Journal of the Royal Statistical Society
SERIES B Statistical Methodology

J. R. Statist. Soc. B (2011)
73, *Part* 1, *pp.* 81–98

Local and omnibus goodness-of-fit tests in classical measurement error models

Yanyuan Ma and Jeffrey D. Hart,

Texas A&M University, College Station, USA

Ryan Janicki

US Census Bureau, Suitland, USA

and Raymond J. Carroll

Texas A&M University, College Station, USA

[Received August 2009. Final revision May 2010]

Summary. We consider functional measurement error models, i.e. models where covariates are measured with error and yet no distributional assumptions are made about the mismeasured variable. We propose and study a score-type local test and an orthogonal series-based, omnibus goodness-of-fit test in this context, where no likelihood function is available or calculated—i.e. all the tests are proposed in the semiparametric model framework. We demonstrate that our tests have optimality properties and computational advantages that are similar to those of the classical score tests in the parametric model framework. The test procedures are applicable to several semiparametric extensions of measurement error models, including when the measurement error distribution is estimated non-parametrically as well as for generalized partially linear models. The performance of the local score-type and omnibus goodness-of-fit tests is demonstrated through simulation studies and analysis of a nutrition data set.

Keywords: Efficient estimation; Efficient testing; Errors in variables; Goodness-of-fit tests; Local alternatives; Measurement error; Score testing; Semiparametric models

1. Introduction

Measurement error models have received much attention recently, with many new estimation procedures having been developed in various linear and non-linear models as well as in non-parametric and semiparametric models (Carroll *et al.*, 2006). It is well known that, in the measurement error context, parametric models are much more favourable than non-parametric models in terms of drawing inferences, accuracy, power, etc. Here, by a parametric model, we mean that the response variable Y depends on the mismeasured covariate X through a parametric function with some unknown parameters, whereas, by a non-parametric model, we mean that Y depends on X via an unspecified smooth function. Parameter estimation and unknown regression function estimation require totally different treatments, and the subsequent asymptotic properties are vastly different. Thus, a reasonable practice is to try to adopt a parametric model, provided that the model is sufficient to capture the features of the data.

Address for correspondence: Yanyuan Ma, Department of Statistics, Texas A&M University, College Station, TX 77845, USA.
E-mail: ma@stat.tamu.edu

 1369–7412/11/73081

This naturally leads to the model testing problem in measurement error models. For example, we might be interested in testing whether a certain covariate needs to be included in the model, or we might be interested in testing whether a certain parametric model, say a linear model, is sufficient to describe the data. Although testing problems in measurement error models are important and basic, they have surprisingly been untouched except for very special cases such as polynomial models (Cheng and Kukush, 2004; Hall and Ma, 2007a). The difficulty lies in the fact that functional measurement error models, which make no assumptions about the distribution of the mismeasured covariates, are semiparametric models, even when the relationship between the response and mismeasured covariates is completely parametric. Likelihood functions for these functional models are either not available or can only be calculated via complex deconvolution or minimum distance procedures which have very slow rates of convergence.

The intention of this paper is to fill this gap by proposing both local and omnibus goodness-of-fit test procedures. The crucial part of our work is in recognizing that tests which are similar to that of a score type can be constructed in semiparametric models, even when the score itself cannot be calculated. Based on this idea, local tests are proposed in measurement error models, which then serve as a foundation for further development in omnibus tests. The testing procedures are proposed for a wide range of functional measurement error models. As such, they are applicable in almost all measurement error models where root n estimation rates can be obtained. In addition, the testing procedures have the property that, among tests based on the same estimating equation and under the same type I error, maximum power is achieved. Finally, our testing procedure is much less computationally burdensome than a full semiparametric treatment using a Wald-type test based on estimating equations (referred to as a Wald-type test hereafter); see Tsiatis and Ma (2004) and Ma and Tsiatis (2006) for evidence of the computational complexity. Only estimation under the null model is required, which makes this an extremely favourable method computationally since alternative models are usually more complex, especially when the alternative model is non-parametric and a completely different approach is required in estimation.

We now describe the testing problem in more detail. Suppose that the main problem of interest involves a response Y and predictors $(\mathbf{X}, \mathbf{Z})$. For a local test, under the full model, the data generation procedure for Y given $(\mathbf{X}, \mathbf{Z})$ is governed by a likelihood function $p_{Y|X,Z}(Y, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma})$. Interest focuses on testing the null hypothesis $\boldsymbol{\gamma} = 0$. For an omnibus test, the null model is $p_{Y|X,Z}(Y, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta})$, and interest focuses on testing whether $\mathbf{X}$, or $\mathbf{Z}$ or both need to enter the model as an unspecified smooth function in place of the parametric form that is specified in the null model. In the measurement error context, $\mathbf{X}$ is not observable and instead we observe a surrogate for it, $\mathbf{W}$.

If we assume that we know the distribution of the measurement error for $\mathbf{W}$ given $(\mathbf{X}, \mathbf{Z})$, $p_{W|X,Z}(\mathbf{W}, \mathbf{X}, \mathbf{Z})$, then we have a semiparametric problem where the unknown infinite dimensional nuisance parameter is the distribution of $\mathbf{X}, \mathbf{Z}$: $p_{X,Z}(\mathbf{X}, \mathbf{Z})$. Tsiatis and Ma (2004) showed how to estimate $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ without having to estimate the infinite dimensional nuisance function $p_{X,Z}(\mathbf{X}, \mathbf{Z})$ directly. The method involves specifying a distribution for $\mathbf{X}$ given $\mathbf{Z}$, which will enter the estimation procedure, and leaving p_Z unspecified. The resulting estimator is efficient if the latent variable distribution $p_{X|Z}$ is correctly specified, and it is consistent even if the distribution is incorrectly specified: this is a so-called functional method that does not rely on a correct specification of the distribution of $(\mathbf{X}, \mathbf{Z})$ and is the estimation method that we use throughout this paper. The method has also been extended to the case where the likelihood or mean includes an additional unknown function $g(R)$ for an exactly observed scalar covariate R (Ma and Carroll, 2006), and where the measurement error distribution has unknown parameters or, in some cases, is completely unknown (Hall and Ma, 2007b). These models are not within the

scope of Tsiatis and Ma (2004) and, in them, the estimation of the non-parametric components $g(R)$ and the error distribution must be carried out non-parametrically.

On the basis of the Wald test idea, if $\hat{\gamma}$ is the estimate and its root n asymptotic covariance matrix is $\mathbf{V}$ with estimate $\hat{\mathbf{V}}$, then the Wald-type test statistic is to reject the hypothesis if $n\hat{\gamma}^{\mathrm{T}}\hat{\mathbf{V}}^{-1}\hat{\gamma}$ exceeds a χ^2-percentile, where the number of degrees of freedom is the dimension of γ. Thus, the Wald-type test seems to be a straightforward testing procedure that can be applied in measurement error models. Despite this, for testing a local hypothesis, there are reasons to avoid the Wald-type test, and these reasons are primarily computational. Specifically, the methodology that was described above requires solving for the roots of $p_\beta + p_\gamma$ estimating equations, where p_β and p_γ are the dimensions of $\boldsymbol{\beta}$ and γ respectively. The difficulty is with the estimating equations, because to compute them we must solve integral equations of dimension $p_\beta + p_\gamma$. In an iterative procedure to estimate $(\boldsymbol{\beta}, \gamma)$, these integral equations must be solved for each iteration. Even in the simple case that γ is scalar, the increase in dimensionality can lead to difficult issues of computational stability. The problem is not merely computational when we perform omnibus testing. In fact, the alternative model of an omnibus test is non-specific; hence it can often be expressed as containing some non-parametric components in addition to the null model. Thus the estimation under the alternative usually only has a $\log(n)$ rate of convergence and cannot be sufficiently precise to construct an effective Wald-type test.

A good example to illustrate the above points is when Y is binary, X is scalar, $\mathbf{Z}$ is a vector that includes a value of 1.0 for the intercept, logistic regression is used and we want to test whether the effect of the covariates is linear in $(X, \mathbf{Z})$. Let $H(\cdot)$ be the logistic distribution function. Thus, for example, the null model would be

$$\mathrm{pr}(Y = 1 | X, \mathbf{Z}) = H\{(X, \mathbf{Z}^{\mathrm{T}})\boldsymbol{\beta}\}, \tag{1}$$

whereas, for a local test, an alternative model allowing for quadratic departures from linearity might be

$$\mathrm{pr}(Y = 1 | X, \mathbf{Z}) = H\{(X, \mathbf{Z}^{\mathrm{T}})\boldsymbol{\beta} + X^2\gamma\}. \tag{2}$$

For an omnibus test, an alternative allowing non-linear departures of X from linearity might be

$$\mathrm{pr}(Y = 1 | X, \mathbf{Z}) = H\{(X, \mathbf{Z}^{\mathrm{T}})\boldsymbol{\beta} + \theta(X)\} \tag{3}$$

for an unspecified function $\theta(\cdot)$ that is orthogonal to X. Assuming the classical measurement error model $W = X + U$ with $U \sim N(0, \sigma_u^2)$, model (1) is easy to fit by the method of sufficient scores (Stefanski and Carroll, 1987). However, fitting model (2) in a functional manner, i.e. without correctly specifying the distribution of $(X, \mathbf{Z})$, is not trivial: it involves repeatedly solving n integral equations of dimension $1 + p_\beta$, where n is the sample size, and attempting to find the roots of equations of size $1 + p_\beta$, which is a non-trivial task especially if the measurement error is large. Estimation under model (3) is even more challenging. In fact, we are not aware of any estimation procedures for model (3) and, even if they can be constructed, the rate of convergence of $\hat{\theta}$ will not be sufficient for any Wald-type test (Fan and Truong, 1993).

Owing to these considerations, we propose here score-type tests, i.e. fitting the model only under the null hypothesis to obtain estimates of $\boldsymbol{\beta}$, and then constructing a test based on the estimating functions of Tsiatis and Ma (2004), Ma and Carroll (2006) and Hall and Ma (2007b). We do not estimate parameters or functions in the full model, only in the reduced model, and it is in this way that our approach is similar to score tests.

However, *there is a crucial difference from the ordinary score test from likelihood theory*: the structure of the general functional semiparametric measurement error model does not involve

estimating the non-parametric components directly; no profile semiparametric likelihood is available on which to base a score test. It is also more general than the test in Small and Wang (2003), page 226, proposition 6.7, in that the Godambe efficiency that was required there is not satisfied in the measurement error context. A similar idea in linear regression appeared in Sen (1982), and in the general estimating equation setting in Janicki (2009). Our main insight is to construct a score-type test, with the following properties.

(a) It is much less computationally burdensome than the semiparametric Wald-type test for measurement error problems and applies to all the problems that were described above, and more.
(b) It has the same power as the semiparametric Wald test based on the same estimating equation.
(c) When the semiparametric Wald-type test is efficient against local alternatives, so is our approach.
(d) Our methodology can be readily adapted to construct an omnibus goodness-of-fit test when the alternative is non-specific.

The paper is organized as follows. In Section 2, we describe the problem in much more generality than that given above and we make the first step by considering parametric alternatives in the case that the distribution of the latent variable $\mathbf{X}$ is unknown, whereas the measurement error distribution is known and there are otherwise no unknown functions. Section 2 serves the purpose of introducing our novel score-type test in these general semiparametric problems. This approach can be used in many other contexts besides measurement error problems. In Section 3, we show how to extend the parametric methods to the cases in which either the measurement error distribution is unknown or there is an additional unknown function. In Section 4, we take up the issue of omnibus goodness-of-fit testing. Section 5 contains empirical examples and simulations. Concluding remarks are given in Section 6. All technical details are given in Appendix A.

2. Score-type testing

2.1. General remarks and definitions

In Section 1, we described the data-generating model $p_{Y|X,Z}(Y|\mathbf{X},\mathbf{Z})$, the measurement error model $p_{W|X,Z}(\mathbf{W}|\mathbf{X},\mathbf{Z})$ and the latent variable model $p_{X|Z}(\mathbf{X}|\mathbf{Z})$ for the error prone covariate. We assume that $\mathbf{W}$ is a surrogate for $\mathbf{X}$, i.e. the distribution of Y is independent of $\mathbf{W}$ given $(\mathbf{X},\mathbf{Z})$. We shall assume in this section that the measurement error model is known, but we shall show how to avoid this assumption in Section 3.

We write the data generation process as

$$p_{Y|X,Z}(Y|\mathbf{X},\mathbf{Z}) = p_{Y|X,Z}\{Y, f(\mathbf{X},\mathbf{Z},\boldsymbol{\beta}) + h(\mathbf{X},\mathbf{Z},\boldsymbol{\gamma})\}, \tag{4}$$

where $f(\cdot,\boldsymbol{\beta})$ is known up to the parameter $\boldsymbol{\beta}$, and $h(\cdot,\boldsymbol{\gamma})$ is a discrepancy from the simpler model $f(\cdot,\boldsymbol{\beta})$ with the property that, under the null hypothesis that $\boldsymbol{\gamma}=0$, $h(\mathbf{X},\mathbf{Z},\boldsymbol{\gamma}) = h(\mathbf{X},\mathbf{Z},0) \equiv 0$. We write the discrepancy in the additive form so it is easier to link with model (3) later on. Non-additive discrepancies can be handled similarly. When it is of interest to test H_0 against a specific parametric alternative, h will have a known form with either known or unknown parameter $\boldsymbol{\gamma}$. When the test is non-parametric, we wish to detect virtually any deviation from the parametric null hypothesis. In this case h is allowed to be an arbitrary linear combination of known, orthogonal basis functions. In both cases, to ensure identifiability, we assume that the space of all $f+h$ is strictly larger than that of all f.

2.2. Parametric tests: level

As described above, we wish to test the hypothesis that $\gamma = 0$. In the measurement error problem, and many other contexts, estimating equations exist (see Tsiatis and Ma (2004)) for (β, γ) and can be written as

$$0 = \sum_{i=1}^{n} \phi_\beta(W_i, Z_i, Y_i, \beta, \gamma), \tag{5}$$

$$0 = \sum_{i=1}^{n} \psi_\gamma(W_i, Z_i, Y_i, \beta, \gamma), \tag{6}$$

where ϕ_β and ψ_γ have the same dimensions as β and γ respectively. We have used different symbols $\phi(\cdot)$ and $\psi(\cdot)$, because these estimating functions are *not* derivatives of some version of a profile likelihood, since no profile likelihood exists in our semiparametric framework.

Since $h(\cdot)$ vanishes at $\gamma = 0$, the estimating equation under the null model is simply

$$\sum_{i=1}^{n} \phi_\beta(W_i, Z_i, Y_i, \beta, 0) = 0,$$

and we call its root $\hat{\beta}$. The existence of $\hat{\beta}$ is a result of the identifiability of the model under the null hypothesis, which we assume. For more general discussion on the existence of the roots of estimating equations, see Heyde (1997). Then, by analogy with the score test, we propose to base our test on the estimated 'score'

$$\hat{U} = n^{-1/2} \sum_{i=1}^{n} \psi_\gamma(W_i, Z_i, Y_i; \hat{\beta}, 0).$$

The analysis of $\hat{U}$ is straightforward. Make the following definitions, with all expectations done under the null hypothesis:

$$\begin{aligned}
A_1 &= E\{\partial \phi_\beta(W, Z, Y, \beta, 0)/\partial \beta^{\mathrm{T}}\};\\
A_2 &= E\{\partial \psi_\gamma(W, Z, Y, \beta, 0)/\partial \beta^{\mathrm{T}}\};\\
A_3 &= E\{\partial \phi_\beta(W, Z, Y, \beta, \gamma)/\partial \gamma^{\mathrm{T}}\};\\
A_4 &= E\{\partial \psi_\gamma(W, Z, Y, \beta, \gamma)/\partial \gamma^{\mathrm{T}}\};\\
B_{11} &= E\{\phi_\beta(W, Z, Y, \beta, 0)\, \phi_\beta(W, Z, Y, \beta, 0)^{\mathrm{T}}\};\\
B_{22} &= \mathrm{cov}\{\psi_\gamma(W, Z, Y, \beta_0, 0)\};\\
B_{12} &= \mathrm{cov}\{\phi_\beta(W, Z, Y, \beta_0, 0), \psi_\gamma^{\mathrm{T}}(W, Z, Y, \beta_0, 0)\};\\
V_\beta &= A_1^{-1} B_{11} (A_1^{-1})^{\mathrm{T}};\\
\Sigma_0 &= \mathrm{cov}\{\psi_\gamma(\cdot, \beta, 0) - A_2 A_1^{-1}\, \phi_\beta(\cdot, \beta, 0)\}.
\end{aligned} \tag{7}$$

All these quantities can be estimated by replacing expectations and covariance matrices by their sample versions. We shall denote the resulting estimate of Σ_0 by $\hat{\Sigma}_0$.

The test statistic with nominal level α that we propose is to reject the hypothesis if $T = \hat{U}^{\mathrm{T}} \hat{\Sigma}_0^{-1} \hat{U}$ exceeds the $(1-\alpha)$-quantile of the χ^2-distribution with p_γ degrees of freedom. Of course, T does not involve estimating γ. That this test has asymptotic level α follows from standard Taylor series calculations, yielding the following result.

Theorem 1. Under the null hypothesis, $n^{1/2}(\hat{\beta} - \beta) \to N(0, +\mathbf{V}_\beta)$ and $\hat{U} \to N(0, \Sigma_0)$. Hence $T = \hat{U}^{\mathrm{T}} \hat{\Sigma}_0^{-1} \hat{U}$ is asymptotically χ^2 with p_γ degrees of freedom.

2.3. Parametric tests: power and the Wald-type test

We now study the power of our score-type test against root n local alternatives and compare its local power with that of the Wald-type test. Define

$$\mathcal{A}=\begin{pmatrix} A_1 & A_3 \\ A_2 & A_4 \end{pmatrix}, \qquad \mathcal{B}=\begin{pmatrix} B_{11} & B_{12} \\ B_{12}^{\mathrm{T}} & B_{22} \end{pmatrix}, \qquad \mathcal{V}=\begin{pmatrix} V_{11} & V_{12} \\ V_{12}^{\mathrm{T}} & V_{22} \end{pmatrix},$$

where $\mathcal{V}=\mathcal{A}^{-1}\mathcal{B}(\mathcal{A}^{-1})^{\mathrm{T}}$. If we were actually to solve both equation (5) and equation (6), then the resulting estimates $(\hat{\boldsymbol{\beta}}_{\mathrm{compr}}, \hat{\boldsymbol{\gamma}}_{\mathrm{compr}})$ satisfy $n^{1/2}\{(\hat{\boldsymbol{\beta}}_{\mathrm{compr}}-\boldsymbol{\beta})^{\mathrm{T}}, (\hat{\boldsymbol{\gamma}}_{\mathrm{compr}}-\boldsymbol{\gamma})^{\mathrm{T}}\}^{\mathrm{T}} \to N(0, \mathcal{V})$. Here and after, the subscript 'compr' (short for 'comprehensive') indicates a result that is obtained from solving the full system of estimating equations. The Wald-type test then rejects the null hypothesis if $n\hat{\boldsymbol{\gamma}}_{\mathrm{compr}}^{\mathrm{T}}\hat{V}_{22}^{-1}\hat{\boldsymbol{\gamma}}_{\mathrm{compr}}$ exceeds the $(1-\alpha)$-quantile of the χ^2-distribution with p_γ degrees of freedom, where $\hat{V}_{22}$ is a consistent estimate of V_{22}.

Suppose that the local alternative is of the form $\boldsymbol{\gamma}_n=\mathbf{c}n^{-1/2}$. In Appendix A, we sketch the following general result.

Theorem 2. Under the alternative hypothesis that $\boldsymbol{\gamma}_n=\mathbf{c}n^{-1/2}$, the test statistic T is asymptotically non-central χ^2 with p_γ degrees of freedom and non-centrality parameter

$$\tfrac{1}{2}\mathbf{c}^{\mathrm{T}}(A_4-A_2A_1^{-1}A_3)^{\mathrm{T}}\Sigma_0^{-1}(A_4-A_2A_1^{-1}A_3)\mathbf{c}=\tfrac{1}{2}\mathbf{c}^{\mathrm{T}}V_{22}^{-1}\mathbf{c}.$$

In addition, the test which rejects when T exceeds $\chi^2_{p_\beta,\alpha}$ has the same local power as the computationally more intensive Wald-type test.

Theorem 2 shows that the Wald-type and score-type tests based on the same estimating equation have Pitman relative efficiency 1. Thus, when either is fully efficient, so is the other. We know that, when the estimates $(\hat{\boldsymbol{\beta}}_{\mathrm{compr}}, \hat{\boldsymbol{\gamma}}_{\mathrm{compr}})$ that are obtained by solving both equation (5) and equation (6) are efficient (have the smallest estimation variance), the corresponding Wald-type test is efficient; therefore our score-type test would also have efficient local power in this case.

2.4. Semiparametric setting and measurement error

Although the equivalence between the score and Wald tests has been established in parametric models, in the measurement error context there is no profile likelihood and hence there is no such general result. Instead, as described in Section 1, Tsiatis and Ma (2004) provided estimating functions $\phi_\beta(\mathbf{W}, \mathbf{Z}, Y, \boldsymbol{\beta}, \boldsymbol{\gamma})$ and $\psi_\gamma(\mathbf{W}, \mathbf{Z}, Y, \boldsymbol{\beta}, \boldsymbol{\gamma})$ that have mean 0 when evaluated at the true parameters.

As a consequence of theorem 2, we have the following result.

Theorem 3. Consider a semiparametric model $f(\mathcal{Y}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \nu)$, where $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ are independent observations, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are finite dimensional parameters and ν represents the infinite dimensional nuisance parameters. Suppose that there exist equations (5) and (6) that define root n consistent and asymptotically normally distributed estimators of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$. Then our score-type test is asymptotically equivalent to the Wald-type test against alternatives of the form $\boldsymbol{\gamma}=\mathbf{c}n^{-1/2}$, in that the asymptotic distributions of the two test statistics are identical.

In theorem 3, although ν appears in the model, it does not appear in the estimating equations (5) and (6) on which the tests are based. Such estimating equations do not necessarily exist for an arbitrary semiparametric model. For the measurement error models that we are considering, these estimating equations do exist.

Estimating equations such as equations (5) and (6) that do not involve a direct estimate of the infinite dimensional nuisance parameter ν also exist in other problems, such as the restricted

moment models (Tsiatis (2006), chapter 4), skew elliptical distribution models (Ma *et al.*, 2005) and generalized linear latent variable models (Ma and Genton, 2010). For these problems theorem 3 applies.

3. Extensions

3.1. Overview

In Section 2, the non-parametric nuisance parameter ν, which is the distribution of the unobservable X in measurement error models, is bypassed in the estimation procedure owing to the structure of these models and is not estimated directly. However, there are important extensions of measurement error models, where additional infinite dimensional nuisance parameters are included in the model and in the estimation procedure. In this section, we describe two such extensions:

(a) to cases in which the measurement error distribution is estimated and
(b) to cases where the underlying regression model has a non-parametric component.

Because the measurement error distribution in extension (a) or the unknown function in extension (b) needs to be estimated non-parametrically to estimate the main parameter, the testing procedure in Section 2 cannot be directly applied without adaptation. As preparation for extension (a), we start with a simpler parametric version.

3.2. Estimated measurement error distributions

3.2.1. Parametric measurement error distributions

In some cases, we may be willing to assume that the measurement error distribution is known up to a finite dimensional parameter $\boldsymbol{\eta}$ and is written as $p_{W|X,Z}(\mathbf{W}|\mathbf{X}, \mathbf{Z}, \boldsymbol{\eta})$. In such cases, the unknown parameter $\boldsymbol{\eta}$ is typically estimated either from another experiment, or from units in the study having replicated $\mathbf{W}$-values. In either case, methods that were discussed in Carroll *et al.* (2006) show that the net result is to obtain an estimating equation for estimating $\boldsymbol{\eta}$, which is called $\phi_\eta(\mathbf{W}, \mathbf{Z}, \boldsymbol{\eta})$, and consistent estimates are formed by solving

$$0 = \sum_{i=1}^{n} \phi_\eta(\mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\eta}). \tag{8}$$

Testing whether $\gamma = 0$ is done by combining the estimating equations (5) and (8). Thus, if $\mathcal{B} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\eta}^{\mathrm{T}})^{\mathrm{T}}$, we replace $\phi_\beta(\cdot)$ in equation (5) by

$$\phi_{\mathcal{B}}(\mathbf{W}, \mathbf{Z}, \mathbf{Y}, \mathcal{B}, \gamma, \boldsymbol{\eta}) = \{\phi_\beta^{\mathrm{T}}(\mathbf{W}, \mathbf{Z}, \mathbf{Y}, \boldsymbol{\beta}, \gamma), \phi_\eta^{\mathrm{T}}(\mathbf{W}, \mathbf{Z}, \boldsymbol{\eta})\}^{\mathrm{T}},$$

and then all the results of Section 2 go through with this simple change in notation.

3.2.2. Non-parametric measurement error distributions

A second extension is to cases where the measurement error distribution is completely unknown. Here, for simplicity, we assume that the covariate subject to error is univariate, and we discuss the case in which multiple measurements W_{ij}, $j = 1, \ldots, m_i$, are available. It is assumed that, possibly after transformation, $W_{ij} = X_i + U_{ij}$; hence $p_{W|X,\mathbf{Z}}(W|X, \mathbf{Z}) = p_U(W - X)$. Here, $U_{i1}, \ldots, U_{im_i}$ are independent of $(X_i, \mathbf{Z}_i, Y_i)$ and have an unknown, symmetric probability density function $p_U(\cdot)$.

Hall and Ma (2007b) proposed a non-parametric estimate of the unknown measurement error distribution $p_U(\cdot)$; call it $\hat{p}_U(\cdot)$. They then proposed to modify equations (5) and (6) to

$$0 = \sum_{i=1}^{n} \phi_\beta(W_i, \mathbf{Z}_i, Y_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \hat{p}_U); \tag{9}$$

$$0 = \sum_{i=1}^{n} \psi_\gamma(W_i, \mathbf{Z}_i, Y_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \hat{p}_U). \tag{10}$$

The key to the analysis of testing when using equations such as equations (9) and (10) is to show that they are asymptotically equivalent to equations that have the same form as equations (5) and (6), but for which estimating the error distribution $p_U(\cdot)$ has no asymptotic effect. Let 'ede' stand for 'error distribution estimated'. Then one needs to construct estimating equations $\phi_{\beta,\mathrm{ede}}(\cdot)$ and $\phi_{\gamma,\mathrm{ede}}(\cdot)$ such that

$$o_p(1) = n^{-1/2} \sum_{i=1}^{n} \{\phi_\beta(W_i, \mathbf{Z}_i, Y_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \hat{p}_U) - \phi_{\beta,\mathrm{ede}}(W_i, \mathbf{Z}_i, Y_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \hat{p}_U)\},$$

$$o_p(1) = n^{-1/2} \sum_{i=1}^{n} \{\psi_\gamma(W_i, \mathbf{Z}_i, Y_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \hat{p}_U) - \psi_{\gamma,\mathrm{ede}}(W_i, \mathbf{Z}_i, Y_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \hat{p}_U)\},$$

$$o_p(1) = n^{-1/2} \sum_{i=1}^{n} \{\phi_{\beta,\mathrm{ede}}(W_i, \mathbf{Z}_i, Y_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \hat{p}_U) - \phi_{\beta,\mathrm{ede}}(W_i, \mathbf{Z}_i, Y_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, p_U)\},$$

$$o_p(1) = n^{-1/2} \sum_{i=1}^{n} \{\psi_{\gamma,\mathrm{ede}}(W_i, \mathbf{Z}_i, Y_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \hat{p}_U) - \psi_{\gamma,\mathrm{ede}}(W_i, \mathbf{Z}_i, Y_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, p_U)\}$$

hold uniformly in a neighbourhood of the true parameter value $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$. The first two of these equations mean that $\phi_{\beta,\mathrm{ede}}$ and $\psi_{\gamma,\mathrm{ede}}$ are asymptotically equivalent versions of ϕ_β and ψ_γ to use in equations (9) and (10), whereas the latter two establish that these asymptotically equivalent versions are not affected by estimating the error distribution.

Hall and Ma (2007b) derived the computable and asymptotically equivalent estimating equations $\phi_{\beta,\mathrm{ede}}(\cdot)$ and $\psi_{\gamma,\mathrm{ede}}(\cdot)$, as $\phi_\beta(\cdot) - S_\beta(\cdot)$ and $\psi_\gamma(\cdot) - S_\gamma(\cdot)$ respectively, where $(S_\beta(\cdot)^{\mathrm{T}}, S_\gamma(\cdot)^{\mathrm{T}})^{\mathrm{T}}$ is given as S_2^* in their appendix A.3. Specifically, they have the form

$$\begin{aligned}(S_\beta^{\mathrm{T}}, S_\gamma^{\mathrm{T}})^{\mathrm{T}} = {} & E^*(af/p_U|W, Y, \mathbf{Z}) - \{E^*(a|W, Y, \mathbf{Z}) + b_1\} E^*(f/p_U|W, Y, \mathbf{Z}) \\ & - E^*(fS_\beta^{F*}/p_U|W, Y, \mathbf{Z}) + S_\beta^* E^*(f/p_U|W, Y, \mathbf{Z}) + b_1 + b_2,\end{aligned}$$

where an asterisk stands for quantities or operations that are calculated under a specified distributional model of X given $\mathbf{Z}$, S_β^{F*} and S_β^* are the scores of $\boldsymbol{\beta}$ in the $(X, Y, \mathbf{Z})$ and $(W, Y, \mathbf{Z})$ spaces respectively, p_U is the probability density function of U,

$$f = h^{-1} K\{(V-u)/h\} - p_U(u) - \tfrac{1}{2} h^2 p_U'' \int t^2 K(t)\, \mathrm{d}t,$$

$b_1 = E^*(b_0|W, Y, \mathbf{Z})$, $b_2 = E^*(b_0 \tilde{r}/p_U|W, Y, \mathbf{Z})$, $b_0 = \hat{a} - a$ and $\tilde{r}(\cdot) = n^{-1} \sum_{i=1}^{n} f(V_i, \cdot, h)$. The function a is the solution of

$$E[E^*\{a(X, \mathbf{Z})|W, Y, \mathbf{Z}\}|X, \mathbf{Z}] = E\{S_\beta^*(W, Y, \mathbf{Z})|X, \mathbf{Z}\},$$

where the calculation is performed by using p_U, and $\hat{a}$ solves the same equation with p_U replaced by $(nh)^{-1} \sum_{i=1}^{n} K\{(V_i - u)/h\}$. When $m_i = 2$, $V_i = (W_{i1} - W_{i2})/2$, whereas for $m_i > 2$, definitions of V_i may be found in Hall and Ma (2007b).

Now suppose that

(a) $\hat{U} = n^{-1/2} \sum_{i=1}^{n} \psi_\gamma(W_i, \mathbf{Z}_i, Y_i, \hat{\boldsymbol{\beta}}, 0, \hat{p}_U)$,
(b) the quantities that were defined in Section 2 use $\phi_{\beta,\mathrm{ede}}(\cdot, p_U)$ and $\phi_{\gamma,\mathrm{ede}}(\cdot, p_U)$ in place of ϕ_β and ϕ_γ respectively and

(c) estimates of $\phi_{\beta,\mathrm{ede}}(\cdot, p_U)$ and $\phi_{\gamma,\mathrm{ede}}(\cdot, p_U)$ are defined by using $\phi_{\beta,\mathrm{ede}}(\cdot, \hat{p}_U)$ and $\phi_{\gamma,\mathrm{ede}}(\cdot, \hat{p}_U)$.

It then follows that results analogous to all those in Section 2 hold; for example $\hat{U}^{\mathrm{T}}\hat{\Sigma}_0^{-1}\hat{U}$ is non-central χ^2 and the Wald-type and our score-type tests are asymptotically equivalent.

3.3. A non-parametric component in the regression model

Similar derivations can be carried out for the models containing an unspecified function of a scalar covariate R, say $g(R)$, in the regression model itself. For any given $(\boldsymbol{\beta}, \boldsymbol{\gamma})$, let $\hat{g}(R, \boldsymbol{\beta}, \boldsymbol{\gamma})$ be an estimate of $g(R)$. In this case, the estimating equations (5)–(6) are modified to

$$0 = n^{-1/2} \sum_{i=1}^{n} \phi_\beta\{W_i, Z_i, Y_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \hat{g}(R, \boldsymbol{\beta}, \boldsymbol{\gamma})\}, \tag{11}$$

$$0 = n^{-1/2} \sum_{i=1}^{n} \psi_\gamma\{W_i, Z_i, Y_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \hat{g}(R, \boldsymbol{\beta}, \boldsymbol{\gamma})\}. \tag{12}$$

With 'ufe' meaning 'unknown function estimated', the key again is to show that, for suitable estimators of $\hat{g}(R, \boldsymbol{\beta}, \boldsymbol{\gamma})$, there are computable, asymptotically equivalent versions of equations (11) and (12), say $\phi_{\beta,\mathrm{ufe}}\{\cdot, g(R)\}$ and $\psi_{\gamma,\mathrm{ufe}}\{\cdot, g(R)\}$, whose asymptotic distributions are unaffected by estimating $g(R)$.

Ma and Carroll (2006) derived these asymptotically equivalent versions when $\hat{g}$ is a local constant estimator, the simplest of the local polynomial estimators. The following results assume that this local constant estimator is used. Denote $\mathcal{B} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$ and $\mathcal{L}_{\mathcal{B}} = (\phi_\beta^{\mathrm{T}}, \psi_\gamma^{\mathrm{T}})^{\mathrm{T}}$. Replacing the function $g(R)$ by an unknown constant a, we would be able to obtain a set of estimating equations for $(\mathcal{B}, a)$. Let $\Sigma_{i=1}^n \Psi_g$ be the component of the estimating equation corresponding to a. Hall and Ma (2007b) showed that

$$\begin{pmatrix} \phi_{\beta,\mathrm{ufe}} \\ \psi_{\gamma,\mathrm{ufe}} \end{pmatrix} = \mathcal{L}_{\mathcal{B}}(\mathcal{Y}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, g_0) - \Psi_g(\mathcal{Y}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, g_0)\, \mathcal{U}(R_i),$$

where $\mathcal{U}(R) = E\{\mathcal{L}_{\mathcal{B}g}(\cdot)|R\}/\Omega(R)$, $g_{\mathcal{B}}(R) = -E\{\Psi_{g\mathcal{B}}(\cdot)|R\}/\Omega(R)$, $\mathcal{L}_{\mathcal{B}\mathcal{B}}$ is the partial derivative of $\mathcal{L}_{\mathcal{B}}$ with respect to $\mathcal{B}$, $\mathcal{L}_{\mathcal{B}g}$ is the partial derivative of $\mathcal{L}_{\mathcal{B}}$ with respect to g, Ψ_{gg} is the partial derivative of Ψ_g with respect to g, $\Psi_{g\mathcal{B}}$ is the partial derivative of Ψ_g with respect to $\mathcal{B}$ and the argument '$(\cdot)$' here stands for $\{\mathcal{Y}, \mathcal{B}_0, g_0(R)\}$, and $\Omega(R) = E\{\Psi_{gg}(\cdot)|R\}$.

Now suppose that

(a) $\hat{U} = n^{-1/2} \Sigma_{i=1}^n \psi_\gamma\{W_i, Z_i, Y_i, \hat{\boldsymbol{\beta}}, 0, \hat{g}(R_i, \hat{\boldsymbol{\beta}}, 0)\}$,
(b) the quantities that were defined in Section 2 use $\phi_{\beta,\mathrm{ufe}}\{\cdot, g(R_i)\}$ and $\phi_{\gamma,\mathrm{ufe}}\{\cdot, g(R_i)\}$ in place of ϕ_β and ϕ_γ respectively and
(c) estimates of $\phi_{\beta,\mathrm{ufe}}\{\cdot, g(R_i)\}$ and $\phi_{\gamma,\mathrm{ufe}}\{\cdot, g(R_i)\}$ are defined by using $\phi_{\beta,\mathrm{ufe}}\{\cdot, \hat{g}(R_i, \hat{\boldsymbol{\beta}}, 0)\}$ and $\phi_{\gamma,\mathrm{ufe}}\{\cdot, \hat{g}(R_i, \hat{\boldsymbol{\beta}}, 0)\}$.

It then follows that results analogous to all those in Section 2 hold; for example $\hat{U}^{\mathrm{T}}\hat{\Sigma}_0^{-1}\hat{U}$ is non-central χ^2 and the Wald-type and our score-type tests are asymptotically equivalent.

To handle the situation where both the measurement error parameters need to be estimated and an unspecified function $g(R)$ is included in the regression model, we need to combine the results in Sections 3.2 and 3.3. A fully detailed technical argument showing how to do this is not pursued here.

4. Omnibus goodness-of-fit testing

The omnibus goodness-of-fit test that we construct is based on the idea that a smooth function can be approximated arbitrarily well by a linear combination of sufficiently many basis functions. For any fixed system of basis functions, a departure from the null model in the direction of a given basis function is a problem of local testing, and the procedure that was developed in Sections 2–3 can be applied. However, because departures along different directions need to be considered, to avoid multiple testing, these tests need to be intelligently combined. In the context of regression without measurement error, Hart (2009) proposed a goodness-of-fit test that is a hybrid of Bayesian and frequentist ideas. Here, we use an analogous statistic and adapt it to the general measurement error framework.

Consider a set of basis functions $h_1(X, Z), h_2(X, Z), \ldots, h_m(X, Z)$, which are arranged from lowest to highest frequency. We may construct m different tests of the null hypothesis

$$H_0 : p_{Y|X,Z}(Y|X, Z) = p_{Y|X,Z}\{Y, f(X, Z, \beta)\}.$$

For $j = 1, \ldots, m$, the jth alternative is $H_{j1}: \gamma_j \neq 0$, where γ_j is such that

$$p_{Y|X,Z} = p_{Y|X,Z}\{Y, f(X, Z, \beta) + \gamma_j h_j(X, Z)\}.$$

Now, let $\hat{T}_j^2 = \hat{u}_j^2/\hat{\sigma}_j^2$ denote the test statistic that was defined in Section 2 for testing H_0 against $H_{j1}: \gamma_j \neq 0$. Then the omnibus test statistic that we propose is $\hat{\mathcal{T}}_m = \Sigma_{j=1}^m j^{-2} \exp(\hat{T}_j^2/2)$.

To motivate $\hat{\mathcal{T}}_m$, we consider the canonical regression model

$$Y_j = r(x_j) + \varepsilon_j, \qquad j = 1, \ldots, n, \tag{13}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent $N(0, \sigma^2)$ random variables, and no measurement error is involved. Suppose that we wish to test the null hypothesis that r is a linear combination of finitely many known functions. A novel approach to doing so is to compute a posterior probability $\mathcal{P}_0$ of the null hypothesis, and then to use $\mathcal{P}_0$ in a *frequentist fashion*. The null hypothesis is rejected at level of significance α if the observed value of $\mathcal{P}_0$ is smaller than the αth percentile of the null sampling distribution of $\mathcal{P}_0$.

In the canonical regression model (13), Hart (2009) showed that tests based on $\mathcal{P}_0$ and $\hat{\mathcal{T}}_n$ are related as follows. Let $\tilde{\mathcal{P}}_0$ be a version of $\mathcal{P}_0$ that is based on a non-informative prior for model parameters and a weakly informative prior for the number of orthogonal functions that are needed to represent the true r. Then the statistic $\hat{\mathcal{T}}_n$ is a monotone transformation of a Laplace approximation to $\tilde{\mathcal{P}}_0$. In the regression without measurement error context, the extensive simulations of Hart (2009), section 7, demonstrated that $\hat{\mathcal{T}}_n$ had statistically significantly larger power than that of either a data-driven Neyman smooth test or a regression analogue of the Cramér–von Mises statistic in a majority of the many cases that were considered. Furthermore, the ratio of competitor power to power of $\hat{\mathcal{T}}_n$ was never larger than 1.61, whereas in several cases the power of $\hat{\mathcal{T}}_n$ was at least five times that of both competitors.

In the measurement error context, the statistic $\hat{\mathcal{T}}_m$ is omnibus in the pure sense of the term if m tends to ∞ with n. In that case the corresponding test will be consistent against virtually any alternative. However, to avoid making our paper overly technical, we avoid the issue of determining an appropriate rate at which m should tend to ∞. Practically speaking, this is relatively unimportant since the class of alternatives against which $\hat{\mathcal{T}}_m$ is consistent is much larger than the space of functions that is spanned by $h_1, \ldots, h_m$. In fact, as long as θ is such that γ_j is non-zero for at least one of the m bases, then the falsity of H_0 can be detected. The reader may better appreciate the last couple of remarks by considering a familiar problem in simple regression. Suppose that we test for no effect of the predictor x by using an F-test based on a

straight line alternative. This test is consistent against every polynomial alternative except those for which the best straight line approximation has slope 0. The space that is spanned by the alternative in this case is all straight lines, but obviously the 'consistency class' is enormously larger than all straight lines. In the same way the consistency class of $\hat{T}_m$ is much larger than the alternatives that are spanned by $h_1, \ldots, h_m$. In our experience, fixing m at, say, 10 suffices for most cases that are encountered in practice. Nonetheless, since we do not study the asymptotic distribution of $\hat{T}_m$ as m increases with n, it is perhaps more precise to call our fixed m test 'well justified with a known asymptotic distribution' rather than 'omnibus'.

A perhaps more important issue is the ordering of the basis functions. Since we order $h_1, \ldots, h_m$ by their complexity, and $\hat{T}_m$ weights the contribution of h_j by j^{-2}, the test will have better power against low than against high frequency alternatives. Ordering the basis functions as we have is not arbitrary, since low frequency functions are clearly more prevalent in practice than are high frequency functions. However, if prior knowledge about the type of alternative is available, then a reordering of the basis functions may lead to a more powerful test. For a more complete discussion of the ordering issue, the reader is referred to section 7.8 of Hart (1997).

The calculation of $\hat{T}_m$ is straightforward. Below we shall describe its asymptotic distribution under the null hypothesis as $n^{1/2}(\hat{T}_1, \ldots, \hat{T}_m)^{\mathrm{T}} \to N(0, \Sigma)$, and we shall calculate an estimate $\hat{\Sigma}$. To approximate the p-value of the test, we use the following algorithm.

(a) For some large B, generate independent vectors $(T_{11}, \ldots, T_{m1}), \ldots, (T_{1B}, \ldots, T_{mB})$ from the m-variate $N(0, \hat{\Sigma})$ distribution, and define

$$\mathcal{T}_b = \sum_{j=1}^{m} j^{-2} \exp(T_{jb}^2/2)$$

for $b = 1, \ldots, B$.

(b) The p-value is then approximated by $B^{-1} \sum_{b=1}^{B} I(\mathcal{T}_b > \hat{T}_m)$.

We now show that $n^{1/2}(\hat{T}_1, \ldots, \hat{T}_m)^{\mathrm{T}} \to N(0, \Sigma)$, and we give the form of Σ. This means that we must compute the joint limit distribution of the terms $\hat{u}_j/\hat{\sigma}_j$. Marginally, $\hat{u}_j/\hat{\sigma}_j \to N(0,1)$ in distribution. For any constants $c_1, \ldots, c_m$, we have

$$\mathcal{S}_n = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{c_j}{\sigma_j} \{\psi_{\gamma_j}(W_i, Z_i, Y_i, ; \boldsymbol{\beta}_0, 0) - A_{j2} A_1^{-1} \phi_\beta(W_i, Z_i, Y_i, ; \boldsymbol{\beta}_0, 0)\} + o_p(1).$$

Here A_{j2} is defined in the same manner as A_2 in Section 2 but with the function $h_j(\cdot)$. It follows from the central limit theorem that $\mathcal{S}_n$ converges to a normal distribution. Hence $\hat{u}_1/\hat{\sigma}_1, \ldots, \hat{u}_m/\hat{\sigma}_m$ are asymptotically jointly normal. The asymptotic covariance between $\hat{u}_j/\hat{\sigma}_j$ and $\hat{u}_k/\hat{\sigma}_k$ is obtained as follows:

$$\begin{aligned}
E\left(\frac{\hat{u}_j}{\hat{\sigma}_j}\frac{\hat{u}_k}{\hat{\sigma}_k}\right) &= \frac{1}{\sigma_j \sigma_k} n^{-1} E\left[\sum_{i=1}^{n} \{\psi_{\gamma_j}(W_i, Z_i, Y_i, ; \boldsymbol{\beta}_0, 0) - A_{j2} A_1^{-1} \phi_\beta(W_i, Z_i, Y_i, ; \boldsymbol{\beta}_0, 0)\}\right. \\
&\qquad \left. \times \sum_{i=1}^{n} \{\psi_{\gamma_k}(W_i, Z_i, Y_i, ; \boldsymbol{\beta}_0, 0) - A_{k2} A_1^{-1} \phi_\beta(W_i, Z_i, Y_i, ; \boldsymbol{\beta}_0, 0)\}\right] + o(1) \\
&= \frac{1}{\sigma_j \sigma_k} E\{\psi_{\gamma_j}(\cdot; \boldsymbol{\beta}_0, 0) - A_{j2} A_1^{-1} \phi_\beta(\cdot; \boldsymbol{\beta}_0, 0)\}\{\psi_{\gamma_k}(\cdot; \boldsymbol{\beta}_0, 0) - A_{k2} A_1^{-1} \phi_\beta(\cdot; \boldsymbol{\beta}_0, 0)\} + o(1) \\
&= \frac{1}{\sigma_j \sigma_k} [A_{j2} A_1^{-1} B_{11} A_1^{-1\mathrm{T}} A_{k2}^{\mathrm{T}} - A_{j2} A_1^{-1} B_{k12} - A_{k2} A_1^{-1} B_{j12} + E\{\psi_{\gamma_j}(\cdot; \boldsymbol{\beta}_0, 0)\, \psi_{\gamma_k}(\cdot; \boldsymbol{\beta}_0, 0)\}] \\
&\quad + o(1).
\end{aligned}$$

The power of the omnibus test against the local alternative $\gamma_j = c_j n^{-1/2}$, $1 \leqslant j \leqslant m$, can be

obtained from the fact that $(\hat{u}_1/\hat{\sigma}_1, \ldots, \hat{u}_m/\hat{\sigma}_m)^{\mathrm{T}}$ has an asymptotically multivariate normal distribution, $N(\mu, \Sigma)$, where $\mu_j = -(A_{j4} - A_{j2}A_1^{-1}A_{j3})c_j$, $j = 1, \ldots, m$. This result implicitly determines the asymptotic power of our omnibus test against root n local alternatives. In practice, this power can be approximated by first generating numerous samples from the $N(\mu, \Sigma)$ distribution, and then proceeding in an obvious way. Specifically,

(a) For some large B, generate independent vectors $(T_{11}, \ldots, T_{m1}), \ldots, (T_{1B}, \ldots, T_{mB})$ from the m-variate $N(\mu, \hat{\Sigma})$ distribution, and define

$$\mathcal{T}_b = \sum_{j=1}^{m} j^{-2} \exp(T_{jb}^2/2)$$

for $b = 1, \ldots, B$.

(b) The power is then approximated by $B^{-1} \sum_{b=1}^{B} I(\mathcal{T}_b > c)$, where c is the estimated critical value.

5. Simulations and an empirical example

5.1. General framework

To investigate the finite sample performance of the testing procedures, we conducted two simulation studies, the first of which concerns local testing and the second our omnibus test. The null model in both cases is the logistic regression model $\mathrm{pr}(Y = 1|X) = H(\beta_0 + \beta_1 X)$, where $H(\cdot)$ is the logistic distribution function. The covariate X follows a normal distribution with mean -0.5 and variance 1, and the measurement error is additive, i.e. $W = X + U$, where U follows a normal distribution with mean 0 and standard deviation 0.4. The true values of β_0 and β_1 are both 1.0. The alternative in both studies is the quadratic regression model $\mathrm{pr}(Y = 1|X) = H(\beta_0 + \beta_1 X + \beta_2 X^2)$. Although the specific null and alternative models seem very simple, the estimating equations still do not have an explicit form. Procedures described between expressions (8) and (9) in Tsiatis and Ma (2004) need to be followed to obtain equations (5) and (6).

We use the locally efficient score method to perform estimation under hypothesis H_0 and to form ψ_γ, where we used both the correct normal model and an incorrect uniform $[-8.5, -1.5]$ model for the distribution p_X of the unobservable covariate X. In both tests, we based our decision on the asymptotic distribution of the corresponding statistic under the null hypothesis. We considered sample sizes ranging from 100 to 1000 in steps of 100. Each simulation result is based on 1000 experiments.

5.2. Simulations for local test

Results on the level of the local test are presented in Table 1. At small sample sizes the test is slightly conservative but essentially has the correct size for $n \geqslant 200$. The level consistency of the tests does not depend on the correctness of the model that we propose for p_X, which is a direct consequence of the consistency of the estimating equations.

To study power, we generated data from the alternative quadratic regression model, taking $\beta_2 = 10/n^{1/2}$ for each sample size n. The empirical powers of the test at nominal levels 0.01, 0.05 and 0.10 are presented in Table 2. Overall, the test is quite powerful for the alternative that was considered. One interesting observation is that we had expected the power to be better when the correct model for p_X is used in constructing the test, but our results show that the empirical power essentially did not depend on which p_X-model was assumed. Note that, when we correctly model p_X as normal, the test that is used is optimal in its class. Our simulation

Table 1. Level of the local test†

	Results for the following values of n:									
	100	*200*	*300*	*400*	*500*	*600*	*700*	*800*	*900*	*1000*
Normal p_X^*										
$\alpha = 0.01$	0.002	0.005	0.006	0.011	0.004	0.007	0.004	0.010	0.005	0.008
Confidence interval	0.006	0.009	0.010	0.013	0.008	0.010	0.008	0.012	0.009	0.011
$\alpha = 0.05$	0.023	0.039	0.039	0.044	0.043	0.043	0.046	0.054	0.037	0.043
Confidence interval	0.019	0.024	0.024	0.025	0.025	0.025	0.026	0.028	0.023	0.025
$\alpha = 0.1$	0.074	0.098	0.097	0.097	0.090	0.089	0.106	0.105	0.091	0.089
Confidence interval	0.032	0.037	0.037	0.037	0.036	0.035	0.038	0.038	0.036	0.035
Uniform p_X^*										
$\alpha = 0.01$	0.002	0.005	0.006	0.011	0.004	0.006	0.004	0.010	0.005	0.009
Confidence interval	0.006	0.009	0.010	0.013	0.008	0.010	0.008	0.012	0.009	0.012
$\alpha = 0.05$	0.023	0.039	0.039	0.045	0.040	0.042	0.045	0.054	0.036	0.042
Confidence interval	0.019	0.024	0.024	0.026	0.024	0.025	0.026	0.028	0.023	0.025
$\alpha = 0.1$	0.074	0.099	0.095	0.097	0.091	0.089	0.105	0.103	0.091	0.090
Confidence interval	0.019	0.024	0.024	0.026	0.024	0.025	0.026	0.028	0.023	0.025

†Nominal levels are $\alpha = 0.01, 0.05, 0.1$. Proportions of rejections among the 1000 tests and 95% confidence interval lengths are given.

Table 2. Power of the local test†

	Results for the following values of n:									
	100	*200*	*300*	*400*	*500*	*600*	*700*	*800*	*900*	*1000*
Normal p_X^*										
$\alpha = 0.01$	0.628	0.837	0.887	0.917	0.934	0.939	0.949	0.953	0.949	0.955
Confidence interval	0.0599	0.046	0.039	0.034	0.031	0.030	0.027	0.026	0.027	0.026
$\alpha = 0.05$	0.841	0.938	0.957	0.971	0.979	0.982	0.987	0.991	0.989	0.994
Confidence interval	0.045	0.030	0.025	0.021	0.018	0.017	0.014	0.012	0.013	0.010
$\alpha = 0.1$	0.903	0.963	0.980	0.986	0.991	0.989	0.994	0.997	0.996	0.999
Confidence interval	0.037	0.023	0.017	0.015	0.012	0.013	0.010	0.007	0.008	0.004
Uniform p_X^*										
$\alpha = 0.01$	0.629	0.838	0.888	0.917	0.934	0.939	0.948	0.951	0.950	0.955
Confidence interval	0.060	0.046	0.039	0.034	0.031	0.030	0.028	0.027	0.027	0.026
$\alpha = 0.05$	0.840	0.938	0.958	0.970	0.978	0.983	0.986	0.992	0.989	0.994
Confidence interval	0.045	0.030	0.025	0.021	0.018	0.016	0.015	0.011	0.013	0.010
$\alpha = 0.1$	0.903	0.964	0.979	0.986	0.991	0.989	0.993	0.998	0.996	0.999
Confidence interval	0.037	0.023	0.018	0.015	0.012	0.013	0.010	0.006	0.008	0.004

†Nominal levels are $\alpha = 0.01, 0.05, 0.1$. Proportions of rejections among the 1000 tests and 95% confidence interval lengths are given.

results thus suggest that a near optimal test can be obtained even when using the wrong model for p_X, although this may not be a general phenomenon.

5.3. Simulations for omnibus test

A second simulation was conducted to study the level and power of our omnibus lack-of-fit

Table 3. Level of the omnibus test†

	Results for the following values of n:									
	100	*200*	*300*	*400*	*500*	*600*	*700*	*800*	*900*	*1000*
Normal p_X^*										
$\alpha = 0.01$	0.004	0.013	0.015	0.011	0.011	0.014	0.010	0.014	0.015	0.011
Confidence interval	0.008	0.014	0.015	0.013	0.014	0.015	0.012	0.015	0.015	0.013
$\alpha = 0.05$	0.033	0.059	0.052	0.063	0.063	0.050	0.051	0.059	0.048	0.049
Confidence interval	0.022	0.029	0.028	0.030	0.025	0.027	0.027	0.029	0.027	0.027
$\alpha = 0.1$	0.093	0.115	0.112	0.114	0.114	0.101	0.112	0.111	0.088	0.110
Confidence interval	0.036	0.040	0.039	0.039	0.037	0.037	0.039	0.039	0.035	0.039
Uniform p_X^*										
$\alpha = 0.01$	0.004	0.014	0.015	0.010	0.012	0.014	0.009	0.015	0.013	0.009
Confidence interval	0.008	0.015	0.015	0.012	0.014	0.015	0.012	0.015	0.014	0.012
$\alpha = 0.05$	0.034	0.058	0.053	0.059	0.043	0.047	0.051	0.058	0.048	0.048
Confidence interval	0.023	0.029	0.028	0.029	0.025	0.026	0.027	0.029	0.027	0.027
$\alpha = 0.1$	0.097	0.121	0.109	0.110	0.096	0.096	0.105	0.110	0.089	0.111
Confidence interval	0.037	0.040	0.039	0.039	0.037	0.037	0.038	0.039	0.035	0.039

†Nominal levels are $\alpha = 0.01, 0.05, 0.1$. Proportions of rejections among the 1000 tests and 95% confidence interval lengths are given.

Table 4. Power of the omnibus test†

	Results for the following values of n:									
	100	*200*	*300*	*400*	*500*	*600*	*700*	*800*	*900*	*1000*
Normal p_X^*										
$\alpha = 0.01$	0.559	0.773	0.848	0.877	0.911	0.901	0.916	0.914	0.917	0.924
Confidence interval	0.062	0.052	0.045	0.041	0.035	0.037	0.034	0.035	0.034	0.033
$\alpha = 0.05$	0.782	0.917	0.945	0.953	0.970	0.976	0.976	0.983	0.976	0.985
Confidence interval	0.051	0.034	0.028	0.026	0.021	0.019	0.019	0.016	0.019	0.015
$\alpha = 0.1$	0.870	0.956	0.974	0.975	0.982	0.987	0.989	0.992	0.992	0.995
Confidence interval	0.042	0.025	0.020	0.019	0.017	0.014	0.013	0.011	0.011	0.009
Uniform p_X^*										
$\alpha = 0.01$	0.521	0.750	0.819	0.855	0.892	0.883	0.896	0.896	0.905	0.898
Confidence interval	0.062	0.054	0.048	0.044	0.039	0.040	0.038	0.038	0.036	0.038
$\alpha = 0.05$	0.757	0.905	0.936	0.947	0.963	0.970	0.969	0.978	0.972	0.982
Confidence interval	0.053	0.036	0.030	0.028	0.023	0.021	0.022	0.018	0.020	0.017
$\alpha = 0.1$	0.852	0.948	0.968	0.973	0.978	0.982	0.986	0.989	0.988	0.995
Confidence interval	0.044	0.028	0.022	0.020	0.018	0.017	0.015	0.013	0.014	0.009

†Nominal levels are $\alpha = 0.01, 0.05, 0.1$. Proportions of rejections among the 1000 tests and 95% confidence interval lengths are given.

test, where the omnibus test statistic $\mathcal{T}$ that was introduced in Section 4 was implemented. The basis functions used are the trigonometric functions $\cos(x)$, $\sin(x)$, $\cos(2x)$ and $\sin(2x)$. The level and power results are presented in Tables 3 and 4. The null sampling distribution of the test statistic was approximated by drawing 10000 independent samples from the statistic's null asymptotic distribution. Although the alternative is not in the space that is spanned by the

four basis functions that were used, our goodness-of-fit test still yields acceptable power at each level. In fact, only at $n = 100$ is the power markedly lower than that of the optimal test that was used in the first study.

5.4. Empirical example

We also implemented the method on a set of nutrition data. The variable of interest is the percentage of calories that come from fat in the diet, which is a variable that is called *fat density*. The ordinary method of computing an individual's fat density is the food frequency questionnaire, and this is the response Y. Typically, it is assumed that the regression of the food frequency questionnaire on true dietary fat density is linear; see for example Kipnis *et al.* (2003). This is the null hypothesis. The alternative hypothesis is that the relationship has curvature.

The data that we use are from the calibration substudy of the National Institutes of Health–American Association of Retired Persons Diet and Health Study (Schatzkin *et al.*, 2001). Instead of observing true dietary fat density, we observed two 24-h recalls on each individual. We summarize the recalled dietary fat densities as W_{i1} and W_{i2}, and assume that $W_{ij} = X_i + U_{ij}$, $j = 1, 2$, where X_i denotes the unobserved true dietary fat density and U_{i1} and U_{i2} are independent and identically distributed, and symmetrically distributed. We found that the distribution of $U_{i1} - U_{i2}$ is very close to being normal; Fig. 1. Hence, we form $W_i = (W_{i1} + W_{i2})/2$ as our

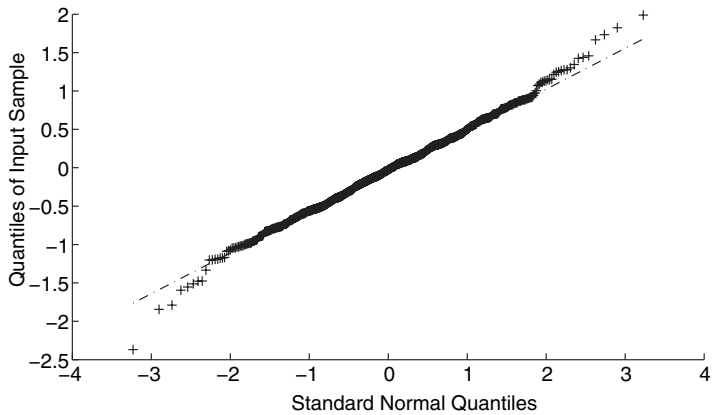

Fig. 1. *QQ*-plot of $W_1 - W_2$

Table 5. Estimates and standard errors under hypothesis H_0 in the nutrition example

Parameter	*Results for normal* p_X^*		*Results for uniform* p_X^*		*Results for estimated* p_X^*	
	Estimate	*Standard error*	*Estimate*	*Standard error*	*Estimate*	*Standard error*
var(ε)	0.3449	0.0299	0.3449	0.0181	0.3192	0.0366
β_0	0.2143	0.1516	0.2132	0.1125	0.3886	0.1692
β_1	0.1193	0.0566	0.0784	0.0535	−0.2215	0.0792
β_2	0.8709	0.0659	0.8809	0.0489	0.8786	0.0824

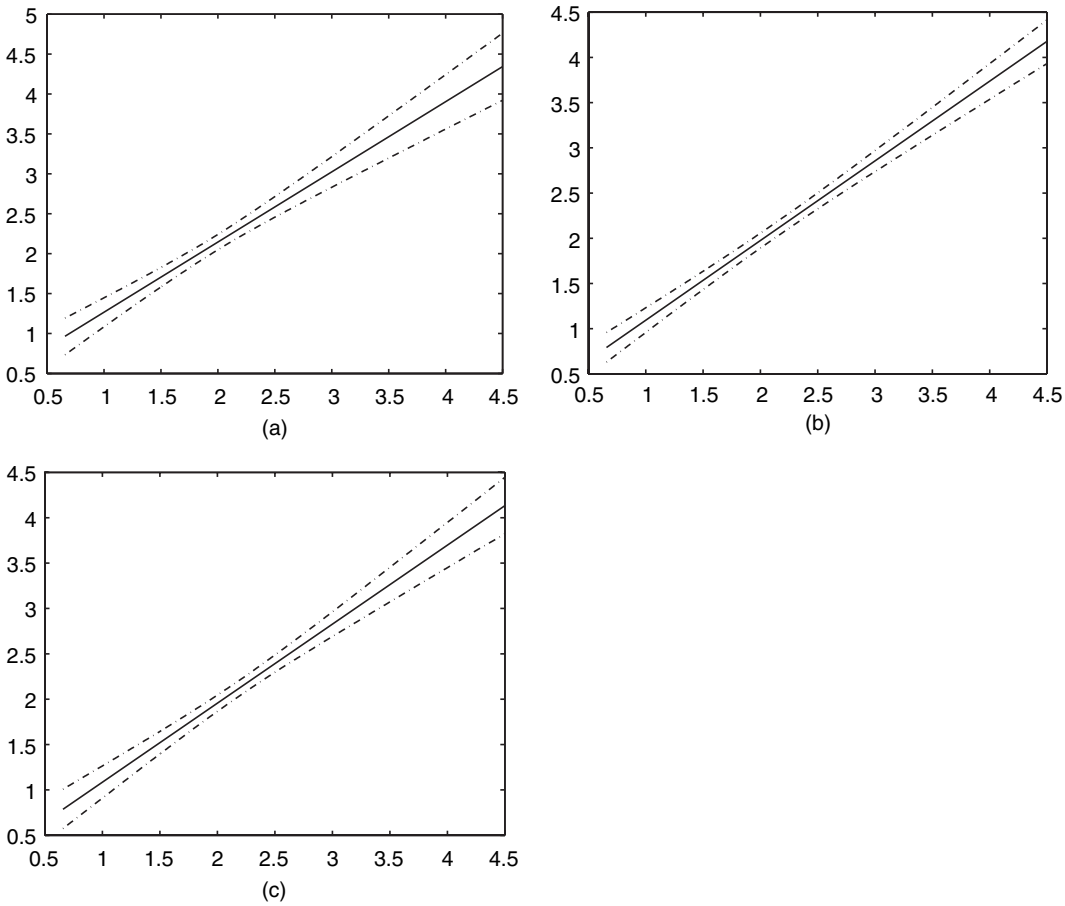

Fig. 2. Curve estimation and pointwise 95% confidence band: (a) normal p_X; (b) uniform p_X; (c) p_X estimated through minimum distance

measurement and use the samples of $(W_1 - W_{i2})/2$ to estimate the error standard deviation, which is approximately 0.554. The data set contains 659 women. The null hypothesis in this problem is $Y = \beta_0 + \beta_1 Z + \beta_2 X + \varepsilon$, where Z is a dummy variable for smoking status. We tested a quadratic local alternative $Y = \beta_0 + \beta_1 Z + \beta_2 X + \beta_3 X^2 + \varepsilon$. Because there is no clear alternative model, we also performed an omnibus test with 10 basis functions. The basis functions that we used are the cosine functions of the form $\cos\{k\pi(X - x_l)/(x_r - x_l)\}$ for $k = 1, \ldots, 10$, where x_l and x_r are the estimated left-hand and right-hand limit of the support of the distribution of X.

Three different models for the distribution of the true intake fat density X were tried. These were normal, uniform and a non-parametric estimate obtained via a minimum distance approach, as used in Claeskens and Hart (2009). The p-values from the local test are 0.7808, 0.6885 and 0.9849, indicating that a quadratic function is not favoured over a linear function. The resulting p-values of the omnibus test are 0.1785, 0.8424 and 0.1842 respectively. So, here also there is not strong evidence to conclude that the linear model is inadequate for describing the relationship between X and Y, after adjusting for smoking status. We present the estimates under hypothesis H_0 in Table 5 and the resulting fitted line at $z = 0$ and the 95% pointwise confidence bands in Fig. 2. A shift of these curves will give the result for $z = 1$.

6. Discussion

Hypothesis testing in measurement error models has not been studied systematically. This by no means implies that the topic is unimportant. The semiparametric nature of the model and the impossibility of calculating the likelihood or forming residuals in functional measurement error models presents a major difficulty in using classical testing procedures.

Taking advantage of the existence of estimating equations in a general measurement error model setting, we have proposed a local testing procedure which has characteristics of the classical score test. The procedure is computationally simple and has the same statistical properties as those of the computationally unwieldy Wald-type test. The procedure is not only valid for parametric measurement error models but also applies to measurement error models that contain unspecified functions of observable covariates or unknown error distributions. The structure of the local test proposed permits a relatively painless adaptation in constructing an omnibus goodness-of-fit test. In the generality of the measurement error models that we consider, both the local and the omnibus tests are the only tests that are available in the literature.

Acknowledgements

Carroll and Ma's research was supported by a grant from the National Cancer Institute (CA57030). Carroll and Hart's work was also supported by award KUS-C1-016-04, made by King Abdullah University of Science and Technology. Ma's work is supported by National Science Foundation grant DMS-0906341, and Hart's was partially supported by National Science Foundation grant DMS-0604801. Janicki's work was done while at the University of Maryland. He thanks his adviser Professor Abram Kagan for his advice and support.

Appendix A: Sketch of technical arguments for theorem 2

Assume that the measurement error model under consideration is identifiable under the null and alternative hypotheses. Let $\boldsymbol{\beta}_+ = \boldsymbol{\beta}_{+n}$ solve $0 = n^{1/2}\, E\{\phi_\beta(\cdot, \boldsymbol{\beta}_+, 0)\}$ under the alternative that $\boldsymbol{\gamma}_n = \mathbf{c} n^{-1/2}$. Then it is readily seen by Taylor expansion that $n^{1/2}(\boldsymbol{\beta}_+ - \boldsymbol{\beta}) \to A_1^{-1} A_3 c$, and

$$n^{1/2}\, E\{\psi_\gamma(\cdot, \boldsymbol{\beta}_+, 0)\} \to -(A_4 - A_2 A_1^{-1} A_3)c.$$

Therefore

$$\hat{U} = n^{-1/2} \sum_{i=1}^{n} \psi_\gamma(W_i, Z_i, Y_i; \hat{\boldsymbol{\beta}}, 0) \to N\{-(A_4 - A_2 A_1^{-1} A_3)c, \Sigma_0\},$$

and hence under the alternative $\hat{T} = \hat{U} \hat{\Sigma}_0^{-1} \hat{U}$ has an asymptotic non-central χ^2-distribution with p_γ degrees of freedom and non-centrality $\frac{1}{2} c^{\mathrm{T}} (A_4 - A_2 A_1^{-1} A_3)^{\mathrm{T}} \Sigma_0^{-1} (A_4 - A_2 A_1^{-1} A_3) c$.

To complete the argument, set $\mathcal{V} = \mathcal{A}^{-1} \mathcal{B} (\mathcal{A}^{-1})^{\mathrm{T}}$, which entails that $\mathcal{B} = \mathcal{A} \mathcal{V} \mathcal{A}^{\mathrm{T}}$. It follows that

$$\begin{aligned}
B_{11} &= A_1 V_{11} A_1^{\mathrm{T}} + A_3 V_{12}^{\mathrm{T}} A_1^{\mathrm{T}} + A_1 V_{12} A_3^{\mathrm{T}} + A_3 V_{22} A_3^{\mathrm{T}}, \\
B_{12} &= A_1 V_{11} A_2^{\mathrm{T}} + A_3 V_{12}^{\mathrm{T}} A_2^{\mathrm{T}} + A_1 V_{12} A_4^{\mathrm{T}} + A_3 V_{22} A_4^{\mathrm{T}}, \\
B_{22} &= A_2 V_{11} A_2^{\mathrm{T}} + A_4 V_{12}^{\mathrm{T}} A_2^{\mathrm{T}} + A_2 V_{12} A_4^{\mathrm{T}} + A_4 V_{22} A_4^{\mathrm{T}}.
\end{aligned}$$

Straightforward calculation then gives

$$\Sigma_0 = (A_4 - A_2 A_1^{-1} A_3) V_{22} (A_4 - A_2 A_1^{-1} A_3)^{\mathrm{T}},$$

and hence

$$\Sigma_0^{-1} = \{(A_4 - A_2 A_1^{-1} A_3)^{-1}\}^{\mathrm{T}} V_{22}^{-1} (A_4 - A_2 A_1^{-1} A_3)^{-1}.$$

This shows that the non-centrality parameter for our score-type test is $\frac{1}{2}c^{\mathrm{T}}V_{22}^{-1}c$. Of course, for the Wald-type test, $n^{1/2}(\hat{\gamma}_{\mathrm{compr}} - \gamma_n) \to N(0, V_{22})$, and hence its local power is also $\frac{1}{2}c^{\mathrm{T}}V_{22}^{-1}c$, which establishes the result.

References

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006) *Measurement Error in Nonlinear Models: a Modern Perspective*, 2nd edn. Boca Raton: Chapman and Hall–CRC Press.

Cheng, C. L. and Kukush, A. (2004) A goodness-of-fit test in a polynomial errors-in-variables model. *Ukr. Math. J.*, **56**, 641–661.

Claeskens, G. and Hart, J. D. (2009) Goodness-of-fit tests in mixed models (with discussion). *Test*, **18**, 213–239.

Fan, J. and Truong, Y. K. (1993) Nonparametric regression with errors in variables. *Ann. Statist.*, **21**, 1900–1925.

Hall, P. and Ma, Y. (2007a) Testing the suitability of polynomial models in errors-in-variables problems. *Ann. Statist.*, **35**, 2620–2638.

Hall, P. and Ma, Y. (2007b) Semiparametric estimators of functional measurement error models with unknown error. *J. R. Statist. Soc.* B, **69**, 429–446.

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

Hart, J. D. (2009) Frequentist-Bayes lack-of-fit tests based on Laplace approximations. *J. Statist. Theor. Pract.*, **3**, 681–704.

Heyde, C. C. (1997) *Quasi-likelihood and Its Application: a General Approach to Optimal Parameter Estimation*. New York: Springer.

Janicki, R. (2009) Statistical inference based on estimating functions in exact and misspecified models. *PhD Thesis*. University of Maryland, College Park.

Kipnis, V., Subar, A. F., Midthune, D., Freedman, L. S., Ballard-Barbash, R., Troiano, R. Bingham, S., Schoeller, D. A., Schatzkin, A. and Carroll, R. J. (2003) The structure of dietary measurement error: results of the OPEN biomarker study. *Am. J. Epidem.*, **158**, 14–21.

Ma, Y. and Carroll, R. J. (2006) Locally efficient estimators for semiparametric models with measurement error. *J. Am. Statist. Ass.*, **101**, 1465–1474.

Ma, Y. and Genton, M. G. (2010) Explicit estimating equations for semiparametric generalized linear latent variable models. *J. R. Statist. Soc.* B, **72**, 475–495.

Ma, Y., Genton, M. G. and Tsiatis, A. A. (2005) Locally efficient semiparametric estimators for generalized skew-elliptical distribution. *J. Am. Statist. Ass.*, **100**, 980–989.

Ma, Y. and Tsiatis, A. A. (2006) Closed form semiparametric estimators for measurement error models. *Statist. Sin.*, **16**, 183–193.

Schatzkin, A., Subar, A. F., Thompson, F. E., Harlan, L. C., Tangrea, J., Hollenbeck, A. R., Hurwitz, P. E., Coyle, L., Schussler, N., Michaud, D. S., Freedman, L. S., Brown, C. C., Midthune, D. and Kipnis, V. (2001) Design and serendipity in establishing a large cohort with wide dietary intake distributions: the National Institutes of Health-American Association of Retired Persons Diet and Health Study. *Am. J. Epidem.*, **154**, 1119–1125.

Sen, P. K. (1982) On M tests in linear models. *Biometrika*, **69**, 245–248.

Small, C. G. and Wang, J. (2003) *Numerical Methods for Nonlinear Estimating Equations*. London: Oxford University Press.

Stefanski, L. A. and Carroll, R. J. (1987) Conditional scores and optimal scores in generalized linear measurement error models. *Biometrika*, **74**, 703–716.

Tsiatis, A. A. (2006) *Semiparametric Theory and Missing Data*. New York: Springer.

Tsiatis, A. A. and Ma, Y. (2004) Locally efficient semiparametric estimators for functional measurement error models. *Biometrika*, **91**, 835–848.