Journal of the Korean
Statistical Society
Vol. 24, No. 2, 1995

Finite-Sample, Small-Dispersion Asymptotic Optimality of the Non-Linear Least Squares Estimator †

Beong-Soo So[1]

ABSTRACT

We consider the following type of general *semi-parametric* non-linear regression model : $y_i = f_i(\theta) + \epsilon_i, i = 1, \ldots, n$ where $\{f_i(\cdot)\}$ represents the set of non-linear functions of the unknown parameter vector $\theta' = (\theta_1, \ldots, \theta_p)$ and $\{\epsilon_i\}$ represents the set of measurement errors with unknown distribution. Under suitable *finite-sample, small-dispersion* asymptotic framework, we derive a general lower bound for the asymptotic mean squared error (AMSE) matrix of the *Gauss-consistent* estimator of θ. We then prove the fundamental result that the general non-linear least squares estimator (NLSE) is an *optimal* estimator within the class of all regular Gauss-consistent estimators irrespective of the type of the distribution of the measurement errors.

KEYWORDS: Small-dispersion asymptotics, Gauss-consistency, Non-linear least-squares, Semi-parametric non-linear model.

†This research was supported by NON-DIRECTED RESEARCH FUND, Korea Research Foundation, 1993.

[1]Department of Statistics, Ewha Womans University, Seoul, 120–750, Korea.

1. INTRODUCTION

We consider the following type of general non-linear regression model :

$$y_i = f_i(\theta) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where $\{f_i(\theta)\}$ represents the set of non-linear functions of the unknown parameter vector $\theta' = (\theta_1, \ldots, \theta_p)$ of known functional form and $\{\epsilon_i\}$ represents the set of uncorrelated measurement errors with mean zero and constant variance but with otherwise *unknown* distribution. One of the most important problems in this model is to find the efficient method of estimating regression parameters θ which minimizes the influence of the unavoidable measurement errors. The research on this important problem has been conducted in the literature along the two different directions depending on the asymptotic framework used in evaluating the performances of alternative estimators.

In the usual *large-sample* asymptotic framework, we assume that the sample size n is large and try to find the asymptotic distribution of the reasonable estimator such as least-squares estimator. Typical results in this direction are Hartley and Booker (1965), Jennrich (1969), Wu (1981). Essentially they established consistency and the asymptotic normality of the non-linear least-squares estimator as sample size n tends to infinity under suitable regularity conditions. As closely related researches, we can also mention the important recent works on the Quasi-likelihood estimators in the GLM (Generalized Linear Model) by, among others, Wedderburn (1974) and McCullagh (1983). Especially McCullagh (1983) showed that Quasi-likelihood estimators are optimal in the class of asymptotically unbiased estimators under appropriate regularity conditions as n tends to infinity.

On the other hand, in the so-called *finite-sample small-dispersion* asymptotic framework, we assume that the sample size n is *finite* but instead suppose that the variance of the measurement errors $\sigma^2 = \text{Var}(\epsilon_i)$ is small. These conditions are typically satisfied for the experimental data from the fields such as astronomy, survey, geophysics, chemistry, and physics where high cost of the experiment does not allow large number of experiments but each experiment can be conducted by the instrument with relatively high precision. As a first related work in this area we can mention Villegas (1969) who considered the problem of efficient estimation in the non-linear functional relation model when there exist replicated observations. Recently Jorgensen (1987) also made it clear that two different asymptotics are possible in the so-called exponential dispersion model depending on the validity of the two different assumptions on the magnitude of the sample size n and that of the dispersion parameter σ^2.

In this paper we consider the problem of optimal estimation of the Euclidean parameter θ of the semi-parametric non-linear regression model (1.1) from the finite-sample small-dispersion asymptotic viewpoint and try to establish a new small-sample optimality of the non-linear least squares estimator which is completely independent of the type of the measurement errors as long as they are uncorrelated. This small-sample optimality result is in sharp contrast with the classical large-sample result of optimal estmator which depends heavily on the type of error distribution.

Specifically, motivated by the equivalence of the unbiasedness and the *Gauss-consistency* in the Gauss-Markov theorem and by the usefulness of the *Fisher-consistency* in the multinomial estimation problem, we introduce the fundamental concept of Gauss-consistency as a most natural non-linear generalization of the unbiasedness concept of the linear model.

We then derive a general lower bound for the AMSE (Asymptotic Mean Squared Error) matrix of the regular Gauss-consistent estimator. We also show that non-linear least-squares estimator (NLSE) is Gauss-consistent and has the smallest AMSE matrix among all regular Gauss-consistent estimators irrespective of the type of the error distribution. We also discuss some examples including small-sample calibration problem as applications of our optimality results.

2. MAIN RESULTS

As a most natural non-linear extension of the linear-unbiased estimator in the linear model, we first introduce the fundamental concept of Gauss-consistency as follows :

Definition 1. We call an estimator $h(y)$ of θ is *Gauss-consistent* if

$$h[f(\theta)] = \theta \quad \text{for all } \theta \in \Theta, \tag{2.1}$$

where $y' = (y_1, \ldots, y_n)$, $f(\theta)' = (f_1(\theta), \ldots, f_n(\theta))$ and Θ is a parameter space in R^p.

Remark 1. In the linear model, Gauss-consistency is equivalent to the unbiasedness for linear estimators.

Remark 2. In the multinomial set up, Gauss-consistency reduces to the usual Fisher-consistency. See Neyman (1949) and Bemis and Bhapkar (1983) for more details on this topics.

Remark 3 (Invariance Property). In contrast to the unbiasedness, Gauss-consistency is preserved under arbitrary non-linear transformation of the parameters.

Remark 4. On the insightful discussion of the history of the least-squares method and the relevance of the Gauss-consistency instead of the unbiasedness as a fundamental requirement of the reasonable estimator, see Sprott (1983).

We also need the definition of the regular estimators in the following.

Definition 2. We say a Gauss-consistent estimator $h(y)$ of θ *regular* Gauss-consistent if $h(y)$ is a *continuously differentiable* function of y in some neighborhood $N \subset R^n$ of the set $S = \{f(\theta) \in R^n; \theta \in \Theta\}$.

Since we are interested in the asymptotic distribution of the Gauss-consistent estimator as the dispersion parameter σ gets small i.e. as $\sigma \rightarrow 0$ for fixed finite sample size n, we will assume the following regularity conditions in this paper.

$\boldsymbol{A_1}$: Θ is an open set in R^p.

$\boldsymbol{A_2}$: Let $\epsilon_i = \sigma Z_i$, $i = 1, \ldots, n$. Then

$$\mathcal{L}[Z|\theta] \to \mathcal{L}[U|\theta] \text{ as } \sigma \to 0,$$

where $Z' = (Z_1, \ldots, Z_n)$, $U' = (U_1, \ldots, U_n)$ and $E[UU'] = I = [\delta_{ij}]$ is an $n \times n$ identity matrix.

$\boldsymbol{A_3}$: The mapping $f(\theta)$ from Θ into R^n is *homeomorphic* (that is one-to-one and bicontinuous) and continuously differentiable.

Now we derive a general lower bound for the AMSE matrix of the regular Gauss-consistent estimator of θ.

Theorem 1. (Lower Bound for AMSE) Let the conditions A_1, A_2, A_3 be satisfied. Let $X(\theta) = Df(\theta) = [\partial f_i/\partial\theta_j]$ be the $n \times p$ Jacobian matrix of $f(\theta)$ with a full rank $p \leq n$. Let $h(\cdot)$ be a regular Gauss-consistent estimator of θ with a $p \times n$ Jacobian matrix $H(f(\theta)) = [h_{ij}(f(\theta))] = [\partial h_i/\partial y_j]$. Then we have :

$$AMSE_\theta[h(\cdot)] = H(f(\theta))H'(f(\theta)) \geq [X'(\theta)X(\theta)]^{-1} \tag{2.2}$$

and the equality holds if and only if $H = (X'X)^{-1}X'$ where $A \geq B$ means $A - B$ is positive semidefinite matrix.

Proof. By the Gauss-consistency and the regularity of $h(\cdot)$, we have :

$$h(y) - \theta = h(y) - h(f(\theta)) = H(f(\theta) + \lambda\epsilon)[y - f(\theta)] \text{ for some } 0 \leq \lambda \leq 1. \tag{2.3}$$

Multiplying both sides of (2.3) by σ^{-1}, we obtain :

$$\sigma^{-1}[h(y) - \theta] = H(f(\theta) + \lambda\epsilon)\sigma^{-1}[y - f(\theta)] \xrightarrow{d} H(f(\theta))U \text{ as } \sigma \to 0.$$

Since

$$AMSE[h(\cdot)] = E[(HU)(HU)'] = HE[UU']H' = HH',$$

we get the first part of (2.2). As for the proof of the second part of (2.2), we start with the identity (2.1) :

$$h[f(\theta)] = \theta \text{ for all } \theta.$$

Differentiating above identity with respect to θ, we obtain the following basic identity :

$$H(f(\theta))X(\theta) = I_p. \tag{2.4}$$

If we let $\bar{H} = (X'X)^{-1}X'$, then we have following identity :

$$HH' = \bar{H}\bar{H}' + (H - \bar{H})(H - \bar{H})'. \tag{2.5}$$

Here we used the fact that :

$$\bar{H}(H - \bar{H})' = (X'X)^{-1}X'(H - \bar{H})' = 0$$

which follows directly from (2.4). Identity (2.5) together with the fact $\bar{H}\bar{H}' = [X'X]^{-1}$ completes the proof.

Next we will construct a regular Gauss-consistent estimator of θ whose AMSE matrix is the same as the lower bound of (2.2). Essentially we will show that non-linear least-squares estimator (NLSE) is a regular Gauss-consistent estimator and its AMSE matrix attains the lower bound.

Now we introduce a non-linear least-squares estimator (NLSE) $\hat{\theta}(y)$ which is defined formally as the unique solution of the following system of the equations:

$$F(y, \theta) = X'(\theta)[y - f(\theta)] = 0 \tag{2.6}$$

if there exists an unique solution and is defined arbitrarily otherwise. Now we will show that $\hat{\theta}(y)$ will be an optimal Gauss-consistent estimator of θ.

Definition 4. We call a regular Gauss-consistent estimator $h(y)$ of θ to be *optimal* if its AMSE matrix attains the lower bound (2.2).

We also assume the following regularity condition :

$\boldsymbol{A_4}$: $X(\theta)$ is a continuously differentiable function of θ in Θ.

Theorem 2. Under the assumption A_1 through A_4 there exists a neighborhood N of the set $S = \{f(\theta) \in R^n; \theta \in \Theta\}$ and an unique function $\hat{\theta}(y)$

from R^n to R^p continuous in N such that

$$\hat{\theta}[f(\theta)] = \theta \text{ for all } \theta \in \Theta$$

and

$$F(y, \hat{\theta}(y)) = X'(\hat{\theta}(y))[y - f(\hat{\theta}(y))] = 0 \text{ for } y \in N.$$

Moreover

$$\mathcal{L}[\sigma^{-1}(\hat{\theta}(y) - \theta)|\theta] \to \mathcal{L}[V|\theta] \text{ as } \sigma \to 0, \tag{2.7}$$

where V is a $p \times 1$ random vector with $E[VV'] = (X'X)^{-1}$.

Proof. The proof closely parallels that of the theorem 1 in Section 3.2 of Ferguson (1958). By the lemma 3.1 of Ferguson (1958) we obtain the first part of the theorem immediately if we let $F(y, \theta) = X'(\theta)[y - f(\theta)]$. Then expanding $F(y, \theta)$ about the point $\hat{\theta}(y)$ to first-order, we have the formula :

$$F(y, \theta) = F(y, \hat{\theta}(y)) + \left[\int_0^1 F_\theta(y, \hat{\theta} + \lambda[\theta - \hat{\theta}])d\lambda\right]\left[\theta - \hat{\theta}(y)\right], \tag{2.8}$$

where $F_\theta(y, \theta) = [\partial F_i(y, \theta)/\partial \theta_j]$ is the $p \times p$ Jacobian matrix of $F(y, \theta)$ with respect to θ. Multiplying (2.8) by σ^{-1} and letting $\sigma \to 0$, we have

$$\mathcal{L}[\sigma^{-1}(\hat{\theta}(y) - \theta)|\theta] \to \mathcal{L}[(X'X)^{-1}X'U|\theta].$$

Here we have used the fact that as $\sigma \to 0$, $\hat{\theta}(y) \to \theta$ and thus

$$\int_0^1 F_\theta(y, \hat{\theta}(y) + \lambda(\theta - \hat{\theta}))d\lambda \xrightarrow{P} \int_0^1 F_\theta(f(\theta), \theta)d\lambda = -X'X.$$

This completes the proof of (2.7) if we let $V = (X'X)^{-1}X'U$.

Remark 5. If we let $\sigma = n^{-1/2}$ and if the random vector U has the normal distribution $N_n(0, I)$, then it can be easily shown that the Gauss-consistency is equivalent to the usual consistency in probability as $n \to \infty$ and above results reduce to the corresponding results of the BAN theory developed by Neyman (1949) and further extended by Ferguson (1958), Bemis and Bhapkar (1983).

Remark 6. If we have correlated measurement errors and if their covariance matrix $\Sigma = E[UU']$ does not depend on θ, then $\hat{\theta}(y)$ can be defined as a weighted least squares estimator (WLSE) of θ :

$$\hat{\theta}(y) = \arg\min_{\theta}[y - f(\theta)]'\Sigma^{-1}[y - f(\theta)].$$

In this case, we can also obtain similar optimality result by simple linear transformation.

3. EXAMPLES AND DISCUSSIONS

In this section we give some examples which illustrate the wide range of applications of the new optimality results developed in section 2. We also discuss some possible extensions of the finite-sample small-dispersion optimality results to more general class of models such as GLMs.

Example 1 (Errors-in-variables model). Consider the following model:

$$\begin{aligned} y_i &= \alpha + \beta u_i + \epsilon_i \\ x_i &= u_i + \delta_i, \end{aligned} \tag{3.1}$$

where $\{\epsilon_i\}$, $\{\delta_i\}$ are uncorrelated measurement errors with $E(\epsilon_i) = E(\delta_i) = 0$, $\text{Var}(\epsilon_i) = \sigma_\epsilon^2, \text{Var}(\delta_i) = \sigma_\delta^2$, $i = 1, \ldots, n$ and $\sigma_\epsilon/\sigma_\delta =$ fixed number. Here we note that $\theta' = (\alpha, \beta, u_1, \ldots, u_n)$ and the regression functions $\{f_i(\theta)\}$ are non-linear functions of θ because of the presence of non-linear terms $\{\beta u_i\}$ in (3.1). Under these assumptions we can show that the usual least squares estimators of α and β are optimal estimators among the class of the regular Gauss-consistent estimators irrespective of the type of the distribution of the measurement errors $\{\epsilon_i\}$ and $\{\delta_i\}$ as $\sigma \to 0$.

Example 2 (Calibration). Here we consider the following calibration model :

$$\begin{aligned} y_i &= \alpha + \beta x_i + \epsilon_i \\ y &= \alpha + \beta x + \epsilon, \end{aligned} \tag{3.2}$$

where measurement errors $\{\epsilon_i\}, \epsilon$ are mutually uncorrelated with $E(\epsilon_i) = E(\epsilon) = 0$, $\text{Var}(\epsilon_i) = \text{Var}(\epsilon) = \sigma^2$, $i = 1, \ldots, n$.

Note that here we have $\theta = (\alpha, \beta, x)$ and we again have a non-linear problem because of the existence of non-linear term $\{\beta x\}$ in (3.2). Then it can be shown that two commomly used estimators of x, the *ordinary* estimator :

$$\hat{x} = \bar{x} + (y - \bar{y})/b$$

and the *inverse* estimator :

$$\tilde{x} = \bar{x} + (y - \bar{y})/b^*$$

where $b = S_{xy}/S_{xx}$ and $b^* = S_{xy}/S_{yy}$ with $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$, $S_{xy} = \sum_{1=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$, are both *optimal* regular Gauss-consistent estimators of x irrespective of the type of the distribution of the measurement errors as $\sigma \to 0$. See So (1994) for more details on this problem.

Remark 7. When the covariance matrix of the measurement errors $\Sigma(\theta) = E[UU']$ depends on θ as is usually assumed in GLM and if there exists a consistent estimator $\bar{\Sigma}$ of $\Sigma(\theta)$, we can modify the definition of the regular Gauss-consistent estimator suitably and still get the similar optimality results. See Bemis and Bhapkar (1983) for this version of the optimality results in the asymptotically normal case. This topic and other related topics will be pursued in the subsequent paper in more detail.

Remark 8. Note also that we can use somewhat weaker set of regularity conditions than A_1 through A_4 and still get the similar optimality results. See Bemis and Bhapkar (1983) for the weaker set of conditions in the asymptotically normal case.

REFERENCES

(1) Bemis, K.G. and Bhapkar, V.P. (1983). On BAN Estimator for Chi-Squared Criteria. *The Annals of Statistics*, **11**, 183–196.

(**2**) Ferguson, T.S. (1958). A Method of Generating BAN Estimates with Application to the Estimation of Bacterial Densities. *The Annals of Mathematical Statistics*, **29**, 1046–1062.

(**3**) Hartley, H.O. and Booker, A. (1965). Non-Linear Least Squares Estimation. *The Annals of Mathematical Statistics*, **36**, 638–650.

(**4**) Jennrich, R.I. (1969). Asymptotic Properties of Non-Linear Least Squares Estimators. *The Annals of Mathematical Statistics*, **40**, 633–643.

(**5**) Jorgensen, B. (1987). Exponential Dispersion Models. *Journal of the Royal Statistical Society*, **B 49**, 127–162.

(**6**) McCullagh, P. (1983). Quasi-Likelihood Functions. *The Annals of Statistics*, **11**, 59–67.

(**7**) Nelder, J.A. and McCullagh, P. (1989). *Generalized Linear Models, 2nd ed.* Chapman and Hall, London and New York.

(**8**) Neyman, J. (1949). Contribution to the Theory of the χ^2 Test. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 239–273. University of California Press.

(**9**) So, B.S. (1994). A Note on the Small Sample Calibration . *Journal of the Korean Society for Quality Management*, **22–2**, 89–97.

(**10**) Sprott, D.A. (1983). *Encyclopedia of Statistics*, Vol. **3**, 305–309. Wiley, New York.

(**11**) Villegas, C. (1969). On the Least Squares Estimation of Non-Linear Relations. *The Annals of Statistics*, **40**, 462–466.

(**12**) Wedderburn, R.W.M. (1974). Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika*, **61**, 439–447.

(**13**) Wu, C.F. (1981). Asymptotic Theory of the Non-Linear Least Squares Estimation. *The Annals of Statistics*, **9**, 502–512.