MEASUREMENT ERROR MODELS

Alexander Kukush [1]

Professor, National Taras Shevchenko University of Kyiv, Ukraine

A (nonlinear) measurement error model (MEM) consists of three parts: (1) a *regression model* relating an observable regressor variable z and unobservable regressor variable ξ (the variables are independent and generally vector valued) to a response variable y, which is considered here to be observable without measurement errors; (2) a *measurement model* relating the unobservable ξ to an observable surrogate variable x; and (3) a *distributional model* for ξ.

1. PARTS OF MEM

The *regression model* can be described by conditional distribution of y given (z, ξ) and given an unknown parameter vector θ. As usual this distribution is represented by a probability density function $f(y|z, \xi; \theta)$ with respect to some underlying measure on the Borel σ-field of $\boldsymbol{R}$. We restrict our attention to distributions that belong to the exponential family, i.e., we assume f to be of the form

$$f(y|z, \xi; \beta, \varphi) = \exp(\frac{y\eta - c(\eta)}{\varphi} + a(y, \varphi)) \tag{1}$$

with

$$\eta = \eta(z, \xi; \beta). \tag{2}$$

Here β is the regression parameter vector, φ a scalar dispersion parameter such that $\theta = (\beta^T, \varphi)^T$, and a, c, and η are known functions. This class comprises the class of generalized linear models, where $\eta = \eta(\beta_0 + z^T\beta_z + \xi^T\beta_\xi)$, $\beta = (\beta_0, \beta_x^T, \beta_\xi^T)^T$.

The *classical measurement model* assumes that the observed variable x differs from the latent ξ by a measurement error variable δ which is independent of z, ξ, and y :

$$x = \xi + \delta \tag{3}$$

with $\mathbf{E}\delta = 0$. Here we assume that $\delta \sim N(0,\ \Sigma_\delta)$ with Σ_δ known. The observable data are independent realizations of the model $(x_i, y_i),\ \ i = 1, \ldots, n$.

[1] Dr Alexander Kukush is Professor, Department of Mechanics and Mathematics, National Taras Shevchenko University of Kyiv, Ukraine. He is an Elected member of the International Statistical Institute (2004). He has authored and co-authored more than 100 papers on statistics and a book: *Theory of Stochastic Processes With Applications to Financial Mathematics and Risk Theory* (with D. Gusak, A. Kulik, Yu. Mishura, and A. Pilipenko, Problem Books in Mathematics, Springer, 2009). Professor Kukush has received the Taras Shevchenko award for a cycle of papers on regression (National Taras Shevchenko University of Kyiv, 2006).

Under the *Berkson measurement model*, the latent variable ξ differs from the observed x by a centered measurement error δ which is independent of z, x and y :

$$\xi = x + \delta. \tag{4}$$

Thus, the values of x are fixed in advance, whereas the unknown true values, ξ are fluctuating.

The *distributional model* for ξ either states that the ξ are unknown constants (*functional case*) or that ξ is a random variable (*structural case*) with a distribution given by a density $h(\xi; \gamma)$, where γ is a vector of nuisance parameters describing the distribution of ξ. In the structural case, we typically assume that

$$\xi \sim N(\mu_\xi,\ \Sigma_\xi), \tag{5}$$

although sometimes it is assumed that ξ follows a mixture of normal distributions. In the sequel, for the structural case we assume γ to be known. If not, it can often be estimated in advance (i.e., pre-estimated) without considering the regression model and the data y_i. For example, if ξ is normal, then μ_ξ and Σ_ξ can be estimated by $\overline{x}$ and $S_x - \Sigma_\delta$, respectively, where $\overline{x}$ and S_x are the empirical mean vector and empirical covariance matrix of the data x_i.

The goal of measurement error modeling is to obtain nearly unbiased estimates of the regression parameter β by fitting a model for y in terms of (z, x). Attainment of this goal requires careful analysis. Substituting x for ξ in the model (1) – (2), but making no adjustments in the usual fitting methods for this substitution, leads to estimates that are biased, sometimes seriously.

In the structural case, the *regression calibration* (RC) estimator can be constructed by substituting $\mathbf{E}(\xi|x)$ for unobservable ξ. In both functional and structural case, another, the *SIMEX* estimator, becomes very popular. Those estimators are not consistent in general, although they often reduce the bias significantly, see Carroll et al. (2006).

2. POLYNOMIAL AND POISSON MODEL

We mention two important examples of the classical MEM (1) – (3) where for simplicity the latent variable is scalar and the observable regressor z is absent. The *polynomial model* is given by

$$y = \beta_0 + \beta_1 \xi + ... + \beta_k \xi^k + \varepsilon,$$

where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ and ε is independent of ξ. Here

$$\eta = \sum_{r=0}^{k} \beta_r \xi^r,\ c(\eta) = \frac{1}{2}\eta^2,$$

and $\varphi = \sigma_\varepsilon^2$. Both cases are possible: (a) the measurement error variance σ_δ^2 is known, and (b) the ratio $\sigma_\varepsilon^2/\sigma_\delta^2$ is known; for the latter case see Shklyar (2008). In

particular case $k = 1$ we obtain the *linear model*; an overview of methods in this MEM is given in Cheng and Van Ness (1999).

In the *loglinear Poisson model* we have $y \sim Po(\lambda)$ with $\lambda = \exp(\beta_0 + \beta_1 \xi)$; then $\eta = \log \lambda,\ c(\eta) = e^\eta$, and $\varphi = 1$.

3. METHODS OF CONSISTENT ESTIMATION IN CLASSICAL MEM

Now, we deal with general model (1) – (3). We distinguish between two types of estimators, functional and structural ones. The latter make use of the distribution of ξ, which therefore must be given, at least up to the unknown parameter vector γ. The former does not need the distribution of ξ and works even when ξ is not random (functional case).

3.1. Functional method: Corrected Score

If the variable ξ were observable, one could estimate β (and also φ) by the method of maximum likelihood (ML). The corresponding likelihood-score function for β is given by

$$\psi(y, z, \xi; \beta, \varphi) = \frac{\partial \log f(y|z, \xi; \beta, \varphi)}{\partial \beta} = \frac{y - c'(\eta)}{\varphi} \frac{\partial \eta}{\partial \beta}$$

We want to construct an unbiased estimating function for β in the observed variables. For this purpose, we need to find functions g_1 and g_2 of z, x, and β such that

$$\mathbf{E}[g_1(z, x; \beta)|z, \xi] = \frac{\partial \eta}{\partial \beta}, \quad \mathbf{E}[g_2(z, x; \beta)|z, \xi] = c'(\eta) \frac{\partial \eta}{\partial \beta}.$$

Then

$$\psi_C(y, z, x; \beta) = y g_1(z, x; \beta) - g_2(z, x; \beta)$$

is termed the corrected score function. The *Corrected Score* (CS) estimator $\hat{\beta}_C$ of β is the solution to

$$\sum_{i=1}^{n} \psi_C(y_i, z_i, x_i; \hat{\beta}_C) = 0.$$

The functions g_1 and g_2 do not always exist. Stefanski (1989) gives the conditions for their existence and shows how to find them if they exist. The CS estimator is consistent in both functional and structural cases. It was first proposed by Stefanski (1989) and Nakamura (1990).

An alternative functional method, particularly adapted to generalized linear models, is the Conditional Score method, see Stefanski and Carroll (1987).

3.2. Structural methods: Quasi-Likelihood and Maximum Likelihood

The conditional mean and conditional variance of y given (z, ξ) are, respectively,

$$\mathbf{E}(y|z, \xi) = m^*(z, \xi; \beta) = c'(\eta),\ \mathbf{V}(y|z, \xi) = v^*(z, \xi; \beta) = \varphi c''(\eta).$$

Then the conditional mean and conditional variance of y given the observable variables are

$$m(z,x;\beta) = \mathbf{E}(y|z,x) = E[m^*(z,\xi;\beta)|x],$$
$$v(z,x;\beta) = \mathbf{V}(y|z,x) = \mathbf{V}[m^*(z,\xi;\beta)|x] + \mathbf{E}[v^*(z,\xi;\beta)|x].$$

For the Quasi-Likelihood (QL) estimator, we construct the quasi-score function

$$\psi_Q(y,z,x;\beta) = [y - m(z,x;\beta)]v(z,x;\beta)^{-1}\frac{\partial m(z,x;\beta)}{\partial \beta}.$$

Here we drop the parameter φ considering it to be known. We also suppress the nuisance parameter γ in the argument of the functions m and v, although m and v depend on γ. Indeed, in order to compute m and v, we need the conditional distribution of ξ given x, which depends on the distribution of ξ with its parameter γ. For instance, assume (5) where the elements of μ_ξ and Σ_ξ make up the components of the parameter vector γ. Then $\xi|x \sim N(\mu(x), T)$ with

$$\mu(x) = \mu_\xi + \Sigma_\xi(\Sigma_\xi + \Sigma_\delta)^{-1}(x - \mu_\xi), \quad T = \Sigma_\delta - \Sigma_\delta(\Sigma_\xi + \Sigma_\delta)^{-1}\Sigma_\delta.$$

The QL estimator $\hat{\beta}_Q$ of β is the solution to

$$\sum_{i=1}^{n} \psi_Q(y_i, z_i, x_i; \hat{\beta}_C) = 0.$$

The equation has a unique solution for large n, but it may have multiple roots if n is not large. Heyde and Morton (1998) develop methods to deal with this case.

Maximum likelihood is based on the conditional joint density of x, y given z. Thus while QL relies only on the error-free mean and variance functions, ML relies on the whole error-free model distribution. Therefore, ML is more sensitive than QL with respect to a potential model misspecification because QL is always consistent as long as at least the mean function (along with the density of ξ) has been correctly specified. In addition, the likelihood function is generally much more difficult to compute than the quasi score function. This often justifies the use of the relatively less efficient QL instead of the more efficient ML method.

3.3. Efficiency comparison

For CS and QL, $\hat{\beta}$ is asymptotically normal with asymptotic covariance matrix (ACM) Σ_C and Σ_Q respectively. In the structural model, it is natural to compare the relative efficiencies of $\hat{\beta}_C$ and $\hat{\beta}_Q$ by comparing their ACMs. In case there are no nuisance parameters, it turns out that

$$\Sigma_C \geq \Sigma_Q \tag{6}$$

in the sense of the Loewner order for symmetric matrices. Moreover, under mild conditions the strict inequality holds.

These results hold true if the nuisance parameters γ are known. If, however, they have to be estimated in advance, (6) need not be true any more. For the Poisson and polynomial structural models, Kukush et al. (2007) prove that (6) still holds even if the nuisance parameters are pre-estimated. Recently Kukush et al. (2009) have been shown that QL can be modified so that, in general, $\Sigma_C \geq \Sigma_Q$; for this purpose the γ must be estimated together with β, and not in advance.

4. ESTIMATION IN BERKSON MODEL

Now, we deal with the model (1), (2), and (4). Substituting x for ξ in the regression model (1) – (2) is equivalent to RC. Therefore, it leads to estimates with typically small bias.

A more precise method is ML. The conditional joint density of x and y given z has simpler form compared with the classical MEM. That is why ML is more reliable in Berkson model.

5. NONPARAMETRIC ESTIMATION

We mention two nonparametric problems overviewed in Carroll et al. (2006), Ch.12: the estimation of the density ρ of a random variable ξ, and the nonparametric estimation of a regression function f, both when ξ is measured with error. In these problems under normally distributed measurement error, the best mean squared error of an estimator of $\rho(x_0)$ or $f(x_0)$ converges to 0 at a rate no faster than the exceedingly slow rate of logarithmic order. However, under more heavy-tailed measurement error, estimators can perform well for reasonable sample size.

References

[1] Carroll, R. J., D, Ruppert, L. A. Stefanski, C. M. Crainiceanu (2006). *Measurement Error in Nonlinear Models* (2nd edition). Chapman and Hall, London.

[2] Cheng, C. L., J. W. Van Ness (1999). *Statistical Regression with Measurement Error*. Arnold, London.

[3] Heyde, C. C., R. Morton (1998). Multiple roots in general estimating equations. *Biometrika*, **85**, 967–972.

[4] Kukush, A., A. Malenko, H. Schneeweiss (2007). Comparing the efficiency of estimates in concrete errors-in-variables models under unknown nuisance parameters. *Theory of Stochastic Processes*, **13** (29), 4, 69–81.

[5] Kukush, A., A. Malenko, H. Schneeweiss (2009). Optimality of the quasi score estimator in a mean-variance model with applications to measurement error models. *Journal of Statistical Planning and Inference*, **139**, 3461–3472.

[6] Shklyar, S. V. (2008). Consistency of an estimator of the parameters of a polynomial regression with a known variance relation for errors in the measurement of the regressor and the echo. *Theory Probability and Mathematical Statistics*, **76**, 181–197.

[7] Stefanski, L. A. (1989). Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Communications in Statistics, Part A – Theory and Methods,* **18**, 4335–4358.

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

[9] Nakamura, T. (1990). Corrected score functions for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, **77**, 127–137.