CHAPTER 3

LINEAR REGRESSION AND ATTENUATION

3.1 Introduction

This chapter summarizes some of the known results about the effects of measurement error in linear regression and describes some of the statistical methods used to correct for those effects. Our discussion of the linear model is intended only to set the stage for our main topic, nonlinear measurement error models, and is far from complete. A comprehensive account of linear measurement error models can be found in Fuller (1987).

3.2 Bias Caused by Measurement Error

Many textbooks contain a brief description of measurement error in linear regression, usually focusing on simple linear regression and arriving at the conclusion that the effect of measurement error is to bias the slope estimate in the direction of zero. Bias of this nature is commonly referred to as *attenuation* or *attenuation to the null*.

In fact, though, even this simple conclusion must be qualified, because it depends on the relationship between the measurement, $\mathbf{W}$, and the true predictor, $\mathbf{X}$, and possibly other variables in the regression model as well. In particular, the effect of measurement error depends on the model under consideration and on the joint distribution of the measurement error and the other variables. In linear regression, the effects of measurement error vary depending on (i) the regression model, be it simple or multiple regression; (ii) whether or not the predictor measured with error is univariate or multivariate; and (iii) the presence of bias in the measurement. The effects can range from the simple attenuation described above to situations where (a) real effects are hidden; (b) observed data exhibit relationships that are not present in the error-free data; and (c) even the signs ($\pm$) of estimated coefficients are reversed relative to the case with no measurement error.

The key point is that the measurement error distribution determines the effects of measurement error, and thus appropriate methods for cor-

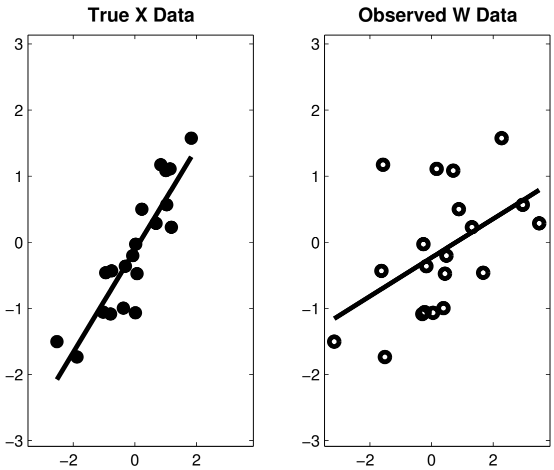

Figure 3.1 *Illustration of additive measurement error model. The left panel displays the true* $(\mathbf{Y}, \mathbf{X})$ *data, while the right panel displays the observed* $(\mathbf{Y}, \mathbf{W})$ *data. Note how the true* $\mathbf{X}$ *data plot has less variability and a more obvious nonzero effect.*

recting for the effects of measurement error depend on the measurement error distribution.

3.2.1 Simple Linear Regression with Additive Error

The basic effects of classical measurement error on simple linear regression can be seen in Figures 3.1 and 3.2. These effects are the double whammy of measurement error described in Section 1.1, namely loss of power when testing and bias in parameter estimation. The third whammy, masking of features, occurs only in nonlinear models, since obviously a straight line has no features to mask.

The left panel of Figure 3.1 displays error-free data $(\mathbf{Y}, \mathbf{X})$ generated from the linear regression model $\mathbf{Y} = \beta_0 + \beta_x \mathbf{X} + \epsilon$, where $\mathbf{X}$ has mean $\mu_x = 0$ and variance $\sigma_x^2 = 1$, the intercept is $\beta_0 = 0$, the slope is $\beta_x = 1$, and the error about the regression line ϵ is independent of $\mathbf{X}$, has mean zero and variance $\sigma_\epsilon^2 = 0.25$. The right panel displays the error-contaminated data $(\mathbf{Y}, \mathbf{W})$ where $\mathbf{W} = \mathbf{X} + \mathbf{U}$, and $\mathbf{U}$ is independent of $\mathbf{X}$, has mean zero, and variance $\sigma_u^2 = 1$. This is the classical additive measurement error model; see Section 1.2. Note how the $(\mathbf{Y}, \mathbf{X})$ data are more tightly grouped around a well delineated line, while the error-prone

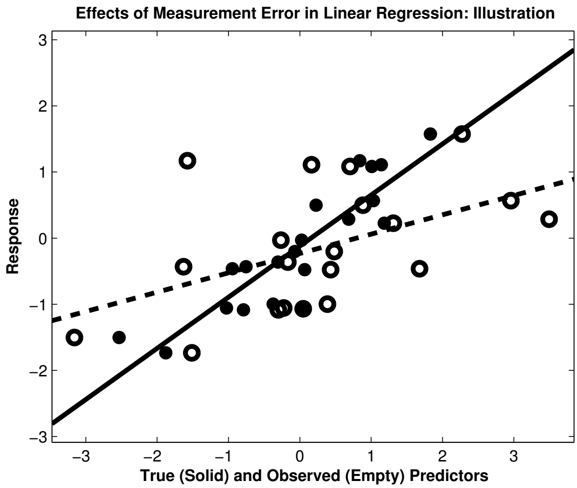

Figure 3.2 *Illustration of additive measurement error model. Here we combine the data in Figure 3.1 and add in least squares fitted lines: The solid line and solid circles are for the true* $\mathbf{X}$ *data, while the dashed line and empty circles are for the observed, error-prone* $\mathbf{W}$ *data. Note how the slope to the true* $\mathbf{X}$ *data is steeper, and the variability about the line is much smaller.*

$(\mathbf{Y}, \mathbf{W})$ data have much more variability about a much less obvious line. This is the loss of power through additional variability.

In Figure 3.2 we combine the data sets: The solid circles and solid line are the $(\mathbf{Y}, \mathbf{X})$ data and least squares fit, while the empty circles and dashed line are the $(\mathbf{Y}, \mathbf{W})$ data and their least squares fit. Here we see the bias in the least squares line due to classical measurement error.

We can understand the phenomena in Figures 3.1–3.2 through some theoretical calculations. For example, it is well known that an ordinary least squares regression of $\mathbf{Y}$ on $\mathbf{W}$ is a consistent estimate not of β_x, but instead of $\beta_{x*} = \lambda\beta_x$, where

$$\lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} < 1. \tag{3.1}$$

Thus ordinary least squares regression of $\mathbf{Y}$ on $\mathbf{W}$ produces an estimator that is attenuated to zero. The attenuating factor, λ, is called the *reliability ratio* (Fuller, 1987). This attenuation is particularly pronounced in Figures 3.1–3.2.

One would expect that because $\mathbf{W}$ is an error-prone predictor, it has a weaker relationship with the response than does $\mathbf{X}$, as seen in Figure 3.1. This can be seen both by the attenuation and also by the fact that the residual variance of this regression of $\mathbf{Y}$ on $\mathbf{W}$ is

$$\mathrm{var}(\mathbf{Y}|\mathbf{W}) = \sigma_\epsilon^2 + \frac{\beta_x^2 \sigma_u^2 \sigma_x^2}{\sigma_x^2 + \sigma_u^2} = \sigma_\epsilon^2 + \lambda \beta_x^2 \sigma_u^2. \tag{3.2}$$

This facet of the problem is often ignored, but it is important. Measurement error causes a double whammy: Not only is the slope attenuated, but the data are more noisy, with an increased error about the line.

It is not surprising that measurement error, as another source of error, increases variability about the line. Indeed, we can substitute $\mathbf{X} = \mathbf{W} - \mathbf{U}$ into the regression model to obtain the model $\mathbf{Y} = \beta_0 + \beta_x \mathbf{W} + (\epsilon - \beta_x \mathbf{U})$, with error $(\epsilon - \beta_x \mathbf{U})$ that has variance $\sigma_\epsilon^2 + \beta_x^2 \sigma_u^2 > \sigma_\epsilon^2$ and covariate $\mathbf{W}$. What may be surprising is that this additional error causes bias. However, the error and the covariate have a common component $\mathbf{U}$, which causes them to be correlated. The correlation between the error and covariate is the source of the bias.

In light of the effects of classical measurement error discussed above, one might expect that the least squares estimate of slope calculated from measured $(\mathbf{Y}, \mathbf{W})$ is more variable than the slope estimator calculated from the true $(\mathbf{Y}, \mathbf{X})$ data. This is not always the case. Buzas, Stefanski, and Tosteson (2004) pointed out that the naive estimate of slope can be *less variable* than the true data estimator. In fact, for the classical error model, the variance of the naive estimator is less than the variance of the true-data estimator asymptotically if and only if $\beta_x^2 \sigma_x^2 / (\sigma_x^2 + \sigma_u^2) < \sigma_\epsilon^2 / \sigma_x^2$, which is possible when σ_ϵ^2 is large, or σ_u^2 is large, or β_x^2 is small. So, relative to the case of no measurement error, classical errors can result in more precise estimates of the wrong, that is, biased, quantity. This phenomenon explains, in part, why naive-analysis confidence intervals often have disastrous coverage probabilities; not only are they centered on the wrong value, but they sometimes have shorter length than would be obtained with the true data. This phenomenon cannot occur with Berkson errors, for which the variance of the naive estimator is never less than the variance of the true-data estimator asymptotically.

3.2.2 Regression Calibration: Classical Error as Berkson Error

There is another way of looking at the bias that will give further insight, namely that by a simple mapping, classical measurement error can be made into a Berkson model. Define $\mathbf{W}_{\mathrm{blp}} = (1 - \lambda)\mu_x + \lambda \mathbf{W}$, the best linear predictor of $\mathbf{X}$ based on $\mathbf{W}$. Then, by (A.8) of Appendix A,

$$\mathbf{X} = \mathbf{W}_{\mathrm{blp}} + \mathbf{U}^*, \tag{3.3}$$

where $\mathbf{U}^*$ is uncorrelated with $\mathbf{W}$, and var$U^* = \lambda\sigma_u^2$. Compare (3.3) with the formal definition of a Berkson error model (1.2) in Section 1.4. Effectively, we have a formal transformation of the classical error model into a Berkson error model, where the observed predictor is now the best linear predictor of $\mathbf{X}$ from $\mathbf{W}$. The calculation leading to (3.3) is at the heart of the regression calibration method of Chapter 4.

Equation (3.3) has important consequences in fitting the linear regression model and correction for the bias due to classical measurement error: Little (generally nothing) can be done to eliminate the loss of power. Substituting (3.3) for $\mathbf{X}$ into the regression model, we have

$$\begin{aligned} \mathbf{Y} &= \beta_0 + \beta_x(1-\lambda)\mu_x + \beta_x\lambda\mathbf{W} + (\epsilon + \beta_x\mathbf{U}^*) \\ &= \beta_0 + \beta_x\mathbf{W}_{\text{blp}} + \epsilon + \beta_x\mathbf{U}^*. \end{aligned} \tag{3.4}$$

In (3.4) the error $\epsilon + \beta_x\mathbf{U}^*$ is uncorrelated with the regressor $\mathbf{W}_{\text{blp}}$ and has variance $\sigma_\epsilon^2 + \lambda\beta_x^2\sigma_u^2$ in agreement with (3.2). Moreover, the regression of $\mathbf{Y}$ on $\mathbf{W}$ has intercept $\beta_0 + \beta_x(1-\lambda)\mu_x$ and slope $\lambda\beta_x$, which explains the attenuation of the slope and the additive bias of the intercept.

However, these considerations show a way to eliminate bias. By (3.4), we have $\mathbf{Y} = \beta_0 + \beta_x\mathbf{W}_{\text{blp}} + \epsilon + \beta_x\mathbf{U}^*$, so if we replace the unknown $\mathbf{X}$ by $\mathbf{W}_{\text{blp}}$, which is known since it depends only on $\mathbf{W}$, then we have a regression model with intercept equal to β_0, slope equal to β_x, and error uncorrelated with the regressor. Therefore, regressing $\mathbf{Y}$ on $\mathbf{W}_{\text{blp}}$ gives unbiased estimates of β_0 and β_x. In fact, regressing $\mathbf{Y}$ on $\mathbf{W}_{\text{blp}}$ is equivalent to the method-of-moments correction for attenuation discussed in Section 3.4.1. Replacing $\mathbf{X}$ with its predictor $\mathbf{W}_{\text{blp}}$ is the key idea behind the technique of regression calibration discussed in Chapter 4. Of course, $\mathbf{W}_{\text{blp}}$ is "known" only if we know λ and μ_x. In practice, these parameters need to be estimated.

3.2.3 Simple Linear Regression with Berkson Error

Suppose that we have linear regression, $\mathbf{Y}_i = \beta_0 + \beta_x\mathbf{X}_i + \epsilon_i$, with unbiased Berkson error, that is, $\mathbf{X}_i = \mathbf{W}_i + \mathbf{U}_i$. Then $E(\mathbf{X}_i|\mathbf{W}_i) = \mathbf{W}_i$ so that $E(\mathbf{Y}_i|\mathbf{W}_i) = \beta_0 + \beta_x\mathbf{W}_i$. As a consequence, the naive estimator that regresses $\mathbf{Y}_i$ on $\mathbf{W}_i$ is unbiased for β_0 and β_x. This unbiasedness can be seen in Figure 3.3 which illustrates linear regression with Berkson errors. In the figure, $(\mathbf{Y}_i, \mathbf{X}_i)$ and $(\mathbf{Y}_i, \mathbf{W}_i)$ are plotted, as well as fits to both $(\mathbf{Y}_i, \mathbf{X}_i)$ and $(\mathbf{Y}_i, \mathbf{W}_i)$. The $\mathbf{W}_i$ are equally spaced on $[-1, 3]$, $\mathbf{X}_i = \mathbf{W}_i + \mathbf{U}_i$, $\mathbf{U}_i = \text{Normal}(0,1)$, $\epsilon_i = \text{Normal}(0, 0.5)$, $n = 50$, $\beta_0 = 1$, and $\beta_x = 1$.

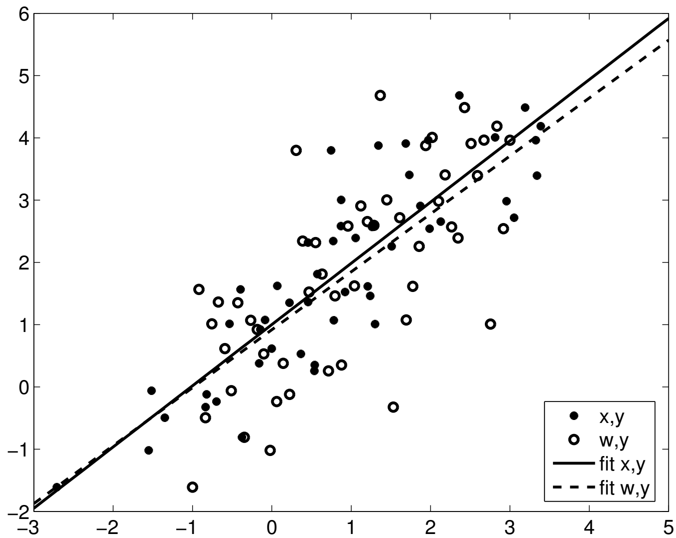

Figure 3.3 *Simple linear regression with unbiased Berkson errors. Theory shows that the fit of* $\mathbf{Y}_i$ *to* $\mathbf{W}_i$ *is unbiased for the regression of* $\mathbf{Y}_i$ *on* $\mathbf{X}_i$*, and the two fits are, in fact, similar.*

3.2.4 Simple Linear Regression, More Complex Error Structure

Despite admonitions of Fuller (1987) and others to the contrary, it is a common perception that the effect of measurement error is always to attenuate the line. In fact, attenuation depends critically on the classical additive measurement error model. In this section, we discuss two deviations from the classical additive error model that do not lead to attenuation.

We continue with the simple linear regression model, but now we make the error structure more complex in two ways. First, we will no longer insist that $\mathbf{W}$ be unbiased for $\mathbf{X}$. The intent of studying this departure from the classical additive error model is to study what happens when one pretends that one has an unbiased surrogate, but in fact the surrogate is biased.

A second departure from the additive model is to allow the errors in the linear regression model to be correlated with the errors in the predictors. This is differential measurement error; see Section 2.5. One example where this problem arises naturally is in dietary calibration studies (Freedman et al., 1991). In a typical dietary calibration study, one is interested in the relationship between a self-administered food frequency questionnaire (FFQ, the value of $\mathbf{Y}$) and usual (or long-term) dietary intake (the value of $\mathbf{X}$) as measures of, for example, the percentage of calories from fat in a person's diet. FFQs are thought to be biased for usual intake, and in a calibration study researchers will obtain a second measure (the value of $\mathbf{W}$), typically either from a food diary or from an interview in which the study subject reports his or her diet in the previous 24 hours. In this context, it is often assumed that the diary or recall is unbiased for usual intake. In principle, then, we have simple linear regression with an additive measurement error model, but in practice a complication can arise. It is often the case that the FFQ and the diary or recall are given very nearly contemporaneously in time, as in the Women's Health Trial Vanguard Study (Henderson et al., 1990). In this case, it makes little sense to pretend that the error in the relationship between the FFQ ($\mathbf{Y}$) and usual intake ($\mathbf{X}$) is uncorrelated with the error in the relationship between a diary or recall ($\mathbf{W}$) and usual intake. This correlation has been demonstrated (Freedman, Carroll, and Wax, 1991), and in this section we will discuss its effects.

To express the possibility of bias in $\mathbf{W}$, we write the model as $\mathbf{W} = \gamma_0 + \gamma_1 \mathbf{X} + \mathbf{U}$, where $\mathbf{U}$ is independent of $\mathbf{X}$ and has mean zero and variance σ_u^2. To express the possibility of correlated errors, we will write the correlation between ϵ and $\mathbf{U}$ as $\rho_{\epsilon u}$. The classical additive measurement error model sets $\gamma_0 = 0$, $\rho_{\epsilon u} = 0$, and $\gamma_1 = 1$, so that $\mathbf{W} = \mathbf{X} + \mathbf{U}$.

If $(\mathbf{X}, \epsilon, \mathbf{U})$ are jointly normally distributed, then the regression of $\mathbf{Y}$ on $\mathbf{W}$ is linear with intercept

$$\beta_{0*} = \beta_0 + \beta_x \mu_x - \beta_{x*}(\gamma_0 + \gamma_1 \mu_x),$$

and slope

$$\beta_{x*} = \frac{\beta_x \gamma_1 \sigma_x^2 + \rho_{\epsilon u}\sqrt{\sigma_\epsilon^2 \sigma_u^2}}{\gamma_1^2 \sigma_x^2 + \sigma_u^2}. \tag{3.5}$$

Examination of (3.5), shows that if $\mathbf{W}$ is biased ($\gamma_1 \neq 1$) or if there is significant correlation between the measurement error and the error about the true line ($\rho_{\epsilon u} \neq 0$), it is possible for $|\beta_{x*}| > |\beta_x|$, an effect exactly the opposite of attenuation. Thus, correction for bias induced by measurement error clearly depends on the nature, as well as the extent, of the measurement error.

For purposes of completeness, we note that the residual variance of the linear regression of $\mathbf{Y}$ on $\mathbf{W}$ is

$$\mathrm{var}(\mathbf{Y}|\mathbf{W}) = \sigma_\epsilon^2 + \frac{\beta_x^2 \sigma_u^2 \sigma_x^2 - \rho_{\epsilon u}^2 \sigma_\epsilon^2 \sigma_u^2 - 2\beta_x \gamma_1 \sigma_x^2 \rho_{\epsilon u}\sqrt{\sigma_\epsilon^2 \sigma_u^2}}{\gamma_1^2 \sigma_x^2 + \sigma_u^2}.$$

3.2.4.1 Diagnostic for Correlation of Errors in Regression and Measurement Errors

In some cases, there is a simple graphical diagnostic to check whether the errors in the regression are correlated with the classical measurement errors. The methods are related to the graphical diagnostics used to detect whether the additive error model is reasonable; see Section 1.7.

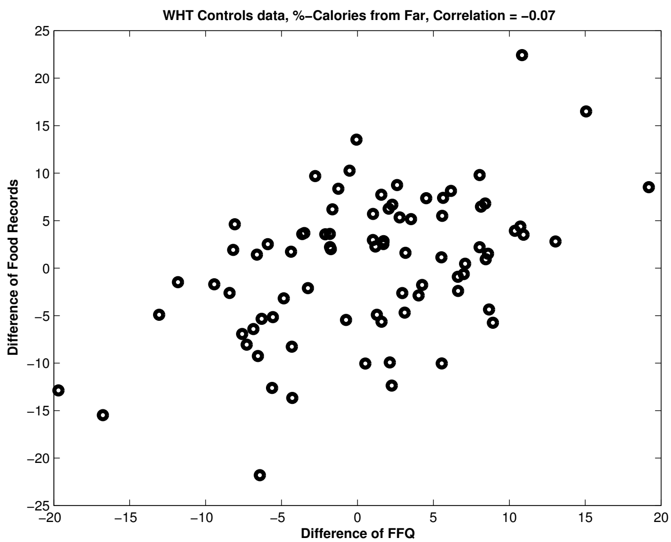

Figure 3.4 *Women's Health Trial Vanguard Study Data. This is a plot for % calories from fat of the differences of food records and the differences of food frequency questionnaires. With replicated* $\mathbf{Y}$ *and* $\mathbf{W}$*, this plot is a diagnostic for whether errors in a regression are correlated with the classical measurement errors.*

Specifically, suppose that the error-prone instrument is replicated, so that we observe $\mathbf{W}_{ij} = \gamma_0 + \gamma_1 \mathbf{X}_i + \mathbf{U}_{ij}$. The difference $\mathbf{W}_{i1} - \mathbf{W}_{i2} = \mathbf{U}_{i1} - \mathbf{U}_{i2}$ is "pure" error, unrelated to $\mathbf{X}_i$. Suppose further that the response is replicated, so that we observe $\mathbf{Y}_{ij} = \beta_0 + \beta_x \mathbf{X}_i + r_i + \epsilon_{ij}$, where r_i is person-specific bias or equation error; see Section 1.5. Then differences $\mathbf{Y}_{i1} - \mathbf{Y}_{i2} = \epsilon_{i1} - \epsilon_{i2}$ are the model errors. A plot of the two sets of differences will help reveal whether the regression errors and the measurement errors are correlated. This is illustrated in Figure 3.4, where there appears to be a very strong correlation between the model errors and the measurement errors. A formal test can be performed by regressing one set of differences on the other and testing the null hypothesis that the slope is zero. This plotting method and the test assume that the covariances of errors separated in time are small. This assumption seems reasonable if the time separation is at all large.

3.2.5 Summary of Simple Linear Regression

Before continuing with a discussion of the effects of measurement error in multiple linear regression, we summarize the primary effects of measurement error in simple linear regression for various types of error models that we study throughout the book. Table 3.1 displays the important error-model parameters and linear regression model parameters for the case that $(\mathbf{Y}, \mathbf{X}, \mathbf{W})$ are multivariate normal for a hierarchy of error model types. In all cases, the underlying regression model is

$$\mathbf{Y} = \beta_0 + \beta_x \mathbf{X} + \epsilon, \tag{3.6}$$

where $\mathbf{X}$ and ϵ are independent and ϵ has mean zero and variance σ^2.

3.2.5.1 Differential Error Measurement

The least restrictive type of error model is one in which $\mathbf{W}$ is not unbiased and the error is differential. This is also the most troublesome type of error in the sense that correcting for bias requires the most additional information or data. The first row in Table 3.1 shows how the parameters in the regression of $\mathbf{Y}$ on $\mathbf{W}$ depend on the true-data regression model parameters, $\beta_0,\ \beta_x,\ \sigma^2$, in this case. Note that to recover β_x from the regression of $\mathbf{Y}$ on $\mathbf{W}$ one would have to know or be able to estimate the covariances, σ_{xw} and $\sigma_{\epsilon w}$. Also, with a differential-error measurement it is possible for the residual variance in the regression of $\mathbf{Y}$ on $\mathbf{W}$ to be *less than* σ^2.

3.2.5.2 Surrogate Measurement

As defined in Section, 2.5, a surrogate measurement is one for which the conditional distribution of $\mathbf{Y}$ given $(\mathbf{X}, \mathbf{Z}, \mathbf{W})$ depends only on $(\mathbf{X}, \mathbf{Z})$. In this case, $\mathbf{W}$ is also said to be a *surrogate*. The second row of Table 3.1 shows how the parameters in the regression of $\mathbf{Y}$ on $\mathbf{W}$ depend on $\beta_0,\ \beta_x,\ \sigma^2$ when $\mathbf{W}$ is a surrogate, with no additional assumptions about the type of error model. With a surrogate, it is apparent that knowledge of or estimability of σ_{xw} is enough to recover β_x from the regression of $\mathbf{Y}$ on $\mathbf{W}$. The residual variance in the regression of $\mathbf{Y}$ on $\mathbf{W}$ is *always greater than* σ^2 when $\mathbf{W}$ is a surrogate. In this sense, a surrogate is always less informative than $\mathbf{X}$.

Error Model	ρ^2_{xw}	Intercept	Slope	Residual Variance
Differential	ρ^2_{xw}	$\beta_0 + \beta_x\mu_x - \frac{\beta_x\sigma_{xw}+\sigma_{\epsilon w}}{\sigma^2_w}\mu_w$	$\beta_x\left(\frac{\sigma_{xw}}{\sigma^2_w}\right) + \frac{\sigma_{\epsilon w}}{\sigma^2_w}$	$\sigma^2_\epsilon + \beta^2_x\sigma^2_x - \frac{(\sigma_{xw}\beta_x+\sigma_{\epsilon w})^2}{\sigma^2_w}$
Surrogate	ρ^2_{xw}	$\beta_0 + \beta_x\mu_x - \frac{\beta_x\sigma_{xw}}{\sigma^2_w}\mu_w$	$\beta_x\left(\frac{\sigma_{xw}}{\sigma^2_w}\right)$	$\sigma^2_\epsilon + \beta^2_x\sigma^2_x(1-\rho^2_{xw})$
Classical	$\frac{\sigma^2_x}{\sigma^2_x+\sigma^2_{u_c}}$	$\beta_0 + \beta_x\mu_x\,(1-\rho^2_{xw})$	$\beta_x\left(\frac{\sigma^2_x}{\sigma^2_x+\sigma^2_{u_c}}\right)$	$\sigma^2_\epsilon + \beta^2_x\sigma^2_x(1-\rho^2_{xw})$
B/C mixture	$\frac{\sigma^4_L(\sigma^2_L+\sigma^2_{u_b})^{-1}}{(\sigma^2_L+\sigma^2_{u_c})}$	$\beta_0 + \beta_x\mu_x\left(1-\frac{\sigma^2_L}{\sigma^2_L+\sigma^2_{u_c}}\right)$	$\beta_x\left(\frac{\sigma^2_L}{\sigma^2_L+\sigma^2_{u_c}}\right)$	$\sigma^2_\epsilon + \beta^2_x\sigma^2_x(1-\rho^2_{xw})$
Berkson	$\frac{\sigma^2_x-\sigma^2_{u_b}}{\sigma^2_x}$	β_0	β_x	$\sigma^2_\epsilon + \beta^2_x\sigma^2_x(1-\rho^2_{xw})$
No error	1	β_0	β_x	σ^2_ϵ

Table 3.1: Table entries are error model squared correlations, and intercepts, slopes and residual variances of the linear model relating $\mathbf{Y}$ to $\mathbf{W}$ when $(\mathbf{Y},\mathbf{X},\mathbf{W})$ is multivariate normal for the cases $\mathbf{W}$ is: a general differential measurement, a general surrogate, an unbiased classical-error measurement, an unbiased classical/Berkson mixture error measurement, an unbiased Berkson measurement, and the case of no error ($\mathbf{W}=\mathbf{X}$). Classical error variance, $\sigma^2_{u_c}$; Berkson error variance, $\sigma^2_{u_b}$; B/C mixture error model, $\mathbf{X} = \mathcal{L} + \mathbf{U}_b$, $\mathbf{W} = \mathcal{L} + U_c$, $\mathbf{Y} = \beta_0 + \beta_x\mathbf{X} + \epsilon$.

3.2.5.3 Classical Error Model

In the classical error model, $\mathbf{W}$ is a surrogate and $E(\mathbf{W} \mid \mathbf{X}) = \mathbf{X}$, and we can write $\mathbf{W} = \mathbf{X} + \mathbf{U}_c$ where $\mathbf{U}_c$ is a measurement error. Here we use the subscript c to emphasize that the error is classical and to avoid confusion with the two error models discussed below. We have already discussed this model in detail elsewhere, for example, in Sections 1.2 and 2.2. It is apparent from the third row of Table 3.1 that if the reliability ratio, $\lambda = \sigma_x^2/(\sigma_x^2 + \sigma_{u_c}^2)$ is known or can be estimated, then β_x can be recovered from the regression of $\mathbf{Y}$ on $\mathbf{W}$.

3.2.5.4 Berkson Error Model

In the Berkson error model, $\mathbf{W}$ is a surrogate and $E(\mathbf{X} \mid \mathbf{W}) = \mathbf{W}$, and we can write $\mathbf{X} = \mathbf{W} + \mathbf{U}_b$ where $\mathbf{U}_b$ is a Berkson error. This model has been discussed in detail elsewhere, for example, Sections 1.4 and 2.2. It is apparent from the fifth row of Table 3.1 that the regression parameters are not biased by Berkson measurement error. However, note that the residual variance in the regression of $\mathbf{Y}$ on $\mathbf{W}$ is greater than σ^2, a consequence of the fact that for this model $\mathbf{W}$ is a surrogate. Both the unbiasedness and increased residual variation are well illustrated in Figure 3.3.

3.2.5.5 Berkson/Classical Mixture Error Model

We now consider an error model that was encountered previously (see Section 1.8.2) on the log-scale, and is discussed again at length in Section 8.6. Here we consider the additive version. The defining characteristic is that the error model contains both classical and Berkson components. Specifically, it is assumed that

$$\mathbf{X} = \mathcal{L} + \mathbf{U}_b, \tag{3.7}$$
$$\mathbf{W} = \mathcal{L} + \mathbf{U}_c. \tag{3.8}$$

When $\mathbf{U}_b = 0$, $\mathbf{X} = \mathcal{L}$ and the classical error model is obtained, whereas the Berkson error model results when $\mathbf{U}_c = 0$, since then $\mathbf{W} = \mathcal{L}$. We denote the variances of the error terms by $\sigma_{u_c}^2$ and $\sigma_{u_b}^2$. This error model has features of both the classical and Berkson error models. Note that there is bias in the regression parameters when $\sigma_{u_c}^2 > 0$, as in the classical model. The inflation in the residual variance has the same form as the other nondifferential error models in terms of ρ_{xw}^2, but ρ_{xw}^2 depends on both error variances for this model.

The error models in Table 3.1 are arranged from most to least problematic in terms of the negative effects of measurement error. Although we discussed the Berkson/classical mixture error model last, in the hierarchy of error models its place is between the classical and Berkson error models.

3.3 Multiple and Orthogonal Regression

3.3.1 Multiple Regression: Single Covariate Measured with Error

In multiple linear regression, the effects of measurement error are more complicated, even for the classical additive error model.

We now consider the case where $\mathbf{X}$ is scalar, but there are additional covariates $\mathbf{Z}$ measured without error. The linear model is now

$$\mathbf{Y} = \beta_0 + \beta_x \mathbf{X} + \beta_z^t \mathbf{Z} + \epsilon, \tag{3.9}$$

where $\mathbf{Z}$ and β_z are column vectors, and β_z^t is a row vector. In Appendix B.2 it is shown that if $\mathbf{W}$ is unbiased for $\mathbf{X}$, and the measurement error $\mathbf{U}$ is independent of $\mathbf{X}$, $\mathbf{Z}$ and ϵ, then the least squares regression estimator of the coefficient of $\mathbf{W}$ consistently estimates $\lambda_1 \beta_x$, where

$$\lambda_1 = \frac{\sigma_{x|z}^2}{\sigma_{w|z}^2} = \frac{\sigma_{x|z}^2}{\sigma_{x|z}^2 + \sigma_u^2}, \tag{3.10}$$

and $\sigma_{w|z}^2$ and $\sigma_{x|z}^2$ are the residual variances of the regressions of $\mathbf{W}$ on $\mathbf{Z}$ and $\mathbf{X}$ on $\mathbf{Z}$, respectively. Note that λ_1 is equal to the simple linear regression attenuation, $\lambda = \sigma_x^2/(\sigma_x^2 + \sigma_u^2)$, only when $\mathbf{X}$ and $\mathbf{Z}$ are uncorrelated. Otherwise, $\sigma_{x|z}^2 < \sigma_x^2$ and $\lambda_1 < \lambda$, showing that collinearity increases attenuation.

The problem of measurement error–induced bias is not restricted to the regression coefficient of $\mathbf{X}$. The coefficient of $\mathbf{Z}$ is also biased in general, unless $\mathbf{Z}$ is independent of $\mathbf{X}$ (Carroll, Gallo, and Gleser, 1985; Gleser, Carroll, and Gallo, 1987). In Section B.2 it is shown that for the model (3.9), the naive ordinary least squares estimates not β_z but rather

$$\beta_{z*} = \beta_z + \beta_x(1 - \lambda_1)\Gamma_z, \tag{3.11}$$

where Γ_z^t is the coefficient of $\mathbf{Z}$ in the regression of $\mathbf{X}$ on $\mathbf{Z}$, that is, $E(\mathbf{X} \mid \mathbf{Z}) = \Gamma_0 + \Gamma_z^t \mathbf{Z}$.

This result has important consequences when interest centers on the effects of covariates measured *without* error. Carroll et al. (1985) and Carroll (1989) showed that in the two-group analysis of covariance where $\mathbf{Z}$ is a treatment assignment variable, naive linear regression produces a consistent estimate of the treatment effect only if the design is balanced, that is, $\mathbf{X}$ has the same mean in both groups and is independent of treatment. With considerable imbalance, the naive analysis may lead to the conclusions that (i) there is a treatment effect when none actually exists; and (ii) the effects are negative when they are actually positive,

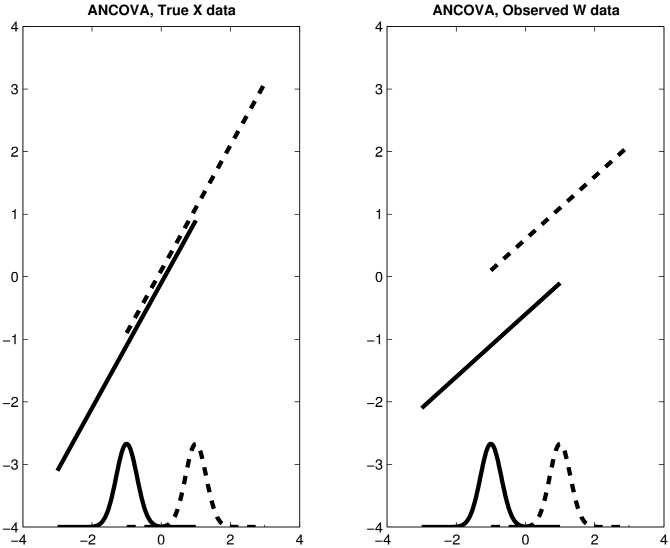

Figure 3.5 *Illustration of the effects of measurement error in an unbalanced analysis of covariance. The left panel shows the actual* $(\mathbf{Y}, \mathbf{X})$ *fitted functions, which are the same, indicating no treatment effect. The density function of* $\mathbf{X}$ *in the two groups are very different, however, as can be seen in the schematic density functions of* $\mathbf{X}$ *at the bottom. The right panel shows what happens when there is measurement error in the continuous covariate: Now the observed data suggest a large treatment effect.*

or vice versa. Figure 3.5 illustrates this process schematically. In the left panel, we show linear regression fits in the analysis of covariance model when there is no effect of treatment, that is, the two lines are the same. At the bottom of this panel, we draw schematic density functions for $\mathbf{X}$ in the two groups: The solid lines are the treatment group with smaller $\mathbf{X}$. The effect of measurement error in this problem is attenuation *around the mean in each group*, leading to the right panel, where the linear regression fits to the observed $\mathbf{W}$ are given. Now note that the lines are not identical, indicating that we would observe a treatment effect, even though it does not exist.

3.3.2 Multiple Covariates Measured with Error

Now suppose that there are covariates $\mathbf{Z}$ measured without error, that $\mathbf{W}$ is unbiased for $\mathbf{X}$, which may consist of multiple predictors, and that the linear regression model is $\mathbf{Y} = \beta_0 + \beta_x^t \mathbf{X} + \beta_z^t \mathbf{Z} + \epsilon$. If we write Σ_{ab} to be the covariance matrix between random variables $\mathbf{A}$ and $\mathbf{B}$, then naive ordinary linear regression consistently estimates not (β_x, β_z) but rather

$$\begin{aligned}
\begin{pmatrix} \beta_{x*} \\ \beta_{z*} \end{pmatrix} &= \begin{pmatrix} \Sigma_{xx} + \Sigma_{uu} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix}^{-1} \\
&\qquad \left\{ \begin{pmatrix} \Sigma_{xy} \\ \Sigma_{zy} \end{pmatrix} + \begin{pmatrix} \Sigma_{u\epsilon} \\ 0 \end{pmatrix} \right\} \qquad (3.12) \\
&= \begin{pmatrix} \Sigma_{xx} + \Sigma_{uu} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix}^{-1} \\
&\qquad \left\{ \begin{pmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix} \begin{pmatrix} \beta_x \\ \beta_z \end{pmatrix} + \begin{pmatrix} \Sigma_{u\epsilon} \\ 0 \end{pmatrix} \right\}.
\end{aligned}$$

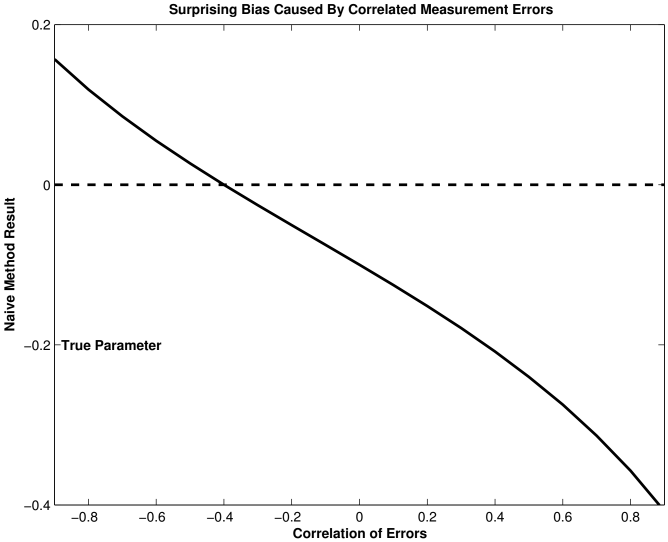

Figure 3.6 *Illustration of the effects of correlated measurement error with two variables measured with error. The true variables are actually uncorrelated, while the errors are correlated, with correlations ranging from* -0.9 *to* 0.9. *Displayed is a plot of what least squares estimates against the correlation of the measurement errors. The true value of the parameter of interest is* -0.2.

Thus, ordinary linear regression is biased. In Section 3.4, we take up the issue of bias correction. However, before doing so, it is worth taking a minute to explore the bias result (3.12). Consider a case of a regression on two error prone covariates, where the coefficient β_x in the regression of $\mathbf{Y}$ on $\mathbf{X}$ is $(1.0, -0.2)^t$, and where the components of $\mathbf{X}$ are independent so that Σ_{xx} is the identity matrix. Let the variance of the measurement errors $\mathbf{U}$ both $= 1.0$, and let their correlation ρ vary from -0.9 to 0.9. In Figure 3.6 we graph what least squares ignoring measurement error is really estimating in the second component (-0.2) of β_x as ρ varies. When the correlation between the measurement error is large but negative, least squares actually suggests that the coefficient is *positive* when it really is negative. Equally surprising, if the correlation between the measurement errors is large and positive, least squares actually suggests a more negative effect than actually exists.

3.4 Correcting for Bias

As we have just seen, the ordinary least squares estimator is typically biased under measurement error, and the direction and magnitude of the bias depends on the regression model, the measurement error distribution, and the correlation between the true predictor variables. In this section, we describe two common methods for eliminating bias.

3.4.1 Method of Moments

In simple linear regression with the classical additive error model, we have seen in (3.1) that ordinary least squares is an estimate of $\lambda\beta_x$, where λ is the reliability ratio. If the reliability ratio were known, then one could obtain an unbiased estimate of β_x simply by dividing the ordinary least squares slope $\widehat{\beta}_{x*}$ by the reliability ratio.

Of course, the reliability ratio is rarely known in practice, and one has to estimate it. If $\widehat{\sigma}_u^2$ is an estimate of the measurement error variance (this is discussed in Section 4.4), and if $\widehat{\sigma}_w^2$ is the sample variance of the $\mathbf{W}$s, then a consistent estimate of the reliability ratio is $\widehat{\lambda} = (\widehat{\sigma}_w^2 - \widehat{\sigma}_u^2)/\widehat{\sigma}_w^2$. The resulting estimate is $\widehat{\beta}_{x*}/\widehat{\lambda}$.

In small samples, the sampling distribution of $\widehat{\beta}_{x*}/\widehat{\lambda}$ is highly skewed, and in such cases a modified version of the method-of-moments estimator is recommended (Fuller, 1987; Section 2.5.1). Fuller's modification depends upon a tuning parameter α. Fuller does not give explicit advice about choosing α, but in his simulations $\alpha = 2$ produced more accurate estimates than the unmodified estimator. As an example, in Figure 3.7, in the top panel, we plot the histogram of the corrected estimate when $n = 20$, $\mathbf{X}$ is standard normal, the reliability ratio $= 0.5$, and the error about the line in the regression model is 0.25: The skewness is clear. In the bottom panel, we plot the histogram of Fuller's corrected estimator: It is slightly biased downwards, but very much more symmetric. In this figure, Fuller's method was defined as follows. Let $\widehat{\sigma}_{yw}$ and $\widehat{\sigma}_y^2$ be the sample covariance between $\mathbf{Y}$ and $\mathbf{W}$ and the sample variance of $\mathbf{Y}$, respectively. Define $\widehat{\kappa} = (\widehat{\sigma}_w^2 - \widehat{\sigma}_{yw}^2/\widehat{\sigma}_y^2)/\widehat{\sigma}_u^2$.

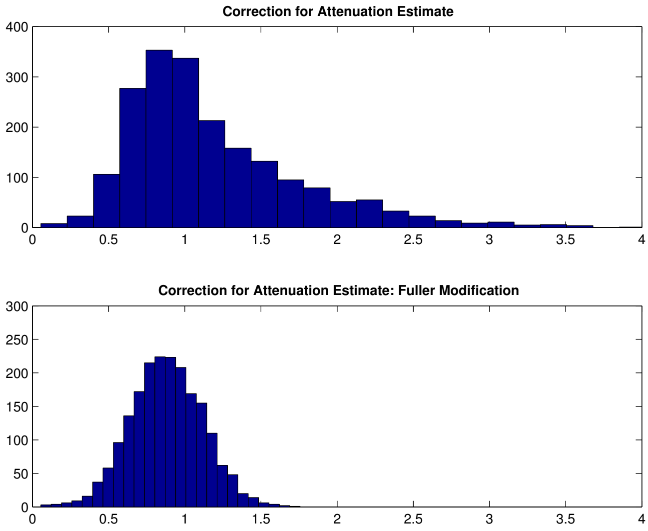

Figure 3.7 *Illustration of the small-sample distribution of the method-of-moments estimator of the slope in simple linear regression when* $n = 20$ *and the reliability ratio* $\lambda = 0.5$. *The top panel is the usual method-of-moments estimate, while the bottom panel is Fuller's correction to it.*

Then define $\widehat{\sigma}_x^2 = \widehat{\sigma}_w^2 - \widehat{\sigma}_u^2$ if $\widehat{\kappa} \geq 1 + (n-1)^{-1}$, while otherwise $\widehat{\sigma}_x^2 = \widehat{\sigma}_w^2 - \widehat{\sigma}_u^2\{\widehat{\kappa} - (n-1)^{-1}\}$. Then Fuller's corrected estimate with his $\alpha = 2$ is given as $(\widehat{\beta}_{x*}\widehat{\sigma}_w^2)/\{\widehat{\sigma}_x^2 + 2\widehat{\sigma}_u^2/(n-1)\}$.

The algorithm described above is called the *method-of-moments* estimator. The terminology is apt, because ordinary least squares and the reliability ratio depend only on moments of the observed data.

The method-of-moments estimator can be constructed for the general linear model, not just for simple linear regression. Suppose that $\mathbf{W}$ is unbiased for $\mathbf{X}$, and consider the general linear regression model with $\mathbf{Y} = \beta_0 + \beta_x^t\mathbf{X} + \beta_z^t\mathbf{Z} + \epsilon$.

The ordinary least squares estimator is biased even in large samples because it estimates (3.12). When Σ_{uu} and $\Sigma_{u\epsilon}$ are known or can be estimated, (3.12) can be used to construct a simple method-of-moments estimator that is commonly used to correct for the bias. Let S_{ab} be the sample covariance between random variables $\mathbf{A}$ and $\mathbf{B}$. The method-of-moments estimator that corrects for the bias in the case that Σ_{uu} and $\Sigma_{u\epsilon}$ are known is

$$\begin{pmatrix} S_{ww} - \Sigma_{uu} & S_{wz} \\ S_{zw} & S_{zz} \end{pmatrix}^{-1} \begin{pmatrix} S_{wy} - \Sigma_{u\epsilon} \\ S_{zy} \end{pmatrix}, \tag{3.13}$$

In the case that Σ_{uu} and $\Sigma_{u\epsilon}$ are estimated, the estimates replace the known values in (3.13). It is often reasonable to assume that $\Sigma_{u\epsilon} = 0$, in which case (3.13) simplifies accordingly.

In the event that $\mathbf{W}$ is biased for $\mathbf{X}$, that is, $\mathbf{W} = \gamma_0 + \gamma_x \mathbf{X} + \mathbf{U}$, that is, the error calibration model, the method-of-moments estimator can still be used, provided estimates of (γ_0, γ_x) are available. The strategy is to calculate the estimators above using the error-calibrated variate $\mathbf{W}_* = \widehat{\gamma}_x^{-1}(\mathbf{W} - \widehat{\gamma}_0)$.

3.4.2 Orthogonal Regression

Another well publicized method for linear regression in the presence of measurement error is *orthogonal regression*; see Fuller (1987, Section 1.3.3). This is sometimes known as the linear statistical relationship (Tan and Iglewicz, 1999) or the linear functional relationship. However, for reasons given below, we are skeptical about the general utility of orthogonal regression, in large part because it is so easily misused. Although it is not fundamental to understanding later material on nonlinear models, we take the opportunity to discuss orthogonal regression at length here in order to emphasize the potential pitfalls associated with it. The work appeared as Carroll and Ruppert (1996), but the message is worth repeating. This section can be skipped by those who are interested only in estimation for nonlinear models or who plan never to use orthogonal regression.

Let $\mathbf{Y} = \beta_0 + \beta_x \mathbf{X} + \epsilon$ and $\mathbf{W} = \mathbf{X} + \mathbf{U}$, where ϵ and $\mathbf{U}$ are uncorrelated. Whereas the method-of-moments estimator (Section 3.4) requires knowledge or estimability of the measurement error variance σ_u^2, orthogonal regression requires the same for the ratio $\eta = \sigma_\epsilon^2/\sigma_u^2$.

The orthogonal regression estimator minimizes the orthogonal distance of $(\mathbf{Y}, \mathbf{W})$ to the line $\beta_0 + \beta_x \mathbf{X}$, weighted by η, that is, it minimizes

$$\sum_{i=1}^{n} \left\{ (\mathbf{Y}_i - \beta_0 - \beta_x x_i)^2 + \eta\,(\mathbf{W}_i - x_i)^2 \right\} \tag{3.14}$$

in the unknown parameters $(\beta_0, \beta_x, x_1, \ldots, x_n)$.

In fact, (3.14) is the sum of squared orthogonal distances between the points $(\mathbf{Y}_i, \mathbf{W}_i)_1^n$ and the line $y = \beta_0 + \beta_x x$, only in the special case that $\eta = 1$. However, the term orthogonal regression is used to describe the method regardless of the value of $\eta < \infty$.

The orthogonal regression estimator is the functional maximum likelihood estimator (Sections 2.1 and 7.1) assuming that $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ are unknown fixed constants, and that the errors $(\epsilon, \mathbf{U})$ are independent and normally distributed.

Orthogonal regression has the appearance of greater applicability than

$\mathbf{W}_i$	$\mathbf{Y}_{i1}$	$\mathbf{Y}_{i2}$
−1.8007	−0.5558	−0.9089
−0.7717	0.2076	0.6499
−0.4287	−1.7365	−1.8542
−0.0857	−0.9018	0.2040
0.2572	−0.2312	−0.3097
0.6002	0.2967	0.5072
0.9432	0.5928	1.5381
1.2862	1.2420	1.2599

Table 3.2 *Orthogonal regression example with replicated response.*

method-of-moments estimation in that only the ratio, η, of the error variances need be known or estimated. However, it is our experience that in the majority of problems η cannot be specified or estimated correctly, and use of orthogonal regression with an improperly specified value of η often results in an unacceptably large *overcorrection* for attenuation due to measurement error.

We illustrate the problem with some data from a consulting problem (Table 3.2). The data include two measurements of a response variable, $\mathbf{Y}_{i1}$ and $\mathbf{Y}_{i2}$, and one predictor variable with true value $\mathbf{X}_i$, $i = 1, \ldots, 8$. The data are proprietary and we cannot disclose the nature of the application. Accordingly, all of the variables have been standardized to have sample means and variances 0 and 1, respectively.

We take as the response variable to be used in the regression analysis, $\mathbf{Y}_i = (\mathbf{Y}_{i1} + \mathbf{Y}_{i2})/2$, the average of the two response measurements.

Using an independent experiment, it had been estimated that $\sigma_u^2 \approx 0.0424$, also after standardization. Because the sample standard deviation of $\mathbf{W}$ is 1.0, measurement error induces very little bias here. The estimated reliability ratio is $\widehat{\lambda} = 1 - 0.0424 \approx 0.96$, and so attenuation is only about 4%. The ordinary least squares estimated slope from regressing the average of the responses on $\mathbf{W}$ is 0.65, while the method-of-moments slope estimate is $\widehat{\lambda}^{-1}0.65 \approx 0.68$.

In a first analysis of these data, our client thought that orthogonal regression was an appropriate method for these data. A components-of-variance analysis resulted in the estimate 0.0683 for the response measurement error variance. If η is estimated by $\widehat{\eta} = 0.0683/0.0424 \approx 1.6118$, then the resulting orthogonal regression slope estimate is 0.88.

The difference in these two estimates, $|0.88 - 0.68|$, is larger than would be expected from random variation alone. Clearly, something is amiss. The method-of-moments correction for attenuation is only $\widehat{\lambda}^{-1} \approx 1.04$, whereas,orthogonal regression in effect, produces a correction for attenuation of approximately $1.35 \approx 0.88/0.65$.

The problem lies in the nature of the regression model error ϵ, which is typically the sum of two components: (i) ϵ_M, the measurement error in determination of the response; and (ii) ϵ_L, the *equation error*, that is, the variation about the regression line of the true response in the absence of measurement error. See Section 1.5 for another example of equation error, which in nutrition is called *person-specific bias.*

If we have replicated measurements, $\mathbf{Y}_{ij}$, of the true response, then $\mathbf{Y}_{ij} = \beta_0 + \beta_x \mathbf{X}_i + \epsilon_{L,i} + \epsilon_{M,ij}$, and of course their average is $\overline{\mathbf{Y}}_{i\cdot} = \beta_0 + \beta_x \mathbf{X}_i + \epsilon_{L,i} + \overline{\epsilon}_{M,i\cdot}$. Here and throughout the book, a subscript "dot" and overbar means averaging. For example, with k replicates,

$$\overline{\mathbf{Y}}_{i\cdot} = k^{-1} \sum_{j=1}^{k} \mathbf{Y}_{ij}; \quad \overline{\epsilon}_{M,i\cdot} = k^{-1} \sum_{j=1}^{k} \epsilon_{M,ij}.$$

The components of variance analysis estimates *only* the variance of the average measurement error $\overline{\epsilon}_{M,i\cdot}$ in the responses, but completely ignores the variability, $\epsilon_{L,i}$, about the line. The net effect is to underestimate η and thus overstate the correction required of the ordinary least squares estimate, because $\mathrm{var}(\overline{\epsilon}_{M,i\cdot})/\sigma_u^2$ is used as the estimate of η instead of the larger, appropriate value $\left\{\mathrm{var}(\overline{\epsilon}_{M,i\cdot}) + \mathrm{var}(\epsilon_{L,i})\right\}/\sigma_u^2$.

The naive use of orthogonal regression on the data in Table 3.2 has assumed that there is no additional variability about the line in addition to that due to measurement error in the response, that is, $\epsilon_{L,i} = 0$. To check this, refer to Figure 3.8. Each replicated response is indicated by a solid and filled circle. Remember that there is little measurement error in $\mathbf{W}$. In addition, the replication analysis suggested that the standard deviation of the replicates was less than 10% of the variability of the responses. Thus, in the absence of equation error we would expect to see the replicated pairs falling along a clearly delineated straight line. This is far from the case, suggesting that the equation error $\epsilon_{L,i}$ is a large part of the variability of the responses. Indeed, while the replication analysis suggests that $\mathrm{var}(\overline{\epsilon}_{M,i\cdot}) \approx 0.0683$, a method-of-moments analysis suggests $\mathrm{var}(\epsilon_{L,i}) \approx 0.4860$.

Fuller (1987) was one of the first to emphasize the importance of equation error. In our experience, outside of some special laboratory validation studies, equation error is almost always important in linear regression. In the majority of cases,orthogonal regression is an inappro-

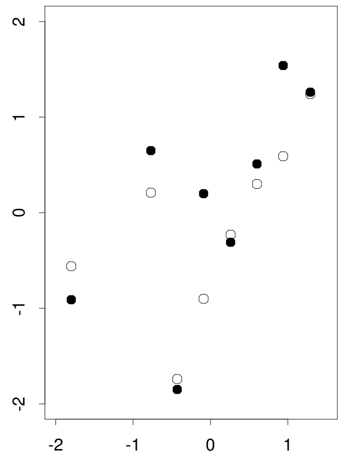

Figure 3.8 *Illustration where the assumptions of orthogonal regression appear violated. The filled and empty circles represent replicated values of the response. Note the evidence of equation error because the replicate responses are very close to each other, indicating little response measurement error, but the circles do not fall on a line, indicating some type of response error.*

priate technique, unless estimation of both the response measurement error and the equation error is possible.

In some cases, $\mathbf{Y}$ and $\mathbf{W}$ are measured in the same way, for example, if they are both blood pressure measurements taken at different times. Here, it is often entirely reasonable to *assume* that the variance of ϵ_M equals σ_u^2, and then there is a temptation to ignore equation error and hence set $\eta = 1$. Almost universally, this is a mistake: Equation error generally exists. This temptation is especially acute when replicates are absent, so that σ_u^2 cannot be estimated and the method-of-moments estimator cannot be used.

3.5 Bias Versus Variance

Estimates which do not account for measurement error are typically biased. Correcting for this bias entails what is often referred to as a *bias versus variance* tradeoff. What this means is that, in most problems, the very nature of correcting for bias is that the resulting corrected estimator will be more variable than the biased estimator. Of course, when an estimator is more variable, the confidence intervals associated with it are longer.

Later in this section we will describe theory, but it is instructive to consider an extreme case, using the same simulated data as in Figure 3.7 and Section 3.4.1. In this problem, the sample size is $n = 20$, the true slope is $\beta_x = 1.0$ and the reliability ratio is $\lambda = 0.5$. The top panel of Figure 3.9 gives the histogram of Fuller's modification of the method-of-moments estimator, while the bottom panel gives the histogram of the naive method that ignores measurement error. Note how the naive estimator is badly biased: Indeed, we know it estimates $\lambda\beta_x = 0.5$, and it is tightly bunched around this (wrong) value. The method-of-moments estimator is roughly unbiased, but this correction for bias is at the cost of a much greater variability (2.7 times greater in the simulation).

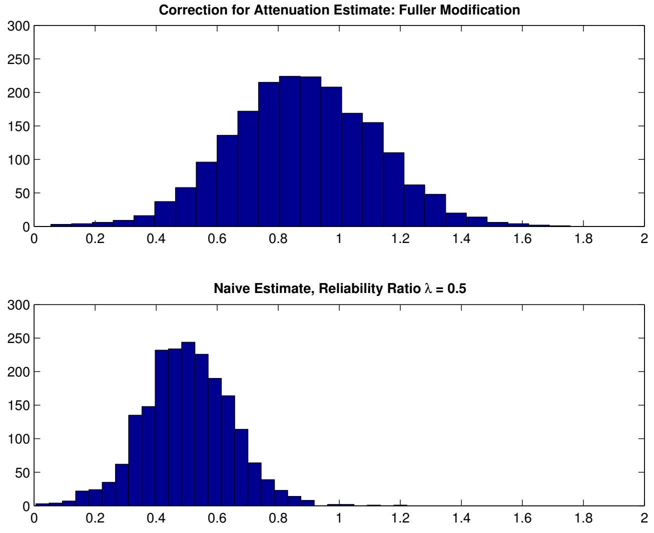

Figure 3.9 *Bias versus variance tradeoff in estimating the slope in simple linear regression. This is an extreme example of simple linear regression, with a sample size of* $n = 20$ *and a reliability ratio of* $\lambda = 0.5$*. The true value of the slope is* $\beta_x = 1$*. The top panel is Fuller's modification of the correction for attenuation estimate; the bottom is the naive estimate that ignores measurement error. The former is much more variable; the latter is very badly biased.*

3.5.1 Theoretical Bias–Variance Tradeoff Calculations

In this section, we will illustrate the bias versus variance tradeoff theoretically in simple linear regression. This material is somewhat technical, and readers may skip it without any loss of understanding of the main points of measurement error models.

Consider the simple linear regression model, $\mathbf{Y} = \beta_0 + \beta_x \mathbf{X} + \epsilon$, with additive independent measurement error, $\mathbf{W} = \mathbf{X} + \mathbf{U}$, under the simplifying assumption of joint normality of $\mathbf{X}$, $\mathbf{U}$, and ϵ. Further, suppose that the reliability ratio λ in (3.1) is known. We make this assumption only to simplify the discussion in this section. Generally, in applications it is seldom the case that this parameter is known, although there are exceptions (Fuller, 1987).

Let $\widehat{\beta}_{x*}$ denote the least squares estimate of slope from the regression of $\mathbf{Y}$ on $\mathbf{W}$. We know that its mean is $E(\widehat{\beta}_{x*}) = \lambda\beta_x$. Denote its variance by σ_*^2.

The method-of-moments estimator of β_x, is $\widehat{\beta}_{x,mm} = \lambda^{-1}\widehat{\beta}_{x*}$ and has mean $E(\widehat{\beta}_{x,mm}) = \beta_x$, and variance $\mathrm{Var}(\widehat{\beta}_{x,mm}) = \lambda^{-2}\sigma_*^2$.

Because $\lambda < 1$, it is clear that while the correction-for-attenuation in $\widehat{\beta}_{x,mm}$ reduces its bias to zero, there is an increase in variability relative to the variance of the biased estimator $\widehat{\beta}_{x*}$. The variability is inflated even further if an estimate $\widehat{\lambda}$ is used in place of λ.

The price for reduced bias is increased variance. This phenomenon is not restricted to the simple model and estimator in this section, but occurs with almost universal generality in the analysis of measurement error models. In cases where the absence of bias is of paramount importance, there is usually no escaping the increase in variance. In cases where some bias can be tolerated, consideration of mean squared error is necessary.

In the following material, we indicate that there are compromise estimators that may outperform both uncorrected and corrected estimators, at least in small samples. Surprisingly, outside of the work detailed in Fuller (1987), such compromise estimators have not been much investigated, especially for nonlinear models.

Remember that mean squared error (MSE) is the sum of the variance plus the square of the bias. This is an interesting criterion to use, because uncorrected estimators have more bias but smaller variance than corrected estimators, and the bias versus variance tradeoff is transparent. Note that

$$\begin{aligned} \mathrm{MSE}(\widehat{\beta}_{x*}) &= \sigma_*^2 + (1-\lambda)^2\beta_x^2; \text{ and} \\ \mathrm{MSE}(\widehat{\beta}_{x,mm}) &= \lambda^{-2}\sigma_*^2. \end{aligned} \tag{3.15}$$

It follows that

$$\mathrm{MSE}\left(\widehat{\beta}_{x,mm}\right) < \mathrm{MSE}\left(\widehat{\beta}_{x*}\right)$$

if and only if

$$\sigma_*^2 < \frac{\lambda^2(1-\lambda)\beta_x^2}{1+\lambda}.$$

Because σ_*^2 decreases with increasing sample size, we can conclude that in sufficiently large samples it is always beneficial, in terms of mean squared error, to correct for attenuation due to measurement error.

Consider now the alternative estimator $\widehat{\beta}_{x,a} = a\beta_{x*}$ for a fixed constant a. The mean squared error of this estimator is $a^2\sigma_*^2 + (a\lambda - 1)^2\beta_x^2$, which is minimized when $a = a_* = \lambda\beta_x^2/(\sigma_*^2 + \lambda^2\beta_x^2)$. Ignoring the fact that a_* depends on unknown parameters, we consider the "estimator" $\widehat{\beta}_{x,*} = a_*\beta_{x*}$, which has smaller mean squared error than either $\widehat{\beta}_{x,mm}$ or $\widehat{\beta}_{x*}$. Note that as $\sigma_*^2 \to 0$, $a_* \to \lambda^{-1}$.

The estimator $\widehat{\beta}_{x,*}$ achieves its mean-squared-error superiority by making a partial correction for attenuation in the sense that $a_* < \lambda^{-1}$. This simple exercise illustrates that estimators that make only partial corrections for attenuation can have good mean-squared-error performance.

Although we have used a simple model and a somewhat artificial estimator to facilitate the discussion of bias and variance, all of the conclusions made above hold, at least to a very good approximation, in general for both linear and nonlinear regression measurement error models.

3.6 Attenuation in General Problems

We have already seen that, even in linear regression with multiple covariates, the effects of measurement error are complex and not easily described. In this section, we provide a brief overview of what happens in nonlinear models.

Consider a scalar covariate **X** measured with error, and suppose that there are no other covariates. In the classical error model for simple linear regression, we have seen that the bias caused by measurement error is always in the form of attenuation, so that ordinary least squares preserves the sign of the regression coefficient asymptotically, but is biased towards zero. Attenuation is a consequence then of (i) the simple linear regression model; and (ii) the classical additive error model. Without (i) and (ii), the effects of measurement error are more complex; we have already seen that attenuation may not hold if (ii) is violated.

In logistic regression when **X** is measured with additive error, attenuation does not always occur (Stefanski and Carroll, 1985), but it is typical. More generally, in most problems with a scalar **X** and no covariates **Z**, the underlying *trend* between **Y** and **X** is preserved under nondifferential measurement error, in the sense that the correlation between **Y** and **W** is positive whenever both $E(\mathbf{Y}|\mathbf{X})$ and $E(\mathbf{W}|\mathbf{X})$ are increasing functions of **X** (Weinberg, Umbach, and Greenland, 1993). Technically, this follows because with nondifferential measurement error, **Y** and **W** are uncorrelated given **X**, and hence the covariance between **Y** and **W** is just the covariance between $E(\mathbf{Y}|\mathbf{X})$ and $E(\mathbf{W}|\mathbf{X})$.

Positively, this result says that for the very simplest of problems (scalar $\mathbf{X}$, no covariates $\mathbf{Z}$ measured without error), the general trend in the data is typically unaffected by nondifferential measurement error. However, the result illustrates only part of a complex picture, because it describes only the *correlation* between $\mathbf{Y}$ and $\mathbf{W}$ and says nothing about the structure of this relationship.

For example, one might expect that if the regression, $E(\mathbf{Y}|\mathbf{X})$, of $\mathbf{Y}$ on $\mathbf{X}$ is nondecreasing in $\mathbf{X}$, and if $\mathbf{W} = \mathbf{X} + \mathbf{U}$ where $\mathbf{U}$ is independent of $\mathbf{X}$ and $\mathbf{Y}$, then the regression of $\mathbf{Y}$ on $\mathbf{W}$ would also be nondecreasing. But Hwang and Stefanski (1994) have shown that this need not be the case, although it is true in linear regression with normally distributed measurement error. However, these results show that making inferences about details in the relationship of $\mathbf{Y}$ and $\mathbf{X}$, based on the observed relationship between $\mathbf{Y}$ and $\mathbf{W}$, is a difficult problem in general.

There are other practical reasons why ignoring measurement error is not acceptable. First, estimating the direction of the relationship between $\mathbf{Y}$ and $\mathbf{X}$ correctly is nice, but as emphasized by MacMahon et al. (1990) we can be misled if we severely underestimate its magnitude. Second, the result does not apply to multiple covariates, as we have noted in Figure 3.5 for the analysis of covariance and in Figure 3.6 for correlated measurement errors. Indeed, we have already seen that in multiple linear regression under the additive measurement error model, the observed and underlying trends may be entirely different. Finally, it is also the case (Sectiob 10.1) that, especially with multiple covariates, one can use error modeling to improve the power of inferences. In large classes of problems, then, there is simply no alternative to careful consideration of the measurement error structure.

Bibliographic Notes

The linear regression problem has a long history and continues to be the subject of research. Excellent historic background can be found in the papers by Lindley (1953), Lord (1960), Cochran (1968) and Madansky (1959). Furthermore technical analyses are given by Fuller (1980), Carroll and Gallo (1982, 1984), and Carroll et al. (1985). Diagnostics are discussed by Carroll and Spiegelman (1986, 1992) and Cheng and Tsai (1992). Robustness is discussed by Ketellapper and Ronner (1984), Zamar (1988, 1992), Cheng and van Ness (1988), and Carroll et al. (1993). Ganse, Amemiya, and Fuller (1983) discussed an interesting prediction problem. Hwang (1986) and Hasenabeldy et al. (1989) discuss problems with unusual error structure. Boggs et al. (1988) discussed computational aspects of orthogonal regression in nonlinear models.