Generalized *p*-Values for Comparing Regression Lines

Tsai-Yu Lin[1,3], and Chi-Rong Li[2]
[1] Department of Applied Mathematics, Feng-Chia University, Taiwan
[2] School of Nursing, Chung Shan Medical University, Taiwan
[3] Corresponding author: Tsai-Yu Lin, e-mail: linty@fcu.edu.tw

Abstracts

One frequently encountered problem in scientific research is to assess whether two regression models for two groups or two treatments, etc., are the same or not. However, the regression models rarely hold over the whole space of the covariates. Therefore testing the difference between the two models over a restricted region or an arbitrary interval one is interested is therefore useful. In this paper, we develop testing procedures to detect if one regression line is strictly greater than the other for given a specific region. Our testing procedures are developed mainly based on the concept of generalized p-value (GPV), which has been successfully used to provide small sample solutions for many hypothesis testing problems when nuisance parameters are present. Moreover, detailed statistical simulation studies will be conducted to evaluate the effectiveness by using their empirical size and power of the proposed procedures.

Keywords: Generalized pivotal quantities, Generalized test variables, Over a finite interval

1. Introduction

Often we are interested in the difference between the two regression models only over a restricted region of the covariates. Therefore comparing of two regression lines over the whole space of covariates is likely to be neither efficient if only the differences over a restricted region are of interest, nor appropriate if the models hold only over certain restricted region. For example, one may want to determine if the regression of *Y* on *x* for one set of observations is uniformly greater than that of the other over an arbitrary finite interval $\mathrm{I} = [x_{\mathrm{L}}, x_{\mathrm{U}}]$ when given two independent sets of bivariate observations, $(x_{11}, Y_{11}), (x_{12}, Y_{12}), \dots, (x_{1\mathrm{n}_1}, Y_{1\mathrm{n}_1})$, and $(x_{21}, Y_{21}), (x_{22}, Y_{22})$, $\dots, (x_{2\mathrm{n}_2}, Y_{2\mathrm{n}_2})$. In fact, this problem has been widely discussed in various practical applications, such as biological applications, pharmaceutical bioequivalence, educational or psychological issues, etc. It was tackled by Tsutakawa and Hewett (1978) (henceforth abbreviated as T-H) under the following assumption. Given the values of the independent variables, x_{ij}, the two sets of observations, $\boldsymbol{Y}_{\mathrm{i}}$, are independent and normal with $\mathrm{E}(Y_{\mathrm{ij}}|X_{\mathrm{ij}}=x_{\mathrm{ij}}) = \alpha_{\mathrm{i}} + \beta_{\mathrm{i}} x_{\mathrm{ij}}$ and $\mathrm{Var}(Y_{\mathrm{ij}}|X_{\mathrm{ij}}=x_{\mathrm{ij}}) = \sigma^2$, where $(\alpha_{\mathrm{i}}, \beta_{\mathrm{i}})(\mathrm{i} = 1,2; \mathrm{j} = 1,2, \dots, \mathrm{n}_{\mathrm{i}})$ and $\sigma^2 > 0$ are unknown parameters. They first considered the joint distribution of the distances between the two regression lines at x_{L} and x_{U} to solve this problem. The study was motivated by a set of experiments conducted by Soler-Argilaga and Heimberg (1976), which were designed to study the effect of sex on hepatic metabolism of free fatty acid (FFA). One of the questions asked was whether the amount of coleic acid incorporated in triglyceride, *Y*, is strictly higher for female rats than for male rats, regardless of the FFA uptake, *x*, over a specified region. They especially noted that there was a physical limitation to the amount of *x* that the rats could tolerate and over which the experiment was meaningful. Therefore, the issue for comparing of two regression lines over a finite interval is very important. Some related literatures proposed testing procedures for comparing two regression lines, for instance, e.g., see Hewett and Lababidi (1980), Spurrier *et al*. (1982), Hill and Padmanabhan (1984), etc. Note that they all need the assumption of

homogeneous variance for each regression line.

As indicated earlier, T-H procedure utilizes the joint distribution of the distances between the two regression lines at x_{L} and x_{U} and assumes normality with homogeneous variance. However, they actually didn't derive the sampling distribution of the order statistic which is required in their study. Since it is very complicated, this makes establishing the exact test very difficult. In this paper, we shall make an attempt to provide alternative testing procedures based on a specific region of the covariate by the GPV. Tsui and Weerahandi (1989) suggest the use of the GPV for the construction of hypothesis testing procedures. Together with the generalized confidence interval (GCI), the GPV has been successfully applied to provide inferences for tolerance intervals by Liao *et al*. (2005), population and individual bioequivalence by McNally, *et al*. (2003), variance components by Mathew and Webb (2005), multivariate analysis of variance by Gamage *et al*. (2004), and non-inferiority test by Li *et al*. (2008). As a result, we propose to apply the concept of the GPV to construct a testing procedure for comparing the regression lines over a finite interval. It is worth to say that the proposed test based on the GPV without the assumption of identical variance are much more practical than tests with this assumption. In the next section, the proposed test using the GPV is given. Section 3 presents the results of simulation studies. Finally, the conclusion and some remarks are provided in the last section.

2. The Test Based on GPV

Suppose that $\mathbf{X}$ is a random sample whose distribution depends on a vector of unknown parameters $\boldsymbol{\varsigma} = (\theta, \boldsymbol{\eta})$, where θ is a parameter of interest and $\boldsymbol{\eta}$ is a vector of nuisance parameters. Let $\mathbf{x}$ be the observed value of $\mathbf{X}$. A test variable $T(\mathbf{X}; \mathbf{x}, \boldsymbol{\zeta})$ is said to be a generalized test variable (GTV) if it satisfies the following three conditions:

(a) $T(\mathbf{x}; \mathbf{x}, \boldsymbol{\zeta})$ does not depend on unknown parameters.

(b) For a specified θ, the distribution of $T(\mathbf{X}; \mathbf{x}, \boldsymbol{\zeta})$ is free of $\boldsymbol{\eta}$.

(c) For fixed $\mathbf{x}$ and $\boldsymbol{\eta}$, $P[T(\mathbf{X}; \mathbf{x}, \boldsymbol{\zeta}) \leq t|\theta]$ is monotonic in θ.

Denote $T = T(\mathbf{X}; \mathbf{x}, \boldsymbol{\zeta})$ and $t_{\mathrm{obs}} = T(\mathbf{x}; \mathbf{x}, \boldsymbol{\zeta})$. Without loss of generality, we consider the one-sided test that $H_0: \theta \leq \delta_0$ *versus* $H_1: \theta > \delta_0$. If the test function T is stochastically increasing in θ, then the GPV is defined as

$$P = \sup_{\theta \leq \delta_0}[T \geq t_{\mathrm{obs}}|\theta] = P[T \geq t_{\mathrm{obs}}|\delta_0].$$

For the test with a nominal significance level α, one typically rejects the null hypothesis if $p < \alpha$.

To set the notation, consider the following statistical model:

$$y_{\mathrm{ij}} = \alpha_{\mathrm{i}} + \beta_{\mathrm{i}} x_{\mathrm{ij}} + \varepsilon_{\mathrm{ij}},$$

where y_{ij} is the *j*th observed response for the *i*th treatment group, α_{i} is the *i*th intercept, x_{ij} is the *j*th covariate response for the *i*th treatment group, β_{i} is the slope parameter associated with the covariate for the *i*th treatment group, and $\varepsilon_{\mathrm{ij}}$ is unobservable random error, where $\mathrm{i} = 1,2$ and $\mathrm{j} = 1,2,\ldots,\mathrm{n_i}$. It is usually assumed that $\varepsilon_{\mathrm{ij}} \sim N(0, \sigma_{\mathrm{i}}^2)$ where σ_{i}^2s are unknown. More specifically, let $\hat{\alpha}_{\mathrm{i}}$ and $\hat{\beta}_{\mathrm{i}}$ be the least squares estimates of α_{i} and β_{i}, respectively. And let $MSE_{\mathrm{i}} = \sum_{\mathrm{j}=1}^{\mathrm{n_i}}(\mathrm{y_{ij}} - \hat{y}_{\mathrm{ij}})^2/(\mathrm{n_i} - 2)$ be the estimate of σ_{i}^2, where $\hat{y}_{\mathrm{ij}} = \hat{\alpha}_{\mathrm{i}} + \hat{\beta}_{\mathrm{i}} x_{\mathrm{ij}}$, and $\mathrm{n} = \sum_{\mathrm{i}=1}^{2} \mathrm{n_i}$. Given the interval $\mathrm{I} = [x_{\mathrm{L}}, x_{\mathrm{U}}]$, let θ_{L} and θ_{U} be the distance between the two regression lines at x_{L} and x_{U}, respectively. Namely,

$$\begin{cases} \theta_{\mathrm{L}} = (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)\mathrm{X_L} \\ \theta_{\mathrm{U}} = (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)\mathrm{X_U} \end{cases}.$$

In this paper, we propose an alternative testing procedure other than T-H method to determine whether one regression line is strictly greater than the other for a specific region. Typically, the hypothesis for one-sided test is considered as follows:

$$\begin{cases} H_0 \in \{(\theta_{\mathrm{L}}, \theta_{\mathrm{U}}): [\theta_{\mathrm{L}} \leq 0 \text{ and } \theta_{\mathrm{U}} \geq 0], [\theta_{\mathrm{L}} \geq 0 \text{ and } \theta_{\mathrm{U}} \leq 0] \text{ or } [\theta_{\mathrm{L}} < 0 \text{ and } \theta_{\mathrm{U}} < 0]\} \\ H_1 \in \{(\theta_{\mathrm{L}}, \theta_{\mathrm{U}}): \theta_{\mathrm{L}} > 0 \text{ and } \theta_{\mathrm{U}} > 0\} \end{cases}.$$

The sampling distributions of the estimated regression coefficients $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\alpha}_2$ and $\hat{\beta}_2$ are as follows:

$$\begin{bmatrix}\hat{\alpha}_1\\ \hat{\beta}_1\end{bmatrix} \sim N\left(\begin{bmatrix}\alpha_1\\ \beta_1\end{bmatrix}, \begin{bmatrix}\frac{1}{n_1}+\frac{\bar{X}_1^2}{\sum_{i=1}^{n_1}(X_{1i}-\bar{X}_1)^2} & \frac{-\bar{X}_1}{\sum_{i=1}^{n_1}(X_{1i}-\bar{X}_1)^2}\\ \frac{-\bar{X}_1}{\sum_{i=1}^{n_1}(X_{1i}-\bar{X}_1)^2} & \frac{1}{\sum_{i=1}^{n_1}(X_{1i}-\bar{X}_1)^2}\end{bmatrix}\sigma_1^2\right),$$

$$\begin{bmatrix}\hat{\alpha}_2\\ \hat{\beta}_2\end{bmatrix} \sim N\left(\begin{bmatrix}\alpha_2\\ \beta_2\end{bmatrix}, \begin{bmatrix}\frac{1}{n_2}+\frac{\bar{X}_2^2}{\sum_{i=1}^{n_2}(X_{2i}-\bar{X}_2)^2} & \frac{-\bar{X}_2}{\sum_{i=1}^{n_2}(X_{2i}-\bar{X}_2)^2}\\ \frac{-\bar{X}_2}{\sum_{i=1}^{n_2}(X_{2i}-\bar{X}_2)^2} & \frac{1}{\sum_{i=1}^{n_2}(X_{2i}-\bar{X}_2)^2}\end{bmatrix}\sigma_2^2\right).$$

It is assumed that observable statistics $\hat{\alpha}_1, \hat{\beta}_1, \hat{\alpha}_2, \hat{\beta}_2, MSE_1$ and MSE_2 are available such that

$$U_1 = \left\{\begin{bmatrix}\hat{\alpha}_1\\ \hat{\beta}_1\end{bmatrix} - \begin{bmatrix}\alpha_1\\ \beta_1\end{bmatrix}\right\}\left\{\begin{bmatrix}\frac{1}{n_1}+\frac{\bar{X}_1^2}{\sum_{i=1}^{n_1}(X_{1i}-\bar{X}_1)^2} & \frac{-\bar{X}_1}{\sum_{i=1}^{n_1}(X_{1i}-\bar{X}_1)^2}\\ \frac{-\bar{X}_1}{\sum_{i=1}^{n_1}(X_{1i}-\bar{X}_1)^2} & \frac{1}{\sum_{i=1}^{n_1}(X_{1i}-\bar{X}_1)^2}\end{bmatrix}\sigma_1^2\right\}^{\frac{-1}{2}} \sim N\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right),$$

$$U_2 = \left\{\begin{bmatrix}\hat{\alpha}_2\\ \hat{\beta}_2\end{bmatrix} - \begin{bmatrix}\alpha_2\\ \beta_2\end{bmatrix}\right\}\left\{\begin{bmatrix}\frac{1}{n_2}+\frac{\bar{X}_2^2}{\sum_{i=1}^{n_2}(X_{2i}-\bar{X}_2)^2} & \frac{-\bar{X}_2}{\sum_{i=1}^{n_2}(X_{2i}-\bar{X}_2)^2}\\ \frac{-\bar{X}_2}{\sum_{i=1}^{n_2}(X_{2i}-\bar{X}_2)^2} & \frac{1}{\sum_{i=1}^{n_2}(X_{2i}-\bar{X}_2)^2}\end{bmatrix}\sigma_2^2\right\}^{\frac{-1}{2}} \sim N\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right),$$

$$U_3 = \frac{(n_1-2)MSE_1}{\sigma_1^2} \sim \chi^2_{n_1-2},\ U_4 = \frac{(n_2-2)MSE_2}{\sigma_2^2} \sim \chi^2_{n_2-2}.$$

Subsequently, we may adopt the following procedure to make decision.

1. Obtain the generalized pivotal quantities (GPQs) of $\sigma_1^2, \sigma_2^2, \alpha_1, \beta_1, \alpha_2, \beta_2$ denoted by $R_{\sigma_1^2}, R_{\sigma_2^2}, R_{\alpha_1}, R_{\beta_1}, R_{\alpha_2}$ and R_{β_2}, respectively. From these GPQs, we can separately obtain the GPQ of θ_{L} and θ_{U}, $R_{\theta_{\mathrm{L}}}$ and $R_{\theta_{\mathrm{U}}}$. Then the GTV of $min(\theta_{\mathrm{L}}, \theta_{\mathrm{U}})$ for the test is given by
$$T_{min\{\theta_{\mathrm{L}},\theta_{\mathrm{U}}\}} = R_{min\{\theta_{\mathrm{L}},\theta_{\mathrm{U}}\}} - min\{\theta_{\mathrm{L}}, \theta_{\mathrm{U}}\}.$$
2. Next the desired GPV of $min(\theta_{\mathrm{L}}, \theta_{\mathrm{U}})$,
$$P\left(T_{min\{\theta_{\mathrm{L}},\theta_{\mathrm{U}}\}} \le t_{min\{\theta_{\mathrm{L}},\theta_{\mathrm{U}}\}} | H_0\right) = P\left(R_{min\{\theta_{\mathrm{L}},\theta_{\mathrm{U}}\}} \le min\{\theta_{\mathrm{L}}, \theta_{\mathrm{U}}\} | H_0\right),$$
can be computed by using Monte-Carlo algorithm. Note that the GPVP is abbreviate from the pooled variance and GPVI is from individual variance.
3. If α is the specified level of significance, we can conclude that line 1 lies uniformly above line 2 over $[x_{\mathrm{L}},\ x_{\mathrm{U}}]$ when GPV$<\alpha$.

3. Simulation Studies

To evaluate the performance of the proposed procedure, the following simulation study is conducted based on Bishop's example on kidney in T-H. We first specify the values of $\alpha = 0.05$. For fixed $\alpha_1 = 400$, $\beta_1 = 20$, $\sigma_1 = 100$, and specified values of n_1, n_2, $\theta_{\mathrm{L}}/\sigma_1$ and $\theta_{\mathrm{U}}/\sigma_1$, we generate a normal random deviate X from the distribution $N(39,13)$ and $\varepsilon_{\mathrm{ij}} \sim N(0, \sigma_{\mathrm{i}})$ with some ratios of σ_1 to σ_2. In order to give some insight on the effect of specific region, the intervals $\mathrm{I}_1 = [26,52], \mathrm{I}_2 = [34,52]$, $\mathrm{I}_3 = [26,44], \mathrm{I}_4 = [34,44]$ are considered. We fixed *K*=2,500 in the Monte Carlo algorithm in computing the desired GPV. The procedure is repeated 2,500 times for each parameter combinations. The partial results are displayed in Tables 1 and 2.

Table 1 demonstrates how the power increases as $\theta_{\mathrm{L}}/\sigma_1$ or $\theta_{\mathrm{U}}/\sigma_1$ increase for a given interval. For I_1 we can see that the power approaches 5% when either $\theta_{\mathrm{L}}/\sigma_1$ or $\theta_{\mathrm{U}}/\sigma_1$ equals 0. This supports the fact that the test has size 5% when equal sample size. Table 1 and Table 2 also illustrate that when we have nested intervals, the power increases as the width of the interval decreases. From Table 2, the power increases as equal sample size increase for a given region. When equal sample size, the three methods have similar performance. However, if different variance and unequal sample size both exist, the GPVI method is proposed.

Table 1 Powers of size 5% tests for four intervals by three methods with $\sigma_1:\sigma_2$=1:1 and $n_1 = n_2 = 25$.

$n_1 = n_2 = 25$			θ_U/σ_1											
			-1			0			1			2		
method			T-H	GPVP	GPVI	T-H	GPVP	GPVI	T-H	GPVP	GPVI	T-H	GPVP	GPVI
$\frac{\theta_L}{\sigma_1}$	-1	I_1							#	#	#	#	#	#
		I_2							0.001	0.003	0.002	0.017	0.028	0.024
		I_3							#	#	#	#	#	#
		I_4							0.001	0.003	0.002	0.017	0.028	0.024
	0	I_1				# *	# *	# *	0.028*	0.029*	0.028*	0.040*	0.050*	0.046*
		I_2				0.006*	0.006*	0.005*	0.204	0.200	0.188	0.608	0.626	0.610
		I_3				0.003*	0.003*	0.003*	0.035*	0.039*	0.036*	0.040*	0.050*	0.047*
		I_4				0.012*	0.014*	0.012*	0.223	0.229	0.218	0.609	0.626	0.612
	1	I_1	#	#	#	0.028*	0.026*	0.024*	0.551	0.544	0.520	0.746	0.764	0.750
		I_2	#	#	#	0.037*	0.043*	0.038*	0.706	0.712	0.695	0.988	0.991	0.990
		I_3	0.002	0.004	0.004	0.202	0.201	0.189	0.701	0.707	0.670	0.748	0.766	0.754
		I_4	0.002	0.004	0.004	0.222	0.236	0.218	0.870	0.877	0.870	0.991	0.992	0.992
	2	I_1	#	#	#	0.042*	0.053*	0.049*	0.756	0.772	0.760	0.993	0.994	0.994
		I_2	#	#	#	0.042*	0.054*	0.050*	0.760	0.778	0.766	0.996	0.997	0.998
		I_3	0.023	0.033	0.028	0.612	0.630	0.616	0.988	0.992	0.990	0.997	0.997	0.997
		I_4	0.023	0.033	0.029	0.612	0.632	0.618	0.992	0.995	0.994	1.000	1.000	1.000

is <0.001 and * is empirical size.

Table 2 Powers of the tests for four intervals by three methods with σ_1: σ_2=1:2, $\theta_L/\sigma_1 = 1$ and θ_U/σ_1=2 when two lines tend to be of the megaphone type.

Method	n_1	25			50		100	
interval	n_2	25	50	100	25	50	25	100
$I_1 = [26,52]$	T-H	0.375	0.442	0.497	0.560	0.706	0.681	0.920
	GPVP	0.372	0.429	0.487	0.567	0.708	0.688	0.923
	GPVI	0.350	0.585	0.774	0.394	0.705	0.436	0.920
$I_2 = [34,52]$	T-H	0.762	0.872	0.941	0.877	0.976	0.940	1.000
	GPVP	0.765	0.866	0.932	0.883	0.976	0.947	1.000
	GPVI	0.748	0.939	0.991	0.770	0.976	0.820	1.000
$I_3 = [26,44]$	T-H	0.413	0.456	0.505	0.583	0.708	0.698	0.920
	GPVP	0.424	0.456	0.509	0.594	0.717	0.705	0.923
	GPVI	0.409	0.600	0.776	0.438	0.713	0.480	0.920
$I_4 = [34,44]$	T-H	0.812	0.895	0.952	0.904	0.980	0.958	1.000
	GPVP	0.820	0.899	0.952	0.920	0.981	0.964	1.000
	GPVI	0.808	0.958	0.992	0.825	0.979	0.866	1.000

4. Conclusions

This study provides the testing procedure to determine if one regression line is strictly greater than the other for given a specific region by using the GPV method. It can be applicable to the heterogeneous error variance and is much more practical than T-H method. According to the simulation results, the three methods are about the same performance when equal sample size. The GPVI method is proposed when different variance and unequal sample size. Therefore, it is shown that the use of GPV allow the testing fairly straightforward and satisfactory. We note that at $(\theta_L, \theta_U) = (0, 0)$, the power could be well below 5%, in contrast to procedures designed to test the equality of two regression lines.

References

Gamage, J., Mathew, T. and Weerahandi, S. (2004) "Generalized *p*-values and generalized confidence regions for the multivariate Behrens-fisher problem and MANOVA," *Journal of Multivariate Analysis*, 88, 177-189.

Hewett, J. E. and Lababidi, Z. (1980) "Comparison of two populations with multivariate data," *Biometrics*, 36, 671-675.

Hill, N. J. and Padmanabhan, A. R. (1984) "Robust comparison of two regression lines and biomedical applications," *Biometrics*, 40, 985-994.

Li, C. R., Liao, C. T., and Liu, J. P. (2008) "A non-inferiority test for diagnostic accuracy based on the paired partial areas under ROC curves," *Statistics in Medicine*, 27(10), 1762-1776.

Liao, C. T., Lin, T. Y. and Iyer, H. K. (2005) "One- and two-sided tolerance intervals for general balanced mixed models and unbalanced one-way random models," *Technometrics*, 47(3), 323-335.

Mathew, T. and Webb, D.W. (2005) "Generalized *p* values and confidence intervals for variance components：applications to army test and evaluation," *Technometrics*, 47(3), 312-322.

McNally, R. J., Iyer, H. K. and Mathew, T. (2003) "Tests for individual and population bioequivalence based on generalized *p*-values," *Statistics in Medicine*, 22, 31-53.

Soler-Argilaga and Heimberg, M. (1976) "Comparison of metabolism of free fatty acid by isolated perfused livers from male and female rates," *Journal of Lipid Research*, 17, 605-615.

Spurrier, J. D., Hewett, J. E. and Lababidi, Z. (1982) "Comparison of two regression

lines over a finite interval," *Biometrics*, 38, 827-836.
Tsui, K. and Weerahandi, S. (1989) "Generalized *p*-values in significance testing of hypotheses in the presence of nuisance parameters," *Journal of American Statistical Association*, 84, 602-607.
Tsutakawa, R. K. and Hewett, J. E. (1978) "Comparison of two regression lines over a finite interval," *Biometrics*, 34, 391-398.