribution (4.3) of Theorem 4.3 will be nonvanishing and needs to be estimated [cf. Lemma 5.3 in Müller and Stadtmüller (1987)]. The additional assumption (K6) which is used exclusively in this section implies $d^* = 0$ in (K5), and it therefore corresponds to further undersmoothing of the nonparametric curve estimates for the purpose of estimating the parametric part β. Bandwidths (3.10) and corresponding bandwidth estimates achieve the necessary degree of undersmoothing.

The null hypothesis corresponding to homoscedasticity in (M7) is

$$\overline{H}_0\colon \theta_1 = \cdots = \theta_{p-1} = 0. \tag{5.1}$$

Under (K6), we have $\mu = 0$, and (4.3) becomes $\sqrt{n}(\hat{\beta} - \beta) \to_{\mathscr{L}} \mathscr{N}(0, \Sigma)$. Define the $(p-1) \times p$ matrix $\Lambda = (\lambda_{ij})_{0 \le i \le p-2; 0 \le j \le p-1}$, where $\lambda_{ij} = 1$ if $j = i+1$, and $\lambda_{ij} = 0$ otherwise. Then the null hypothesis (5.1) can be recast as

$$\overline{H}_0\colon \Lambda\beta = 0,$$

and it follows that, under $\overline{H}_0$,

$$\sqrt{n}\,\Lambda\hat{\beta} \to_{\mathscr{L}} \mathscr{N}(0, \Lambda\Sigma\Lambda^T). \tag{5.2}$$

To find a data-based test statistic, we still need to estimate $\Sigma = \Sigma_0^{-1}\Sigma_1\Sigma_0^{-1}$ as defined in (4.3), (4.5) and (4.6). Considering estimates

$$\hat{\Sigma}_0 = (\hat{\rho}_{\kappa,\lambda})_{0 \le \kappa, \lambda \le p-1}, \tag{5.3}$$

$$\hat{\Sigma}_1 = (\hat{\tau}_{\kappa,\lambda})_{0 \le \kappa, \lambda \le p-1}, \tag{5.4}$$

we note that the corresponding quantities $\rho_{\kappa,\lambda}$ and $\tau_{\kappa,\lambda}$ [see (4.5) and (4.6)] contain the unknowns g, $\sigma^2(\cdot)$, $\mu_3(\cdot)$, $\mu_4(\cdot)$ and θ_j. We estimate the integrals by corresponding sample-based sums; due to averaging n terms, we expect parametric rates of convergence for these estimates, which are given by

$$\hat{\rho}_{\kappa,\lambda} = \frac{1}{n}\sum_{i=1}^{n} H_\kappa(\hat{g}(t_i))H_\lambda(\hat{g}(t_i))q(t_i), \qquad 0 \le \kappa, \lambda \le p-1, \tag{5.5}$$

$$\begin{aligned}\hat{\tau}_{\kappa,\lambda} = \frac{1}{n}\sum_{i=1}^{n} & H_\kappa(\hat{g}(t_i))H_\lambda(\hat{g}(t_i))q^2(t_i) \\ & \times \Bigg\{ \big[G'(\hat{\sigma}^2(t_i))\big]^2 \Big[\hat{\mu}_4(t_i) - \big(\hat{\sigma}^2(t_i)\big)^2\Big] \\ & \quad + 2G'(\hat{\sigma}^2(t_i)) \left[\sum_{j=0}^{p-1} \hat{\theta}_j H_j'(\hat{g}(t_i))\right] \hat{\mu}_3(t_i) \\ & \quad + \left[\sum_{j=0}^{p-1} \hat{\theta}_j H_j'(\hat{g}(t_i))\right]^2 \hat{\sigma}^2(t_i) \Bigg\}.\end{aligned} \tag{5.6}$$

Here, $\hat{g}$ and $\hat{\sigma}^2(\cdot)$ are curve estimates as given in (3.2) and (3.3); $(\hat{\theta}_0, \dots, \hat{\theta}_{p-1})$ is the parameter estimate (3.4); and estimates $\hat{\mu}_3(\cdot)$ and $\hat{\mu}_4(\cdot)$ of the moment functions of the errors $\mu_3(\cdot)$ and $\mu_4(\cdot)$ [see (M4)] are given by

$$\begin{aligned} \hat{\mu}_3(t) &= \sum_{i=1}^{n} W_i(t) y_i^3 - \{\hat{g}(t)\}^3 - 3\hat{g}(t)\hat{\sigma}^2(t), \\ \hat{\mu}_4(t) &= \sum_{i=1}^{n} W_i(t) y_i^4 - \{\hat{g}(t)\}^4 - 4\hat{g}(t)\hat{\mu}_3(t) - 6(\hat{g}(t))^2\hat{\sigma}^2(t). \end{aligned} \tag{5.7}$$

We note that the quality of the proposed inference procedures depends to a large extent on the feasibility of estimates $\hat{\mu}_3(\cdot)$ and $\hat{\mu}_4(\cdot)$. While these estimates are shown to be consistent in the following lemma, satisfactory estimation in practice may require large sample sizes. In some specific applications, one may be able to assume symmetric errors [i.e., $\mu_3(\cdot) = 0$], so that only $\mu_4(\cdot)$ remains to be estimated.

Given estimates $\hat{\Sigma}_0$ and $\hat{\Sigma}_1$ [(5.3) and (5.4)], we set

$$\hat{\Sigma} = \hat{\Sigma}_0^{-1}\hat{\Sigma}_1\hat{\Sigma}_0^{-1}. \tag{5.8}$$

Observe the following.

LEMMA 5.1. *Under* (M1)–(M9) *and* (K1)–(K5),

$$\hat{\Sigma} \to_P \Sigma. \tag{5.9}$$

The proof, which is omitted, is based on Lemmas 7.1 and 7.3 and establishes the convergences $\hat{\Sigma}_0 \to_P \Sigma_0$, $\hat{\Sigma}_1 \to_P \Sigma_1$ and $\hat{\mu}_3(t) \to_P \mu_3(t)$, $\hat{\mu}_4(t) \to_P \mu_4(t)$, uniformly in t.

Note that under $\overline{H}_0$, $\sum_{j=0}^{p-1} \theta_j H_j'(g(t)) = 0$ for all t, since $H_0' \equiv 0$. Let Ξ be an $(m \times p)$-matrix of rank m, $m \le p$, and let ζ_0 and ζ_{1n} be m-vectors. To test the hypothesis

$$\tilde{H}_0\colon \Xi\beta = \zeta_0,$$

consider the test statistic

$$T_n = n\big(\Xi\hat{\beta} - \zeta_0\big)^T \big(\Xi\hat{\Sigma}\Xi^T\big)^{-1}\big(\Xi\hat{\beta} - \zeta_0\big). \tag{5.10}$$

Note that, for the special case of testing $\overline{H}_0$, corresponding to homoscedasticity, this becomes

$$T_n' = n\big(\Lambda\hat{\beta}\big)^T \big(\Lambda\hat{\Sigma}\Lambda^T\big)^{-1}\big(\Lambda\hat{\beta}\big), \tag{5.11}$$

with Λ as defined before (5.2). Consider a family of simple alternative hypotheses,

$$\tilde{H}_{1n}\colon \Xi\beta = \zeta_{1n}.$$

THEOREM 5.1. *Assuming* (M1)–(M9) *and* (K1)–(K6), *it holds that, under the null hypothesis* $\tilde{H}_0$,

$$T_n \to_{\mathscr{L}} \chi_m^2, \tag{5.12}$$

where χ_m^2 *has a central* χ^2 *distribution with m degrees of freedom. Assume that the alternatives* $\tilde{H}_{1n}$ *satisfy*

$$n(\zeta_{1n} - \zeta_0)^T (\Xi\Sigma\Xi^T)^{-1}(\zeta_{1n} - \zeta_0) \to \rho^2, \tag{5.13}$$

for a fixed real constant ρ. *Then, under* $\tilde{H}_{1n}$,

$$T_n \to_{\mathscr{L}} \chi_m^2(\rho^2), \tag{5.14}$$

where $\chi_m^2(\rho^2)$ *has a noncentral* χ^2 *distribution with m degrees of freedom and noncentrality parameter* ρ^2.

The proof is in Section 7. The application to the construction of a level-α test is immediate: reject $\tilde{H}_0$ if $T_n > \chi^2_{m;\alpha}$, the $100(1-\alpha)\%$ quantile of the corresponding χ^2 distribution; power calculations for specific sequences of alternatives follow from (5.14). Choosing $m = p$, $\Xi = \text{id}$ and $\zeta_0 = \beta$, an asymptotic $100(1-\alpha)\%$ confidence region for β is obtained as $\{\beta: n(\hat{\beta}-\beta)^T \times\hat{\Sigma}^{-1}(\hat{\beta}-\beta) \le \chi^2_{p;1-\alpha}\}$.

It is clear that tests for $\tilde{H}_0$ include tests for many simple null hypotheses, besides homoscedasticity, which may be of interest in various models (M7). Such null hypotheses may correspond to statements like "no overdispersion," "quasi-Negative Binomial model" or "quasi-Gamma model" (see the examples given in the Introduction).

6. Auxiliary results and proofs of Theorems 4.1 and 4.2. Throughout this section, we will use the notation $\rho_n = (\log n/nb)^{1/2}$ and $B = ((-1)^k/k!)\int K(x)x^k\,dx$, and in this and the following section repeatedly make use of the following facts. According to (M3), there exist constants c_1 and c_2, $0 < c_1 \le c_2 < \infty$, such that

$$\frac{c_1}{n} \le \inf_i |t_i - t_{i-1}| \le \sup_i |t_i - t_{i-1}| \le \frac{c_2}{n}. \tag{6.1}$$

Furthermore, according to (M2) and (M3), by an application of the mean value theorem,

$$s_i - s_{i-1} = \big(nf(t_i)\big)^{-1}(1 + O(n^{-1})), \tag{6.2}$$

and for any Lipschitz-continuous function h defined on $[0, 1]$, Riemann sum approximation yields

$$\sum_{i=1}^{n} (s_i - s_{i-1}) f(t_i) h(t_i) = \int_0^1 f(u)h(u)\,du + O\left(\frac{1}{n}\right).$$

Therefore, for any such function,

$$\frac{1}{n}\sum_{i=1}^{n} h(t_i) = \int_0^1 f(u)h(u)\,du + O\left(\frac{1}{n}\right). \tag{6.3}$$

Observing that the number of nonzero summands in $\sum K((x - t_i)/b)$ is $O(nb)$, uniformly in $x \in [0, 1]$, we obtain analogously

$$\frac{1}{n}\sum_{i=1}^{n} h(t_i) K\left(\frac{x - t_i}{b}\right) = \int_0^1 f(u)h(u)K\left(\frac{x-u}{b}\right) du + O\left(\frac{1}{n}\right). \tag{6.4}$$

We first prove Theorems 4.1 and 4.2, which will be needed to establish Theorem 4.3. The proofs require the following auxiliary results, Lemmas 6.1–6.3.

LEMMA 6.1. *Under* (M1)–(M3), (M6), (K1) *and* (K2), *for* $t \in (0,1)$,

$$\hat{g}(t) - g(t) = [b^k B g^{(k)}(t) + O([nb]^{-1}) + o(b^k)] + \sum_{i=1}^{n} W_i(t)\delta_i. \tag{6.5}$$

The proof is standard by a Taylor expansion for the bias [cf., e.g., Müller (1988), (4.9) and (4.13)].

LEMMA 6.2. *Under* (M1)–(M6) *and* (K1)–(K4), *for* $t \in (0,1)$,

$$\begin{aligned}\hat{\sigma}^2(t) - \sigma^2(t) = &\Big[b^k B\Big[\big(g^2(t) + \sigma^2(t)\big)^{(k)} - 2g(t)g^{(k)}(t)\Big] \\ &+ O\big([nb]^{-1}\big) + o(b^k)\Big] + \left[\sum_{i=1}^{n} W_i(t)\{\delta^2(t_i) - \sigma^2(t_i)\}\right] \\ &+ O_p\left(\left\{\left[\frac{\log n}{nb}\right]^{1/2} + b^k\right\}^2\right) + o_p(n^{-1/2}).\end{aligned} \tag{6.6}$$

PROOF. The proof of Theorem 4.1 given below does not require Lemma 6.2, and we may use (4.1), as well as (6.5). Therefore, noting that $\delta_i = \delta(t_i)$ and

$$\begin{aligned}&\hat{\sigma}^2(t) - \sigma^2(t) \\ &\quad= \sum W_i(t)\big(g^2(t_i) + \sigma^2(t_i)\big) \\ &\qquad+ 2\sum W_i(t)g(t_i)\delta(t_i) + \sum W_i(t)\{\delta^2(t_i) - \sigma^2(t_i)\} \\ &\qquad-\big(g^2(t) + \sigma^2(t)\big) - 2g(t)\big[b^k B_k g^{(k)}(t) + O\big([nb]^{-1}\big) + o(b^k)\big] \\ &\qquad- 2g(t)\sum W_i(t)\delta(t_i) + O_p\big((\rho_n + b^k)^2\big).\end{aligned}$$

Observing that

$$\begin{aligned}&\sum W_i(t)\big(g^2(t_i) + \sigma^2(t_i)\big) - \big(g^2(t) + \sigma^2(t)\big) \\ &\qquad= b^k B\big(g^2(t) + \sigma^2(t)\big)^{(k)} + O\big([nb]^{-1}\big) + o(b^k)\end{aligned}$$

and $\sum W_i(t)(g(t_i) - g(t))\delta(t_i) = o_p(n^{-1/2})$ completes the proof. □

LEMMA 6.3. *Let* $I \subset [0,1]$ *be a compact interval, where* $I = [0,1]$ *can be chosen in the case that boundary modifications are employed for estimators* (3.2) *and* (3.3) *or that* g *and* $\sigma^2(\cdot)$ *are cyclical. Otherwise, assume* $I \subset (0,1)$. *Let* $\eta(t_1), \dots, \eta(t_n)$ *be independent random variables with* $E\eta(t_i) = 0$ *and* $E\eta^2(t_i) = \varphi(t_i)$, *for a function* φ *satisfying* $0 < \inf|\varphi(t)| \le \sup|\varphi(t)| < \infty$.

Furthermore, for some $s > 2$, *let* $E(|\eta(t_i)|^s) < c < \infty$. *Then, assuming* (M2), (M3) *and* (K1)–(K4),

$$\sup_{t\in I}\left|\sum_{i=1}^{n} W_i(t)\eta(t_i)\right| = O_p\left(\left[\frac{\log n}{nb}\right]^{1/2}\right). \tag{6.7}$$

PROOF. Defining $\varepsilon_i = \eta(t_i)/[\varphi(t_i)]^{1/2}$, it is sufficient to show

$$\sup_{t\in I}|\Sigma W_i'(t)\varepsilon_i| = O_p(\rho_n), \tag{6.8}$$

where $W_i'(t) = W_i(t)\varphi(t_i)^{1/2}$. Consider an equidistant grid $\tau_{jn} \in [0, 1]$, $j \geq 1$, with $\max_j|\tau_j - \tau_{j-1}| = n^{-3/\zeta}$, $\tau_1 = 0$ and $\tau_n = 1$, where ζ is as in (K3). Define $\tau(t) = \arg\min_{\tau_j}|\tau_j - t|$ and $\bar{\varepsilon}_i = \varepsilon_i I\{|\varepsilon_i| \leq (in)^{1/r}\}$, where r is as in (K4), and $I(A)$ denotes the indicatior of a set A. Then

$$\begin{aligned}\sup_{t\in I}\left|\sum W_i'(t)\varepsilon_i\right| &\leq \sup_{t\in I}\left|\sum (W_i'(t) - W_i'(\tau(t)))\varepsilon_i\right| + \sup_{t\in I}\left|\sum W_i'(\tau(t))(\varepsilon_i - \bar{\varepsilon}_i)\right| \\ &\quad + \sup_{t\in I}\left|\sum W_i'(\tau(t))\bar{\varepsilon}_i\right| = \mathrm{I} + \mathrm{II} + \mathrm{III}.\end{aligned}$$

For I, observe that, by (6.1), $\sup_i|s_i - s_{i-1}| = O(n^{-1})$, so that there exist constants $c > 0$ with

$$\begin{aligned}&\sup_{t\in I}\sum (W_i'(t) - W_i'(\tau(t)))^2 \\ &\quad \leq \frac{c}{nb}\sup_{t\in I}\int\left|K(v) - K\left(v + \frac{\tau(t) - t}{b}\right)\right| dv = O(n^{-4}b^{-2}).\end{aligned}$$

Applying the Cauchy–Schwarz inequality and the weak law of large numbers, it follows that $\mathrm{I} = O_p(\rho_n)$.

As in the proof of Lemma 5.2 and the following remark in Müller and Stadtmüller (1987), one shows that $\mathrm{II} = O_p(n^{1/r}\sup_{t\in I, 1\leq i\leq n}|W_i'(t)|)$, and this is bounded by $O_p(n^{1/r}\sup_{t\in I, 1\leq i\leq n}|W_i(t)|)$, which is seen to be $O_p(\rho_n)$ by (K4).

For III, we proceed as in Müller and Stadtmüller (1987), requiring the bound

$$\sup_{t\in I}\left(\sum W_i'^2(t)\log n\right)^{1/2} \leq C\sup_{t\in I}\left(\sum W_i^2(t)\log n\right)^{1/2} = O(\rho_n). \qquad \square$$

PROOF OF THEOREM 4.1. Noting that the remainder terms in (6.5) are uniform in t, Lemma 6.1 implies $\sup_t|E\hat{g}(t) - g(t)| = O(b^k)$ for the bias part. For the stochastic part, apply Lemma 6.3, choosing $\eta(t_i) = \delta(t_i)$ and $\varphi(\cdot) = \sigma^2(\cdot)$. $\square$

PROOF OF THEOREM 4.2. For the bias part we use Lemma 6.2, noting the uniformity of all remainder terms. For the stochastic part, apply Lemma 6.3

with $\varphi(\cdot) = \mu_4(\cdot) - \sigma^4(\cdot)$ and errors $\eta_i = \delta_i^2 - E\delta_i^2$. For these errors η_i, (M5) holds with $s' = s/2 > 2$ and (K4) is used for r with $2 < r < s'$. □

7. Proofs of Theorems 4.3 and 5.1. For the proof of Theorem 4.3 we need some additional notation. Let $\gamma_n = (\log n/nb)^{1/2} + b^k$. Observe that by (M8), Taylor expansions and Theorems 4.1 and 4.2,

$$G(\hat{\sigma}^2(t)) - G(\sigma^2(t)) = G'(\sigma^2(t))[\hat{\sigma}^2(t) - \sigma^2(t)] + O_p(\gamma_n^2) \tag{7.1}$$

and

$$H_j(\hat{g}(t)) - H_j(g(t)) = H_j'(g(t))[\hat{g}(t) - g(t)] + O_p(\gamma_n^2), \quad 0 \le j \le p-1. \tag{7.2}$$

For a matrix $A = [a_{ij}]_{0 \le i \le m_1 - 1, 0 \le j \le m_2 - 1}$ with $a_{ij} = O_p(\gamma_n)$, we write $A = [O_p(\gamma_n)]_{m_1 \times m_2}$.

For the matrices X, $\hat{X}$, Z and $\hat{Z}$, defined before (3.5), we obtain, from (7.1) and (7.2),

$$\hat{Z} - Z = S + [O_p(\gamma_n^2)]_{n \times 1}, \tag{7.3}$$

$$\hat{X} - X = V + [O_p(\gamma_n^2)]_{n \times p}, \tag{7.4}$$

where

$$S = (G'(\hat{\sigma}^2(t_1))(\sigma^2(t_1) - \sigma^2(t_1)), \dots, G'(\sigma^2(t_n))(\hat{\sigma}^2(t_n) - \sigma^2(t_n)))^T$$

and

$$V = (v_{ij})_{1 \le i \le n, 1 \le j \le p} \quad \text{with } v_{ij} = H_{j-1}'(g(t_i))(\hat{g}(t_i) - g(t_i)).$$

Observing (6.2) and $S = [O_p(\gamma_n)]_{n \times 1}$, $V = [O_p(\gamma_n)]_{n \times p}$, $V^T Q^{-1} S = n[O_p(\gamma_n^2)]_{p \times 1}$ and $V^T Q^{-1} V = n[O_p(\gamma_n^2)]_{p \times p}$, we obtain the following auxiliary results.

LEMMA 7.1. *As $n \to \infty$,*

$$\frac{1}{n} X^T Q^{-1} X \to \Sigma_0, \tag{7.5}$$

$$\frac{1}{n} \hat{X}^T Q^{-1} \hat{X} = \frac{1}{n} X^T Q^{-1} X + \frac{1}{n} X^T Q^{-1} V + \frac{1}{n} V^T Q^{-1} X + [O_p(\gamma_n^2)]_{p \times p}, \tag{7.6}$$

$$\frac{1}{n} \hat{X}^T Q^{-1} \hat{X} \to_p \Sigma_0, \tag{7.7}$$

where Σ_0 is given in (4.5).

LEMMA 7.2.

$$\sqrt{n}\,(\hat{\beta}-\beta) = \left(n^{-1}\hat{X}^T Q^{-1}\hat{X}\right)^{-1} \times \left[n^{-1/2}X^T Q^{-1}S - n^{-1/2}X^T Q^{-1}V\beta + \left[o_p(1)\right]_{p\times 1}\right] \tag{7.8}$$

PROOF OF THEOREM 4.3. According to Lemmas 7.1 and 7.2, it is sufficient to show that

$$n^{-1/2}X^T Q^{-1}S - n^{-1/2}X^T Q^{-1}V\beta \to_{\mathscr{L}} \mathscr{N}(\eta, \Sigma_1), \tag{7.9}$$

where η and Σ_1 are defined as in (4.4) and (4.6). Observe that

$$n^{-1/2}X^t Q^{-1}S - n^{-1/2}X^T Q^{-1}V\beta = (\zeta_0, \dots, \zeta_{p-1})^T, \tag{7.10}$$

where

$$\zeta_j = n^{-1/2}\sum_{i=1}^{n} H_j(g(t_i))q(t_i)G'(\sigma^2(t_i))(\hat{\sigma}^2(t_i) - \sigma^2(t_i)) - n^{-1/2}\sum_{i=1}^{n} H_j(g(t_i))q(t_i)\left[\sum_{l=0}^{p-1}\theta_l H_l'(g(t_i))\right](\hat{g}(t_i) - g(t_i)), \quad 0 \le j \le p-1.$$

Applying the decompositions for $\hat{g}(t_i) - g(t_i)$ and $\hat{\sigma}^2(t_i) - \sigma^2(t_i)$ provided in Lemmas 6.1 and 6.2, observing (K5) we obtain

$$\zeta_j = \zeta_j' + \zeta_j'' + o_p(1), \quad 0 \le j \le p-1,$$

where

$$\zeta_j' = n^{-1/2}b^k B\left\{\sum_{i=1}^{n} H_j(g(t_i))q(t_i)G'(\sigma^2(t_i)) \times\left[\left(g^2(t_i) + \sigma^2(t_i)\right)^{(k)} - 2g(t_i)g^{(k)}(t_i)\right] - \sum_{i=1}^{n} H_j(g(t_i))q(t_i)\left[\sum_{l=0}^{p-1}\theta_l H_l'(g(t_i))\right]g^{(k)}(t_i)\right\},$$

$$\zeta_j'' = n^{-1/2}\left\{\sum_{i=1}^{n} H_j(g(t_i))q(t_i)G'(\sigma^2(t_i))\left[\sum_{\lambda=1}^{n} W_\lambda(t_i)\{\delta^2(t_\lambda) - \sigma^2(t_\lambda)\}\right] - \sum_{i=1}^{n} H_j(g(t_i))q(t_i)\left[\sum_{l=0}^{p-1}\theta_l H_l'(g(t_i))\right]\left[\sum_{\lambda=1}^{n} W_\lambda(t_i)\delta(t_\lambda)\right]\right\}, \quad 0 \le j \le p-1.$$

Note that for the nonrandom part, by (6.2) and (K5), $\zeta_j' = \eta_j + o(1)$, $0 \le j \le p-1$, where η_j is as in (4.4). Thus (7.10) follows from

$$(\zeta_0'', \dots, \zeta_{p-1}'')^T \to_{\mathscr{L}} \mathscr{N}(0, \Sigma_1). \tag{7.11}$$

To prove (7.11), we observe that, by (6.2) and (6.4),

$$\begin{aligned}
&\sum_{i=1}^{n} W_\lambda(t_i) q(t_i) H_j(g(t_i)) G'(\sigma^2(t_i)) \\
&\quad = \frac{1}{f(t_\lambda)}\left(1 + O\left(\frac{1}{n}\right)\right) \\
&\quad\quad \times \left\{\frac{1}{b}\int_0^1 f(u) q(u) H_j(g(u)) G'(\sigma^2(u)) K\left(\frac{t_\lambda - u}{b}\right) du + O\left(\frac{1}{nb}\right)\right\} \\
&\quad = H_j(g(t_\lambda)) G'(\sigma^2(t_\lambda)) q(t_\lambda) + O(b) + O((nb)^{-1}), \\
&\qquad\qquad 1 \le \lambda \le n,\ 0 \le j \le p-1,
\end{aligned} \tag{7.12}$$

where the last step requires (M6), (K1) and a Taylor expansion. Similarly,

$$\begin{aligned}
&\sum_{i=1}^{n} W_\lambda(t_i) H_j(g(t_i)) q(t_i)\left[\sum_{l=0}^{p-1} \theta_l H_l'(g(t_i))\right] \\
&\quad = H_j(g(t_\lambda))\left[\sum_{l=0}^{p-1} \theta_l H_l'(g(t_\lambda))\right] q(t_\lambda) \\
&\quad\quad + O(b) + O((nb)^{-1}), \qquad 1 \le \lambda \le n,\ 0 \le j \le p-1.
\end{aligned} \tag{7.13}$$

Combining (7.12) and (7.13) with

$$n^{-1/2} \sum_{i=1}^{n} \left\{O(b) + O((nb)^{-1})\right\}\left\{\delta^2(t_i) - \sigma^2(t_i)\right\} = o_p(1),$$

$$n^{-1/2} \sum_{i=1}^{n} \left\{O(b) + O((nb)^{-1})\right\}\delta(t_i) = o_p(1),$$

which follows by calculating first and second moments, we obtain

$$\zeta_j'' = \sum_{i=1}^{n} U_{nij} + o_p(1), \qquad 0 \le j \le p-1, \tag{7.14}$$

where

$$\begin{aligned}
U_{nij} = n^{-1/2}\Bigg\{ & H_j(g(t_i)) G'(\sigma^2(t_i)) q(t_i)\left\{\delta^2(t_i) - \sigma^2(t_i)\right\} \\
& - H_j(g(t_i)) q(t_i)\left[\sum_{l=0}^{p-1} \theta_l H_l'(g(t_i))\right]\delta(t_i)\Bigg\}.
\end{aligned}$$

Observe that, by (6.3),

$$\operatorname{Cov}(\zeta_k'', \zeta_\lambda'') \to \int_0^1 f(u) q^2(u) H_k(g(u)) H_\lambda(g(u)) \times \Bigg\{ \big[G'(\sigma^2(u))\big]^2 (\mu_4(u) - \sigma^4(u)) + 2G'(\sigma^2(u)) \left[\sum_{j=0}^{p-1} \theta_l H_l'(g(u))\right] \mu_3(u) + \left[\sum_{l=0}^{p-1} \theta_l H_l'(g(u))\right]^2 \sigma^2(u) \Bigg\}\, du,$$

noting that $E(\delta^2(t_i) - \sigma^2(t_i))^2 = \mu_4(t_i) - \sigma^4(t_i)$ and $E\delta(t_i)(\delta^2(t_i) - \sigma^2(t_i)) = \mu_3(t_i)$.

Now let $a = (a_0, \dots, a_{p-1})^T$ and define $U_{ni} = \sum_{j=0}^{p-1} a_j U_{nij}$. Then it follows that $E(\Sigma\, U_{ni}) = 0$ and $\operatorname{var}(\Sigma\, U_{ni}) \to a^T \Sigma_1 a$, where Σ_1 is as given in (4.6). From (M5),

$$\begin{aligned}\sum_{i=1}^n E|U_{ni}|^{s/2} &\le \sum_{i=1}^n \left| n^{-1/2} \left[\sum_{l=0}^{p-1} a_l H_l(g(t_i)) q(t_i)\right]\right|^{s/2} 2^{s/2} \\ &\quad \times \Bigg\{ |G'(\sigma^2(t_i))|^{s/2} E|\delta^2(t_i) - \sigma^2(t_i)|^{s/2} + \left|\sum_{l=0}^{p-1} \theta_l H_l'(g(t_i))\right|^{s/2} E|\delta(t_i)|^{s/2} \Bigg\} \\ &= O(n^{-(s-4)/4}) \to 0.\end{aligned}$$

Hence the central limit theorem implies

$$a^T(\zeta_0'', \dots, \zeta_{p-1}'')^T \to_{\mathscr{L}} \mathscr{N}(0, a^T \Sigma_1 a),$$

and (7.11) follows by applying the Cramér–Wold device. □

We turn now to the proof of Theorem 5.1. The first step is to prove Lemma 5.1, which requires the following auxiliary result.

LEMMA 7.3. *Under* (M1)–(M9),

$$\sup_{t \in I} |\hat{\mu}_j(t) - \mu_j(t)| = o_p(1) \quad \textit{for } j = 3, 4. \tag{7.15}$$

PROOF. Apply Lemma 6.3, choosing $\eta(t_i) = (\delta(t_i))^j - E((\delta(t_i))^j)$. Note that (M5) implies that the assumptions of Lemma 6.3 are satisfied for $j = 3, 4$. Observing that

$$E(y_i^3) = (g(t_i))^3 + 3g(t_i)\sigma^2(t_i) + \mu_3(t_i),$$
$$E(y_i^4) = (g(t_i))^4 + 6(g(t_i))^2\sigma^2(t_i) + 4g(t_i)\mu_3(t_i) + \mu_4(t_i),$$

one first shows the result for $j = 3$, applying Theorems 4.1 and 4.2. Then this result is used to establish the case $j = 4$. □

PROOF OF THEOREM 5.1. Theorem 4.3 and Lemma 5.1 imply that

$$n^{1/2}(\Xi\hat{\Sigma}\Xi^T)^{-1/2}(\Xi\hat{\beta} - \zeta_0) \to_{\mathscr{L}} \mathscr{N}(0, I),$$

which implies (5.12), whereas under alternatives $\tilde{H}_{1n}$,

$$\begin{aligned} n^{1/2}(\Xi\hat{\Sigma}\Xi^T)^{-1/2}(\Xi\hat{\beta} - \zeta_0) &= n^{1/2}(\Xi\hat{\Sigma}\Xi^T)^{-1/2}(\Xi\hat{\beta} - \zeta_{1n}) \\ &\quad + n^{1/2}(\Xi\hat{\Sigma}\Xi^T)^{-1/2}(\zeta_{1n} - \zeta_0) \\ &\to \mathscr{N}(\mu, I) \quad \text{with } \|\mu\| = \rho, \end{aligned}$$

which implies (5.14). □

Acknowledgments. We thank Dr. J. A. Nelder and three referees for helpful comments on earlier versions of this paper. One referee provided one of the example scenarios in the Introduction and raised several of the issues which are discussed in the text.

REFERENCES

BOX, G. E. P. and HILL, W. J. (1974). Correcting inhomogeneity of variance with power transformation weighting. *Technometrics* **16** 385–389.

BREUSCH, T. S. and PAGAN, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrika* **47** 1287–1294.

CARROLL, R. J. (1982). Adapting for heteroscedasticity in linear models. *Ann. Statist.* **10** 1224–1233.

CARROLL, R. J. and RUPPERT, D. (1981). On robust tests for heteroscedasticity. *Ann. Statist.* **9** 206–210.

CARROLL, R. J. and RUPPERT, D. (1988). *Transformation and Weighting in Regression.* Chapman and Hall, New York.

CHENG, K. F. and LIN, P. E. (1981). Nonparametric estimation of a regression function. *Z. Wahrsch. Verw. Gebiete* **57** 223–233.

COOK, R. D. and WEISBERG, S. (1982). *Residuals and Influence in Regression.* Chapman and Hall, New York.

COOK, R. D. and WEISBERG, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika* **70** 1–10.

DAVIDIAN, M. and CARROLL, R. J. (1987). Variance function estimation. *J. Amer. Statist. Assoc.* **82** 1079–1091.

EUBANK, R. L. (1988). *Spline Smoothing and Nonparametric Regression.* Dekker, New York.

FULLER, W. A. and RAO, J. N. K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *Ann. Statist.* **6** 1149–1158.

GASSER, TH. and MÜLLER, H. G. (1984). Nonparametric estimation of regression functions and their derivatives by the kernel method. *Scand. J. Statist.* **11** 171–185.

HARRISON, M. J. and MCCABE, B. P. M. (1979). A test for heteroscedasticity based on least squares residuals. *J. Amer. Statist. Assoc.* **74** 494–500.

JOBSON, J. D. and FULLER, W. A. (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. *J. Amer. Statist. Assoc.* **75** 176–181.

KOENKER, R. and BASSETT, G. (1981). Robust tests for heteroscedasticity based on regression quantiles. *Econometrika* **50** 43–61.

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models.* Chapman and Hall, New York.

MÜLLER, H. G. (1988). *Nonparametric Regression Analysis of Longitudinal Data.* Springer, New York.

MÜLLER, H. G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika* **78** 221–230.

MÜLLER, H. G. and STADTMÜLLER, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15** 610–625.

PRIESTLEY, M. B. and CHAO, M. T. (1972). Nonparametric function fitting. *J. Roy. Statist. Soc. Ser. B* **34** 384–392.

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616

MERCK RESEARCH LABORATORIES
P.O. BOX 2000
RAHWAY, NEW JERSEY 07065