INVITED

Estimation of Generalised Regression Models by the Grouping Method

Kazumitsu Nawata
Department of Social and International Relations
University of Tokyo

Abstract

This paper considers estimation of the model given by $y_t = h(x_t'\beta + u_t, \alpha)$, $t = 1,2,\ldots,T$, where h is a known transformation function. The model includes various types of important models such as the standard regression, censored regression, qualitative choice, duration, and Box-Cox transformation models. It is shown in this paper that the grouping estimator proposed by Nawata (1990a, b) can be generalised for estimation of the model. Nawata's grouping estimator represents a new class of estimation method, and has the same asymptotic distribution of estimators such as the maximum likelihood (ML) estimator, least squares (LS) estimator, and the least absolute deviations (LAD) estimator. The grouping estimator considered in this paper is defined by the following steps. First, the observations are divided into N different groups based on the values of x_t. Next, a proper location parameter is chosen and the location parameter of y_t (and other parameters if necessary) is estimated in each group. Finally, the parameters are estimated by the ML method based on the asymptotic distribution function of the location parameter estimated in the previous stage.

1. Introduction

This paper considers estimation of the model given by

$$(1)\quad y_t = h(x_t'\beta + u_t, \alpha),\ t = 1,2,\ldots,T,$$

where h is a known transformation function. The model given by (1) includes various types of important models such as the standard regression, censored regression, qualitative choice, duration, and Box-Cox transformation models.

It is shown in this paper that the grouping estimator proposed by Nawata (1990a, b) can be generalised for estimation of the model in (1). Nawata's grouping estimator represents a new class of estimation method, and has the same asymptotic distribution of estimators such as the maximum likelihood (ML) estimator , least squares (LS) estimator, and the least absolute deviations (LAD) estimator.

2. Models and the Definition of the Estimator

The model considered in this paper is given as follows:

$$(2)\quad \begin{aligned} y_t^* &= x_t'\beta + u_t, \\ y_t &= h(y_t^*, \alpha), \quad t = 1,2,\ldots,T, \end{aligned}$$

β is a K-dimensional vector of unknown parameters, $\{x_t\}$ are K-dimensional vectors of random variables, $\{x_t\}$ and $\{y_t\}$ are observable but $\{y_t^*\}$ are not observable, and $\{u_t\}$ are unknown error terms. The following assumptions are made.

Assumption 1

$\{(u_t, x_t')\}$ are i.i.d. random variables, however, u_t and x_t do not have to be independent for each t. The distribution functions of $\{u_t\}$ may depend on the values of x_t but are the same if the values of x_t are the same. The proper location parameter of the distribution of u_t is 0 for any t.

Assumption 2

The estimator of the location parameter has a proper asymptotic distribution.

To clarify the ideas of the grouping estimator proposed in this paper, the following assumption is made for $\{x_t\}$.

Assumption 3

$\{x_t\}$ take finitely many vector values $(\xi_1, \xi_2, \ldots, \xi_N)$.

This assumption can be removed by adjusting the values of $\{y_t\}$ in each group. However, to avoid unnecessary complications, the adjustment method is explained in the appendix.

3. General Principles of the Grouping Estimator

The grouping estimator considered in this paper is defined in the following steps.

Step 1

Divide the observations into N different groups based on the values of x_t. This means that the t-th observation belongs to the i-th group if and only if $x_t = \xi_i$. Let I_i be an index set of the i-th group such that $t \in I_i$ if and only if $x_t = \xi_i$. (Figure 1.)

Step 2

Choose the proper location parameter and estimate the location parameter of y_t (and other parameters, if necessary) in each group. Let $\hat{\theta}' = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_N)$ be estimators of the parameters. Estimate the asymptotic distribution functions of $\hat{\theta}_i$. (Figure 2.)

Step 3

Estimate α and β by the ML method using $\hat{\theta}$ and the estimated asymptotic distribution functions.

As mentioned above, the definition and properties of the grouping estimator depend on the choice and method of calculation of the location parameter. In the next section, several important cases are investigated.

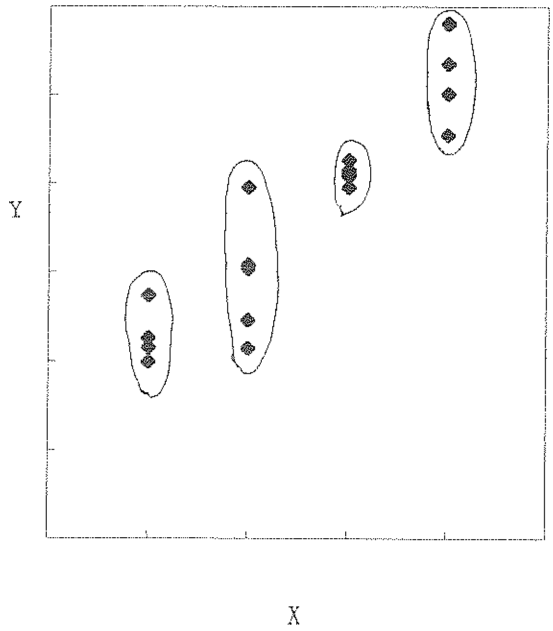

Figure 1 Divide the observations into N different groups based on the values of x_t.

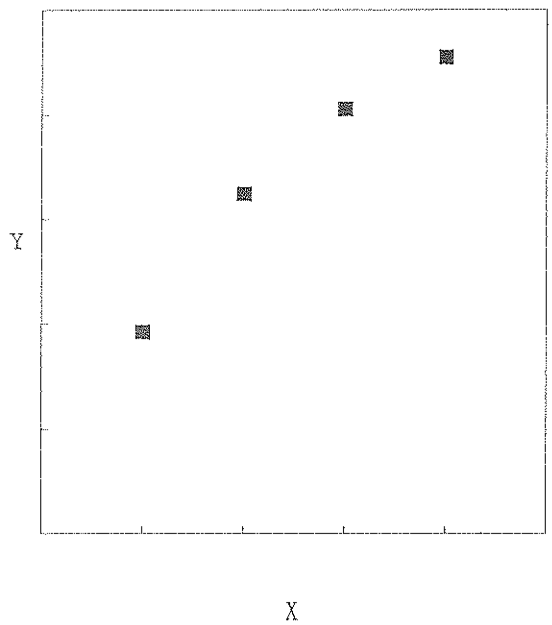

Figure 2 Estimate the location parameter of y_t and apply the ML method to the estimated values.

4. Grouping Estimator in Some Important Cases

In this section, cases where the choice of the location parameter is
i) ML estimator (MLE),
ii) mean,
iii) median, and
iv) other parameters
are considered. In most cases, the grouping estimator has the same asymptotic distribution as the known estimators.

4.1 Grouping Estimator Using the Maximum Likelihood Method

If the distribution functions of $\{u_t\}$ are known, the model can be estimated using the MLE in Step 2. Steps 2 and 3 are as follows in this case.

Step 2

Since h and the distributions of $\{u_t\}$ are known, the distribution functions of $\{y_t\}$ can be calculated. Let $\mu_i = x_i' \beta$. If the t-th observation belongs to i, that is $t \in I_i$, the distribution function of y_t can be written as $G_i(\mu_i, \alpha, \eta)$, where η is a vector of nuisance parameters of the distribution function of u_t. Estimate (μ_i, α', η') and the asymptotic variance of the estimators in each group by the ML method.

Step 3

Under the standard assumptions of ML, the estimators are asymptotically normal and are written as:

$$(3) \quad \begin{aligned} &\hat{\lambda}_i = \lambda_i + \epsilon_i, \\ &\sqrt{n_i}\epsilon_i \to N(0, A_i), \end{aligned} \qquad i = 1, 2, \ldots, N,$$

is an estimator of λ_i, n_i is the number of observations in the i-th cell, and A_i is the asymptotic variance matrix of $\hat{\lambda}_i$.

(α', β', η') are estimated by the generalised least squares (GLS) method, replacing A_i by the estimated variance matrix $\hat{A}_i$, and the asymptotic distribution of the estimator defined here is as efficient as the standard MLE. Note that Berkson's (1944, 1955, 1980) minimum chi-square estimator in binary choice models is just a special case of the grouping estimator.

4.2 Grouping Estimator Using Sample Means

When (2) is written as

$$(4) \quad \begin{aligned} &y_t = h(x_t'\beta, \alpha) + u_t^*, \\ &E(u_t) = 0, \quad V(u_t^*) = \sigma_i^2, \end{aligned}$$

for $t \in I_i$, α and β can be estimated by the grouping estimator using sample means. Steps 2 and 3 are as follows.

Step 2

Estimate the means of $\{y_t\}$ in each group and let ζ_i be the mean of y_t in the i-th group.

Step 3

Under standard assumptions, ζ_i is asymptotically normal and is written as

$$(5) \quad \begin{aligned} &\zeta_i = h(\xi_i'\beta, \alpha) + \epsilon_i, \\ &\sqrt{n_i}\epsilon_i \to N(0, \sigma_i^2), \quad i = 1, 2, \ldots, N. \end{aligned}$$

α and β are estimated by (nonlinear) weighted least squares replacing σ_i^2 by estimated variances. The asymptotic distribution of the estimator is the same as the (nonlinear) GLS estimator of (5).

4.3 Grouping Estimator Using Sample Medians

Suppose that $h(z, a)$ is a monotonically increasing function of z for each value of a and the medians of $\{u_t\}$ are zero. Powell (1984, 1991) proposed a modified LAD estimator which is consistent and asymptotically normal in these cases. The estimator minimises

$$(6) \quad S(b, a) = \sum |y_t - h(x_t'b, a)|.$$

The grouping estimator using sample medians are consistent and asymptotically normal under the same assumptions as in Powell (1991). Steps 2 and 3 are as follows.

Step 2

Estimate the sample medians of y_t in each group. Let ν_i be the sample median in the i-th group.

Step 3

Let

$\Gamma_0 = R^1 - \Gamma_1$, $\Omega_1 = \{h(z, \alpha): z \in \Gamma_1\}$, and

$\Omega_0 = \{h(z, \alpha): z \in \Gamma_0\}$,

where $h_z(z, \alpha) = \partial h(z, \alpha)/\partial z$.

To identify the model, Ω_0 and Ω_1 must be known. Under the assumptions of Powell (1991), it is easy to show that

$$(7)\quad \sqrt{n_i}\epsilon_i \to \begin{cases} N(0,\ 1/4g_i^2(0)), & \text{if } \mu_i \in \Gamma_1 \\ 0 & , \text{if } \mu_i \in \Gamma_0, \end{cases}$$

where

$$\mu_i = \xi_i' \beta,\ \epsilon_i = v_i - h(\mu_i, \alpha),$$
$$g_i(0) = f_i(0)/h_z(\mu_i, \alpha),$$

and f_i is the density function of u_t in the i-th cell.

Define the pseudo-likelihood function as

$$(8)\ L(a,b,s) = \prod_{v_i \in \Omega_0} \Big[\Phi[\sqrt{n_i}\{\overline{m_i} - h(\xi_i' b,\ a)\}/s] - \Phi[\sqrt{n_i}\{\underline{m_i} - h(\xi_i' b,\ a)\}/s]\Big] \cdot \prod_{v_i \in \Omega_1} \phi[\sqrt{n_i}\{v_i - h(\xi_i' b,\ a)\}/s],$$

where

$$\overline{m_i} = inf\{z: z \in \Gamma_1,\ z > v_i\},$$

and

$$\underline{m_i} = sup\{z: z \in \Gamma_1,\ z < v_i\}.$$

ϕ and Φ are the density and distribution functions of the standard normal distribution.

$\theta' = (\beta', \alpha')$ is estimated by maximising (8). The asymptotic distribution of the estimator, $\hat{\theta}$, is the same as Powell's LAD estimator and given by:

$$(9)\ \sqrt{T}(\hat{\theta} - \theta) \to N(0, (1/4)\, D_2^{-1} D_1 D_2^{-1})$$

where

$$D_1 = \sum_{\mu_i \in \Gamma_1} \delta_i\ \omega_i\ \omega_i',$$
$$D_2 = \sum_{\mu_i \in \Gamma_1} \delta_i\ f_i(0)\{h_z(\mu_i, \alpha)\}^{-1}\ \omega_i \omega_i',$$
$$\omega_i' = (h_z(\mu_i, \alpha)x_i', h_a(\mu_i, \alpha)'),$$
$$h_a = \partial h(z, a)/\partial a,$$
$$\mu_i = \xi_i' \beta,$$

and

$$\delta_i = \operatorname*{plim}_{T \to \infty}\ n_i/T.$$

If $h_z(z, \alpha) = 0$ almost everywhere for $z \in (-\infty, \infty)$, which includes discrete choice models, the assumptions of Powell (1991) are violated. However, under certain distributional conditions of x_t, the parameters are estimated up to a multiplicative constant using the LAD method.

Manski (1975, 1985) called the LAD estimator for the binary choice models the maximum score estimator. The grouping estimator can be consistent up to a multiplicative constant under the same conditions as the LAD estimator (Nawata (1994)).

4.4 Grouping Estimator Using Other Location Parameters

It is possible to use other types of location parameters in Step 2, such as trimmed means, L_p estimators, and modes. Note that for the cases in 4.1-4.3, there exist estimators which correspond to the grouping estimator. However, for some location parameters, the corresponding estimators are unknown and the grouping estimator can be potentially asymptotically more efficient than the existing estimators in some occasions.

As an example, consider censored regression models where $h(z) = z$ if $z > 0$ and $h(z) = 0$ if $z \leq 0$. Suppose that the distributions of $\{u_t\}$ are symmetric. The grouping estimator using trimmed means can be more efficient than the LAD estimator or the symmetrically trimmed estimator (Powell (1986b)).

5. Conclusion

In this paper, Nawata's (1990a, b, 1994) grouping method was shown to be applicable for the estimation of generalised regression models. The grouping estimator is asymptotically as efficient as the corresponding estimators, and can be more efficient on some occasions.

Since Nawata's grouping method represents a new class of estimation method, various properties of the estimator are not yet known. Therefore, further investigation of the method is necessary to determine the usefulness of the estimator in finite samples.

Appendix

When x_t contains elements which take infinitely many values, it is necessary to adjust the values of y_t in each group, as suggested in Nawata (1990a, b). (Ritov (1990) independently suggested the similar idea of "adjustment" in the censored regression model.) In this case, the following assumptions are made:

Assumption A.1

$\{x_t\}$ are bounded.

Assumption A.2

$h(z, a)$ is uniformly continuous for any possible value of a.

The estimator is defined as follows (the proof of consistency and asymptotic normality is a simple modification of Nawata (1990a, b)).

Step 1

Divide the sample space of x_t into N non-overlapping groups (Figure 3) so that:

i) the largest distance between two points in the same group goes to zero as the order of T^{-d_1} for some $d_1 > 0$;

ii) for some $d_2 \in (2/3, 1)$, $n_i \cdot T^{-d_2}$ converges to δ_i in probability for any i, where n_i is the number of observations in the i-th group and $\{\delta_i\}$ are bounded away from zero.

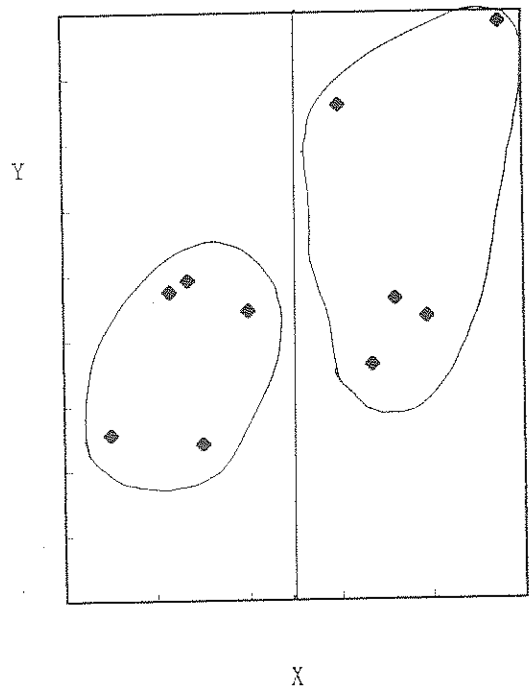

Figure 3 Divide the sample space of x_t into N non-overlapping groups

To obtain the asymptotic distribution of the estimator by the adjustment of y_t, the following assumptions are required in addition to Assumptions A.1 - A.2.

Assumption A.3

The estimator of the location parameter used in the estimation satisfies the conditions given in Appendix A of Nawata (1990a).

Assumption A.4

Let $\hat{\lambda}_i$ be a vector of estimators of the location parameter and other necessary parameters and λ_i be the true value of $\hat{\lambda}_i$. Then there exists c > 0 such that

$$(10) \quad \sup_{\omega} |F_i(\omega) - \Psi_i(\omega)| < c/\sqrt{n_i} ,$$

for any i, where $F_i(\omega)$ and $\Psi_i(\omega)$ are (exact) distribution and asymptotic distribution functions of $\sqrt{n_i}\cdot(\hat{\lambda}_i - \lambda_i)$, respectively. Note that this assumption is satisfied in most cases by the simple modification of the Berry-Esseen Theorem.

Unlike the case where x_t takes finitely many values, there are two potential problems in this case:
i) errors caused by the sizes of groups;
ii) since the number of groups goes to infinity, the difference between the exact and asymptotic distributions may not be ignored.

The first problem is corrected by the adjustment of y_t as in the following steps and, if $d_2 > 2/3$, it is easy to show that the error caused by the second problem can be ignored from Assumption A.4.

Step 2

Let X_i be the sample mean of x_t in the i-th group. Estimate the location parameters of interest of y_t and other necessary parameters in each group.

Step 3

Apply the maximum likelihood methods based on the asymptotic distributions and estimate the first-round estimators $\hat{\alpha}_1$ and $\hat{\beta}_1$ using $\{X_i\}$. It is easy to show that the estimators are consistent estimators of order T^{-c}, where $c = \min\{d_1, 1/2\}$.

Step 4

Adjust the value of y_t by the first round estimators $\hat{\alpha}_1$ and $\hat{\beta}_1$ using the proper methods as the following examples, and let y_t^+ be the adjusted value of y_t. (Figure 4.)

Example A.1 Censored Regression Models

When $h(z) = z$ if $z > 0$ and 0 if $z \leq 0$, the model is censored regression model. Let y_t^0 be given by

$$(11)\quad y_t^0 = y_t - (x_t - X_i)'\hat{\beta}.$$

The censoring values of y_t^0 are not zero. Therefore, let the new censoring value be $C = c_1\cdot T^{-d_3}$, where $c_1 > 0$ and $0 < d_3 < d_1$, and define the adjusted values as

$$(12)\quad y_t^+ = \begin{cases} y_t^0, & if\ y_t^0 > C \\ 0, & otherwise. \end{cases}$$

Note that, after the second round, the censoring value becomes C rather than zero.

Example A.2 Binary Choice Models When the Distribution of u_t is Known

Suppose $h(z) = 1(z > 0)$ where $1(\cdot)$ is an indicator function which takes 1 if $\cdot$ is true and 0 otherwise, and the distribution function of $-u_t$ is given by a known function $F(u)$. The model is a binary choice model, and the adjusted value y_t^+ is defined as

$$(13)\quad y_t^+ = y_t - \{F(x_t'\hat{\beta}_1) - F(X_i'\hat{\beta}_1)\}.$$

Note that the adjusted value y_t^+ may not be in [0, 1].

Example A.3 Box-Cox Transformation Models

In the Box-Cox transformation models, $h(z, a)$ is differentiable with respect to z for any possible value of a. The adjustment values are defined as

$$(14)\quad y_t^+ = y_t - \hat{h}_z\cdot(x_t'\hat{\beta}_1 - X_i'\hat{\beta}_1),$$

where $\hat{h}_z$ is estimated value of $h_z(y_t^*, \alpha)$ using y_t, $\hat{\alpha}_1$, and $\hat{\beta}_1$. After the second round estimation, the adjustment is done using a higher-order Taylor expansion; that is the j-th round adjustment uses the j-th order of the Taylor expansion.

By this adjustment, the error caused by the sizes of cells becomes $O_p\{T^{-(c+d_1)}\}$. Estimate the second-round estimators $\hat{\alpha}_2$ and $\hat{\beta}_2$ using $\{y_t^+\}$.

Step 5

Repeat the adjustment j times such that $j\cdot d_1 > 1/2$ and determine the final estimator. By the repetition of the adjustment procedure, the effect of the sizes of groups becomes $o_p(T^{-1/2})$, and the asymptotic distributions of the grouping estimator are given by the same formulae as in the case where x_t takes finitely many values.

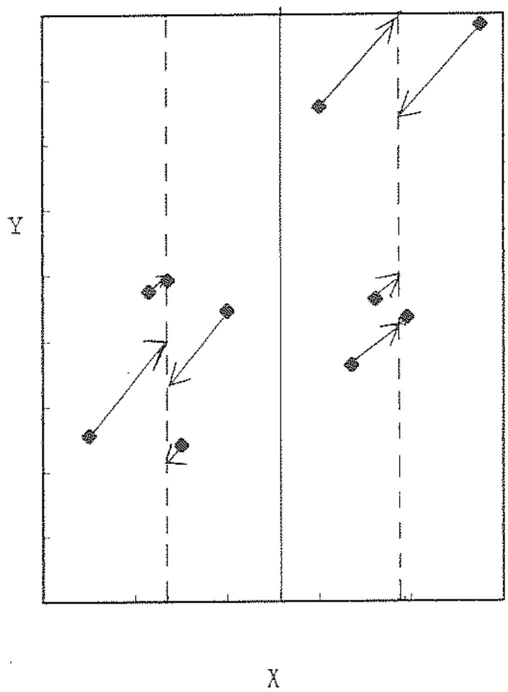

Figure 4 Adjust the values of y_t using the previous round estimator

Acknowledgment

I would like to thank Michael McAleer for helpful comments and suggestions.

References

Berkson J., "Application of the Logistic Function to Bio-Assay," *Journal of the American Statistical Association*, 39, 357-365, 1944.

____,"Maximum Likelihood and Minimum χ^2 Estimates of the Logistic Function," *Journal of the American Statistical Association*, 50, 130-162, 1955,.

____,"Minimum Chi-Square, Not Maximum Likelihood," *Annals of Statistics*, 8, 457-487, 1980.

Manski, C. F., "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205-28, 1975.

____, "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 313-333, 1985.

Nawata, K., "Robust Estimation Based on Grouped-Adjusted Data in Linear Regression Models, " *Journal of Econometrics*, 43, 317-336, 1990a.

____, "Robust Estimation Based on Grouped-Adjusted Data in Censored Regression Models, " *Journal of Econometrics*, 43, 337-362, 1990b.

____, "Estimation of the Boundary of Two-Regions by the Grouping Estimator," *Journal of Japan Statistical Society*, 24, 15-35 , 1994.

Powell, J., "Least Absolute Deviations Estimator for Censored Regression Models," *Journal of Econometrics*, 25, 303-325 , 1984.

____, "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143-155, 1986a.

____, "Symmetrically Trimmed Least Squares Estimation of Tobit Models," *Econometrica*, 54, 1435-1460, 1986b.

____, "Estimation of Monotonic Regression Models under Quantile Restrictions," in W. Barnett, J. Powell, and G. Tauchen , (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge University Press, 1991.

Ritov, Y., "Estimation in a Linear Regression Model with Censored Data," *Annals of Statistics*, 18, 303-328, 1990.