Covariate Measurement Error in Quadratic Regression

Jouni Kuha *
Nuffield College, Oxford OX1 1NF, UK

Jonathan Temple†
Hertford College, Oxford, OX1 3BW, UK
and Institute of Economics and Statistics,
Oxford University, Manor Road, Oxford, OX1 3UL, UK

January 25, 1999

Abstract

We consider quadratic regression models where the explanatory variable is measured with error. The effect of classical measurement error is to flatten the curvature of the estimated function. The effect on the observed turning point depends on the location of the true turning point relative to the population mean of the true predictor. Two methods for adjusting parameter estimates for the measurement error are compared. First, two versions of regression calibration estimation are considered. This approximates the model between the observed variables using the moments of the true explanatory variable given its surrogate measurement. For certain models an expanded regression calibration approximation is exact. The second approach uses moment-based methods which require no assumptions about the distribution of the covariates measured with error. The estimates are compared in a simulation study, and used to examine the sensitivity to measurement error in models relating income inequality to the level of economic development. The simulations indicate that the expanded regression calibration estimator dominates the other estimators when its distributional assumptions are satisfied. When they fail, a small-sample modification of the moment-based estimator performs best.

*Nuffield College, Oxford OX1 1NF, UK. Email jouni.kuha@nuf.ox.ac.uk. Telephone +44 1865 278620
†Hertford College, Oxford, OX1 3BW, UK. Email jon.temple@economics.oxford.ac.uk. Telephone +44 1865 279433.
1 Introduction

Quadratic regression models are one of the simplest ways to explore the presence of non-linearities, and they are used across a wide range of disciplines. In this paper, we examine the effects of classical measurement error on quadratic regression estimates and discuss ways to conduct a sensitivity analysis. Two competing approaches are considered: regression calibration and moment-based methods. We compare their performance in a simulation study, and demonstrate their potential usefulness using data on development and inequality.

Understanding the effects of classical measurement error on quadratic regression estimates is particularly important given the popularity of such models. Quadratic regression is widely used in the social sciences: as an example, labour economists often use a quadratic specification to examine the effect of age or experience on earnings. A quick search through the literature indicates a wide variety of applications in other subjects: for instance astronomy in Barton and David (1960), botany in Causton and Venus (1981), and the earth sciences in Wolter and Fuller (1982).

In most applications, measurement errors are likely to be present. Our results indicate that the attenuation of the slope coefficient in a simple linear regression has a close analogue in a quadratic model. The curvature of the estimated function will be less steep than in the true model, and hence measurement error will tend to hide the presence of a non-linearity, at least in this simple context.

Sometimes, in estimating a quadratic model, the focus of interest is the turning point. We also explore the implications of measurement error for estimates of the turning point. We illustrate the relevance of these results using an important example drawn from development economics: the relation, probably non-linear, between the level of development and income inequality. The estimation approaches we use shed new light on this intensively studied issue, and hence we believe that they could be useful in a wide variety of contexts.

The model of interest that we consider is a quadratic regression model for a response variable $Y$, given by

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i$$  \hspace{1cm} (1.1)
where $X_i$, $i = 1, \ldots, n$, are values of a continuous explanatory variable $X$, and $e_i$ are independent random variables with mean $E(e_i|X_i) = 0$ and variance $\text{var}(e_i|X_i) = \sigma^2_e$. The true $X_i$ are measured with error, so that we only observe a surrogate variable $W_i$ related to $X_i$. In this paper we will assume that the measurement error follows the classical measurement error model

$$W_i = X_i + u_i$$

where $u_i$, $i = 1, \ldots, n$, are independent random variables with $E(u_i|X_i) = 0$ and $\text{var}(u_i|X_i) = \sigma^2_u$. We will further assume that both $e_i$ and $u_i$ are normally distributed and that $\text{cov}(u_i, e_i|X_i) = 0$. The response $Y_i$ and the surrogate $W_i$ are thus conditionally uncorrelated given $X_i$, so that the measurement error in $X_i$ is nondifferential with respect to $Y_i$. The assumptions of normality and nondifferential error can sometimes be relaxed, as discussed in Section 3.

If the model (1.1) is fitted with $X_i$ replaced by $W_i$, the resulting naive estimate $\hat{\beta}^*$ of $\beta = (\beta_0, \beta_1, \beta_2)'$ will be biased. If sufficient information on the measurement error parameters is available, it can be utilized to remove or reduce the bias by using estimators adjusted for the error. There is extensive literature on such estimation methods for linear models with measurement error (Fuller 1987), but much less on quadratic models. The ideas for linear models have been extended to quadratic (or, more generally, polynomial) regression models by Wolter and Fuller (1982), Chan and Mak (1985), Fuller (1987), and Cheng and Schneeweiss (1998). The quadratic model can also be approached by applying more general methods for nonlinear regression models with measurement error (Carroll et al. 1995; Buonaccorsci 1996). Instrumental variables estimation and the identifiability of the parameters in that case have been considered by Amemiya (1985), Hsiao (1989) and Hausman et al. (1991).

Two types of adjusted estimators of $\beta$ are considered in this paper. In Section 2 we apply the general regression calibration and expanded regression calibration estimators (Carroll et al. 1995) to quadratic regression. Section 3 discusses a moment-based estimator proposed by Fuller (1987) and others. A key difference between these approaches is in the assumptions they make about the true explanatory variables $X_i$. The methods described in Section 2 require a model for the mean and variance of $X_i$ given $W_i$. In particular, we consider a model implied by (1.2) when $X_i$ are normally distributed. In contrast, the estimators considered in Section 3 make no assumptions about $X_i$, which can then be regarded either as random variables or fixed constants.
Estimation of the turning point of the quadratic regression function is discussed in Section 4. The estimators are applied to economic data on development and inequality in Section 5 and compared using simulation studies in Section 6. Section 7 gives some conclusions.

2 Regression calibration

The general approaches of regression calibration (RC) and expanded regression calibration (ERC) are described in detail in Carroll, Ruppert and Stefanski (1995, chapter 3). The basic idea of RC estimation is to replace the unobserved true predictors $X$ and $X^2$ by their means given the measured variable $W$, and fit a model for $Y$ given these conditional means. Expanded regression calibration improves this approximation further by adding terms depending on the variance of $X$ given $W$. For quadratic regression, the ERC method yields consistent estimates of $\beta$, and also closed-form expressions for the bias in naive estimates of $\beta$ and functions of it, such as the turning point $\eta = -\beta_1/(2\beta_2)$.

To clarify the ideas, we start with a general model, in which $Y$ given $X$ is characterized by the mean and variance relations

$$\text{E}(Y|X) = f(X; \beta), \quad (2.3)$$

$$\text{var}(Y|X) = \sigma^2 g^2(X; \beta). \quad (2.4)$$

Suppose also that $X$ is a random variable, with its first two moments given $W$ given by

$$\text{E}(X|W) = m(W; \phi), \quad (2.5)$$

$$\text{var}(X|W) = \tau^2 V^2(W; \phi). \quad (2.6)$$

The method requires that consistent estimates of $\phi$ and $\tau^2$ are available or that their values are assumed to be known.

The expanded regression calibration model uses the approximation that

$$\text{E}(Y|W) \approx f[m(W)] \quad (2.7)$$

$$+(1/2)\tau^2 V^2(W)f_{xx}[m(W)],$$

$$\text{var}(Y|W) \approx \sigma^2 g^2[m(W)] \quad (2.8)$$

$$+\tau^2 V^2(W)\{f^2_x[m(W)] + (1/2)\sigma^2 s_{xx}[m(W)]\}.$$
where $f_x$ and $f_{xx}$ are the first and second derivatives of $f$ with respect to $X$, and $s_{xx}$ is the second derivative of $g^2$ with respect to $X$. Here $X$ is assumed to be a scalar variable. The model can be extended to vector-valued variables, but the algebra then becomes more complex (Carroll and Stefanski 1990).

Equations (2.7)–(2.8) specify a model for $Y$ given $W$. This is in general an approximation of the true model, which is in many cases not available in a closed form. Adjusted estimates of $\beta$ and $\sigma_e^2$ are obtained by fitting (2.7)–(2.8) to the data.

A simpler approximate model is the regression calibration approximation, which is obtained by using only the first terms of the right hand sides of (2.7) and (2.8). This is simply the original model with $X$ replaced by $m(W) = \text{E}(X|W)$. Unlike in the ERC case, the same model holds also for vector $X$ with no further complications.

For the quadratic regression model (1.1), the functions in (2.3) and (2.4) are $f(X) = \beta_0 + \beta_1 X + \beta_2 X^2$ and $g^2(X) = 1$, and the expanded regression calibration model is given by

$$
\text{E}(Y|W) = \beta_0 + \beta_1 m(W) + \beta_2 m(W)^2 + \beta_2 \tau^2 V^2(W)
$$

$$
\text{var}(Y|W) \approx \sigma_e^2 + \tau^2 V^2(W)[\beta_1 + 2\beta_2 m(W)]^2
$$

Now (2.9) is actually the true model for $\text{E}(Y|W)$ rather than an approximation, since $\text{E}(Y|W) = \text{E}[\text{E}(Y|X)|W]$. A similar computation for the conditional variance (see Carroll et al. 1995, p. 78) shows that in many cases the expanded regression calibration model also accurately reflects the variance function $\text{var}(Y|W)$. If the distribution of $X$ given $W$ is homoscedastic and symmetric, as it will be for the models considered below, the exact variance function is the same as (2.10), except that $\sigma_e^2$ is replaced by $\sigma_e^2 = \sigma_e^2 + \beta_2^2 (\kappa - \tau^4)$. Here $\kappa = \text{E}\{[X - m(W)]^4|W\}$, which is a constant because of the homoscedasticity; for a normal distribution $\kappa = 3\tau^4$ and $\sigma_e^2 = \sigma_e^2 + 2\beta_2^2 \tau^4$. Since both $\sigma_e^2$ and $\sigma_e^2$ are constants, this replacement is unimportant; the model (2.9)–(2.10) gives consistent estimates for $\beta$ and $\sigma_e^2$, from which a consistent estimate for $\sigma_e^2$ can be recovered. If the distribution of $X$ given $W$ is asymmetric, (2.10) omits a term involving its third moment.

The simple regression calibration approximation for the quadratic model has $\text{E}(Y|W) = \beta_0 + \beta_1 m(W) + \beta_2 m(W)^2$ and $\text{var}(Y|W) = \sigma_e^2$. Thus it is a homoscedastic quadratic regression model for $m(W)$. As seen from (2.9) and (2.10), this ignores the heteroscedasticity in the full
model and omits the last term in (2.9). If \( V^2(W) \) is not a function of \( W \), the simple regression calibration model is correct for \( \beta_1 \) and \( \beta_2 \) but not for \( \beta_0 \).

2.1 Classical measurement error

Suppose that the classical measurement error model (1.2) holds, and that \( X \) is normally distributed with mean \( \mu_x \) and variance \( \sigma_x^2 \). Then \( X \) given \( W \) is also normally distributed with mean and variance

\[
m(W) = E(X|W) = \left( \frac{\sigma_u^2}{\sigma_w^2} \right) \mu_x + \left( \frac{\sigma_x^2}{\sigma_w^2} \right) W
\]

\[
\equiv \alpha_0 + \alpha_1 W
\]

\[
\text{var}(X|W) = \frac{\sigma_u^2 \sigma_x^2}{\sigma_w^2} = \frac{\rho}{(1 + \rho)^2} \sigma_w^2 = \tau^2
\]

(2.12)

where \( \sigma_w^2 = \text{var}(W) = \sigma_u^2 + \sigma_x^2 \) and \( \rho = \sigma_u^2 / \sigma_x^2 \). Thus \( V^2(W; \phi) \equiv 1 \) in (2.6). Substituting these into (2.9) and (2.10), we obtain

\[
E(Y|W) = \beta_0 + \beta_1 m + \beta_2 (m^2 + \tau^2)
\]

(2.13)

\[
= \beta_0^* + \beta_1^* W + \beta_2^* W^2,
\]

(2.14)

\[
\text{var}(Y|W) \approx \sigma_e^2 + \tau^2 (\beta_1 + 2 \beta_2 m)^2
\]

(2.15)

\[
= \sigma_e^2 + (1 + \rho)^2 \tau^2 (\beta_1^* + 2 \beta_2^* W)^2
\]

(2.16)

where \( m = m(W), \beta_0^* = \beta_0 + \beta_1 \alpha_0 + \beta_2 \tau^2 + \beta_2 \alpha_0^2, \beta_1^* = \beta_1 \alpha_1 + 2 \beta_2 \alpha_0 \alpha_1 \) and \( \beta_2^* = \beta_2 \alpha_1^2 \).

Thus the relationship between \( W \) and \( E(Y|W) \) is also quadratic, and the naive estimate of \( \beta \) estimates \( \beta^* = (\beta_0^*, \beta_1^*, \beta_2^*) \).

The model (2.13)–(2.16) was derived using the classical measurement error model (1.2) and assuming that \( X \) is normally distributed. Note, however, that regression calibration makes use of only the first two moments of \( X \) given \( W \). Even when \( X \) is not normal, (2.11) is the best linear approximation of \( E(X|W) \) (Carroll et al. 1995). Thus the model should be approximately correct if the model for \( X \) given \( W \) is approximately linear and homoscedastic. In Section 6 we examine further the behaviour of the estimates when \( X \) is not normally distributed.
2.2 Effects of measurement error

The results of the previous section can be used to examine the bias in the observed mean model under a classical measurement error model with normally distributed $X$. Without loss of generality, we assume that $\beta_2 < 0$ so that the true model for $E(Y|X)$ has a maximum at the turning point $\eta = -\beta_1/2\beta_2$. Note first that the turning point in the observed model is

$$\eta^* = -\frac{\beta_1^*}{2\beta_2^*} = -\frac{\beta_1^* + 2\beta_2^*\alpha_0\alpha_1}{2\beta_2^*\alpha_1^2} = \frac{1}{\alpha_1} \eta - \frac{\alpha_0}{\alpha_1} = (1 + \rho) \eta - \rho \mu_x = \eta + \rho(\eta - \mu_x)$$

(2.17)

We thus obtain a simple formula for the bias of the naive estimate of the turning point. The degree of the bias depends on the amount of measurement error, characterized by the ratio $\rho$, and on the location of the true turning point relative to the population mean of $X$. This location also determines the direction of the bias: $\eta^*$ is biased upwards if $\eta > \mu_x$, and downwards if $\eta < \mu_x$. When $\eta = \mu_x$, the bias is zero irrespective of the amount of measurement error. Also, solving for $\eta$ in the bias formula gives

$$\eta = \frac{1}{1 + \rho} \eta^* + \frac{\rho}{1 + \rho} \mu_x.$$  

(2.18)

This shows that the true turning point is always between $\mu_x$ and $\eta^*$, which are estimated by the sample mean of $W_i$ and $\hat{\eta}^* = -\hat{\beta}_1^*/2\hat{\beta}_2^*$, where $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ are the naive estimates of $\beta_1$ and $\beta_2$.

The effect of measurement error is to ‘flatten’ the observed relationship between $W$ and $Y$. This can be demonstrated by considering the slope of the quadratic response curve $c$ units from its turning point. For the true curve this is given by $f_x(\eta + c) = 2\beta_2(\eta + c) + \beta_1 = 2\beta_2 c$, while for the observed curve it is $f_w(\eta^* + c) = 2\beta_2^* c = 2\alpha_1^2 \beta_2 c < f_x(\eta + c)$ since $\alpha_1 < 1$. Similarly, the expected value of $Y$ at the true turning point $\eta$ is $\mu_{\text{max}} = \beta_0 - \beta_1^2/4\beta_2$, while the maximum of the observed model at $\eta^*$ is the smaller value $\mu_{\text{max}} + \beta_2 \tau^2$.

Compared to the true model for $E(Y|X)$, the observed model for $E(Y|W)$ thus curves less steeply and has a smaller maximum value for $Y$. These effects are illustrated in Figure 1. The flattening of the quadratic relationship is the equivalent of the well-known attenuation of the observed slope induced by classical measurement error in simple linear models.

Similar bias results for $\beta^*$ are obtained by Griliches and Ringstad (1970), who assume that
Figure 1: Illustration of the measurement error bias in quadratic regression. The solid line denotes $E(Y|X)$, and the other lines, in order of decreasing steepleness, show $E(Y|W)$ for $\rho = 0.2, 0.5$ and 1.0 respectively. In both plots $\eta = 1$, while $\mu_x = 1$ on the left and $\mu_x = 2$ on the right.

In this case we obtain the simple result that $\beta_1^* = \alpha_1 \beta_1$ and $\beta_2^* = \alpha_1^2 \beta_2$, which shows clearly that the measurement error bias is most severe for the coefficient of the quadratic term. Griliches and Ringstad also obtain broadly similar results for the case where $X$ is symmetrically distributed.

2.3 Estimation of regression coefficients

Let $(Y_i, W_i), i = 1, \ldots, n$, be independent observations of $(Y, W)$. These data alone are not sufficient to identify $\beta$ and $\sigma^2_e$ from (2.13)–(2.16). In addition, some auxiliary information about the measurement error are required, in the form of either additional data or assumptions about the error parameters. An ideal type of auxiliary data are validation data, which consist of observations $(X_i, W_i)$. These may be available if $X_i$ can be measured, even if at a much higher cost than $Z_i$, in a subset of the data. A validation data allows direct modelling of the moments of $X$ given $W$, so that more general forms than (2.11) and (2.12) can be considered if necessary. Obtaining validation data is, however, often not feasible in practice. An alternative is repeated measurements data of observations $(W_{i1}, \ldots, W_{iJ})$ with $J \geq 2$. For identifiability, these are assumed to be conditionally independent given the underlying $X_i$. The error variance $\sigma^2_u$ can then be estimated using simple components of variance estimates (see, e.g., Carroll
et al. 1995).

If no auxiliary data are available, identifiability requires that the measurement error parameters are regarded as known. It is then essential to carry out sensitivity analyses by computing adjusted estimates of $\beta$ for a range of assumptions. Here we consider this approach, with the ratio $\rho = \sigma_u^2 / \sigma_x^2$ assumed to be known.

For given $\rho$, the simple and expanded regression calibration models involve fitting models for the moments $E(Y|W)$ and $\text{var}(Y|W)$. This approach is known as quasilikelihood or quasilikelihood and variance function (QVF) estimation. It produces consistent estimates of the parameters if the mean and variance function of the response have been correctly specified. A brief description of QVF estimation is given by Carroll et al. (1995), Appendix A.4. A more comprehensive account can be found in Carroll and Ruppert (1988).

We now describe our implementation of QVF estimation. First, estimates of $\alpha_0$, $\alpha_1$ and $\tau^2$ are obtained from (2.11)–(2.12) with $\mu_x$ and $\sigma_w^2$ estimated by $\bar{W}$ and $s_w^2$, the sample mean and variance of $W_i$ respectively. It is then convenient to compute $m_i = \hat{\alpha}_0 + \hat{\alpha}_1 W_i$ and treat these as the data in subsequent fitting of the model given by (2.13) and (2.15).

The simple regression calibration approximation is obtained from (2.13) and (2.15) by omitting the terms depending on $\tau^2$. This is a homoscedastic quadratic model for $Y_i$ given $m_i$, for which the QVF estimates of $\beta$ and $\sigma_e^2$ are given by the ordinary least squares estimates. Denote these estimates by $\hat{\beta}_{rc}$ and $\hat{\sigma}_{e(rc)}^2$ respectively. An alternative way of computing $\hat{\beta}_{rc}$ is suggested by the closed-form expressions for $\beta^* = (\beta_0^*, \beta_1^*, \beta_2^*)$ given by the RC approximation. Inverting these shows that $\hat{\beta}_{rc}$ can be obtained from

$$\hat{\beta}_{rc} = C\hat{\beta}^*,$$  \hspace{1cm} (2.19)

where

$$C = \hat{\alpha}_1^{-2} \begin{bmatrix}
\hat{\alpha}_1^2 & -\hat{\alpha}_0 \hat{\alpha}_1 & \hat{\alpha}_0^2 \\
0 & \hat{\alpha}_1 & -2 \hat{\alpha}_0 \\
0 & 0 & 1
\end{bmatrix}.$$ \hspace{1cm} (2.20)

Equation (2.19) facilitates an easy sensitivity analysis, as $\hat{\beta}_{rc}$ can be computed for different values of $\rho$ without refitting the quadratic model.
Simple regression calibration gives consistent estimates of $\beta_1$ and $\beta_2$. The estimators are, however, not fully efficient because they ignore the heteroscedasticity in (2.15). The RC estimate of $\sigma^2_e$ is equal to the naive estimate, which overestimates the true $\sigma^2_e$.

For the expanded regression calibration model given by (2.13) and (2.15), define $f(m_i; \beta) = \beta_0 + \beta_1 m_i + \beta_2 (m_i^2 + \bar{\tau}^2)$ and $g^2(m_i; \beta, \sigma^2_e) = \sigma^2_e + \bar{\tau}^2 (\beta_1 + 2 \beta_2 m_i)^2 = \sigma^2_e + c(m_i; \beta)$. QVF estimates of $\beta$ and $\sigma^2_e$ are solutions to the conditionally unbiased estimating equations

$$\sum_{i=1}^{n} \frac{y_i - f(m_i; \beta)}{g^2(m_i; \beta, \sigma^2_e)} \frac{\partial}{\partial \beta} f(m_i; \beta) = 0 \quad (2.21)$$

$$\sum_{i=1}^{n} \left\{ \frac{[y_i - f(m_i; \beta)]^2}{g^2(m_i; \beta, \sigma^2_e)} - \frac{n - k}{n} \right\} \frac{\partial}{\partial \sigma^2_e} \log g(m_i; \beta, \sigma^2_e) = 0 \quad (2.22)$$

where $k = \dim(\beta) = 3$. Other conditionally unbiased estimating equations could be used for $\sigma^2_e$ to replace (2.22); see Carroll and Ruppert (1988).

In practice the estimating equations are solved iteratively. The following algorithm is used here:

1. Given estimates $\beta^{(j)}$ and $\sigma^2_e^{(j)}$ from the $j$th round of the algorithm, solve (2.22) for $\sigma^2_e$ treating $\beta^{(j)}$ as known. That is, $\sigma^2_e^{(j+1)}$, an updated estimate of $\sigma^2_e$ is the solution to the equation

$$h(\sigma^2_e) = \sum_{i=1}^{n} \left[ \frac{r_i^2}{\sigma^2_e + c_i} - a \right] \frac{1}{\sigma^2_e + c_i} = 0$$

where $r_i = y_i - f(m_i; \beta^{(j)})$, $c_i = c(m_i; \beta^{(j)})$ and $a = (n - k)/n$. We have used the Newton–Raphson algorithm to solve the equation.

2. Solve (2.21) for $\beta$, treating $\sigma^2_e^{(j+1)}$ as fixed. The updated estimate $\beta^{(j+1)}$ is given by a weighted least squares estimate from regression for $Y_i$ on $m_i$ and $m_i^2 + \bar{\tau}^2$, with weights $w_i = [g^2(m_i; \beta^{(j)}, \sigma^2_e^{(j+1)})]^{-1}$.

The naive estimate $\hat{\beta}_r$ or the simple regression calibration estimate $\hat{\beta}_r$ can be used as an initial value $\beta^{(0)}$ for $\beta$. A simple choice for $\sigma^2_e$ is

$$\sigma^2_e^{(0)} = \frac{1}{n - k} \sum_{i=1}^{n} [Y_i - f(m_i; \beta^{(0)})]^2 - \bar{c} \quad (2.23)$$

where $\bar{c} = n^{-1} \sum c(m_i; \beta^{(0)})$. 

10
Carroll and Ruppert (1988) note that for QVF models it is not usually necessary to iterate the algorithm to convergence. Given a sufficiently accurate starting value, the estimate after any iteration is a consistent estimate of the parameters. In fact, in some cases the estimate \((\beta^{(1)}', \sigma^2_e^{(1)})\) has the best properties. More commonly, \((\beta^{(2)}', \sigma^2_e^{(2)})\) is preferable to the first-round estimate but essentially equivalent to all subsequent ones. This result is supported by our simulations, and we thus use the results after two iterations, denoting them by \(\hat{\beta}_{erc}\) and \(\tilde{\sigma}^2_e(erc)\). As noted in Section 2, \(\tilde{\sigma}^2_e(erc)\) actually estimates \(\sigma^2_e^*\) rather than \(\sigma^2_e\); a consistent estimate of \(\sigma^2_e\) is given by \(\hat{\sigma}^2_e(erc) = \tilde{\sigma}^2_e(erc) - 2\hat{\beta}_{erc}^2\hat{\tau}^4\).

### 2.4 Variance estimation

Variances estimates for the estimates of \(\beta\) are obtained by applying general results for QVF models (see Carroll et al. 1995, Appendix A.4, and Carroll and Ruppert 1988). If an estimate \(\hat{\beta}\) satisfies the estimating equations (2.21) and (2.22), asymptotically \(\sqrt{n}(\hat{\beta} - \beta) \sim N(0, A_n^{-1}B_nA_n^{-1})\), where

\[
A_n = n^{-1}M'D'^{-1}M \\
B_n = n^{-1}M'R^{-1}M,
\]

\(M\) is a matrix with rows \(\partial f(m_i; \beta)/\partial \beta\), \(D\) is a diagonal matrix with diagonal elements \(g^2(m_i; \beta, \sigma^2_e)\) and \(R\) is a diagonal matrix with diagonal elements \(g^4(m_i; \beta, \sigma^2_e)/\{E[Y_i - f(m_i; \beta)]^2\}\). A consistent estimate of var(\(\hat{\beta}\)) is

\[
\text{var}(\hat{\beta}) = n^{-1}\hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-1} \quad (2.24)
\]

where \(\hat{A}_n\) and \(\hat{B}_n\) are obtained by substituting consistent estimates \(\hat{\beta}\) and \(\hat{\sigma}^2_e\) for \(\beta\) and \(\sigma^2_e\) and estimating \(E[Y_i - f(m_i; \beta)]^2\) by \([Y_i - f(m_i; \hat{\beta})]^2\). These results apply to both the RC and ERC estimates, with \(f(\cdot)\) and \(g^2(\cdot)\) defined appropriately.

The estimator (2.24), which is often known as the sandwich estimator of variance, is consistent for the asymptotic variance of \(\hat{\beta}\) even when the variance model \(g^2(\cdot)\) is misspecified. When the variance model is correct, \(A_n = B_n\) and the asymptotic variance matrix is \(n^{-1}A_n^{-1}\). Carroll et al. (1995) note that a variance estimate based on this is preferable when the variance function is at least approximately correct, because \(n^{-1}\hat{A}_n^{-1}\) can be much less variable than the sandwich estimator. Our simulations support this recommendation.
For the simple regression calibration estimator $\hat{\beta}_{rc}$, the matrix $n^{-1} \hat{A}_n$ is the standard variance estimate for estimate of $\beta$ from quadratic regression for $Y_i$ given $m_i$. It can also be computed as $\text{var}(\hat{\beta}_{rc}) = C \text{var}(\hat{\beta}^*) C'$, where $C$ is defined in (2.20) and $\text{var}(\hat{\beta}^*)$ is the standard variance estimate for the naive estimate $\hat{\beta}^*$.

It should be noted that the variance estimates given here will underestimate the true variability of $\hat{\beta}$, since they treat $m_i$ and $\sigma^2$ as known, ignoring the uncertainty in $\bar{W}$ and $s^2_w$ used in computing them. Similarly, if $\rho$ or $\sigma^2_u$ is estimated from auxiliary data rather than assumed known, the variance estimates should also reflect the uncertainty in this estimate. Asymptotic formulas which do this are available but they are complicated (see Carroll et al., 1995, Sec. 3.12.2). An alternative is to use bootstrapping to account for all the sources of variability.

3 Moment-based estimators

If there is no measurement error, $\beta$ is estimated by the ordinary least squares estimator $\hat{\beta}_{OLS}$ which solves the unbiased estimating equation

$$M_{xx} \hat{\beta}_{OLS} = M_{xy}$$

where $M_{xx} = n^{-1} \sum_i Z_i Z_i'$, $M_{xy} = n^{-1} \sum_i Y_i Z_i$ and $Z_i = (1, X_i, X_i^2)'$, $i = 1, \ldots, n$. When $X$ is measured with error, $X_i$ are unknown and the naive estimator which replaces $X_i$ with $W_i$ is inconsistent for $\beta$.

Assume again that $\rho$ is regarded as known, so that an unbiased estimate of the measurement error variance is $\hat{\sigma}^2_u = s^2_w \rho/(1 + \rho)$. A consistent estimator for $\beta$ can be obtained by replacing $M_{xx}$ and $M_{xy}$ in (3.25) with consistent estimators of them. Defining $V_i = (1, W_i, W_i^2 - \hat{\sigma}^2_u)'$, let $M_{vw} = n^{-1} \sum_i V_i V_i'$ and $M_{vy} = n^{-1} \sum_i Y_i V_i$. Then $\text{E}(Y_i V_i | Y_i, X_i) = Y_i Z_i$ and

$$\text{E}(V_i V_i' | Y_i, X_i) = Z_i Z_i' + \begin{bmatrix} 0 & 0 & 0 \\ \hat{\sigma}^2_u & 2X_i \hat{\sigma}^2_u & 4X_i^2 \hat{\sigma}^2_u + 2\hat{\sigma}^4_u \end{bmatrix} = Z_i Z_i' + \Omega_i.$$

An unbiased estimator $\hat{\Omega}_i$ for $\Omega_i$ is obtained by replacing $X_i$ with $W_i$ and $X_i^2$ with $W_i^2 - \hat{\sigma}^2_u$. Defining $\hat{\Omega} = n^{-1} \sum_i \hat{\Omega}_i$, $\hat{M}_{xx} = M_{vw} - \hat{\Omega}$ and $\hat{M}_{xy} = M_{vy}$ are approximately unbiased estimators of $M_{xx}$ and $M_{xy}$ respectively. We can then define an approximately unbiased estimating
equation \( \hat{M}_{xx}\beta = \hat{M}_{xy} \) which involves only the observed variables \( W_i \) and \( Y_i \). Solving it for \( \beta \) gives the consistent estimator

\[
\hat{\beta}_F = \hat{M}_{xx}^{-1} \hat{M}_{xy}
\]

\[
= \begin{bmatrix}
1 & \overline{W} & \overline{W^2 - \hat{\sigma}_u^2} \\
\overline{W^2 - \hat{\sigma}_u^2} & \overline{W^2} - 3\overline{W^2\hat{\sigma}_u^2} & \overline{Y} \\
\overline{W^2 - 3\overline{W^2\hat{\sigma}_u^2}} & \overline{Y\overline{W}} & \overline{Y\overline{W^2}} - \overline{Y\hat{\sigma}_u^2}
\end{bmatrix}^{-1}
\]

where the bars denote averages of the \( n \) observations; e.g., \( \overline{W^2} = n^{-1} \sum_i W_i^2 \).

The estimator \( \hat{\beta}_F \) is described by Fuller (1987). It is also closely related to the estimator considered by Chan and Mak (1985), and is the same as those proposed by Buonaccorsi (1996) and Cheng and Schneeweiss (1998) when applied to the model considered here, that is, a quadratic regression model with \( e_i \) and \( u_i \) normally distributed and conditionally independent given \( Y_i \) and \( X_i \).

The estimator \( \hat{\beta}_F \) makes no assumptions about the distribution of \( X_i \). Its assumptions can also be further relaxed without changing the basic idea of the estimator. If \( e_i \) and \( u_i \) are allowed to be correlated, (3.26) will have further terms involving \( \text{cov}(e_i, u_i | X_i) \). If \( e_i \) and \( u_i \) are not assumed to be normal, terms involving \( \text{E}(u_i^3 | X_i) \) and \( \text{E}(e_i u_i^2 | X_i) \) will be required (Cheng and Schneeweiss 1998). Finally, the covariance matrix of \( (e_i, u_i) \) may depend on \( i \), with obvious adjustments to the formulas (Buonaccorsi 1996).

Fuller (1987) notes that although \( \hat{\beta}_F \) has a limiting normal distribution, it may behave badly in small samples, where it need not have a finite mean and variance. This is caused by the matrix inverse in (3.26). Fuller (1987) proposed modified estimators to improve the small-sample properties of \( \hat{\beta}_F \). These were further refined for polynomial regression by Cheng et al. (1998). First, define \( \hat{\lambda} \) as the smallest root of

\[
|M_s - \lambda \hat{\Omega}_s| = 0 \tag{3.27}
\]

where \( M_s = n^{-1} \sum_i V_s i V_s' \), \( V_s = (Y_s, V_s')' \) and

\[
\hat{\Omega}_s = \begin{bmatrix}
0 & 0' \\
0' & \hat{\Omega}
\end{bmatrix}.
\]

Solving (3.27) can be converted into a standard eigenvalue problem by noting that \( \hat{\lambda} = \hat{\delta}^{-1} \), where \( \hat{\delta} \) is the largest eigenvalue of \( C^{-1} \hat{\Omega}_s (C')^{-1} \) and \( M_s = CC' \) is the Cholesky decomposition.
of the full rank matrix $M$. Cheng et al. (1998) note that in the general case of a polynomial regression model with measurement error (3.27) may have negative solutions, and thus $\lambda$ should be defined as the smallest positive root of the equation. Here, however, it can be shown that $\hat{\Omega}$ is positive definite and thus $C^{-1}\hat{\Omega}(C')^{-1}$ is positive semidefinite, so that all the roots of (3.27) are nonnegative.

The modified estimator of $\beta$ is

$$\hat{\beta}_F(\alpha) = \hat{H}^{-1}_{xx} \hat{M}_{xy}$$

(3.28)

where $\hat{H}_{xx} = M_{vv} - c\hat{\Omega}$,

$$c = \begin{cases} 
(n - \alpha)/n & \text{if } \hat{\lambda} > 1 + n^{-1} \\
\hat{\lambda}(n - \alpha)/(n + 1) & \text{if } \hat{\lambda} \leq 1 + n^{-1},
\end{cases}$$

and $\alpha > 0$ is a fixed number. This definition of $c$ is due to Cheng et al. (1998). Fuller (1987) suggested a similar estimator, but with $c = \hat{\lambda} - (\alpha + 1)/n$ when $\hat{\lambda} \leq 1 + n^{-1}$ rather than $c = \hat{\lambda} - \hat{\lambda}(\alpha + 1)/(n + 1)$ as here. The two estimators are very similar when $\hat{\lambda}$ is close to 1, but in the fairly rare cases where $\hat{\lambda}$ is very small Fuller’s estimator has the disadvantage that $c$ can be negative.

Fuller (1987) notes that for linear measurement error models with the measurement error variances known, the mean squared error of $\hat{\beta}_F(\alpha)$ is, through terms of order $n^{-2}$, smaller for $\alpha = \dim(\beta) + 4$ than for any smaller $\alpha$. In our model this suggests that it is best to choose $\alpha$ to be at least 7. However, since Fuller’s result was obtained in a different context, it is not clear whether it applies in the quadratic case. It also fails to provide any upper bounds for good values of $\alpha$. We will examine the choice of $\alpha$ further in the simulations of Section 6.

Wolter and Fuller (1982) and Fuller (1987) consider similar modified estimators for models with no error in the equation, i.e. where $e_i \equiv 0$ but there may be measurement error in the response variable $Y_i$.

Under suitable regularity conditions both $\hat{\beta}_F$ and $\hat{\beta}_F(\alpha)$ are asymptotically normally distributed with mean $\beta$ and variance matrix which can be consistently estimated by

$$\text{var}(\hat{\beta}) = n^{-1} \hat{A}_n^{-1} \hat{B}_n \hat{A}_n^{-1}$$
where \( \hat{\beta} \) denotes \( \hat{\beta}_F \) or \( \hat{\beta}_F(\alpha) \), \( \hat{A}_n \) is \( \hat{M}_{xx} \) for \( \hat{\beta}_F \) and \( \hat{H}_{xx} \) for \( \hat{\beta}_F(\alpha) \), and

\[
\hat{B}_n = n^{-1} \sum_{i=1}^{n} (H_i \hat{\beta} - Y_i V_i)(H_i \hat{\beta} - Y_i V_i)'
\]

where \( H_i \) is \( V_i V_i' - \hat{\Omega}_i \) for \( \hat{\beta}_F \) and \( V_i V_i' - \hat{c}\hat{\Omega}_i \) for \( \hat{\beta}_F(\alpha) \). These results can be obtained by considering (3.26) specifically (Fuller 1987) or by appealing to general results for unbiased estimating equations (Carroll et al. 1995; Cheng and Schneeweiss 1998, Cheng et al. 1998).

The residual variance \( \sigma_e^2 \) is consistently estimated by

\[
\hat{\sigma}_e^2 = n^{-1} \sum_{i=1}^{n} Y_i^2 - \hat{M}_{y}' \hat{\beta}
\]

(Cheng and Schneeweiss 1998). For \( \hat{\beta}_F \), \( \sigma_e^2 \) from (3.29) may be negative, in which case \( \hat{\sigma}_e^2 = 0 \) is chosen.

### 4 Estimates of the turning point

Often we are interested in estimates of the turning point \( \eta = -\beta_1/2\beta_2 \). Consistent estimates are given by \( \hat{\eta} = -\hat{\beta}_1/2\hat{\beta}_2 \), where \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) are any of the consistent estimates of \( \beta_1 \) and \( \beta_2 \) discussed above. Estimates \( \hat{\eta}_{rc} = -\hat{\beta}_{rc1}/2\hat{\beta}_{rc2} \) based on the simple regression calibration estimates can also be computed from (2.18) for any values of \( \rho \) using only \( \bar{W} \) and \( \hat{\eta}^* \).

An application of the delta method gives a formula for the asymptotic variance of the estimated turning point. This can be estimated by

\[
\text{vár}(\hat{\eta}) = \frac{\hat{\beta}_1^2}{4\hat{\beta}_2^2} \left[ \text{vár}(\hat{\beta}_1) + \frac{\text{vár}(\hat{\beta}_2)}{\hat{\beta}_1^2} - \frac{2\text{cov}(\hat{\beta}_1, \hat{\beta}_2)}{\hat{\beta}_1\hat{\beta}_2} \right]
\]

(4.30)

where the estimated variances and covariances are obtained as described in Sections 2.4 and 3.

The variance formula (4.30) is appropriate in large samples, but may break down badly in smaller ones. The main reason for this is that estimates of \( \eta \) are less stable and have heavier-tailed distributions than estimates of \( \beta \). Symmetric confidence intervals based on (4.30) may thus be misleading in small samples.

Alternative confidence intervals may be obtained by applying Fieller’s theorem (Fieller 1944), which gives confidence intervals for ratios of linear combinations of parameters. The intervals
are exact if both the numerator and denominator are normal and their finite-sample variances and covariances are known; otherwise we have to appeal to asymptotic normality of $\hat{\beta}$ and the Fieller interval is only approximately correct.

Fieller’s theorem is presented in convenient form for the general linear model by Zerbe (1978) and extended to the MLEs in generalized linear models by Cox (1990). Applied to $\eta$, the theorem states that a $100(1 - \alpha)$ per cent confidence interval for $\eta$ is given by values satisfying

$$A\eta^2 + B\eta + C \leq 0,$$  \hspace{1cm} (4.31)

where

$$A = 4[\hat{\beta}_2^2 - t^2 \text{var}(\hat{\beta}_2)],$$

$$B = 4[\hat{\beta}_1\hat{\beta}_2 - t^2 \text{cov}(\hat{\beta}_1, \hat{\beta}_2)],$$

$$C = \hat{\beta}_2^2 - t^2 \text{var}(\hat{\beta}_1),$$

and $t$ is the $1 - \alpha/2$ percentile of the $t$ distribution with $n - 3$ degrees of freedom. The set satisfying (4.31) is a finite interval only when $A > 0$, which implies that $D = B^2 - 4AC > 0$ (Buonaccorsi 1979). The confidence interval is then $\{(-B - \sqrt{D})/2A, (-B + \sqrt{D})/2A\}$. When $A < 0$, the confidence interval is either the complement of this interval ($C > 0$) or the whole real line ($C < 0$).

It is useful to note that $A > 0$ only when $\hat{\beta}_2$ is significantly different from zero (using $t$ as the critical value). This can serve as a warning sign also for the point estimates of $\eta$ and confidence intervals based on (4.30), which can be computed even when $A < 0$. It is clear that any inference on the turning point is likely to be unreliable if it is uncertain whether a quadratic relationship exists in the first place. In the simulation results of Section 6 we report only results for those simulations where the 95 % confidence interval of $\beta_2$ did not include zero. For the rest of the simulations the results where highly unstable and dominated by very large estimates obtained when $\hat{\beta}_2$ was very small.

A further alternative is to use bootstrap methods to obtain confidence intervals for $\eta$ (Zhou et al. 1993). This approach is not considered here.
5 Illustration: estimating the Kuznets curve

We now illustrate how the two estimation approaches above can be used to conduct a sensitivity analysis. Our application is drawn from development economics. Since the seminal paper of Kuznets (1955), there has been much interest in how income inequality is likely to evolve with the level of economic development. Kuznets proposed that, due to migration from rural agriculture to modern urban areas, the level of inequality would first rise and then fall with the level of development: this is the famous ‘Kuznets curve’.

A variety of approaches have been taken to testing this idea, and to investigating the evolution of income inequality more generally. One simple and prominent method has been to carry out a cross-sectional regression of a summary inequality measure on a quadratic in national income per worker. In this context there is much interest in the location of the turning point. Although this is likely to vary across countries, even just a rough estimate of the level of development at which inequality might start falling would often be helpful.

In the application here, we use our estimation methods to show how sensitive such results are to measurement error in the indicator of development. We also use our findings to demonstrate that, in this context, the share of agricultural employment in total employment is probably a more sensible measure of development to use than income per worker.

In our application, we use two measures of inequality: the Gini coefficient and the variance of log income. Both measures are calculated using the quintile share information in Deininger and Squire (1996) for 1990 or, where necessary, a year either side. We make small adjustments to their figures to bring the results of expenditure surveys in line with income surveys, in a manner similar to that outlined by Deininger and Squire.

The data on GDP per worker are drawn from version 5.6 of the Penn World Table; an earlier version of these data is documented in Summers and Heston (1991). Data on the share of the economically active population in agriculture are taken from the 1995 Production Yearbook of the Food and Agricultural Organization of the United Nations (FAO). Both indicators of development are measured in 1990. When the development data are combined with the inequality measures, we have observations on between 62 and 65 countries depending on the choice of development indicator.
Table 1 shows the parameter estimates and estimates of the turning point for a range of values for \( \rho = \sigma_u^2/\sigma_x^2 \). The estimates given are the expanded regression calibration estimate and the moment-based estimate with \( \alpha = 10 \). The 95% symmetric confidence intervals for the turning point based on the delta method variances (4.30) are also shown (with standard errors of \( \hat{\beta}_{erc} \) obtained from the standard, not sandwich, formula), as are 95% Fieller intervals for cases where these were defined. These choices were motivated by the results of the simulations described in the next section.

The first part of Table 1 shows the estimates for the Gini coefficient given GDP per worker. Here the estimates of both \( \beta \) and \( \eta \) appear to be highly sensitive to even small amounts of measurement error. ERC and moment-based estimates also give very different adjusted estimates for the parameters. The sample mean of GDP per worker is $15840, which is some distance away from the estimates of the turning point. As discussed in Section 2.2, bias in the naive estimates of \( \eta \) is expected to be high under such circumstances. Possibly more importantly, any conclusions about the turning point are uncertain because evidence for a quadratic relationship between GDP and the Gini coefficient is weak. While the linear coefficient is highly significant, confidence intervals for \( \beta_2 \) include zero for all of the estimates. Fieller’s intervals could thus not be computed, and the symmetric intervals are too wide to be meaningful.

The results are more stable when agricultural share of employment is used to predict the Gini coefficient. The estimates of the turning point are now within the central range of the data (the sample mean of the agricultural share of employment is 36 %), and the evidence for a quadratic relationship is fairly strong. In particular, the moment-based estimates of the parameters are insensitive to the amount of measurement error. For \( \rho = 0.4 \) and higher, the estimates are essentially unchanged. It seems that here an increase in the error variance \( \sigma_u^2 \) is roughly balanced by a decrease in the shrinkage coefficient \( c \), leading to the observed stability in the results. The expanded regression calibration estimates are slightly more variable, but they too broadly agree about the location of the turning point. Because the ERC estimate of \( \beta_2 \) is not significantly different from zero for large values of \( \rho \), and only barely significant when \( \rho \) is small, the Fieller intervals for \( \hat{\eta}_{erc} \) are either not defined or give very little information about the upper bound of the confidence interval (to give an idea of the width of the intervals, they are shown without truncating at 100 %). Here we would in fact obtain different results for \( \hat{\eta}_{erc} \) if we used the sandwich estimates of the standard errors of \( \hat{\beta}_{erc} \). Because these are
<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>$\hat{\eta}$</th>
<th>95% c.i. for $\eta$</th>
<th>$\hat{\delta}$</th>
<th>Fieller</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for the Gini coefficient given GDP per worker ($n = 62$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-0.022</td>
<td>-0.011</td>
<td>-1.00</td>
<td>(-31.4, 29.4)</td>
<td>-</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.016</td>
<td>-0.018</td>
<td>4.27</td>
<td>(-15.5, 24.0)</td>
<td>-</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.064</td>
<td>-0.028</td>
<td>11.44</td>
<td>(-21.8, 44.7)</td>
<td>-</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.039</td>
<td>-0.028</td>
<td>7.03</td>
<td>(-7.7, 21.8)</td>
<td>-</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.055</td>
<td>-0.026</td>
<td>10.64</td>
<td>(-27.9, 49.2)</td>
<td>-</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.066</td>
<td>-0.038</td>
<td>8.69</td>
<td>(-3.3, 20.7)</td>
<td>-</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.047</td>
<td>-0.024</td>
<td>9.84</td>
<td>(-34.4, 54.0)</td>
<td>-</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.098</td>
<td>-0.050</td>
<td>9.79</td>
<td>(0.3, 19.7)</td>
<td>-</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.041</td>
<td>-0.023</td>
<td>9.05</td>
<td>(-41.1, 59.2)</td>
<td>-</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.134</td>
<td>-0.064</td>
<td>10.57</td>
<td>(1.7, 19.5)</td>
<td>-</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.035</td>
<td>-0.021</td>
<td>8.28</td>
<td>(-48.0, 64.6)</td>
<td>-</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Model for the Gini coefficient given agricultural share of employment ($n = 65$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.517</td>
<td>-0.463</td>
<td>55.9</td>
<td>(41.5, 70.2)</td>
<td>(46.0, 110)</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.650</td>
<td>-0.615</td>
<td>52.8</td>
<td>(39.6, 66.1)</td>
<td>(43.9, 128)</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.473</td>
<td>-1.564</td>
<td>47.1</td>
<td>(42.2, 52.0)</td>
<td>(43.2, 65.3)</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.797</td>
<td>-0.787</td>
<td>50.6</td>
<td>(38.2, 63.0)</td>
<td>(42.5, 184)</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.719</td>
<td>-1.845</td>
<td>46.6</td>
<td>(41.9, 51.2)</td>
<td>(43.0, 79.4)</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.962</td>
<td>-0.984</td>
<td>48.9</td>
<td>(37.3, 60.5)</td>
<td>(41.4, 464)</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.720</td>
<td>-1.842</td>
<td>46.7</td>
<td>(42.1, 51.3)</td>
<td>(43.1, 82.3)</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>1.146</td>
<td>-1.206</td>
<td>47.5</td>
<td>(36.7, 58.3)</td>
<td>-</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.721</td>
<td>-1.838</td>
<td>46.8</td>
<td>(42.1, 51.4)</td>
<td>(43.2, 85.6)</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.345</td>
<td>-1.454</td>
<td>46.4</td>
<td>(36.3, 56.5)</td>
<td>-</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.721</td>
<td>-1.835</td>
<td>46.9</td>
<td>(42.3, 51.5)</td>
<td>(43.3, 89.2)</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Model for the variance of log income given agricultural share of employment ($n = 65$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.275</td>
<td>-1.306</td>
<td>48.8</td>
<td>(36.7, 60.9)</td>
<td>(38.5, 107)</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>1.660</td>
<td>-1.797</td>
<td>46.2</td>
<td>(35.4, 57.0)</td>
<td>(36.6, 145)</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.528</td>
<td>-3.935</td>
<td>44.8</td>
<td>(40.3, 49.4)</td>
<td>(40.2, 58.7)</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>2.095</td>
<td>-2.361</td>
<td>44.4</td>
<td>(34.6, 54.1)</td>
<td>(35.2, 705)</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.781</td>
<td>-5.393</td>
<td>44.3</td>
<td>(40.2, 48.5)</td>
<td>(39.3, 1646)</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>2.582</td>
<td>-3.003</td>
<td>43.0</td>
<td>(34.1, 51.9)</td>
<td>-</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.788</td>
<td>-5.388</td>
<td>44.4</td>
<td>(40.3, 48.5)</td>
<td>-</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>3.125</td>
<td>-3.728</td>
<td>41.9</td>
<td>(33.7, 50.1)</td>
<td>-</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.793</td>
<td>-5.384</td>
<td>44.5</td>
<td>(40.4, 48.6)</td>
<td>-</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>3.726</td>
<td>-4.536</td>
<td>41.1</td>
<td>(33.5, 48.6)</td>
<td>-</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.798</td>
<td>-5.380</td>
<td>44.6</td>
<td>(40.5, 48.7)</td>
<td>-</td>
<td>0.41</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1:** Estimates for three quadratic regression models for measures of income inequality given indicators of economic development, adjusting for different levels of measurement error. In the model with GDP per worker as predictor, results for $\hat{\eta}$ correspond to thousands of dollars, but $\hat{\beta}_1$ and $\hat{\beta}_2$ to tens of thousands. In the models given agricultural share of employment, all results correspond to percents. See the text for further details.
somewhat smaller than the standard estimates used in Table 1, \( \hat{\beta}_{\text{erc2}} \) is then judged significant for all the values of \( \rho \) in the table. The confidence intervals are also much narrower, ranging from \((46.3, 86.4)\) for \( \rho = 0 \) to \((40.5, 83.2)\) for \( \rho = 1 \). The Fieller intervals for the moment-based estimators are much narrower, and defined for all values of \( \rho \). For both estimators, delta-method confidence intervals are much shorter than the Fieller intervals, a warning sign that the symmetric intervals are likely to be too narrow.

The last part of Table 1 shows the results for models with the variance of log income given agricultural share of employment. The results are broadly similar to those obtained with the Gini index as the response, except that now Fieller intervals are undefined for both estimates for large values of \( \rho \). Although there is in general no reason to expect the turning point to be the same for different inequality measures, in this case it is quite similar between these two specifications, with the estimated turning points consistently slightly higher for the Gini index.

In summary, we draw three tentative findings from these results. First, estimates of both the estimated coefficients and turning point in the Kuznets curve when income is the explanatory variable are highly sensitive to measurement error. Second, the turning point estimates are rather more stable when we use data on the agricultural share of employment, and tend to fall in the 40%-50% range. Third, even when employment data are used, the degree of curvature of the estimated Kuznets curve is more sensitive to measurement error than the turning point.

6 Simulation study

A simulation study was conducted to examine the behaviour of the estimators described above. In the first simulation, \( X_i \) and \( W_i \) were generated from the model (1.2) with \( \mu_x = 2, \sigma_x^2 = 1 \) and \( \sigma_u^2 = \rho \sigma_x^2 \) with \( \rho \) as 0.2, 0.5 and 1.0. Values of \( Y_i \) were generated from a normal distribution implied by (1.1) with parameters \( \sigma^2 = 0.25 \) and \( \beta = (0, 1, -0.5) \). The true turning point was thus \( \eta = 1 \). Results are reported for sample sizes \( n = 60 \) and 200. In each case, 10,000 data sets were generated.

All estimators were computed assuming \( \rho \) to be known. For the expanded regression calibration estimator we report results after two iterations. For the moment-based estimator, results are given for the unmodified estimator \( \hat{\beta}_F \) and the modified estimator \( \hat{\beta}_F(\alpha) \) with \( \alpha = 10 \). To select
<table>
<thead>
<tr>
<th>Estimand</th>
<th>$\rho$</th>
<th>$n = 60$</th>
<th>$n = 200$</th>
<th>$n = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_2$ (bias)</td>
<td></td>
<td>10 7 4</td>
<td>10 10 7</td>
<td>20 20 10</td>
</tr>
<tr>
<td>$\beta_2$ (RMSE)</td>
<td></td>
<td>20 20 10</td>
<td>30 30 20</td>
<td>40 40 40</td>
</tr>
<tr>
<td>$\eta$ (bias)</td>
<td></td>
<td>7 7 0</td>
<td>10 7 0</td>
<td>10 7 0</td>
</tr>
<tr>
<td>$\eta$ (RMSE)</td>
<td></td>
<td>7 7 0</td>
<td>20 10 0</td>
<td>20 20 0</td>
</tr>
<tr>
<td>coverage of c.i. for $\eta$ ($\delta$-method)</td>
<td></td>
<td>30 20 20</td>
<td>30 30 4</td>
<td>60 40 80</td>
</tr>
<tr>
<td>coverage of c.i. for $\eta$ (Fieller’s interval)</td>
<td></td>
<td>0*0 0</td>
<td>0*0 0</td>
<td>7 0 0</td>
</tr>
</tbody>
</table>

Table 2: Values of $\alpha$ which minimize the absolute values of the given quantities for the moment-based estimator in the simulations described in the text. The estimators considered were $\hat{\beta}_F(\alpha)$ with $\alpha$ chosen from $\{0, 4, 7, 10, 20, 30, 40\}$ for $n = 60$ and, in addition, from $\{60, 80, 100\}$ for $n = 200$ and 500, and the unmodified estimator $\hat{\beta}_F$, denoted in the table by *. The results for $\eta$ refer only to those simulations were the estimate of $\beta_2$ was significant at the 95 % level.

A good value for $\alpha$, estimators were first computed for values of $\alpha$ in the set $\{0, 4, 7, 10, 20, 30, 40\}$ for $n = 60$ and, in addition, $\{60, 80, 100\}$ for $n = 200$ and 500. Table 2 reports the values which gave the best results according to various criteria. Clearly, the choice of ‘optimal’ $\alpha$ depends on both the estimand and the quantity we wish to minimise. Best values of $\alpha$ were consistently higher for $\beta$ than for the turning point $\eta$, and higher if the aim was to minimise mean squared error rather than bias. In general, the variance of the estimator decreases with increasing $\alpha$, while the bias reaches its minimum for some finite $\alpha$.

We chose to report results for $\alpha = 10$, because it gave reasonable results for all estimands and criteria. We also considered the case where $\alpha$ was chosen separately in each simulation as the value which gave the smallest estimated root mean squared error (RMSE) of $\beta_2$. This is an example of choosing $\alpha$ to optimise some criterion, something which is not possible in practice without further theoretical results. However, the results were not very different from those for $\hat{\beta}_F(10)$, even for the mean squared errors.

Table 3 shows the simulation results for estimates of $\beta_2$ and the turning point $\eta$. Results for $\beta_1$ were similar to those for $\beta_2$. It can be seen that all of the adjusted estimators reduce or remove the bias of the naive estimator. Both regression calibration estimators are essentially unbiased, with $\hat{\beta}_{erc}$ slightly better than $\hat{\beta}_{rc}$. The modified estimator $\hat{\beta}_F(10)$ shows some bias, especially when $\rho$ is large. The unmodified estimator $\hat{\beta}_F$ suffers badly from occasional extreme
values, especially in small samples and when the measurement error variance is large. The ranking of the estimators is generally the same for median absolute errors (MAE) and RMSEs (not shown here). The effect of extreme values of $\hat{\beta}_F$ in some simulations is shown most clearly by the inflated simulation standard deviations and estimated standard errors of the estimator. For the other estimators the standard deviations are fairly similar. These are consistently underestimated by the estimated standard errors, perhaps mainly for the reasons discussed in Section 2.4. However, the underestimation is in many cases not severe; for example, for $\hat{\beta}_{erc}$ the mean of the estimated standard errors is in the worst case 8 per cent smaller than the standard deviation. For the RC and ERC estimators the standard errors are from the standard variance estimator $n^{-1}\hat{A}_n^{-1}$. We also considered the sandwich standard errors based on (2.24). For $\hat{\beta}_{erc}$ these were worse but for $\hat{\beta}_{rc}$ slightly better than the standard errors reported here. In both cases the sandwich standard errors were substantially more variable than the standard ones.

For the turning point $\eta = -\beta_1/2\beta_2$, results are based on only those simulations where the estimate of the quadratic coefficient $\beta_2$ was significantly different from zero at the 5 % level. As discussed in Section 4, it can be argued that any conclusions about the turning point are questionable if the evidence for a quadratic relationship itself is weak. The numbers of simulations included in the computations vary substantially across estimators and simulations. Many simulations were excluded when the sample size was small or the measurement error large, especially for $\hat{\beta}_F$. The estimates of $\eta$ in the cases left out are naturally mostly large in absolute value, since $\hat{\beta}_2$ is close to zero. In the full set of 10000 simulations the standard deviations are thus larger, in some cases dramatically so, than those reported in Table 3. The difference in means and medians is relatively small in comparison.

The adjusted estimates of the turning point perform well in this restricted set, with the regression calibration estimates again consistently better than the moment-based estimates. The unmodified estimator $\hat{\beta}_F$ is now also reasonable, albeit with only a small minority of the 10,000 simulations included in some cases.

Estimated coverage probabilities of 95 % confidence intervals for $\eta$ are also reported in Table 3. The intervals were computed using the delta method and Fieller’s method as described in Section 4. In most cases the coverage is slightly better for Fieller’s intervals. These are also substantially longer than the intervals based on the delta method, with mean length (not shown
Table 3: Simulation results for estimates of $\beta_2$ and $\eta$ when the true explanatory variable $X$ is normally distributed. Here ‘med.’ denotes median of the simulations, ‘mae’ the median absolute error, ‘s.d.’ the sample standard deviation, ‘s.e.’ the mean of the estimated standard errors, ‘Coverage’ the sample coverage ($\times 100$) of 95% confidence intervals (based on the delta method and Fieller’s method), and ‘No.’ the number of simulations included in the results for $\eta$ (those simulations where the estimate of $\beta_2$ was significant at the 95% level; the results for $\beta_2$ are based on all 10,000 simulations). See the text for further details.
here) up to 8 times (for $\rho = 1$, $n = 60$) and median length up to twice as large. Note also that in this restricted set of simulations, which excludes those cases where the Fieller interval is not a single closed interval, the theoretical coverage probability would not be expected to be exactly 95%.

To examine the robustness of the regression calibration estimates, we repeated the simulation with $X_i$ generated as $X_i \sim a \chi^2_\nu + b$, where $a = \sqrt{\sigma^2_x/(2\nu)}$ and $b = \mu - \nu \sqrt{\sigma^2_x/(2\nu)}$. The density of $X$ thus has the shape of a $\chi^2_\nu$ density but mean $\mu$ and variance $\sigma^2_x$. The measurement model was still (1.2) with normal error, and the parameter values were the same as in the first simulation. For these distributions the linear model (2.11) for $E(X|W)$ is essentially correct, but $\text{var}(X|W)$ depends on $W$. Thus the assumption that $V^2(W)$ is a constant, made in deriving the regression calibration approximation (2.13)–(2.16), is violated.

Table 4 gives results for the case $\nu = 3$. The regression calibration estimates of $\beta_2$ (and $\beta_1$) are now clearly biased. The bias is much smaller for $\hat{\beta}_{erc}$ than for $\hat{\beta}_{rc}$, but neither offers an improvement over the naive estimate. As should be expected, the results for the moment-based estimators are roughly the same as in the first simulation. Overall, $\hat{\beta}_F(10)$ performs best in terms of both bias and median absolute error. Differences between the estimators are smaller for the turning point. The moment-based estimators are again preferred in terms of bias, but now the ERC estimate has a smaller median absolute error.

We also conducted simulations with other parameter combinations which are not reported here. For example, when $\mu = \eta = 1$, the naive estimate of the turning point was unbiased as shown in Section 2.2. It did, however, have a higher MAE and RMSE than the measurement-error adjusted estimates. Furthermore, both naive and adjusted estimates of $\nu$ were more precise than in simulations with $\mu \neq \eta$, due to larger portion of the sample being around the turning point. Simulations with $n = 500$ showed that the occasional extreme values of $\hat{\beta}_F$, which were responsible for its very poor performance in smaller samples, were now largely absent, apart from a small number of cases when $\rho = 1$. Even with this sample size, however, $\hat{\beta}_F$ had a higher MAE and RMSE than the small-sample modification $\hat{\beta}_F(10)$. Finally, we varied the degrees of freedom $\nu$ in the simulations where $X$ had a scaled $\chi^2$ distribution. The main result was that the RC and ERC estimators of $\beta$ (but not of $\eta$) were still typically outperformed by the moment-based estimator even when $\nu = 10$, so that the distribution $X$ was only moderately skewed.
<table>
<thead>
<tr>
<th>n = 60, ρ = 0.2</th>
<th>β2</th>
<th>η</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>med.</td>
<td>mae</td>
<td>s.d.</td>
</tr>
<tr>
<td>β*</td>
<td>-0.446</td>
<td>0.082</td>
<td>0.112</td>
</tr>
<tr>
<td>βrc</td>
<td>-0.643</td>
<td>0.155</td>
<td>0.161</td>
</tr>
<tr>
<td>βerc</td>
<td>-0.605</td>
<td>0.118</td>
<td>0.128</td>
</tr>
<tr>
<td>βF</td>
<td>-0.539</td>
<td>0.120</td>
<td>0.334</td>
</tr>
<tr>
<td>βF(10)</td>
<td>-0.516</td>
<td>0.100</td>
<td>0.163</td>
</tr>
<tr>
<td>n = 60, ρ = 0.5</td>
<td>β*</td>
<td>η</td>
<td>Coverage</td>
</tr>
<tr>
<td></td>
<td>med.</td>
<td>mae</td>
<td>s.d.</td>
</tr>
<tr>
<td>β*</td>
<td>-0.360</td>
<td>0.147</td>
<td>0.130</td>
</tr>
<tr>
<td>βrc</td>
<td>-0.810</td>
<td>0.316</td>
<td>0.292</td>
</tr>
<tr>
<td>βerc</td>
<td>-0.692</td>
<td>0.203</td>
<td>0.203</td>
</tr>
<tr>
<td>βF</td>
<td>-0.605</td>
<td>0.118</td>
<td>0.128</td>
</tr>
<tr>
<td>βF(10)</td>
<td>-0.521</td>
<td>0.143</td>
<td>0.228</td>
</tr>
<tr>
<td>n = 60, ρ = 1.0</td>
<td>β*</td>
<td>η</td>
<td>Coverage</td>
</tr>
<tr>
<td></td>
<td>med.</td>
<td>mae</td>
<td>s.d.</td>
</tr>
<tr>
<td>β*</td>
<td>-0.252</td>
<td>0.249</td>
<td>0.131</td>
</tr>
<tr>
<td>βrc</td>
<td>-1.006</td>
<td>0.515</td>
<td>0.524</td>
</tr>
<tr>
<td>βerc</td>
<td>-0.806</td>
<td>0.325</td>
<td>0.335</td>
</tr>
<tr>
<td>βF</td>
<td>-0.363</td>
<td>0.557</td>
<td>6938</td>
</tr>
<tr>
<td>βF(10)</td>
<td>-0.462</td>
<td>0.155</td>
<td>0.092</td>
</tr>
<tr>
<td>n = 200, ρ = 0.2</td>
<td>β*</td>
<td>η</td>
<td>Coverage</td>
</tr>
<tr>
<td></td>
<td>med.</td>
<td>mae</td>
<td>s.d.</td>
</tr>
<tr>
<td>β*</td>
<td>-0.455</td>
<td>0.055</td>
<td>0.064</td>
</tr>
<tr>
<td>βrc</td>
<td>-0.655</td>
<td>0.155</td>
<td>0.092</td>
</tr>
<tr>
<td>βerc</td>
<td>-0.606</td>
<td>0.106</td>
<td>0.065</td>
</tr>
<tr>
<td>βF</td>
<td>-0.512</td>
<td>0.061</td>
<td>0.095</td>
</tr>
<tr>
<td>βF(10)</td>
<td>-0.508</td>
<td>0.059</td>
<td>0.091</td>
</tr>
<tr>
<td>n = 200, ρ = 0.5</td>
<td>β*</td>
<td>η</td>
<td>Coverage</td>
</tr>
<tr>
<td></td>
<td>med.</td>
<td>mae</td>
<td>s.d.</td>
</tr>
<tr>
<td>β*</td>
<td>-0.379</td>
<td>0.122</td>
<td>0.079</td>
</tr>
<tr>
<td>βrc</td>
<td>-0.853</td>
<td>0.353</td>
<td>0.178</td>
</tr>
<tr>
<td>βerc</td>
<td>-0.696</td>
<td>0.106</td>
<td>0.107</td>
</tr>
<tr>
<td>βF</td>
<td>-0.540</td>
<td>0.118</td>
<td>0.630</td>
</tr>
<tr>
<td>βF(10)</td>
<td>-0.521</td>
<td>0.101</td>
<td>0.161</td>
</tr>
<tr>
<td>n = 200, ρ = 1.0</td>
<td>β*</td>
<td>η</td>
<td>Coverage</td>
</tr>
<tr>
<td></td>
<td>med.</td>
<td>mae</td>
<td>s.d.</td>
</tr>
<tr>
<td>β*</td>
<td>-0.272</td>
<td>0.228</td>
<td>0.081</td>
</tr>
<tr>
<td>βrc</td>
<td>-1.090</td>
<td>0.590</td>
<td>0.323</td>
</tr>
<tr>
<td>βerc</td>
<td>-0.803</td>
<td>0.303</td>
<td>0.173</td>
</tr>
<tr>
<td>βF</td>
<td>-0.561</td>
<td>0.246</td>
<td>6851</td>
</tr>
<tr>
<td>βF(10)</td>
<td>-0.520</td>
<td>0.145</td>
<td>0.224</td>
</tr>
</tbody>
</table>

Table 4: Simulation results for estimates of $\beta_2$ and $\eta$ when the true explanatory variable $X$ is distributed as a scaled $\chi^2_3$ variate. See the text and caption of Table 3 for further details.
The computations were carried out as described in previous sections. The naive estimator and the simple RC estimator are trivial to compute, and \( \hat{\beta}_F \) and \( \hat{\beta}_F(\alpha) \) only slightly more complicated. The expanded regression calibration estimator requires the iterative algorithm outlined in Section 2.3. This was unproblematic in most simulations, but there were a few exceptions. Occasionally the algorithm failed to converge when the starting value for \( \sigma^2_e \) was given by (2.23); in such cases convergence was usually achieved by restarting the algorithm with \( \sigma^2_e \) initialised at a small value (here \( \sigma^2_e(0) = 0.1 \) was used). The same approach was used when (2.23) yielded a negative starting value.

In a small number of simulations the algorithm failed to converge from any initial values. This was most common when \( n \) was small and \( \rho \) large, with a maximum of 183 failures out of 10,000 when \( n = 60 \) and \( \rho = 1 \). These cases were recomputed with the Newton–Raphson algorithm to solve the equations (2.21) and (2.22) simultaneously. When both converge, full Newton–Raphson yields the same estimates as iterating the two-step procedure of Section 2.3 to convergence, but it is typically slower and requires more iterations than the two-step algorithm. In the problematic cases Newton–Raphson ‘converged’ to values of \( \sigma^2_e \) of the order \( 10^9 \), indicating that there is no meaningful solution for \( \sigma^2_e \) to the estimating equations (2.21) – (2.22); clearly the equations will be satisfied as \( \sigma^2_e \) tends to infinity. The estimated values of \( \beta \) were then generally very close to the simple regression calibration estimates, a result which can be explained by noting that the weights \( g^2(m_i; \beta, \sigma^2_e) \) are effectively constant when \( \sigma^2_e \) is very large.

7 Conclusions

We have considered two families of measurement error adjustments for quadratic regression models. Within each class, some clear recommendations emerged from our simulations. For regression calibration, the estimator based on the expanded approximation performed consistently better than the simpler RC estimator. The difference was slight when the assumptions of the estimators were satisfied, but substantial when they did not hold. In such cases the ERC estimator was somewhat less sensitive to failures of the assumptions. The main advantage of the simple RC estimator is that it is very flexible and easy to compute, but its apparent lack of robustness is cause for concern.
The basic moment-based estimator \( \hat{\beta}_F \) performed poorly in small and moderate samples, where it has a very heavy-tailed distribution. This meant that the estimator was unstable and its values frequently very large. A small-sample modification corrected this problem in small samples, and also outperformed the unmodified estimator in larger ones. Since the two estimators are also asymptotically equivalent, it seems preferable to use the modified estimator in all cases. A minor complication in its use is that it requires the user-specified tuning constant \( \alpha \), for which few clear guidelines exist. In our simulations the precise choice of the constant was not critical, since most small values of it led to fairly similar results. Nevertheless, further theoretical results on the choice of \( \alpha \) would be useful.

A key difference between regression calibration and moment-based approaches is that the latter makes no assumptions about the distribution of the true predictors \( X \), whereas regression calibration specifies a model for \( X \) given its surrogate \( W \), and, by implication, assumes something about \( X \). When these conditions are satisfied, regression calibration estimates are a more efficient choice, because they utilize more information on the variables. When the assumptions fail, however, regression calibration may give biased results, while moment-based estimators are essentially unaffected by changes in the distribution of \( X \). This is an attractive property, since we are typically not interested in \( X \) and \( W \), only on the model for \( Y \) given \( X \). We would thus want to minimise the extent to which conclusions about this primary feature of the model can be affected by assumptions about a secondary aspect such as the assumed distribution of the true predictor. The observed degree of the sensitivity of RC estimates was in fact somewhat surprising, given the generally good performance of these approximations in a wide range of models.

Here we have considered a classical measurement error model, where the error is additive and unbiased. Combined with normality of \( X \), this implies a homoscedastic linear model for \( X \) given \( W \). As discussed in Section 6, it is the assumption of homoscedasticity in this model that fails in the cases where \( X \) has a skewed distribution, leading to biases in the regression calibration estimates. An obvious remedy would be to allow for heteroscedasticity in the model for \( X \), or, more generally, for some other type of dependence of \( X \) on \( W \) that may occur in practice. In principle this poses no new issues for regression calibration, which is immediately defined for the general case as shown in equations (2.9) and (2.10). The moment-based estimators are somewhat more difficult to generalise to other measurement models, but in special cases this
should also be possible.

In practice the range of measurement models that can realistically be considered depends on the available auxiliary information. In this paper we have concentrated on sensitivity analyses where the effects of measurement error are explored by assuming the error parameters to be known. The classical measurement error model (1.2) is an almost universal default model in such studies, if for no other reason than that it has only one adjustable parameter. In a sensitivity analysis we are of course free to try any measurement model we like; for example, a multiplicative error model which could be used as an analogous default model has been described by Iturria et al. (1999). In practice, however, there is a limit to the number of models that it is feasible or informative to consider. The same is to a large extent true when repeated measurements data are available, because identifiability constraints then typically dictate that at least some of the measurement models have to be additive and unbiased. Because $X$ is not observed, any measurement model will also inevitably involve some unverifiable assumptions.

Only validation data thus give a proper basis for flexible modelling of the measurement error and assessment of any assumptions made about the error model. Since $X$ is observed in the validation data set, modelling it given other observed variables becomes a straightforward task. This provides a powerful reason for trying to obtain validation data in that minority of measurement error problems where it is feasible. It should also be recognised that such data should be drawn from a population where the distributions of both $X$ and the measurement error are the same as in the main study population. Otherwise conditional models for $X$ given $W$ will not be transferable across data sets. The best way to achieve this is to draw the validation data set as a subsample from the main study data.

Some other approaches are also available for adjusting quadratic regression models for measurement error. One interesting possibility is the simulation–extrapolation (SIMEX) algorithm of Cook and Stefanski (1994), which also works without explicit assumptions about the distribution of $X$. Its use for quadratic models is outlined by Carroll et al. (1995). Küchenhoff and Carroll (1995) have applied SIMEX estimation to estimate the changepoint in segmented linear regression, a problem which is somewhat similar to (although harder than) that of estimating the turning point of quadratic regression.

In a fully likelihood-based analysis all of the distributions involved, including that of $X$, would
be specified parametrically. Maximum likelihood estimates of the parameters could then be computed, using numerical methods if necessary. For reasons discussed above, this approach would be sensible only if validation data were available. In such cases, however, the expanded regression calibration approximation provides the correct quasilikelihood model for the data, so the gain from a fully parametric analysis would seem to be negligible.

In order to illustrate the main results, we have concentrated on a simple quadratic model involving only one explanatory variable. There are two kinds of multivariate extensions of this case. First, there may be further explanatory variables which are measured without error. Accommodating such extra predictors in the regression calibration estimators is straightforward: we simply use moments of $X$ given both $W$ and the additional predictors throughout. The moment-based estimators are also essentially unchanged, since they rely on unbiased estimates of powers of $X$. These are unaffected by the presence of additional covariates.

The model may also contain several covariates measured with error, which may enter as linear or quadratic terms (or in some more general form, but that is beyond the scope of this paper). Of the estimators considered here, the simple regression calibration estimator extends immediately to such cases. The extension of the moment-based estimator is also in principle straightforward. The expanded regression calibration estimator, as noted in Section 2, becomes algebraically very complex when more than one variable is measured with error.

We have given special attention to estimates of the turning point in quadratic regression. Some of the findings differ from typical results for the regression coefficients. It was shown that measurement error may induce little or no bias for the observed turning point. The key condition for this to happen is that the data should be obtained from a population where the mean of $X$ is close to the true turning point. In practice it is at least necessary to obtain data from the whole range of $X$, so that there are enough data points on both sides of the turning point. This design also provides most information for estimates of the turning point, both unadjusted and adjusted for measurement error.
References


