Comment 7: Zhenlin YANG

1. INTRODUCTION

The effect of estimating a transformation on the subsequent inferences is an important issue in the applications of the Box–Cox transformation method. Many have made their contributions to this issue, directly or indirectly. Those include, among others, Bickel & Doksum (1981), Carroll & Ruppert (1981), Carroll (1982), Doksum & Wong (1983), Hinkley & Runger (1984), Taylor (1989), Duan (1993) and Yang (1999).

The present authors have studied this issue by concentrating on the ratio of the regression slope and the error standard deviation and provide important results to show that the estimation of this scaled slope is much more stable than the estimation of the slope itself with respect to the transformation estimation. Making use of some of the results of Yang (1999), I am able to provide some similar results under a slightly different set-up. Also, I carry out some Monte Carlo experiments to investigate the finite sample effect of transformation estimation. Both reinforce the authors' findings.

2. THE GENERAL RESULTS

First, I introduce some different notation to reflect the changes in the set-up: $\mathbf{Y}(\lambda_0) = \mathbf{X}\beta_0 + \sigma_0\varepsilon$, $\eta_i = x_i'\beta_0$ and $\delta_i = \lambda_0\sigma_0/(1+\lambda_0\eta_i)$. Thus β_0, and hence $\theta = \beta_0/\sigma_0$, includes the intercept parameter and the definition of δ_i incorporates the values of the regressors. Let $\xi_0 = (\beta_0', \lambda_0, \sigma_0)$ and let $\hat{\xi}$ be an M-estimator of ξ_0 that solves the following estimating equation

$$\frac{1}{n}\sum_{i=1}^{n}\Psi_i(Y_i;\hat{\xi}) = 0_{(p+3)\times 1},$$

where the function Ψ_i is partitioned according to β_0, λ_0, and σ_0. Define

$$\overline{\Psi} = \frac{1}{n}\sum_{i=1}^{n}\Psi_i(Y_i,\xi_0),$$

and $\mathbf{A} = \mathrm{E}\,(\partial\bar{\Psi}/\partial\xi_0)$, both partitioned accordingly. The elements of $\overline{\Psi}$ are denoted by $\overline{\Psi}_i$, $i = 1,2,3$, and the elements of $\mathbf{A}$ by $\mathbf{A}_{ij}$, $i,j = 1,2,3$. Let $\tilde{\beta}$ and $\tilde{\sigma}$ be, respectively, the M-estimate of β_0 and σ_0 when λ_0 is known. Assume that the conditions C1, C2, C3, and C5 of Yang (1999) are satisfied. Following Taylor expansions on the estimating equation, we find

$$\hat{\beta} - \beta_0 = -\mathbf{A}_{11}^{-1}\overline{\Psi}_1 - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\hat{\lambda}-\lambda_0) + O_p(n^{-1}), \qquad (13)$$

$$\hat{\sigma} - \sigma_0 = -\mathbf{A}_{33}^{-1}\overline{\Psi}_3 - \mathbf{A}_{33}^{-1}\mathbf{A}_{32}(\hat{\lambda}-\lambda_0) + O_p(n^{-1}), \qquad (14)$$

$$\hat{\lambda} - \lambda_0 = \frac{\overline{\Psi}_2 - \mathbf{A}_{23}\mathbf{A}_{33}^{-1}\overline{\Psi}_3 - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\overline{\Psi}_1}{\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} - \mathbf{A}_{22} + \mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{A}_{32}} + O_p(n^{-1}). \qquad (15)$$

Equating $\hat{\lambda}$ to λ_0 in (1) and (2) gives

$$\tilde{\beta} - \beta_0 = -\mathbf{A}_{11}^{-1}\overline{\Psi}_1 + O_P(n^{-1})$$

and

$$\tilde{\sigma} - \sigma_0 = -\mathbf{A}_{33}^{-1}\overline{\Psi}_3 + O_P(n^{-1}).$$

Now, considering $\hat{\sigma}^{-1}$ as a function of $\hat{\sigma}$, a first-order Taylor expansion around $\tilde{\sigma}$ gives

$$\hat{\sigma}^{-1} = \tilde{\sigma}^{-1} - \sigma_0^{-2}(\hat{\sigma}-\tilde{\sigma}) + O_p(n^{-1}) = \tilde{\sigma}^{-1} + \sigma_0^{-2}\mathbf{A}_{33}^{-1}\mathbf{A}_{32}(\hat{\lambda}-\lambda_0) + O_p(n^{-1}). \qquad (16)$$

Writing (1) as $\hat{\beta} = \tilde{\beta} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\hat{\lambda} - \lambda_0) + O_p(n^{-1})$ and combining this with (4), we find

$$\hat{\theta} = \tilde{\theta} + (\theta\mathbf{A}_{33}^{-1}\mathbf{A}_{32} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12})(\hat{\lambda} - \lambda_0)\sigma_0^{-1} + O_p(n^{-1}). \tag{17}$$

The second term in the expansion (5) reflects the effect of the estimating transformation. The magnitude of this effect can be studied in detail if the estimating function Ψ_i and the transformation function $Y(\lambda)$ are both specified. This result can be compared with the expansion for $\hat{\beta}$ to see which quantity is more stable with respect to $\hat{\lambda}$.

3. THE BOX–COX POWER TRANSFORMATION AND NEAR NORMAL ERRORS

When the Box–Cox power transformation is used and the errors are approximately normal, the estimating function corresponding to the maximum likelihood estimation takes the form

$$\Psi_i(Y_i, \xi_0) = \begin{cases} \Psi_{1i}(Y_i, \xi_0) &= \sigma_0^{-2} x_i \{Y_i(\lambda_0) - x_i'\beta_0\}, \\ \Psi_{2i}(Y_i, \xi_0) &= \log Y_i - \sigma_0^{-2}\{Y_i(\lambda_0) - x_i'\beta_0\}\dot{Y}_i(\lambda_0), \\ \Psi_{3i}(Y_i, \xi_0) &= \sigma_0^{-3}\{Y_i(\lambda_0) - x_i'\beta_0\}^2 - \sigma_0^{-1}, \end{cases}$$

where

$$\dot{Y}_i(\lambda_0) = \partial Y_i(\lambda_0)/\partial\lambda_0 = \begin{cases} \lambda_0^{-1}\{1 + \lambda_0 Y_i(\lambda_0)\}\log Y_i - \lambda_0^{-1}Y_i(\lambda_0), & \lambda_0 \neq 0, \\ \frac{1}{2}(\log Y_i)^2, & \lambda_0 = 0. \end{cases}$$

With the Box–Cox power transformation, the exact normality of $Y_i(\lambda_0)$ is incompatible with the positivity of Y_i. Hence there is a truncation effect when the above Ψ_i function is used to approximate to the true likelihood estimating function. However, this truncation effect is small if $\delta_0 = \max|\delta_i|$ is small, which is achievable when (i) σ_0 is small, (ii) λ_0 is small, or (iii) $\min|\eta_i|$ is large.

Assuming that the truncation effect is small and using the Ψ_i function given above, one can easily evaluate all the quantities involved in (1)–(3) to simplify the expansions. Further, when $\lambda_0 = 0$, all the expansions can be expressed explicitly in terms of ξ_0 and $\varepsilon_i s$, which allows one to examine the affecting term in detail as well as to find an explicit expression for the variance inflation. When $\lambda_0 \neq 0$, however, an approximation to $\log Y_i$ is necessary. When δ_0 is small, we have,

$$\lambda_0 \log Y_i = \log(1 + \lambda_0\eta_i) + \delta_i\varepsilon_i - \tfrac{1}{2}\delta_i^2\varepsilon_i^2 + O_p(\delta_i^3). \tag{18}$$

I use ε^2 to mean $(\varepsilon_i^2)_{n\times 1}$, etc. Thus,

$$\dot{\mathbf{Y}}(\lambda_0) = \begin{cases} \frac{1}{2}(\eta^2 + 2\eta\sigma_0\varepsilon + \sigma_0^2\varepsilon^2), & \lambda_0 = 0, \\ \dfrac{1}{\lambda_0^2}[(\mathbf{1} + \lambda_0\eta)\#\phi] + \dfrac{\sigma_0}{\lambda_0}\phi\#\varepsilon + \dfrac{\sigma_0}{2\lambda_0}\delta\#\varepsilon^2 - \dfrac{\sigma_0}{2\lambda_0}\delta^2\#\varepsilon^3 + O_p(\delta^3), & \lambda_0 \neq 0. \end{cases}$$

By assuming the first six moments of ε_1 are the same as those of a standard normal random variable and making use of the above approximations, Yang (1999) derived the following explicit expansions:

$$\frac{\hat{\lambda} - \lambda_0}{\sigma_0} = \frac{-\frac{1}{2}(\mathbf{M}\eta^2)'\varepsilon - \sigma_0(\eta - \bar{\eta}\mathbf{1})'\varepsilon^2 + \frac{1}{2}\sigma_0^2\mathbf{1}'(3\varepsilon - \varepsilon^3)}{\frac{1}{4}||\mathbf{M}\eta^2||^2 + 2\sigma_0^2||\eta - \bar{\eta}\mathbf{1}||^2 + \frac{3}{2}n\sigma_0^4} + O_p(n^{-1}), \quad \lambda_0 = 0,$$

$$\frac{\hat{\lambda} - \lambda_0}{\lambda_0} = \frac{-(\delta^{-1}\#\phi + \frac{1}{2}\delta)'\mathbf{M}\varepsilon - (\phi - \bar{\phi})'\varepsilon^2 + \frac{1}{2}(3\varepsilon - \varepsilon^3)}{||\mathbf{M}(\delta^{-1}\#\phi + \frac{1}{2}\delta)||^2 + 2||\phi - \bar{\phi}\mathbf{1}||^2 + \frac{3}{2}||\delta||^2} + O_p(n^{-1}) + O_p(\delta_0^3), \ \lambda_0 \neq 0,$$

where $\phi = \log(1 + \lambda_0\eta)$ and $||\cdot||$ denotes the Euclidean norm. These lead immediately to the approximations to the variance of $\hat{\lambda}$ for large n and small δ_0,

$$\tau^2(\xi_0) \equiv \mathrm{var}(\hat{\lambda}) \approx \sigma_0^2\big\{\tfrac{1}{4}||\mathbf{M}\eta^2||^2 + 2\sigma_0^2||\eta - \bar{\eta}\mathbf{1}||^2 + \tfrac{3}{2}n\sigma_0^4\big\}^{-1}, \quad \lambda_0 = 0,$$

$$\tau^2(\xi_0) \equiv \mathrm{var}(\hat{\lambda}) \approx \lambda_0^2\big\{||\mathbf{M}(\delta^{-1}\#\phi + \tfrac{1}{2}\delta)||^2 + 2||\phi - \bar{\phi}\mathbf{1}||^2 + \tfrac{3}{2}||\delta||^2\big\}^{-1}, \quad \lambda_0 \neq 0.$$

Now,

$$\mathbf{A}_{11} = -\frac{1}{n\sigma_0^2}\mathbf{X}'\mathbf{X}, \quad \mathbf{A}_{12} = \frac{1}{n\sigma_0^2}\mathbf{X}'\mathrm{E}\{\dot{\mathbf{Y}}(\lambda_0)\}, \quad \mathbf{A}_{33} = \frac{-2}{\sigma_0^2}, \quad \mathbf{A}_{32} = \frac{2}{n\sigma_0^2}\,\mathrm{E}\{\varepsilon'\dot{\mathbf{Y}}(\lambda_0)\}.$$

With the approximation (6) and the explicit expansion for $\hat{\lambda}$, the expansion (5) becomes

$$\hat{\theta} = \tilde{\theta} + v(\xi_0)(\hat{\lambda} - \lambda_0)\sigma_0^{-1} + O_p(n^{-1}) + O_p(\theta_0^3), \tag{19}$$

where

$$v(\xi_0) = \begin{cases} \frac{1}{2}\mathbf{X}'\mathbf{X}^{-1}\mathbf{X}'\{(\eta - \bar{\eta}\mathbf{1})^2 - (\bar{\eta}^2 - \sigma_0^2)\mathbf{1}\}, & \lambda_0 = 0, \\ \lambda_0^{-1}\mathbf{X}'\mathbf{X}^{-1}\mathbf{X}'\{\lambda_0^{-1}(\mathbf{1} + \lambda_0\eta)\#\phi - \left(1 + \bar{\phi} - \frac{3}{2}\bar{\delta}^2\right)\eta + \frac{1}{2}\sigma_0\delta\}, & \lambda_0 \neq 0. \end{cases}$$

Finally, considering $\tilde{\sigma}^{-1}$ as a function of $\tilde{\sigma}^2$ and Taylor expanding it around σ_0^2 give

$$\tilde{\sigma}^{-1} = \sigma_0^{-1} + \frac{1}{2\sigma_0^3}(\tilde{\sigma}^2 - \sigma_0^2) + O_p(n^{-1}) = \sigma_0^{-1} + \frac{1}{n\sigma_0}\sum_{i=1}^{n}(\varepsilon_i^2 - 1) + O_p(n^{-1}).$$

This, combined with

$$\tilde{\beta} = \beta_0 - \mathbf{A}_{11}^{-1}\overline{\boldsymbol{\Psi}}_1 + O_p(n^{-1}) = \beta_0 + \sigma_0\mathbf{X}'\mathbf{X}^{-1}\mathbf{X}'\varepsilon + O_p(n^{-1})$$

gives

$$\tilde{\theta} = \theta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon + \frac{\theta}{n}\sum_{i=1}^{n}(\varepsilon_i^2 - 1) + O_p(n^{-1})$$

These lead to approximations for the variances of $\hat{\theta}$ and $\tilde{\theta}$ when n is large and δ_0 is small

$$\mathbf{V}(\hat{\theta}) \approx (\mathbf{X}'\mathbf{X})^{-1} + \frac{1}{2n}\theta\theta' + \frac{1}{\sigma_0^2}v(\xi_0)v(\xi_0)'\tau^2(\xi_0)$$

and $\mathbf{V}(\tilde{\theta}) \approx (\mathbf{X}'\mathbf{X})^{-1} + \frac{1}{2n}\theta\theta'$. Similarly, one obtains the expansions for $\hat{\beta}$ and its variance, namely

$$\hat{\beta} = \tilde{\beta} + w(\xi_0)(\hat{\lambda} - \lambda_0) + O_p(n^{-1})$$

and

$$\mathbf{V}(\hat{\beta}) \approx \sigma_0^2(\mathbf{X}'\mathbf{X})^{-1} + w(\xi_0)w(\xi_0)'\tau^2(\xi_0),$$

where

$$w(\xi_0) = \begin{cases} \frac{1}{2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\eta^2 + \sigma_0^2\mathbf{1}), & \lambda_0 = 0, \\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\{\lambda_0^{-2}(\mathbf{1} + \lambda_0\eta)\#\phi - \lambda_0^{-1}\eta + \frac{1}{2}\sigma_0\lambda_0^{-1}\delta\}, & \lambda_0 \neq 0. \end{cases}$$

The expansions for the $\lambda_0 = 0$ case are the same as those given by the authors, except that the intercept parameter is also included here. The effect of estimating the transformation on the estimation of θ is governed by $v(\xi_0)$, and that on the estimation of β_0 is governed by $w(\xi_0)$. A close examination of the two quantities for the case of $\lambda_0 = 0$ reveals that there are many cases in which the component of $v(\xi_0)$ corresponding to the slope parameters vanishes (see the detailed discussions provided by the authors), but no such cases exist for $w(\xi)$. In the case of $\lambda_0 \neq 0$, the two quantities behave similarly to the case of $\lambda_0 = 0$. This means that the effect on the estimation of θ can be zero or small, but the effect on the estimation of β_0 is generally large. Note that the effect of the estimating transformation on the estimation of the (scaled) intercept parameter is not small in general. When the regressors are centered, $\bar{\eta} = \mu_0$.

4. MONTE CARLO SIMULATION

I now present some Monte Carlo simulation results to show the finite sample effect of the estimating transformation. The model used in the simulation is: $Y_i(\lambda_0) = \mu_0 + \beta_0 x_i + \sigma_0 \varepsilon_i$, $i = 1, \ldots, n$. The x values are centered values of $\exp(2i/n)$, $i = 1, \ldots, n$. I use the relative bias (RB) and the relative efficiency (REF) to measure the effects. The simulation results are summarized in Table 7-1. From the results, we see that there is generally a large effect on the estimation of β_0, but a very small effect on the estimation of β_0/σ_0. It is interesting to note that as the sample size increases, the effect on the estimation of β_0/σ_0 reduces significantly, but the effect on the estimation of β_0 does not seem to change much. Estimating the transformation induces extra bias in the estimation of β_0/σ_0, but only slightly. Additional simulations (not reported here) reveal that the spread in the x values matters. A larger spread gives a smaller effect on the estimation of β_0/σ_0. The magnitude of β_0/σ_0 affects the value of MSE, but affects very little on the relative efficiency.

TABLE 7-1: A summary of simulated RB and REF, where, for example, RB($\hat{\theta}$) $= 100 \times (\hat{\theta} - \theta)/\theta$ and REF($\tilde{\theta}, \hat{\theta}$) $= \mathrm{MSE}(\hat{\theta})/\mathrm{MSE}(\tilde{\theta})$. The MSE stands for the mean squared error.

		Relative bias				Relative efficiency			
ℓ	σ_0	RB($\tilde{\beta}$)	RB($\hat{\beta}$)	RB($\tilde{\theta}$)	RB($\hat{\theta}$)	MSE($\tilde{\beta}$)	REF($\tilde{\beta}, \hat{\beta}$)	MSE($\tilde{\theta}$)	REF($\tilde{\theta}, \hat{\theta}$)
		$n = 20$,	$\mu_0 = 8.0$,	$\beta_0 = 2.0$					
.25	.05	0.00	0.02	10.18	13.77	.000036	81.14	79.6946	1.28
	.1	0.00	0.08	9.63	13.18	.000143	80.52	18.5465	1.32
	.5	0.01	4.49	10.11	13.76	.003639	80.07	0.7933	1.30
.0	.05	0.00	0.02	10.40	13.93	.000035	30.23	80.3997	1.29
	.1	0.00	0.09	10.09	13.60	.000143	30.26	19.4002	1.27
	.5	0.01	2.57	10.08	13.65	.003565	29.34	0.7782	1.28
		$n = 50$,	$\mu_0 = 8.0$,	$\beta_0 = 2.0$					
.25	.05	0.00	0.03	3.64	4.76	.000015	82.74	20.9905	1.11
	.1	0.00	0.14	3.73	4.87	.000061	83.14	5.3149	1.11
	.5	−0.03	1.95	3.53	4.72	.001512	76.03	0.2155	1.13
.0	.05	0.00	0.02	3.68	4.80	.000015	31.16	20.8424	1.11
	.1	−0.01	0.04	3.66	4.83	.000060	32.13	5.2787	1.12
	.5	0.02	1.05	3.71	4.92	.001494	28.64	0.2202	1.13
		$n = 100$,	$\mu_0 = 8.0$,	$\beta_0 = 2.0$					
.25	.05	0.00	0.00	1.76	2.29	.000008	85.55	9.1725	1.05
	.1	0.00	0.04	1.87	2.41	.000030	87.02	2.3389	1.06
	.5	0.02	0.84	1.87	2.42	.000768	73.57	0.0951	1.08
.0	.05	0.00	0.01	1.84	2.37	.000008	31.15	9.3085	1.06
	.1	0.00	0.01	1.73	2.28	.000031	32.75	2.2728	1.06
	.5	−0.02	0.59	1.73	2.30	.000769	27.81	0.0942	1.08

Zhenlin YANG: ecsyzl@nus.edu.sg
Department of Economics, National University of Singapore
1 Arts Link, Singapore 117570