ARTICLE IN PRESS

economics letters

Economics Letters 1 (2002) 000–000

www.elsevier.com/locate/econbase

A score test for Box–Cox functional form

Zhenlin Yang[a,*], Tilak Abeysinghe[b]

[a]School of Economics and Social Sciences, Singapore Management University, Singapore 259756, Singapore
[b]Department of Economics, National University of Singapore, Singapore 117570, Singapore

Received 5 June 2002; accepted 4 September 2002

Abstract

This paper presents two score tests to determine a value for the Box–Cox transformation parameter. The test based on expected information performs better in small samples and is computationally simpler than the one based on observed information; therefore, the former is recommended.

Keywords: Profile score function; Expected information; Observed information; Distributional properties

JEL classification: C12; C21

1. Introduction

Yang and Abeysinghe (2002) presented an explicit formula for the variance of the Box–Cox transformation estimator $\hat{\lambda}$, in a regression model where both endogenous and exogenous variables are subject to a power transformation indexed by an unknown transformation parameter λ. Their results can be used for hypothesis testing about λ. In certain applications the grid-search procedures used to obtain the maximum likelihood estimate of λ may run into difficulties due to poor model specifications, high multicollinearity and other data characteristics. However, if a value is imposed on λ the other regression parameters can be estimated relatively easily. It is, therefore, desirable to have a simple test procedure to determine a value for λ without having to estimate it.

The most commonly tested hypotheses about the functional form are $\lambda = 1$ (linear) and $\lambda = 0$ (loglinear). The most commonly used test for this is the likelihood ratio test (Box and Cox, 1964). When estimating of λ runs into difficulties both the likelihood ratio and Wald tests become un-operational. The objective of this paper is to propose a test that does not require $\hat{\lambda}$. This is

*Corresponding author. Tel.: +65-68-22-0852; fax: +65-68-22-0833.
E-mail address: zlyang@smu.edu.sg (Z. Yang).

0165-1765/02/$ – see front matter
PII: S0165-1765(02)00294-X

obviously a score test. The proposed test is simple and direct as opposed to the one suggested by Davidson and MacKinnon (1985), see also the references therein. The Davidson–MacKinnon test requires an artificial regression and it is designed for testing linear and loglinear versions against a Box–Cox alternative. Our test is designed to test H_0:$\lambda = \lambda_0$ against H_a:$\lambda \neq \lambda_0$. In Section 2 we examine two score tests, one based on expected Fisher information and the other based on observed Fisher information. In Section 3 we present the results of a Monte Carlo experiment that enables us to choose between the two test statistics.

2. The score test

As opposed to a more general transformation that allows for different λs across the regression we consider the more commonly estimated regression given by

$$Y_t(\lambda) = \sum_{j=0}^{p} \beta_j X_{tj}(\lambda) + \epsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

where $X(\lambda)$ is defined by the power transformation

$$X(\lambda) = \begin{cases} (X^{\lambda} - 1)/\lambda & \text{if } \lambda \neq 0, \\ \log X & \text{if } \lambda = 0, \end{cases}$$

and $X_{tj}(\lambda) = X_{tj}$ for untransformed variables such as the constant term ($X_{t0} = 1$) and dummy variables. Assuming normality, the log likelihood function is given by

$$\ell(\beta, \sigma^2, \lambda) \propto -\frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left\{Y_t(\lambda) - \sum_{j=0}^{p}\beta_j X_{tj}(\lambda)\right\}^2 + \log J(\lambda), \tag{2}$$

where $J(\lambda) = |\prod_{t=1}^{T} \partial Y_t(\lambda)/\partial Y_t|$.

In matrix notation model (1) is written as $Y(\lambda) = X(\lambda)\beta + \epsilon$, and for a given λ, $\ell(\beta, \sigma^2, \lambda)$ is maximized at

$$\hat{\beta}(\lambda) = [X'(\lambda)X(\lambda)]^{-1}X'(\lambda)Y(\lambda), \quad \hat{\sigma}^2(\lambda) = n^{-1}\|M(\lambda)Y(\lambda)\|^2,$$

where $M(\lambda) = I_T - X(\lambda)[X'(\lambda)X(\lambda)]^{-1}X'(\lambda)$.

The profile likelihood for λ is

$$\ell_p(\lambda) = \ell[\hat{\beta}(\lambda), \hat{\sigma}^2(\lambda), \lambda] \propto -\frac{n}{2}\log \hat{\sigma}^2(\lambda) + \log J(\lambda). \tag{3}$$

The profile score function for λ is

$$S_p(\lambda) = \frac{\mathrm{d}\ell_p(\lambda)}{\mathrm{d}\lambda} = -\frac{TY'(\lambda)M(\lambda)[\dot{Y}(\lambda) - \dot{X}(\lambda)\hat{\beta}(\lambda)]}{\|M(\lambda)Y(\lambda)\|^2} + 1_T' \log Y, \tag{4}$$

where 1_T is a column vector of 1s, $\dot{Y}(\lambda) = \mathrm{d}Y(\lambda)/\mathrm{d}\lambda$ and similarly for $\dot{X}(\lambda)$ (see Appendix A). The profile score $S_p(\lambda)$ is the key quantity needed for developing the score tests. Maximizing $\ell_p(\lambda)$ or solving $S_p(\lambda)=0$ gives the maximum likelihood estimator (MLE) $\hat{\lambda}$ of λ. Substituting $\hat{\lambda}$ back into $\hat{\beta}(\lambda)$ and $\hat{\sigma}^2(\lambda)$ gives the unrestricted MLEs of β and σ^2.

The score test statistic for testing the null hypothesis, $H_0:\lambda=\lambda_0$ is formally defined as (Cox and Hinkley, 1974, p. 324)

$$T_S(\lambda_0)=\frac{S_p(\lambda_0)}{\varpi[\hat{\beta}(\lambda_0),\,\hat{\sigma}^2(\lambda_0),\,\lambda_0]} \tag{5}$$

where $\varpi^2(\beta,\ \sigma^2,\ \lambda)=I_{\lambda\lambda}-I_{\lambda\psi}I^{-1}_{\psi\psi}I_{\psi\lambda}$ is the asymptotic variance of $S_p(\lambda)$, $\psi=(\beta',\ \sigma^2)'$, the I-quantities are the elements of the expected information matrix (see Appendix A) and $\hat{\beta}(\lambda_0)$ and $\hat{\sigma}^2(\lambda_0)$ are the MLEs of β and σ^2 under H_0.

To operationalize (5) we need explicit expressions for the expected information. This is, however, difficult in general and many turn to the observed information instead. The major drawback of the observed information is that the positivity of the variance estimator is not guaranteed, especially when it is evaluated at a point λ_0 that is far away from $\hat{\lambda}$ (Lawrance, 1987). Yang and Abeysinghe (2002) obtained an explicit expression for the asymptotic variance of $\hat{\lambda}$. Using their result and noting that $\mathrm{Var}[S_p(\lambda)]=1/\mathrm{Var}(\hat{\lambda})$ for large T, we can directly extract the following expressions for the denominator of (5):

$$\varpi^2(\beta,\sigma^2,0)=\frac{1}{\sigma^2}\|M(0)\delta\|^2+2\|\mu(0)-\bar{\mu}(0)\|^2+\frac{3}{2}T\sigma^2, \tag{6}$$

$$\varpi^2(\beta,\sigma^2,\lambda)\approx\frac{1}{\sigma^2}\|M(\lambda)\delta\|^2+\frac{1}{\lambda^2}\left[2\|\phi-\bar{\phi}\|^2-4(\phi-\bar{\phi})'(\theta^2-\overline{\theta^2})+\frac{3}{2}\|\theta\|^2\right], \tag{7}$$

where $\mu(\lambda)=\mathbf{X}(\lambda)\beta$, $\phi=\log(1+\lambda\mu(\lambda))$, $\theta=\lambda\sigma/(1+\lambda\mu(\lambda))$, $\|\cdot\|$ is the Euclidian norm, $\delta=\frac{1}{2}[\mu^2(0)+\sigma^2]-\dot{X}(0)\beta$, for $\lambda=0$; and $1/\lambda^2(1+\lambda\mu(\lambda))\#\phi+(\sigma/2\lambda)\theta-\dot{X}(\lambda)\beta$, for $\lambda\neq 0$. Moreover, for vectors $\mathbf{a}$ and $\mathbf{b}$ of length T and a constant c, $\bar{a}$ is the average, $\mathbf{a}^2=\{a_t^2\}_{T\times 1}$ (similarly for the other functions), $\mathbf{a}\#\mathbf{b}=\{a_t b_t\}_{T\times 1}$, and $\mathbf{a}+c=\{a_t+c\}_{T\times 1}$.

If explicit expressions of the expected information are not available, then $I_{\lambda\lambda}-I_{\lambda\psi}I^{-1}_{\psi\psi}I_{\psi\lambda}$ may be replaced by $J_{\lambda\lambda}-J_{\lambda\psi}J^{-1}_{\psi\psi}J_{\psi\lambda}$, where the J-quantities are the elements of the observed information matrix (see Appendix A). Calculation of this quantity seems to be a burden as it does not appear to be simplifiable to an acceptable form. However, by noting that $I_{\beta\sigma^2}=0$, we can use the following simpler but asymptotically equivalent quantity $J_{\lambda\lambda}-J_{\sigma^2\beta}J^{-1}_{\beta\beta}J_{\beta\sigma^2}-J_{\lambda\sigma^2}J^{-1}_{\sigma^2\sigma^2}J_{\sigma^2\lambda}$ as an approximation, which, after some algebra, is reduced to the following nice form:

$$\begin{aligned}\kappa^2[\beta,\sigma^2,\lambda] &= J_{\lambda\lambda}-J_{\sigma^2\beta}J^{-1}_{\beta\beta}J_{\beta\sigma^2}-J_{\lambda\sigma^2}J^{-1}_{\sigma^2\sigma^2}J_{\sigma^2\lambda}\\ &=\frac{1}{\sigma^2}\left[e'\ddot{e}+\dot{e}'M(\lambda)\dot{e}-e'A(\lambda)e-2\dot{e}B(\lambda)e-\frac{(e'\dot{e})^2}{e'e-\sigma^2T/2}\right]\end{aligned}$$

where $A(\lambda)=\dot{X}(\lambda)[X'(\lambda)X(\lambda)]^{-1}\dot{X}'(\lambda)$, $B(\lambda)=X(\lambda)[X'(\lambda)X(\lambda)]^{-1}\dot{X}'(\lambda)$, $e=e(\lambda,\ \beta)=Y(\lambda)-X(\lambda)\beta$, $\dot{e}=(\partial/\partial\lambda)e(\lambda,\ \beta)$, and $\ddot{e}=(\partial^2/\partial\lambda^2)e(\lambda,\ \beta)$ (see Appendix A for the expressions of $\ddot{Y}(\lambda)$ and $\ddot{X}(\lambda)$). This gives the following score test statistic with observed information:

$$T_S^0(\lambda_0) = \frac{S_p(\lambda_0)}{\kappa[\hat{\beta}(\lambda_0),\, \hat{\sigma}^2(\lambda_0), \lambda_0]}. \tag{8}$$

As noted earlier, the positivity of the variance estimator $\kappa^2[\hat{\beta}(\lambda_0),\, \hat{\sigma}^2(\lambda_0), \lambda_0]$ is not guaranteed, especially when λ_0 is far away from $\hat{\lambda}$. This problem does not arise in the case of $\varpi^2[\hat{\beta}(\lambda_0),\, \hat{\sigma}^2(\lambda_0), \lambda_0]$.

3. Monte Carlo results

The two score tests, one based on expected information and the other based on observed information, are asymptotically equivalent. Both test statistics have the standard normal as their limiting distribution. The performance of the two tests in small samples may, however, differ. We carry out a Monte Carlo simulation to evaluate the small sample performance of the two test statistics. The following model is used in the Monte Carlo experiment.

$$Y_t(\lambda) = \beta_0 + \beta_1 X_t(\lambda) + \sigma e_t, \quad t = 1, 2, \ldots, T, \tag{9}$$

where the log X_t values are selected uniformly from the interval (0, 5]. Several parameter configurations are considered to assess the effect of the parameter values on the performance of the test statistics. To ensure that $P\{1 + \lambda(X_t\beta + \epsilon_t) \leq 0\}$ is negligible, the values for λ and β are chosen to have the same sign.

The basic steps of the experiment are as follows. For a given parameter configuration, (i) generate $\{e_t,\, t = 1, \ldots, T\}$ from the standard normal population, (ii) convert $e_t s$ to $Y_t(\lambda)s$ using (9), (iii) invert $Y_t(\lambda)$ to get $Y_t s$ in the original scale, and (iv) calculate $\hat{\beta}_0(\lambda_0)$, $\hat{\beta}_1(\lambda_0)$, $\hat{\sigma}^2(\lambda_0)$, and then the test statistics. Based on 10,000 replicates, the simulated size, null distribution and power of the test statistics $T_S(\lambda_0)$ and $T_S^0(\lambda_0)$ are obtained.

3.1. Size of the tests

The simulated sizes of the tests are summarized in Table 1. The full set of results are given only for $T = 50$. From the results we see that the size of $T_S(\lambda_0)$ is very close to the nominal level, whereas the size of $T_S^0(\lambda_0)$ is larger than and in certain cases twice as large as the nominal level. The size of $T_S(\lambda_0)$ does not change much with the parameter values, but not so in the case of $T_S^0(\lambda_0)$. The results for $T = 25$ and $T = 100$ show that the sample size has a greater effect on $T_S^0(\lambda_0)$ than on $T_S(\lambda_0)$.

3.2. Null distributions of the tests

Besides the size of the tests, the null behavior of the tests can be further assessed by simulating some important characteristics of the distributions, such as means, standard deviations, and percentage values, to see how much these characteristics differ from those of the standard normal distribution. Table 2 presents a portion of the results, from which we see that the distribution of $T_S(\lambda_0)$ is generally much closer to the standard normal than that of $T_S^0(\lambda_0)$.

Table 1
Simulated sizes of the score tests

T	λ	σ	$\alpha=0.1$		$\alpha=0.05$		$\alpha=0.01$	
			T_S	T_S^0	T_S	T_S^0	T_S	T_S^0
50	−1.0	0.01	0.1015	0.1201	0.0494	0.0696	0.0120	0.0229
	0.1	0.1072	0.1296	0.0514	0.0744	0.0094	0.0256	
	1.0	0.0902	0.1257	0.0456	0.0839	0.0093	0.0421	
	−0.5	0.01	0.1070	0.1235	0.0540	0.0745	0.0111	0.0223
	0.1	0.1085	0.1305	0.0552	0.0744	0.0096	0.0247	
	1.0	0.0995	0.1340	0.0500	0.0815	0.0102	0.0360	
	−0.25	0.01	0.1040	0.1249	0.0516	0.0706	0.0106	0.0218
	0.1	0.1057	0.1258	0.0543	0.0740	0.0100	0.0224	
	1.0	0.0993	0.1195	0.0505	0.0667	0.0118	0.0234	
	0.0	0.01	0.1103	0.1320	0.0565	0.0745	0.0112	0.0251
	0.1	0.0985	0.1183	0.0475	0.0665	0.0096	0.0201	
	1.0	0.0890	0.1020	0.0422	0.0515	0.0094	0.0107	
	0.25	0.01	0.1068	0.1282	0.0527	0.0705	0.0094	0.0222
	0.1	0.1043	0.1244	0.0513	0.0716	0.0089	0.0223	
	1.0	0.1001	0.1263	0.0519	0.0733	0.0087	0.0222	
	0.5	0.01	0.1088	0.1286	0.0540	0.0756	0.0102	0.0224
	0.1	0.1000	0.1217	0.0502	0.0675	0.0087	0.0207	
	1.0	0.1051	0.1268	0.0516	0.0744	0.0097	0.0231	
	1.0	0.01	0.1087	0.1282	0.0555	0.0741	0.0106	0.0236
	0.1	0.1084	0.1275	0.0549	0.0744	0.0085	0.0207	
	1.0	0.1107	0.1314	0.0534	0.0778	0.0104	0.0221	
25	0.0	0.01	0.1196	0.1602	0.0575	0.1028	0.0108	0.0420
	0.1	0.1022	0.1442	0.0541	0.0900	0.0094	0.0354	
	1.0	0.0853	0.1212	0.0442	0.0678	0.0084	0.0153	
	0.25	0.01	0.1173	0.1568	0.0570	0.0995	0.0092	0.0417
	0.1	0.1154	0.1585	0.0575	0.1017	0.0096	0.0432	
	1.0	0.0956	0.1515	0.0445	0.0941	0.0079	0.0365	
100	0.0	0.01	0.1054	0.1159	0.0513	0.0599	0.0098	0.0149
	0.1	0.1024	0.1132	0.0507	0.0597	0.0083	0.0136	
	1.0	0.0961	0.1040	0.0506	0.0529	0.0096	0.0107	
	0.25	0.01	0.1027	0.1110	0.0548	0.0616	0.0109	0.0165
	0.1	0.0973	0.1090	0.0492	0.0573	0.0102	0.0146	
	1.0	0.0945	0.1009	0.0479	0.0534	0.0113	0.0112	

Note: $\beta=(8.0,\ 1.25)$ for $\lambda\geq 0$; and $(-8,\ -1.12)$ for $\lambda<0$.

3.3. Power of the tests

The power of the score tests is simulated over a grid of null λ values. To make a fair comparison, simulated percentage points of the two tests (given in Table 2) are used. This ensures that both tests have comparable sizes. Table 3 summarizes the powers of the 5% tests. The null values (λ_0) are chosen to be r standard deviations bellow or above the true (alternative) values of λ. The simulated standard deviations of $\hat{\lambda}$, the MLE of λ, are given in the last column of Table 3 under the heading of

Table 2
The null distributions of $T_S(\lambda_0)$ and $T_S^0(\lambda_0)$

	Nominal value	$\sigma = 0.01$		$\sigma = 0.1$		$\sigma = 1.0$	
		$T_S(\lambda_0)$	$T_S^0(\lambda_0)$	$T_S(\lambda_0)$	$T_S^0(\lambda_0)$	$T_S(\lambda_0)$	$T_S^0(\lambda_0)$
		$\beta = (8.0,\ 1.25),\ T = 50,\ \lambda_0 = 0.0$					
Mean	0.0000	−0.0103	−0.0119	−0.0146	−0.0148	−0.0029	−0.0030
S.D.	1.0000	1.0196	1.0949	0.9986	1.0699	0.9773	1.0220
$Q_{0.050}$	−1.6449	−1.7091	−1.8251	−1.6633	−1.7847	−1.6288	−1.7123
$Q_{0.950}$	1.6449	1.6567	1.7585	1.6208	1.7151	1.5822	1.6653
$Q_{0.025}$	−1.9600	−2.0155	−2.2129	−1.9811	−2.1393	−1.9561	−2.0535
$Q_{0.975}$	1.9600	1.9518	2.1193	1.9089	2.0781	1.9195	1.9813
$Q_{0.005}$	−2.5758	−2.6151	−3.0547	−2.6168	−2.9991	−2.5793	−2.7129
$Q_{0.995}$	2.5758	2.5492	3.0302	2.4772	2.8594	2.5120	2.5822
		$\beta = (8.0,\ 1.25),\ T = 50,\ \lambda = 0.25$					
Mean	0.0000	0.0228	0.0244	−0.0145	−0.0169	0.0184	0.0259
S.D.	1.0000	1.0225	1.0987	1.0193	1.0955	0.9927	1.0807
$Q_{0.050}$	−1.6449	−1.6596	−1.7617	−1.7114	−1.8271	−1.6360	−1.7600
$Q_{0.950}$	1.6449	1.7102	1.8181	1.6650	1.7551	1.6307	1.7765
$Q_{0.025}$	−1.9600	−1.9866	−2.1659	−2.0037	−2.1941	−1.9501	−2.1235
$Q_{0.975}$	1.9600	2.0447	2.2385	1.9745	2.1300	1.8953	2.1175
$Q_{0.005}$	−2.5758	−2.6247	−3.0763	−2.5208	−2.9394	−2.5549	−2.9420
$Q_{0.995}$	2.5758	2.6280	3.0875	2.5817	2.9972	2.4984	3.0470
		$\beta = (8.0,\ 1.25),\ T = 50,\ \lambda = 0.5$					
Mean	0.0000	0.0081	0.0087	−0.0010	−0.0026	0.0259	0.0260
S.D.	1.0000	1.0129	1.0873	1.0253	1.0989	1.0002	1.0807
$Q_{0.050}$	−1.6449	−1.6730	−1.7759	−1.6883	−1.7975	−1.6315	−1.7381
$Q_{0.950}$	1.6449	1.6744	1.7770	1.6986	1.7933	1.6551	1.7757
$Q_{0.025}$	−1.9600	−2.0026	−2.1852	−2.0195	−2.2127	−1.9511	−2.1425
$Q_{0.975}$	1.9600	2.0015	2.1840	1.9957	2.1713	1.9698	2.1620
$Q_{0.005}$	−2.5758	−2.5825	−3.0130	−2.5863	−3.0298	−2.5624	−3.0142
$Q_{0.995}$	2.5758	2.5677	2.9896	2.5352	2.9448	2.5379	2.8905
		$\beta = (8.0,\ 1.25),\ T = 50,\ \lambda = 1.0$					
Mean	0.0000	−0.0094	−0.0109	0.0026	0.0027	0.0509	0.0496
S.D.	1.0000	1.0171	1.0931	1.0259	1.1009	1.0139	1.0915
$Q_{0.050}$	−1.6449	−1.7027	−1.8119	−1.6776	−1.7779	−1.6246	−1.7313
$Q_{0.950}$	1.6449	1.6536	1.7510	1.7011	1.8076	1.7117	1.8093
$Q_{0.025}$	−1.9600	−2.0278	−2.2207	−1.9917	−2.1723	−1.9390	−2.1192
$Q_{0.975}$	1.9600	1.9425	2.1070	2.0049	2.1804	2.0097	2.1809
$Q_{0.005}$	−2.5758	−2.6206	−3.0812	−2.5002	−2.9040	−2.4954	−3.0017
$Q_{0.995}$	2.5758	2.6029	3.0466	2.5778	2.9994	2.5764	2.9539

Table 3
Powers of $T_S(\lambda_0)$ (upper entry) and $T_S^0(\lambda_0)$ at 5% level, $\beta = (8.0, 1.25)$, $T = 50$

σ	$\lambda_0 = \lambda + r\ sd(\hat{\lambda})$									$sd(\hat{\lambda})$
	$r = -4$	-3	-2	-1	0	1	2	3	4	
	$\lambda = 0.0$									
0.01	0.9773	0.8496	0.5257	0.1711	0.0537	0.1591	0.4914	0.8357	0.9736	0.00487
	0.9659	0.8501	0.5273	0.1711	0.0533	0.1567	0.4892	0.8326	0.9638	
0.1	0.9613	0.8254	0.5212	0.1850	0.0524	0.1661	0.4882	0.8099	0.9530	0.03777
	0.9737	0.8484	0.5345	0.1809	0.0518	0.1691	0.5117	0.8385	0.9734	
1.0	0.9249	0.7705	0.4681	0.1671	0.0479	0.1600	0.4575	0.7478	0.9176	0.05421
	0.9775	0.8658	0.5526	0.1899	0.0482	0.1700	0.5313	0.8386	0.9717	
	$\lambda = 0.25$									
0.01	0.9669	0.8008	0.4510	0.1389	0.0458	0.1584	0.4849	0.8181	0.9711	0.00151
	0.9596	0.8008	0.4511	0.1390	0.0457	0.1585	0.4849	0.8169	0.9632	
0.1	0.9794	0.8434	0.5136	0.1718	0.0548	0.1609	0.4978	0.8357	0.9717	0.01539
	0.9712	0.8480	0.5198	0.1739	0.0551	0.1614	0.4999	0.8371	0.9589	
1.0	0.9689	0.8575	0.5384	0.1838	0.0527	0.1716	0.5021	0.8083	0.9512	0.09973
	0.9805	0.8717	0.5463	0.1822	0.0507	0.1704	0.5165	0.8342	0.9672	
	$\lambda = 0.5$									
0.01	0.9766	0.8396	0.5076	0.1669	0.0506	0.1586	0.4935	0.8322	0.9737	0.00080
	0.9669	0.8386	0.5080	0.1663	0.0505	0.1589	0.4937	0.8314	0.9621	
0.1	0.9749	0.8371	0.5011	0.1631	0.0463	0.1566	0.4715	0.8242	0.9706	0.00793
	0.9673	0.8361	0.5003	0.1629	0.0456	0.1572	0.4725	0.8234	0.9581	
1.0	0.9763	0.8442	0.5169	0.1706	0.0549	0.1712	0.4854	0.7984	0.9478	0.06972
	0.9804	0.8660	0.5308	0.1737	0.0541	0.1684	0.4923	0.8131	0.9501	
	$\lambda = 1.0$									
0.01	0.9771	0.8479	0.5189	0.1757	0.0503	0.1527	0.4951	0.8264	0.9735	0.00038
	0.9681	0.8473	0.5192	0.1759	0.0502	0.1524	0.4939	0.8252	0.9627	
0.1	0.9741	0.8307	0.4978	0.1576	0.0513	0.1545	0.4883	0.8224	0.9726	0.00376
	0.9657	0.8307	0.5007	0.1579	0.0516	0.1552	0.4891	0.8224	0.9608	
1.0	0.9825	0.8470	0.5195	0.1683	0.0449	0.1728	0.4952	0.8144	0.9608	0.03614
	0.9830	0.8533	0.5198	0.1673	0.0475	0.1743	0.4996	0.8199	0.9514	

$sd(\hat{\lambda})$ = Simulated standard deviation of $\hat{\lambda}$ based on 10,000 replicates.

$sd(\hat{\lambda})$. The results summarized in Table 3 suggest that the two tests are generally comparable in terms of powers when their sizes are adjusted to the same level. The results also suggest that the tests are very powerful in the sense that they are able to detect a small change in λ. For example, when $\lambda = 0.25$ (the true value) with $\sigma = 0.1$, the probabilities of rejecting H_0:$\lambda = 0.3116$ (4 *sd*s above 0.25) are 0.9717 and 0.9589, respectively, for $T_S(\lambda_0)$ and $T_S^0(\lambda_0)$.

Appendix A. Scores and information

The score functions are:

$$\frac{\partial \ell(\beta, \sigma^2, \lambda)}{\partial \beta} = \frac{1}{\sigma^2} X'(\lambda)e(\lambda, \beta)$$
$$\frac{\partial \ell(\beta, \sigma^2, \lambda)}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} e'(\lambda, \beta)e(\lambda, \beta)$$
$$\frac{\partial \ell(\beta, \sigma^2, \lambda)}{\partial \lambda} = -\frac{1}{\sigma^2} e'(\lambda, \beta)\dot{e}(\lambda, \beta) + 1_T' \log Y$$

The elements of the observed information matrix are:

$$J_{\beta\beta} = -\frac{\partial^2 \ell}{\partial \beta \partial \beta'} = \frac{1}{\sigma^2} X'(\lambda)X(\lambda)$$
$$J_{\sigma^2\sigma^2} = -\frac{\partial^2 \ell}{\partial (\sigma^2)^2} = -\frac{T}{2\sigma^4} + \frac{1}{\sigma^6} e'(\lambda, \beta)e(\lambda, \beta)$$
$$J_{\lambda\lambda} = -\frac{\partial^2 \ell}{\partial \lambda^2} = \frac{1}{\sigma^2} [\dot{e}'(\lambda, \beta)\dot{e}(\lambda, \beta) + e'(\lambda, \beta)\ddot{e}(\lambda, \beta)]$$
$$J_{\beta\sigma^2} = -\frac{\partial^2 \ell}{\partial \beta \partial \sigma^2} = \frac{1}{\sigma^4} X'(\lambda)e(\lambda, \beta)$$
$$J_{\beta\lambda} = -\frac{\partial^2 \ell}{\partial \beta \partial \lambda} = -\frac{1}{\sigma^2} [X'(\lambda)\dot{e}(\lambda, \beta) + \dot{X}(\lambda)e(\lambda, \beta)]$$
$$J_{\sigma^2\lambda} = -\frac{\partial^2 \ell}{\partial \sigma^2 \partial \lambda} = -\frac{1}{\sigma^4} e'(\lambda, \beta)\dot{e}(\lambda, \beta)$$

The elements of the expected information matrix are:

$$I_{\beta\beta} = \frac{1}{\sigma^2} \mathbf{X}'(\lambda)\mathbf{X}(\lambda), \quad I_{\sigma^2\sigma^2} = \frac{T}{2\sigma^4}, \quad I_{\lambda\lambda} = \frac{1}{\sigma^2} E[\dot{e}'(\lambda, \beta)\dot{e}(\lambda, \beta) + e'(\lambda, \beta) + \ddot{e}(\lambda, \beta)],$$
$$I_{\beta\lambda} = -\frac{1}{\sigma^2} [\mathbf{X}'(\lambda)E[\dot{e}(\lambda, \beta)], \quad I_{\beta\sigma^2} = 0 \quad I_{\sigma^2\lambda} = -\frac{1}{\sigma^4} E[e'(\lambda, \beta)\dot{e}(\lambda, \beta)],$$

The partial derivatives of $Y_t(\lambda)$ are:

$$\dot{Y}_t(\lambda) = \begin{cases} \frac{1}{\lambda} [1 + \lambda Y_t(\lambda)] \log Y_t - \frac{1}{\lambda} Y_t(\lambda), & \lambda \neq 0, \\ \frac{1}{2} (\log Y_t)^2, & \lambda = 0, \end{cases}$$
$$\ddot{Y}_t(\lambda) = \begin{cases} \dot{Y}_t(\lambda)\left(\log Y_t - \frac{1}{\lambda}\right) - \frac{1}{\lambda^2} [\log Y_t - Y_t(\lambda)], & \lambda \neq 0, \\ \frac{1}{3} (\log Y_t)^3, & \lambda = 0, \end{cases}$$

and similarly for the partial derivatives of $X(\lambda)$. The $\dot{X}(\lambda)$ and $\ddot{X}(\lambda)$ corresponding to untransformed Xs are columns of zeros.

References

Box, G.E.P., Cox, D.R., 1964. An analysis of transformation (with discussion). Journal of the Royal Statistical Society B 26, 211–252.
Cox, D.R., Hinkley, D.V., 1974. Theoretical Statistics. Chapman & Hall, London.
Davidson, R., MacKinnon, J.G., 1985. Testing linear and loglinear regressions against Box–Cox alternatives. Canadian Journal of Economics 18, 499–517.
Lawrance, A.J., 1987. The score statistic for regression transformation. Biometrika 74, 275–279.
Yang, Z., Abeysinghe, T., 2002. An explicit variance formula for the Box–Cox functional form estimator. Economics Letters 76, 259–265.