Statistics and Probability Letters 97 (2015) 165–175

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

On Hong–Tamer's estimator in nonlinear errors-in-variable regression models

Jianghong Wu [a], Weixing Song [b,*]

[a] Department of Mathematics, Changzhi College, 046011, China
[b] Department of Statistics, Kansas State University, Manhattan, KS 66503, USA

ARTICLE INFO

Article history:
Received 10 August 2014
Received in revised form 23 November 2014
Accepted 26 November 2014
Available online 4 December 2014

Keywords:
Nonlinear regression model
Errors-in-variable
Laplace distribution
Revised moment conditions
Simulation extrapolation

ABSTRACT

Under some regularity conditions, the paper provides an alternative proof for the revised moment conditions proposed by Hong and Tamer (2003) in the nonlinear least squares regression model, when the covariates are measured with Laplace error. The asymptotic normality of the revised moment estimates is developed. The choice of optimal weight functions is also discussed and a nearly optimal weight function is identified. Moreover, a simulation extrapolation estimation procedure is suggested when the estimating equations based on the revised moment conditions are difficult to solve. Simulation studies are conducted to evaluate the finite performance of the proposed methods.

Published by Elsevier B.V.

1. Introduction

Constructing proper estimating equations is a commonly used and often effective way to find the estimates for the unknown parameters in many econometric models. The estimating equations are usually the sample versions of their population analogs. For example, the parameter of interest, say $\boldsymbol{\beta}$, might be defined by a set of population moment conditions, $Em(\mathbf{X};\boldsymbol{\beta}) = 0$, where $\mathbf{X} = (X_1, \ldots, X_k)^\tau$ is a k-dimensional random vector, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\tau$ is a p-dimensional vector of unknown parameters to be estimated, and $m(\cdot;\cdot)$ is a vector of functions. For any vector or matrix, the superscript τ means transposition. Econometric and statistical literatures are abundant in solving the estimation equations, as well as the discussion of the large sample properties of the resulting estimates. In many real applications, the vector $\mathbf{X}$ may not be observed directly. Instead, a surrogate $\mathbf{Z} = (Z_1, \ldots, Z_k)^\tau$ is available, which related to $\mathbf{X}$ additively through the relationship $\mathbf{Z} = \mathbf{X} + \mathbf{U}$, where $\mathbf{U} = (U_1, \ldots, U_k)^\tau$ is called the measurement error. It is well known that the presence of measurement error often creates some model identification problems. See Fuller (2006) for such examples. To identify the model parameters, we can either impose stronger distributional assumptions on random entities in the model, or seek extra data resources, such as the validation data set and replication measurements. Hong and Tamer (2003) tackled the identifiability problem by assuming the p random variables in $\mathbf{U}$ to be independent and each follows a Laplace distribution with mean 0 and unknown variance. Moreover, under the Laplace measurement error, the moment conditions $Em(\mathbf{X};\beta) = 0$ can be replaced by the revised moment conditions based on the observed variables only.

To be specific, suppose the measurable function $m(\mathbf{x};\boldsymbol{\beta})$ satisfies the Assumption 3 in Hong and Tamer (2003). Then follows a Laplace distribution with characteristic function $\phi_{\mathbf{U}}(\mathbf{t}) = \prod_{j=1}^{k}(1+\sigma_j^2 t_j^2/2)^{-1}$, where $\mathbf{t} = (t_1, \ldots, t_p)^\tau$, then they

* Corresponding author.
E-mail address: weixing@ksu.edu (W. Song).

http://dx.doi.org/10.1016/j.spl.2014.11.021
0167-7152/Published by Elsevier B.V.

showed that

$$Em(\mathbf{X};\boldsymbol{\beta}) = Em(\mathbf{Z};\boldsymbol{\beta}) + \sum_{l=1}^{k}\left(-\frac{1}{2}\right)^{l}\sum\cdots\sum_{j_1<\cdots<j_l}\sigma_{j_1}^2\cdots\sigma_{j_l}^2 E\frac{\partial^{2l}m(\mathbf{X};\beta)}{\partial X_{j_1}^2\cdots\partial X_{j_l}^2}. \tag{1}$$

If the revised moment conditions are sufficient for identifying σ and β, then the estimating equations based on the sample version of the above revised moment conditions can be employed to derive the estimates and show their consistency and asymptotic normality. However, if the revised moment conditions cannot completely identify the unknown parameters, some identified features can be still consistently estimated by the so called modified moment estimation procedure using some side information about the relative magnitude of the measurement error variances. The methodology cannot be easily extended to normal measurement error model case unless the function m has some simpler structures, such as polynomial or exponential. In fact, the similar differentiation idea was already adopted in the Masry and Rice (1992) when dealing with the deconvolution density estimate. Although Masry and Rice (1992) mainly discuss the normal measurement error case, they did mention that the same technique could be used for Laplace measurement error case.

Hong and Tamer (2003) provided a very nice proof for (1) using the deconvolution relationship between the density functions of $\mathbf{X}$ and $\mathbf{Z}$. Another way of proving (1) is to consistently estimate the left hand side of (1) using the deconvolution kernel density estimate, then show the estimate also converges to the right hand side of (1) in probability. To be specific, let K be a symmetric kernel density function and b_n denote a sequence of bandwidth depending on the sample size n, then a consistent deconvolution kernel estimate for the density function of $\mathbf{X}$ can be written as

$$\hat{f}(\mathbf{x}) = \frac{1}{nb_n^k}\sum_{j=1}^{n}L_n\left(\frac{\mathbf{x}-\mathbf{Z}_j}{b}\right), \tag{2}$$

where $L_n(\mathbf{x})$ is defined by

$$L_n(\mathbf{x}) = \prod_{l=1}^{k}\frac{1}{2\pi}\int \exp(-\mathbf{i}t_l x_l)\frac{\psi_K(t_l)}{\psi_{U_l}(t_k/b_n)}dt, \quad \mathbf{i}^2=-1,$$

where $\mathbf{x}=(x_1,\ldots,x_k)$, $\mathbf{Z}_j=(Z_{j,1},\ldots,Z_{j,k})^\tau$, and for any generic random variable or vector X, ψ_X denotes its characteristic function. If we choose K to be the standard normal density function ϕ, then

$$L_n(\mathbf{x}) = \prod_{l=1}^{k}\left[\phi(x_l) - \frac{\sigma_l^2}{2b_n^2}\phi''(x_l)\right].$$

Accordingly, an estimate of $Em(\mathbf{X};\boldsymbol{\beta})$ can be obtained by directly calculating the expectation of $m(\mathbf{X};\boldsymbol{\beta})$ with respect to $\hat{f}(\mathbf{x})$. In fact

$$\begin{aligned}\widehat{Em(\mathbf{X};\boldsymbol{\beta})} &= \int m(\mathbf{x};\boldsymbol{\beta})\frac{1}{nb_n^k}\sum_{j=1}^{n}L_n\left(\frac{\mathbf{x}-\mathbf{Z}_j}{b_n}\right)d\mathbf{x}\\ &= \frac{1}{nb_n^k}\sum_{j=1}^{n}\int m(\mathbf{x};\beta)\prod_{l=1}^{k}\left[\phi\left(\frac{x_l-Z_{j,l}}{b_n}\right) - \frac{\sigma_l^2}{2b_n^2}\phi''\left(\frac{x_l-Z_{j,l}}{b_n}\right)\right]d\mathbf{x}. \end{aligned} \tag{3}$$

We can show that under some regularity conditions, the right hand side of the above equality is a consistent estimator of the right hand side of the revised moment formula. A sketch of the proof can be found in Section 5.

The significance of Theorem 1 of Hong and Tamer (2003) is demonstrated by the fact that in Laplace measurement error case, the estimating equations based on the true variables, which cannot be used in real application due to the lack of observations for mismeasured variables, can be replaced by estimating equations involving second order partial derivatives of the original estimating functions based on the observed variables. Therefore, one can estimate the unknown parameters just like we are facing a classic estimation problem. Similar phenomenon can be found in nonparametric regression with Laplace measurement error. In fact, Example 2 from Fan and Truong (1993) shows that estimating the regression function is equivalent to estimating up to the second order derivative of the regression function of the response variable on the observed surrogates.

In the nonlinear regression setup, Hong and Tamer (2003) provided explicit revised moment conditions and indicated that these conditions could be justified by using (1), see Example 2 in Hong and Tamer (2003). To be specific, suppose the parametric regression model is $E(Y|X=x)=g(x;\boldsymbol{\beta})$, where Y is a scalar response variable and X is a scalar predictor, g is a known twice differentiable function, and $E[Y^2|X=x]$ is finite. For any measurable function $h(\cdot)$ with finite second moment, possibly vector-valued, we have $E[h(X)(Y-g(X;\boldsymbol{\beta}))]=0$. Then based on (1), the authors claimed that the revised moment conditions are

$$E\left[h(Z)R(Y,Z;\boldsymbol{\beta}) - \frac{\sigma^2}{2}\left(h''(\mathbf{Z})R(Y,Z;\boldsymbol{\beta}) - 2h'(Z)g'(Z;\beta) - h(Z)g''(Z;\boldsymbol{\beta})\right)\right] = 0, \tag{4}$$

where $R(Y, Z; \boldsymbol{\beta}) = Y - g(Z; \boldsymbol{\beta})$, $h'(z)$, $h''(z)$ are the first and second order derivatives of h w.r.t. z, and $g'(z; \boldsymbol{\beta})$, $g''(z; \boldsymbol{\beta})$ are the first and second order derivatives of g w.r.t. z. Considering the regression model as a special case of the triplet $Y = y + u_1$, $Z = X + u_2$, and $y = g(X; \boldsymbol{\beta}) + \varepsilon$, with u_1 and u_2 are independently from Laplace distribution with mean 0 and variance σ_1^2, σ_2^2, respectively, the Eq. (4) can be obtained by treating (Y, Z) as Z, (y, x) as X in (1) and letting $\sigma_1^2 \to 0$. In this paper we shall provide a new proof of (4) directly using the regression structure.

The paper is organized as follows. The revised moment conditions (4) will be presented under some regularity conditions in Section 2. The asymptotic normality of the estimates obtained from the estimating equations based on the revised moment conditions will be developed in Section 3, as well as some discussions on the choice of the weight functions and a simulation extrapolation algorithm to obtain the estimates without using the revised moment conditions. Section 4 will present two simulation studies to evaluate the finite sample performance of the proposed estimates, and the proofs of the main results will be postponed to Section 5. For convenience, we shall assume that X is one-dimensional in the following discussion, and β, instead of the bold $\boldsymbol{\beta}$, will be used to denote the unknown parameter vector.

2. Revised moment conditions

Define $\mu_{j,k} = Eh^{(j)}(X; \beta_0)g^{(k)}(X; \beta_0)$ for $j, k = 0, 1, 2, \ldots$, where $h^{(j)}(x; \beta)$, $g^{(k)}(x; \beta)$ denote the jth and k-th derivatives of h, g with respect to x, respectively. The following result restates the revised moment conditions in nonlinear regression under some regularity conditions.

Theorem 1. *Suppose that the regression function $g(x; \beta)$ and the weight function $h(x)$ are infinitely differentiable with respect to x, and*

$$\sum_{j=2}^{\infty}\sum_{k=2}^{\infty}\left(\frac{(j+k)!|\mu_{j,k}|\sigma^{j+k}}{j!k!2^{(j+k)/2}}\right) < \infty. \tag{5}$$

Then the revised moment condition (4) holds.

The regularity conditions stated in the above theorem is much stronger than the ones imposed by Hong and Tamer (2003), they are adopted for the legitimacy of all the derivations conducted below, involving Taylor expansions, changing the order of expectations and summations. Weaker conditions might exist to serve the same purpose. See Remark 2.

Remark 1. Using Stirling formula for the factorials, we can get a sufficient condition for (5), which is slightly easier to verify. For any positive integers j, k, denote $i = (j + k)/2$. By Stirling formula, we have

$$\frac{(j+k)!}{j!k!} \le \frac{(2i)!}{i!i!} \approx \frac{2^{2i}}{\sqrt{\pi i}} = \frac{2^{j+k+1/2}}{\sqrt{\pi(j+k)}}.$$

This implies that

$$\sum_{j=2}^{\infty}\sum_{k=2}^{\infty}\left(\frac{(j+k)!|\mu_{j,k}|\sigma^{j+k}}{j!k!2^{(j+k)/2}}\right) \le \sum_{j=2}^{\infty}\sum_{k=2}^{\infty}\left(\frac{(2\sigma^2)^{(j+k)/2}|\mu_{j,k}|}{\sqrt{j+k}}\right).$$

Therefore, to verify the convergence of the left series, it is enough to check the finiteness of the right series. As an example, consider an errors-in-variables model in which the regression function $E(Y|X = x) = e^x$ with X being standard normal, u following double exponential with mean 0 and variance σ^2, also ε, X and u being independent. For $h(x) = e^x$, certainly $Eh(X)[Y - g(X)] = 0$. Note that $h^{(k)}(x) = e^x$ and $g^{(k)}(x) = e^x$ for any positive integer k, we can directly verify that

$$\begin{aligned}
&E\left(h(Z)[Y - g(Z)] - \frac{\sigma^2}{2}\left[h^{(2)}(Z)(Y - g(Z)) - 2h^{(1)}(Z)g^{(1)}(Z) - h(Z)g^{(2)}(Z)\right]\right)\\
&= E[e^Z(Y - e^Z)] - \frac{\sigma^2}{2}E\left[e^Z(Y - e^Z) - 2e^{2Z} - e^{2Z}\right]\\
&= \left(1 - \frac{\sigma^2}{2}\right)EYe^Z + (2\sigma^2 - 1)Ee^{2Z}\\
&= \left(1 - \frac{\sigma^2}{2}\right)Ee^{2X}Ee^u + (2\sigma^2 - 1)Ee^{2X}Ee^{2u}\\
&= \left(1 - \frac{\sigma^2}{2}\right)e^2Ee^u + (2\sigma^2 - 1)e^2Ee^{2u}.
\end{aligned}$$

Since $Ee^{tu} = (1 - \sigma^2 t^2/2)^{-1}$ for $|t| < \sqrt{2}/\sigma$, so if $\sigma^2 < 1/2$, we have

$$E\left(h(Z)[Y - g(Z)] - \frac{\sigma^2}{2}\left[h^{(2)}(Z)(Y - g(Z)) - 2h^{(1)}(Z)g^{(1)}(Z) - h(Z)g^{(2)}(Z)\right]\right)$$
$$= \left(1 - \frac{\sigma^2}{2}\right)\frac{2e^2}{2 - \sigma^2} + (2\sigma^2 - 1)\frac{e^2}{1 - 2\sigma^2} = 0.$$

In fact, for this example, we can verify that all the conditions in Theorem 1 holds. Note that for all $j, k = 0, 1, \ldots, \mu_{j,k} = Ee^{2X} = e^2$. Therefore, if $\sigma^2 < 1/2$, we have

$$\sum_{j=2}^{\infty}\sum_{k=2}^{\infty}\left(\frac{(2\sigma^2)^{(j+k)/2}|\mu_{j,k}|}{\sqrt{j+k}}\right) \le e^2 \sum_{j=2}^{\infty}\sum_{k=2}^{\infty}\left(\frac{(2\sigma^2)^{(j+k)/2}}{\sqrt{j+k}}\right) < \infty.$$

Remark 2. The requirement of the infinite differentiability on $g(x;\beta)$, $h(x)$ in Theorem 1 can be weakened to the infinite differentiability almost surely with respect to the distribution of X, which certainly covers many practical situations. Consider the following errors-in-variables model as an example:

$$Y = g(X; \boldsymbol{\beta}) + \varepsilon = \beta_3 X^3 I(X < 0) + \beta_2 X^2 + \beta_1 X + \beta_0 + \varepsilon, \qquad Z = X + u,$$

where X follows standard normal, u follows double exponential with mean 0 and variance σ^2, also ε, X and u are independent. For the sake of simplicity, assume that the true regression parameters are $\boldsymbol{\beta}_0 = (\beta_0, \beta_1, \beta_2, \beta_3) = (1, 1, 1, 1)$. It is easy to verify that as a function of x, $g(x;\beta)$ is infinitely differentiable almost surely with respect to the distribution of X. In fact, the infinite differentiability is violated at $x = 0$. Denote the Laplace density function of u as $f(u) = \exp(-\sqrt{2}|u|/\sigma)/(\sqrt{2}\sigma)$. Choose $h(x) = 1$, certainly we have $Eh(X)[Y - g(X;\boldsymbol{\beta}_0)] = 0$. Note that $h^{(1)}(x) = h^{(2)}(x) = 0$, then the right hand side of (4) can be shown as

$$\begin{aligned} Eh(Z)[Y - g(Z;\boldsymbol{\beta}_0)] + \frac{\sigma^2}{2}Eg''(Z;\boldsymbol{\beta}_0) &= E[X^3 I(X<0) + X^2 + X - Z^3 I(Z<0) - Z^2 - Z] + \frac{\sigma^2}{2}E[6ZI(Z<0) + 2] \\ &= E[X^3 I(X<0) - Z^3 I(Z<0)] + 3\sigma^2 E[ZI(Z<0)] \\ &= -2\phi(0) + \int_{-\infty}^{\infty}[(u^2 + 2 - 3\sigma^2)\phi(u) + (3u + u^3 - 3\sigma^2 u)\Phi(u)]f(u)du \\ &= 0. \end{aligned}$$

3. Large sample properties of revised moment estimates

In this section, we shall assume that β is a p-dimensional vector, and h is a $p+1$ vector of functions. It is easy to see that (4) holds if $h(Z)$ is also a function of β, or $h(Z;\beta)$. For the sake of convenience, denote $H(Y, Z; \beta, \sigma^2)$ as

$$h(Z;\beta)R(Y,Z;\beta) - \frac{\sigma^2}{2}\left(h''(Z;\beta)R(Y,Z;\beta) - 2h'(Z;\beta)g'(Z;\beta) - h(Z;\beta)g''(Z;\beta)\right).$$

Based on the revised moment condition (4), we define an estimate $(\hat{\beta}_n, \hat{\sigma}_n^2)$ of (β, σ^2) to be the solution of the following equation

$$\sum_{i=1}^{n} H(Y_i, Z_i; \beta, \sigma^2) = 0. \tag{6}$$

In fact, the estimation procedure based on (6) and the one proposed in Hong and Tamer (2003) by minimizing

$$Q_n(\beta,\sigma^2) = \left(\frac{1}{n}\sum_{i=1}^{n} H(Y_i, Z_i; \beta, \sigma^2)\right)' W_n \left(\frac{1}{n}\sum_{i=1}^{n} H(Y_i, Z_i; \beta, \sigma^2)\right)$$

are indeed equivalent when the sample size is large enough, where W_n is a known positive definite matrix. In fact, by Assumptions 6 and 7 in Hong and Tamer (2003), when n is sufficiently large, the nonsingularity of W_n and $\sum_{i=1}^{n}\partial H(Y_i, Z_i; \beta, \sigma^2)/\partial(\beta, \sigma^2)$ implies the minimizer of the target function is the same as the solution of (6). This also implies that the weighting function W_n practically has no effects on the estimate and its large sample properties. So it might be more practical to discuss the estimation procedure based on (6).

Denote β_0, σ_0^2 as the true values of β and σ^2. For any generic function $f(x;\beta)$, we use the dots to denote the derivative of f w.r.t. β, and primes to denote the derivative of f w.r.t. x. For example, $\ddot{f}'(x;\beta)$ denotes the second derivative of $f'(x;\beta)$ w.r.t. β, and $f'(x;\beta)$ is the first derivative of f w.r.t. x. In order to derive meaningful large sample results of $(\hat{\beta}_n, \hat{\sigma}_n^2)$, the following technical assumptions are needed.

Assumptions:

(A1) The parameter space Θ of β, and Γ of σ^2 are compact;
(A2) As a function of (β, σ^2), $EH(Y, Z; \beta, \sigma^2)$ is continuous on $\Theta \times \Gamma$;
(A3) $EH(Y, Z; \beta, \sigma^2) = 0$ if and only if $\beta = \beta_0, \sigma^2 = \sigma_0^2$;
(A4) For each y and z, $H(y, z; \beta, \theta)$ is continuous function of (β, θ), and there exists a function $B(y, z) \geq 0$, such that $EB(Y, Z) < \infty$ and $|H(y, z; \beta, \theta)| \leq B(y, z)$;
(A5) There exists a measurable function $D(z) \geq 0$, such that $ED(Z) < \infty$, and $|\ddot{g}(z; \beta)|, |\ddot{g}'(z; \beta)|, |\ddot{g}''(z; \beta)|$ are all bounded above by $D(z)$ for all β;
(A6) Denote $g_0(Z) = g(Z; \beta_0), h_0(Z) = h(Z; \beta_0), A = (A_1(Z; \beta_0, \sigma_0^2), A_1(Z; \beta_0, \sigma_0^2))$ is a nonsingular $(p + 1) \times (p + 1)$ matrix, where

$$A_1(Z; \beta_0, \sigma_0^2) = E\left[h_0(Z)\dot{g}_0^\tau(Z) - \sigma_0^2 h_0''(Z)\dot{g}_0^\tau(Z)/2 - \sigma_0^2 h_0^\tau(Z)\dot{g}_0^\tau(Z) - \sigma_0^2 h_0(Z)\dot{g}_0''^\tau(Z)/2\right],$$
$$A_2(Z; \beta_0, \sigma_0^2) = h_0''(Y - g_0(Z))/2 - h_0'(Z)g_0'(Z) - h_0(Z)g_0''(Z)/2.$$

(A7) The matrix $\Sigma = EH(Y, Z; \beta_0, \sigma_0)H^\tau(Y, Z; \beta_0, \sigma_0^2)$ is positive definite.

Assumptions (A1)–(A4) are needed to guarantee the consistency of the estimate, and together with (A5), (A6) and (A7), asymptotic normality of the estimate can be established. Weaker assumptions exist to achieve the same goal but may require more subtle technicality. The above assumptions are structurally simple and widely adopted in the M-estimation theory.

The consistency and asymptotic normality of $(\hat{\beta}_n, \sigma_n^2)$ are summarized in the following theorem.

Theorem 2. *Under the assumptions* (A1)–(A4), *the estimate* $(\hat{\beta}_n, \hat{\sigma}_n^2)$, *which is the solution of* (6), *is consistent. If we further assume that* (A5), (A6) *and* (A7) *hold, then*

$$\sqrt{n}(\hat{\beta} - \beta_0, \hat{\sigma}_n^2 - \sigma_0^2) \Longrightarrow N(0, (A'\Sigma^{-1}A)^{-1}).$$

Often times, the measurement error variance σ^2 is assumed to be known. In this case, one can choose h to be a p-dimensional vector of functions, then under slightly modified conditions, we can show $\sqrt{n}(\hat{\beta}_n - \beta_0) \Longrightarrow N(0, (A_1'\Sigma^{-1} A_1)^{-1})$. In this case, the estimate from (6) has a close connection to the nonlinear least squares estimation procedure. In fact, the nonlinear least squares estimate of β tries to minimize $n^{-1}\sum_{i=1}^n (Y_i - g(X_i; \beta))^2$, and is the solution of $n^{-1}\sum_{i=1}^n (Y_i - g(X_i; \beta))\dot{g}(X_i; \beta) = 0$, which is the empirical version of $E(Y - g(X; \beta))\dot{g}(X; \beta) = 0$. However, $E(Y - g(X; \beta))\dot{g}(X; \beta) = EH(Y, Z; \beta, \sigma^2)$ by Theorem 1 with $h(x) = \dot{g}(x; \beta)$.

An interesting question about the estimation procedure (6) is how to choose the weight function h. The general guideline should be to choose the one to minimize the asymptotic variances of the estimates. For example, one can choose h to minimize the trace of the asymptotic covariance matrix. Given the complexity of the asymptotic covariance matrix, an analytically tractable optimal h seems very hard to find. But if σ^2 is known, then a viable method could be going back to the original moment condition $Eh(X; \beta)(Y - g(X; \beta)) = 0$, treating X to be observable, and looking for a h such that the solution of $\sum_{i=1}^n h(X_i; \beta)(Y_i - g(X_i\beta)) = 0$ has an optimal asymptotic covariance matrix. In fact, in this case, we can show that

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \left[n^{-1}\sum_{i=1}^n h(X_i; \beta_0)\dot{g}(X_i; \beta_0)\right]^{-1} \sqrt{n}\sum_{i=1}^n h(X_i; \beta_0)[Y_i - g(X_i; \beta_0)] + o_p(1).$$

If $Y_i - g(X_i; \beta_0),\ i = 1, \ldots, n$ are i.i.d. with $E(Y_i - g(X_i; \beta_0)|X_i) = 0$, and $E(Y_i - g(X_i; \beta_0)^2|X_i) = \sigma_\varepsilon^2$, then we can show that $\sqrt{n}(\hat{\beta}_n - \beta_0) \Longrightarrow N(0, \Omega)$, where

$$\Omega = \sigma_\varepsilon^2[E(h_0(X)\dot{g}_0^\tau(X))]^{-1}Eh_0(X)h_0^\tau(X)[E(\dot{g}_0(X)h_0^\tau(X))]^{-1}.$$

From Tripathi (1999), for any h, we must have

$$[E(h_0(X)\dot{g}_0^\tau(X))][Eh_0(X)h_0^\tau(X)]^{-1}[E(\dot{g}_0(X)h_0^\tau(X))] \leq [E(\dot{g}_0(X)\dot{g}_0^\tau(X))]^{-1}.$$

So, if we choose $h(x; \beta) = \dot{g}(x; \beta_0)$, the asymptotic covariance matrix of $\hat{\beta}_n$ attains its minimum in the sense that, for two nonnegative matrices M_1, M_2, M_1 is smaller, if $M_2 - M_1 \geq 0$. Therefore, one may try $h(Z; \beta) = \dot{g}(Z; \beta)$ in the revised moment estimation procedure.

Due to its clean mathematical structure and nice asymptotic theory, the revised moment estimation procedure can certainly attract much attention from applied statisticians. However, the relatively complicated revised moment estimating equation (6) might create some computational difficulties when implementing the method. This is extremely undesirable for the applied statisticians, in particular at the beginning stage of study, they might only want to do some exploratory analysis. To satisfy this need, we suggest the following simulation extrapolation (SIMEX) estimation procedure for nonlinear regression when covariate is measured with Laplace error under the assumption that σ^2 is known. The SIMEX procedure is very useful when the estimating equations based on the true covariates are easy to solve. The general theory of SIMEX with Laplace measurement error is well discussed in Koul and Song (2014), so we only introduce how to use this method here without any justification.

Suppose $\hat{\beta}_{\text{TRUE}} = T(\{Y_i, X_i\}_{i=1}^n)$ is the solution of the estimating equation $\sum_{i=1}^n h(X_i; \beta)[Y_i - g(X_i; \beta)]$. $\hat{\beta}_{\text{TRUE}}$ is not a real estimate since X_i, $i = 1, 2, \ldots$ are not observed. The SIMEX estimate for β can be obtained as follows. Fix a $p \geq 0$. In the simulation step, we first generate two sequences of Gamma random variables of length n each, denote them as $\{V_{1,b,i}\}_{i=1}^n$ and $\{V_{2,b,i}\}_{i=1}^n$, which are independent of all other random variables in the model. Then define

$$Z_{b,i}(p) = Z_i + V_{1,b,i} - V_{2,b,i} = X_i + U_i + V_{1,b,i} - V_{2,b,i}.$$

Here, b is a positive integer-valued index number. Let $\hat{\beta}_b(p) = T(\{Y_i, Z_{b,i}(p)\}_{i=1}^n)$, and calculate $\hat{\beta}(p) = E[\hat{\beta}_b(p)|\{Y_i, Z_i\}_{i=1}^n]$. Note that the expectation is with respect to the distribution of $\{V_{1,b,i}, V_{2,b,i}\}_{i=1}^n$ only. This expectation might not have an explicit form, but it can be approximated arbitrarily well by the average of $\hat{\beta}_b(p)$, $b = 1, 2, \ldots, B$, based on B random samples $\{V_{1,b,i}, V_{2,b,i}\}_{i=1}^n$, $b = 1, 2, \ldots, B$. Repeat the above computations for a sequence of p values $0 = p_1 < p_2 < \cdots < p_{k-1} < p_k$ for some positive $k > 1$. The convention we make here is that $p = 0$ corresponding to the generated Gamma data all being 0.

In the extrapolation step, unless the true relationship between $\hat{\beta}(p)$ can be identified, a trend of $\hat{\beta}(p)$ with respect to p should be approximately determined by checking the scatter plot of $(p_j, \hat{\beta}(p_j))$, $j = 1, 2, \ldots, k$. Again, least square procedure might be needed to decide a reasonable analytic form of the extrapolation function. Once the trend is set, then extrapolating the final trend back to $p = -1$ gives the SIMEX estimate.

To estimate the variances of the SIMEX estimates for the regression parameters, let $T_{\text{Var}}(\{Y_i, X_i\})$ denote an asymptotically unbiased estimate of $\text{Var}(\hat{\beta}_{\text{TRUE}})$. Then an estimate of $\text{Var}(\hat{\beta}_{\text{SIMEX}})$ can be constructed via the following steps. First calculate $T_{\text{Var}}(\{Y_i, Z_{b,i}(p)\})$ for $b = 1, 2, \ldots, B$, and denote the average as $\hat{\tau}^2(p)$. If possible, calculate the limit of the average as $B \to \infty$, and denote the limit as $\hat{\tau}^2(p)$; then calculate the sample variance of $\hat{\beta}_b(p)$, $b = 1, 2, \ldots, B$. Denote it as $s_\Delta^2(p)$; finally, extrapolating $\hat{\tau}^2(p) - s_\Delta^2(p)$ to $p = -1$ to get an estimator of the variance of $\hat{\beta}_{\text{SIMEX}}$. If the exact extrapolant is used, then the resulting estimator is asymptotically unbiased.

4. Numerical study

In this section, we conduct two simulation studies. Simulation 1 considers the log model discussed in Hong and Tamer (2003), we shall compare the finite sample performance of the revised moment estimates with four different sets of weight functions. Simulation 2 considers the logit model discussed in Hong and Tamer (2003), the finite sample performance of the revised moment estimates, SIMEX estimates will be compared. For simplicity, and also in order to use SIMEX, σ^2 will be assumed to be known.

Simulation 1 (log model): The log model is generally used in the estimation of Engle curves. It relates the response variable Y and the scalar latent variable X through $Y = \beta_0 + \beta_1 \log(X^2 + 1) + \varepsilon$. For any measurable function $h(x)$, it is easy to see that $E[h(X)(Y - \beta_0 - \beta_1 \log(X^2 + 1))] = 0$. We choose four different weight functions h to see their effects on the estimation of β_0 and β_1.

$$h_1(x) = \begin{pmatrix} 1 \\ x \end{pmatrix}, \qquad h_2(x) = \begin{pmatrix} 1 \\ x^3 \end{pmatrix}, \qquad h_3(x) = \begin{pmatrix} x \\ x^3 \end{pmatrix}, \qquad h_4(x) = \begin{pmatrix} 1 \\ \log(x^2 + 1) \end{pmatrix}.$$

The components in the first three simple weight functions are used in Hong and Tamer (2003) to estimate β_0, β_1 when σ^2 is assumed to be unknown, the weight function $h_4(x) = \dot{g}(x; \beta)$. With such choices of h, the empirical versions of the revised moment conditions have explicit solutions. For example, when h_4 is used, the revised moment conditions (4) take the form of

$$E[Y - \beta_0 - \beta_1 B_0(Z)] = 0, \qquad E[B(Y, Z) - \beta_0 B_1(Z) - \beta_1 B_2(Z)] = 0,$$

where

$$B_0(Z) = \log(Z^2 + 1) + \frac{\sigma^2(Z^2 - 1)}{(Z^2 + 1)^2},$$

$$B(Y, Z) = Y \log(Z^2 + 1) + \frac{\sigma^2 ZY}{(Z^2 + 1)^2}, \qquad B_1(Z) = \log(Z^2 + 1) + \frac{Z\sigma^2}{(Z^2 + 1)^2}$$

and

$$B_2(Z) = \log^2(Z^2 + 1) + \frac{Z\sigma^2 \log(Z^2 + 1)}{(Z^2 + 1)^2} - \frac{2Z\sigma^2}{(Z^2 + 1)^2} - \frac{\sigma^2(1 - Z^2)\log(Z^2 + 1)}{(Z^2 + 1)^2}.$$

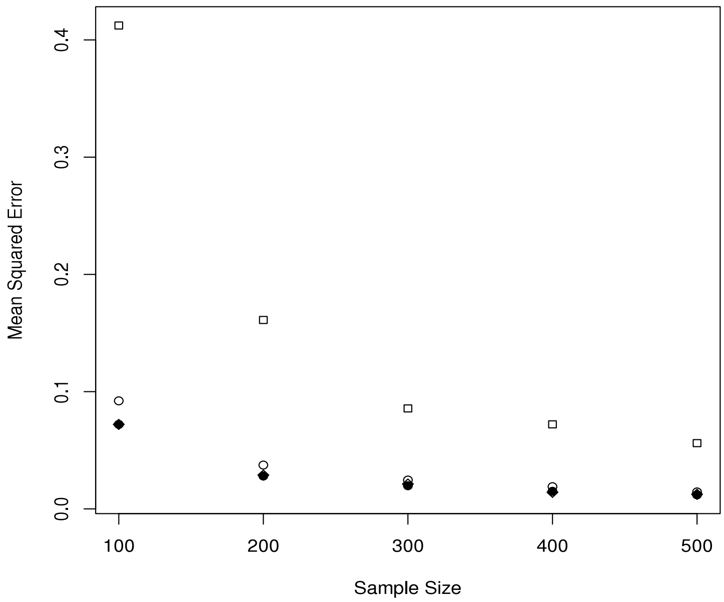

Fig. 1. Mean squared errors of estimates for β_0.

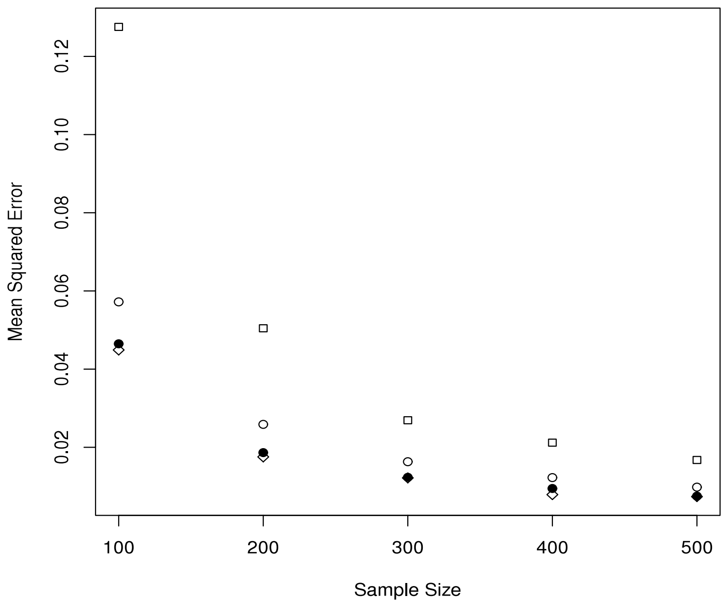

Fig. 2. Mean squared errors of estimates for β_1.

The empirical version of revised moment conditions are

$$\begin{pmatrix} n & \sum_{i=1}^{n} B_0(Z_i) \\ \sum_{i=1}^{n} B_1(Z_i) & \sum_{i=1}^{n} B_2(Z_i) \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} B(Y_i, Z_i) \end{pmatrix}.$$

In the simulation, choose $\beta_0 = \beta_1 = \sigma^2 = 1$, and the latent variable $X \sim \chi_1^2 + U(0, 1)$, and $\varepsilon \sim N(0, 1)$. The sample size n is chosen to be 100, 200, 300, 400 and 500, and the simulation is replicated 500 times for each sample size. Since the revised moment equations have an explicit solution, so there is no need to use SIMEX. The mean squared errors (MSE) of 500 estimates are used to evaluate the performance of the revised moment estimation procedures with different weight functions. Fig. 1 is the simulation results for estimates of β_0 and Fig. 2 is for β_1. In both figures, the diamonds, circles,

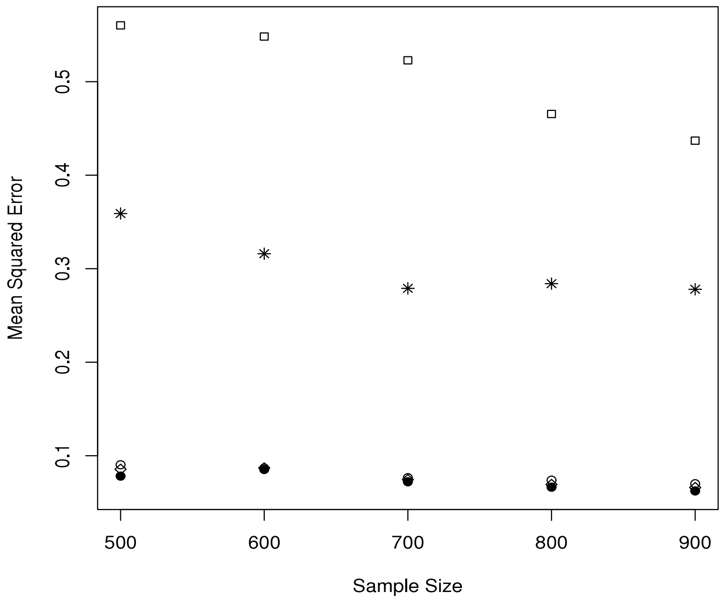

Fig. 3. Mean squared errors of estimates for β_0.

squares and solid circles correspond to the MSEs with weight function $h_1(x)$, $h_2(x)$, $h_3(x)$ and $h_4(x)$, respectively. Clearly, the estimates with the weight function $h_4(x)$ are comparable or outperforms the other three estimates.

Simulation 2 (logistic model): In this simulation study, we consider the model $Y = I[\beta_0 + \beta_1 X + \varepsilon \geq 0]$, where ε follows a standard logistic distribution. The same model is also discussed in Hong and Tamer (2003), with the exception of σ^2 being known here. For any measurable function $h(x)$, we have $E[h(X)(Y - F(\beta_0 + \beta_1 X))]$, where F is the logistic CDF. In this case, the revised moment equations have complicated forms. The following four different weight functions h are used to see their effects on the estimation of β_0 and β_1,

$$h_1(x) = \begin{pmatrix} 1 \\ x \end{pmatrix}, \qquad h_2(x) = \begin{pmatrix} 1 \\ x^3 \end{pmatrix}, \qquad h_3(x) = \begin{pmatrix} x^2 \\ x^3 \end{pmatrix}, \qquad h_4(x) = \begin{pmatrix} \dfrac{\exp(\beta_0 + \beta_1 x)}{[1 + \exp(\beta_0 + \beta_1 x)]^2} \\ \dfrac{x \exp(\beta_0 + \beta_1 x)}{[1 + \exp(\beta_0 + \beta_1 x)]^2} \end{pmatrix}.$$

Note that here $h_4(x) = \dot{g}(x; \beta) = \dot{F}(\beta_0 + \beta_1 x)$. The values of the true parameters β_0, β_1 and σ^2 are chosen to be 1's, and the logistic distribution has location parameter 0 and scale parameter 1. The sample size n is chosen to be 500, 600, 700, 800 and 900, and the simulation is replicated 200 times for each sample size. As in Simulation 1, the MSE is used to evaluate the performance of the revised moment estimation procedures. Fig. 3 is the simulation results for estimates of β_0 and Fig. 4 is for β_1. In both figures, the diamonds, circles, squares and solid circles correspond to the MSEs with weight function $h_1(x)$, $h_2(x)$, $h_3(x)$ and $h_4(x)$, respectively. The stars denote the MSEs of the SIMEX estimates based on the algorithm proposed in Section 3. Again, the estimates with the weight function $h_4(x)$ behaves better than all other three estimates in most cases.

Although SIMEX estimates have larger MSEs, it still provides reasonable good estimates for the parameters. Since SIMEX only needs a solution for the empirical version of the moment condition $Eh(X)(Y - g(X; \beta))$, which is usually much simpler than the one based on the revised moment condition, so it remains a competitive estimation procedure over the revised moment estimation methods, in particular, if the empirical version of the revised moment condition is hard to solve.

5. Proofs of main results

We begin with showing (1) by using the deconvolution kernel estimate of $f_X(x)$. Without loss of generality, we shall assume that $k = 1$. The following technical assumptions are adopted to facilitate the proof: (i) for any sufficiently small h, $\iiint |m(w + v + xh)|\phi(x) f_X(w) f_u(v) dx dw dv < \infty$; (ii) the second derivative of $f_x(x)$, $m(x; \beta)$ with respected to x are bounded. Assumption (i) ensures that the order of multiple integration is interchangeable, and Assumption (ii) is commonly used when discussing the consistency in kernel smoothing. Alternative and weaker conditions might exist to serve the same purpose, but we will not explore this possibility here.

On the one hand, by Assumption (i) and (3), we have

$$E(\widehat{Em(X; \beta)}) = \frac{1}{h} \iiint m(x; \beta) L_n \left(\frac{x - w - v}{h} \right) f_x(w) f_u(v) dx dw dv.$$

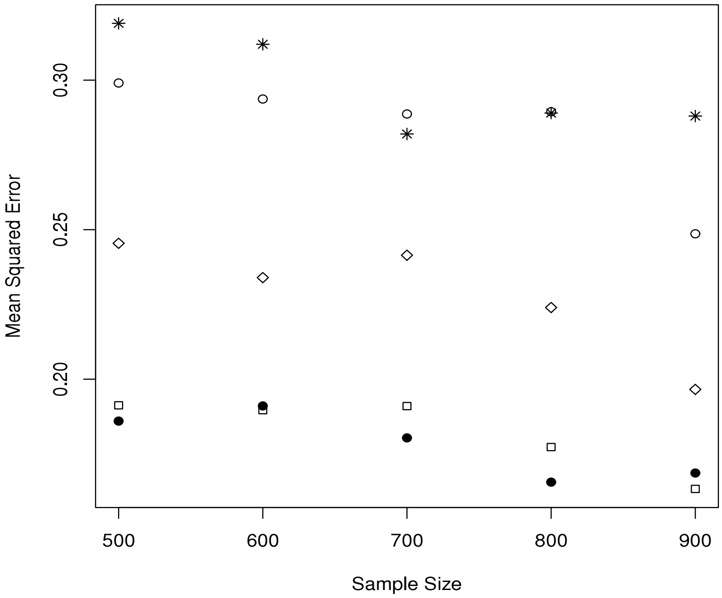

Fig. 4. Mean squared errors of estimates for β_1.

From the definition of L_n in (2), the right hand side can be written as

$$\begin{aligned}
&\frac{1}{h}\iiint m(x;\beta)\frac{1}{2\pi}\int \exp\left(\frac{-\mathbf{i}t(x-w-v)}{h}\right)\frac{\phi_K(t)}{\phi_u(t/h)}dtf_x(w)f_u(v)dxdwdv\\
&=\frac{1}{h}\iiint m(x;\beta)\frac{1}{2\pi}\int \exp\left(\frac{\mathbf{i}tv}{h}\right)f_u(v)dv\exp\left(\frac{-\mathbf{i}t(x-w)}{h}\right)\frac{\phi_K(t)}{\phi_u(t/h)}f_x(w)dtdxdw\\
&=\frac{1}{h}\iiint m(x;\beta)\frac{1}{2\pi}\exp\left(\frac{-\mathbf{i}t(x-w)}{h}\right)\phi_K(t)f_x(w)dtdxdw\\
&=\frac{1}{h}\iint m(x;\beta)\frac{1}{2\pi}\int \exp\left(\frac{-\mathbf{i}t(x-w)}{h}\right)\phi_K(t)dtf_x(w)dxdw\\
&=\frac{1}{h}\iint m(x;\beta)\phi\left(\frac{x-w}{h}\right)f_x(w)dwdx=\iint m(x;\beta)\phi(w)f_x(x-wh)dxdw,
\end{aligned}$$

which converges to $Em(X;\beta)$ by Assumption (ii).

On the other hand, the expectation of the right hand side in (3) has the form of

$$\begin{aligned}
&\frac{1}{h}E\int m(x;\beta)\left[\phi\left(\frac{x-Z}{h}\right)-\frac{\sigma^2}{2h^2}\phi''\left(\frac{x-Z}{h}\right)\right]dx\\
&=\frac{1}{h}E\int m(x;\beta)\phi\left(\frac{x-Z}{h}\right)dx-\frac{\sigma^2}{2h^3}E\int m(x;\beta)\phi''\left(\frac{x-Z}{h}\right)dx.
\end{aligned}$$

By Assumption (ii), the first term converges to $Em(Z;\beta)$, and the second term, using integration by parts, equals to

$$-\frac{\sigma^2}{2h}E\int m''(x;\beta)\phi\left(\frac{x-Z}{h}\right)dx$$

which tends to $-\sigma^2Em''(Z;\beta)/2$ as $h\to\infty$ by the dominated convergence theorem.

To finish the proof, we shall show that the variance of the right hand side in (3) converges to 0 as $n\to\infty$. In fact, since the variance is bounded above by the second moment, using the elementary inequality $(a+b)^2\le 2a^2+2b^2$, we obtain

$$\begin{aligned}
\mathrm{Var}((\widehat{Em(X;\beta)}))&\le\frac{2}{nh^2}E\left[\int m(x;\beta)\phi\left(\frac{x-Z}{h}\right)dx\right]^2+\frac{\sigma^4}{2nh^6}E\left[\int m(x;\beta)\phi''\left(\frac{x-Z}{h}\right)dx\right]^2\\
&=\frac{2}{n}E\left[\int m(Z+xh;\beta)\phi(x)dx\right]^2+\frac{\sigma^4}{2n}E\left[\int m''(Z+xh;\beta)\phi(x)dx\right]^2,
\end{aligned}$$

which is the order of $O(1/n)$ by Assumption (ii). This completes the proof.

Proof of Theorem 1. For convenience, denote $h_k := h^{(k)}(X)$, $g_k := g^{(k)}(X, \beta)$. Note that ε, X, u are independent, and $E\varepsilon u^k = 0$ for any integer $k > 0$, we obtain

$$Eh(Z)[Y - g(Z,\beta)] = Eh(X+u)[Y - g(X+u,\beta)] = -E\left(\sum_{j=0}^{\infty}\sum_{k=1}^{\infty}\frac{h_j g_k}{j!k!}u^{j+k}\right),$$

$$Eh_2(Z)[Y - g(Z,\beta)] = Eh_2(X+u)[Y - g(X+u,\beta)] = -E\left(\sum_{j=0}^{\infty}\sum_{k=1}^{\infty}\frac{h_{2+j} g_k}{j!k!}u^{j+k}\right),$$

$$Eh_1(Z)g_1(Z,\beta) = E\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{h_{1+j} g_{1+k}}{j!k!}u^{j+k}\right), \qquad Eh(Z)g_2(Z,\beta) = E\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{h_j g_{2+k}}{j!k!}u^{j+k}\right).$$

It is well known that if $u \sim$ Laplace distribution with mean 0 and variance σ^2, then

$$Eu^k = \frac{k!\sigma^k}{2^{k/2}}, \quad \text{when } k \text{ is even; 0 otherwise.}$$

For convenience, denote $\mu_{j,k} = E(h_j g_k)$. Therefore,

$$Eh(Z)[Y - g(Z,\beta)] = -\sum_{j=0}^{\infty}\sum_{k=1}^{\infty}\frac{(j+k)!\mu_{j,k}\sigma^{j+k}}{j!k!2^{(j+k)/2}}$$

$$-\frac{\sigma^2}{2}Eh_2(Z)[Y - g(Z,\beta)] = \sum_{j=0}^{\infty}\sum_{k=1}^{\infty}\frac{(j+k)!\mu_{j+2,k}\sigma^{j+k+2}}{2j!k!2^{(j+k)/2}} = \sum_{j=2}^{\infty}\sum_{k=1}^{\infty}\frac{(j+k-2)!\mu_{j,k}\sigma^{j+k}}{2(j-2)!k!2^{(j+k-2)/2}}$$

$$\sigma^2 Eh_1(Z)g_1(Z,\beta) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{(j+k)!\mu_{1+j,1+k}\sigma^{j+k+2}}{j!k!2^{(j+k)/2}} = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\frac{(j+k-2)!\mu_{j,k}\sigma^{j+k}}{(j-1)!(k-1)!2^{(j+k-2)/2}}$$

$$\frac{\sigma^2}{2}Eh(Z)g_2(Z,\beta) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{(j+k)!\mu_{j,2+k}\sigma^{j+k+2}}{2j!k!2^{(j+k)/2}} = \sum_{j=0}^{\infty}\sum_{k=2}^{\infty}\frac{(j+k-2)!\mu_{j,k}\sigma^{j+k}}{2j!(k-2)!2^{(j+k-2)/2}}.$$

After some algebra, we can show that

$$\begin{aligned}
Eh(Z)[Y - g(Z,\beta)] &= -\left[\sum_{j=0}^{1}\sum_{k=1}^{1} + \sum_{j=0}^{1}\sum_{k=2}^{\infty} + \sum_{j=2}^{\infty}\sum_{k=1}^{1} + \sum_{j=2}^{\infty}\sum_{k=2}^{\infty}\right]\left(\frac{(j+k)!\mu_{j,k}\sigma^{j+k}}{j!k!2^{(j+k)/2}}\right) \\
&= -\sigma^2\mu_{1,1} - \sum_{k=2}^{\infty}\frac{\mu_{0,k}\sigma^k}{2^{k/2}} - \sum_{k=2}^{\infty}\frac{\mu_{1,k}(k+1)\sigma^{k+1}}{2^{(k+1)/2}} \\
&\quad - \sum_{j=2}^{\infty}\frac{\mu_{j,1}(j+1)\sigma^{j+1}}{2^{(j+1)/2}} - \sum_{j=2}^{\infty}\sum_{k=2}^{\infty}\frac{(j+k)!\mu_{j,k}\sigma^{j+k}}{j!k!2^{(j+k)/2}} \\
&= -T_{11} - T_{12} - T_{13} - T_{14} - T_{15}.
\end{aligned}$$

$$\begin{aligned}
-\frac{\sigma^2}{2}Eh_2(Z)[Y - g(Z,\beta)] &= \left[\sum_{j=2}^{\infty}\sum_{k=1}^{1} + \sum_{j=2}^{\infty}\sum_{k=2}^{\infty}\right]\left(\frac{(j+k-2)!\mu_{j,k}\sigma^{j+k}}{2(j-2)!k!2^{(j+k-2)/2}}\right) \\
&= \sum_{j=2}^{\infty}\frac{\mu_{j,1}(j-1)\sigma^{j+1}}{2\cdot 2^{(j-1)/2}} + \sum_{j=2}^{\infty}\sum_{k=2}^{\infty}\frac{(j+k-2)!\mu_{j,k}\sigma^{j+k}}{2(j-2)!k!2^{(j+k-2)/2}} \\
&= T_{21} + T_{22}.
\end{aligned}$$

$$\begin{aligned}
\sigma^2 Eh_1(Z)g_1(Z,\beta) &= \left[\sum_{j=1}^{1}\sum_{k=1}^{1} + \sum_{j=1}^{1}\sum_{k=2}^{\infty} + \sum_{j=2}^{\infty}\sum_{k=1}^{1} + \sum_{j=2}^{2}\sum_{k=2}^{2}\right]\left(\frac{(j+k-2)!\mu_{j,k}\sigma^{j+k}}{(j-1)!(k-1)!2^{(j+k-2)/2}}\right) \\
&= \sigma^2\mu_{1,1} + \sum_{k=2}^{\infty}\frac{\mu_{1,k}\sigma^{k+1}}{2^{(k-1)/2}} + \sum_{j=2}^{\infty}\frac{\mu_{j,1}\sigma^{j+1}}{2^{(j-1)/2}} + \sum_{j=2}^{\infty}\sum_{k=2}^{\infty}\frac{(j+k-2)!\mu_{j,k}\sigma^{j+k}}{(j-1)!(k-1)!2^{(j+k-2)/2}} \\
&= T_{31} + T_{32} + T_{33} + T_{34}.
\end{aligned}$$

$$
\begin{aligned}
\frac{\sigma^2}{2} Eh(Z)g_2(Z,\beta) &= \left[\sum_{j=0}^{1}\sum_{k=2}^{\infty} + \sum_{j=2}^{\infty}\sum_{k=2}^{\infty}\right]\left(\frac{(j+k-2)!\mu_{j,k}\sigma^{j+k}}{2j!(k-2)!2^{(j+k-2)/2}}\right) \\
&= \sum_{k=2}^{\infty}\frac{\mu_{0,k}\sigma^k}{2\cdot 2^{(k-2)/2}} + \sum_{k=2}^{\infty}\frac{\mu_{1,k}(k-1)\sigma^{k+1}}{2\cdot 2^{(k-1)/2}} + \sum_{j=2}^{\infty}\sum_{k=2}^{\infty}\frac{(j+k-2)!\mu_{j,k}\sigma^{j+k}}{2j!(k-2)!2^{(j+k-2)/2}} \\
&= T_{41} + T_{42} + T_{43}.
\end{aligned}
$$

Eventually, we can verify that

$$
\begin{aligned}
&E\left\{h(Z)[Y-g(Z,\beta)] - \frac{\sigma^2}{2}\left(h_2(Z)[Y-g(Z,\beta)] - 2h_1(Z)g_1(Z,\beta) - h(Z)g_2(Z,\beta)\right)\right\} \\
&= -T_{11} - T_{12} - T_{13} - T_{14} - T_{15} + T_{21} + T_{22} + T_{31} + T_{32} + T_{33} + T_{34} + T_{41} + T_{42} + T_{43} \\
&= (T_{31} - T_{11}) + (T_{41} - T_{12}) + (T_{32} + T_{42} - T_{13}) + (T_{21} + T_{33} - T_{14}) + (T_{22} + T_{33} + T_{43} - T_{15}) \\
&= 0. \quad \square
\end{aligned}
$$

Acknowledgment

The second author's research is supported by the grant NSF DMS 1205276.

References

Fan, J., Truong, Y.K., 1993. Nonparametric regression with errors in variables. Ann. Statist. 21, 1900–1925.
Fuller, W.A., 2006. Measurement Error Models. Wiley-Interscience.
Hong, H., Tamer, E., 2003. A simple estimator for nonlinear error in variable models. J. Econometrics 117, 1–19.
Koul, H.L., Song, W.X., 2014. Simulation extrapolation estimation in parametric models with Laplace measurement error. Electron. J. Statist. 8, 1973–1995.
Masry, E., Rice, J.A., 1992. Gaussian deconvolution via differentiation. Canad. J. Statist. 20, 9–21.
Tripathi, G., 1999. A matrix extension of the Cauchy–Schwarz inequality. Econom. Lett. 63, 1–3.