

Journal of
Research Methods and Methodological Issues

Volume 2, Issue 1, 2008

Analyzing Data from a Regression Discontinuity Study: A Research Note.

Howard B. Lee, Professor of Psychology, California State University, Northridge,
howard.lee@csun.edu

Abstract

The Chow Test (Chow, 1960) is a method well known in econometrics. It was originally designed to analyze the same variables obtained in two different data sets to determine if they were similar enough to be pooled together. The method, however, could be used to determine if two regression lines are different from one another. This article discusses the use of the Chow Test on data obtained in a regression discontinuity study.

Introduction

A seldom used quasi-experimental research design that is covered in a number of publications on research methodology such as Christensen (2006) and Kerlinger & Lee (2000) is the regression discontinuity design. A number of popular texts, however, do not mention this method (Graziano & Raulin, 1993; McGuigan, 1997). There are however, a number of documents available on the Internet discussing and using the regression discontinuity design.

This design starts with a selection criterion that separates entities (people) into two groups on the basis of some measurement such as achievement or anxiety. This design then tests whether some intervention, treatment or program change alters the relation between the selection criteria and the outcome measure.

Recent articles written by Zuckerman, Lee, and Wutoh (2006); Gormley, Gayer, & Phillips (2005); Braden and Bryant (1990) shows the utility of the regression-discontinuity method in pharmaceutical health services research; studying the effects of preschool on cognitive development and effects of gifted students' program in school psychology research. Cook and Campbell (1979) and Shadish, Cook, & Campbell, (2001) demonstrated its usefulness in studies that does not have the constraints of laboratory research, e.g. field studies. Braden & Bryant (1990) used both real and fictitious data from the regression-discontinuity design that were analyzed with multiple regression or analysis of covariance using dummy variables. Braden and Bryant (1990) provides an example of a two-dimensional application where the predictor variable is group membership (i.e., selected vs. excluded), the outcome variable was achievement scores and the covariate was IQ scores.

This paper presents a method of analysis that is a variation of the analysis of covariance that does not use dummy variables directly in the computations of the test statistic. Plus this method is general enough that it could be easily adapted for cases where there are more than one independent variable, more than one covariate and nonlinear (e.g. quadratic, curvilinear) predictor variables. This method was originally developed by G.C. Chow (1960) and later explored, expanded and enhanced by Fisher (1970), Toyoda (1974), Schmidt & Sickles (1977). It has appeared in textbooks such as Johnston (1972) and Johnston & Dinardo (1996). This method is well-known in economics and econometrics. The *Chow Test* as it is called, was created to statistically determine if two sets of observations with the same variables could be regarded as belonging to the same regression model. This test allows testing

whether m additional observations are from the same regression as the first sample of n observations. Trochim (1984) briefly mentioned this method in analyzing data from regression discontinuity studies. Trochim gave a few references where the Chow Test was discussed in more detail. However these references were either unpublished dissertations or difficult to obtain reports. It is the intention of this paper to show in greater detail where the Chow Test can be applied to regression discontinuity problems with clear-cut classifications.

The paper by Chow examined financial data that were taken from two different time periods. The two data sets had the same variables. Chow showed that his method could help the researcher determine whether or not newly collected data exhibit the same relationship between dependent and independent variables as the previously collected data. However, this method could be applied to situations where observations are from two different samples that have the same variables, such as those found in regression-discontinuity designs to determine if a change has occurred between the two samples collected over time.

According to Chow (1960), to test the equality between sets of coefficients in two linear regressions, one starts with the assumption that both are equal. All regression coefficients were computed using the method of ordinary least squares available on any commercially available statistical analysis software, e.g. SPSS. A regression equation is fitted to the combined set of observations, i.e., excluded and selected, and the residual sum-of-squares ($\mathbf{e}'\mathbf{e}$) is computed. Next, a regression equation is fitted to the data without assuming the sets are equal. Likewise, the residual sum-of-squares ($\mathbf{e}'_1\mathbf{e}_1$) is obtained. Chow shows that the ratio of the difference between these two sums of squares ($\mathbf{e}'\mathbf{e} - \mathbf{e}'_1\mathbf{e}_1$) to the latter sums of squares ($\mathbf{e}'_1\mathbf{e}_1$), adjusted for the corresponding degrees of freedom, will be distributed as an F-ratio under the null hypothesis. Chow presents two variations of his method.

Depending on which situation, a different F-ratio is computed. One situation occurs when one sample has more observations than regression parameters or weights estimated ($n > p$) but the second sample does not have enough observations ($m < p$) to compute a regression equation. Here, the sum of squares is computed for the sample of n observations where $n > p$ (number of regression weights). Another regression equation is computed using the combination of first and second sample. The Chow Test can be computed using the following steps:

Step 1: For the first n observations, fit the least squares equation:

$$\mathbf{Y}_1 = \mathbf{X}_1\beta_1 + \mathbf{e}_1,$$

Step 2: Compute the residual sum of squares, $\mathbf{e}'_1\mathbf{e}_1$

Step 3: pool the $n + m$ sample observations to give \mathbf{Y} and \mathbf{X} and fit the least squares regression: $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$

Step 4: compute the residual sum of squares, $\mathbf{e}'\mathbf{e}$.

Step 5: The test of the null hypothesis that the m additional observations obey the same relation as the first is given by:

$$F = \frac{\left[\frac{\mathbf{e}'\mathbf{e} - \mathbf{e}'_1\mathbf{e}_1}{m} \right]}{\left[\frac{\mathbf{e}'_1\mathbf{e}_1}{(n-p)} \right]}$$

that is distributed as F with m and $n - p$ degrees of freedom.

The other situation is when both samples have enough observations to compute a regression equation. That is, the number of observations exceeds the number of regression parameters

estimated. For this situation, the following steps of the Chow Test would be

Step 1: To the first n observations, fit the least squares equation:

$$Y_1 = X_1\beta_1 + e_1$$

Step 2: Compute the residual sum of squares, $e_1'e_1$.

Step 3: To the second m observations, fit the least squares equation:

$$Y_2 = X_2\beta_2 + e_2$$

Step 4: Compute the residual sum of squares, $e_2'e_2$.

Step 5: Pool the $n + m$ sample observations to give Y and X and fit the least squares regression:

$$Y = X\beta + e$$

Step 6: Compute the residual sum of squares, $e'e$.

Step 7: The test of the null hypothesis that the m additional observations obey the same relation as the first is given by:

$$F = \frac{\left[e'e - e_1'e_1 - e_2'e_2 \right] / p}{\left[e_1'e_1 + e_2'e_2 \right] / (n + m - 2p)}$$

which is distributed as F with p and $n + m - 2p$ degrees of freedom.

Examples

Using the data from Braden & Bryant (1990), one regression equation would be for the selected group, another would be for the excluded group and a third equation is fitted to a combination of both groups. This can be done because the number of observations in both excluded and selected groups are greater than the number of regression weights to be estimated. Braden & Bryant (1990) used regression discontinuity to study children who were placed or not placed in gifted educational programs and the effect it had on achievement.

Without the actual data used by Braden & Bryant (1990), the author visually estimated each data point from the graphs presented in Braden & Bryant.¹ There were two data sets. The first Braden & Bryant data set consisted of 60 data points. The second data set consisted of 90 data points. The second data set consisted of the data points from the first data set plus 30 fictitious data points to create a significant regression discontinuity. However, in the visual estimation process, only 88 data points were distinguishable and were used in illustrating the method in this paper. The results obtained in using the Chow Test were essentially the same. The null hypothesis could not be rejected at the $\alpha = .05$ level for the first data set but was rejected for the second data set.

The analyses were done using SPSS for Windows with three separate executions of the regression subprogram. From the output, the residual sum of squares presented in each of the three ANOVA summary tables was used in the computations.

Braden & Bryant Example 1

$$n = 30; \quad m = 30 \quad p = 2$$

$$n + m = 60 \quad e_1'e_1 = 2575.563 \quad e_2'e_2 = 1408.964 \quad e'e = 4029.183$$

¹ All three data sets used in this article are available upon request from the author.

$$F = \frac{\left[\frac{\mathbf{e}'\mathbf{e} - \mathbf{e}'_1\mathbf{e}_1 - \mathbf{e}'_2\mathbf{e}_2}{p} \right]}{\left[\frac{\mathbf{e}'_1\mathbf{e}_1 + \mathbf{e}'_2\mathbf{e}_2}{(n+m-2p)} \right]} = \frac{(4029.183 - 2575.563 - 1408.964) / 2}{(2575.563 + 1408.964) / (60 - 4)}$$

$$= \frac{44.656 / 2}{3984.527 / 56} = \frac{22.328}{153.251} = 0.146$$

Critical F value ($\alpha = .05$, $df = 2, 56$) = 3.17.

Since $F = 0.146 < 3.17$, the hypothesis of equality is not rejected. There is insufficient evidence that being in a gifted program led to different (higher) achievement than those who were not in a gifted program. Figure 1 shows the regression lines for the two groups.

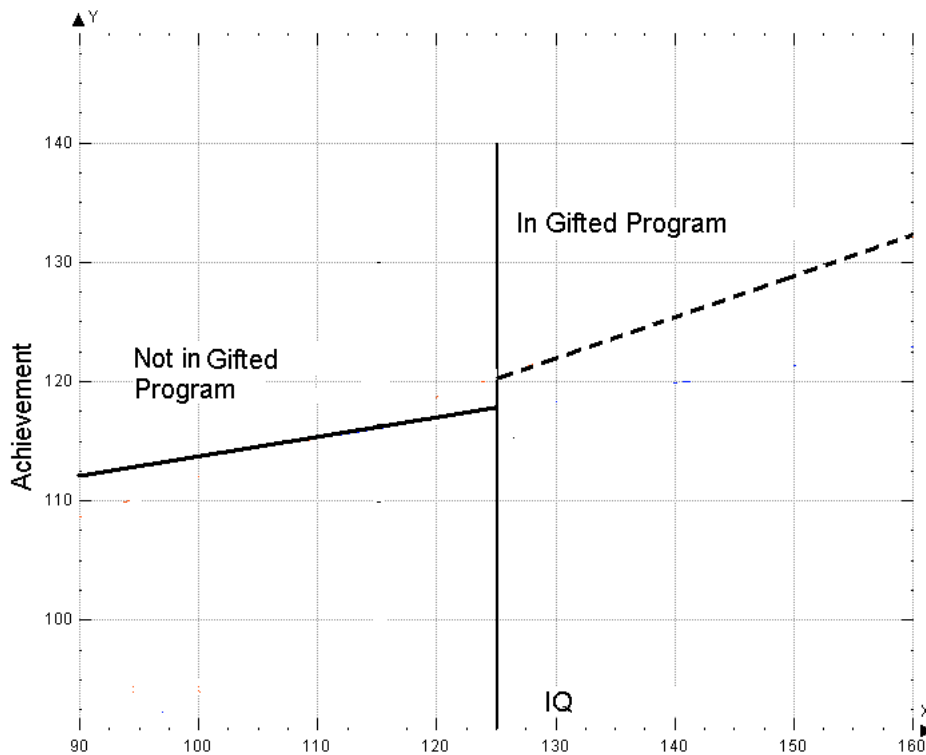


Figure 1. Regression lines for two groups using Braden & Bryant data, $N = 60$.

Braden & Bryant Example 2

$n = 58$; $m = 30$ $p = 2$
 $n + m = 88$ $\mathbf{e}'_1\mathbf{e}_1 = 4937.961$ $\mathbf{e}'_2\mathbf{e}_2 = 1408.964$ $\mathbf{e}'\mathbf{e} = 6851.371$

$$F = \frac{\left[\frac{\mathbf{e}'\mathbf{e} - \mathbf{e}'_1\mathbf{e}_1 - \mathbf{e}'_2\mathbf{e}_2}{p} \right]}{\left[\frac{\mathbf{e}'_1\mathbf{e}_1 + \mathbf{e}'_2\mathbf{e}_2}{(n+m-2p)} \right]} = \frac{(6851.371 - 4937.961 - 1408.964) / 2}{(4937.961 + 1408.964) / (88 - 4)}$$

$$= \frac{504.446 / 2}{6346.925 / 84} = \frac{252.223}{75.559} = 3.338$$

Critical F value ($\alpha = .05$, $df = 2, 84$) = 3.13.

Since $F = 3.338 > 3.13$, the hypothesis of equality is rejected. There is evidence that being in a gifted program led to different (higher) achievement than those who were not in a gifted program.

Figure 2 shows a plot of the two regression lines.

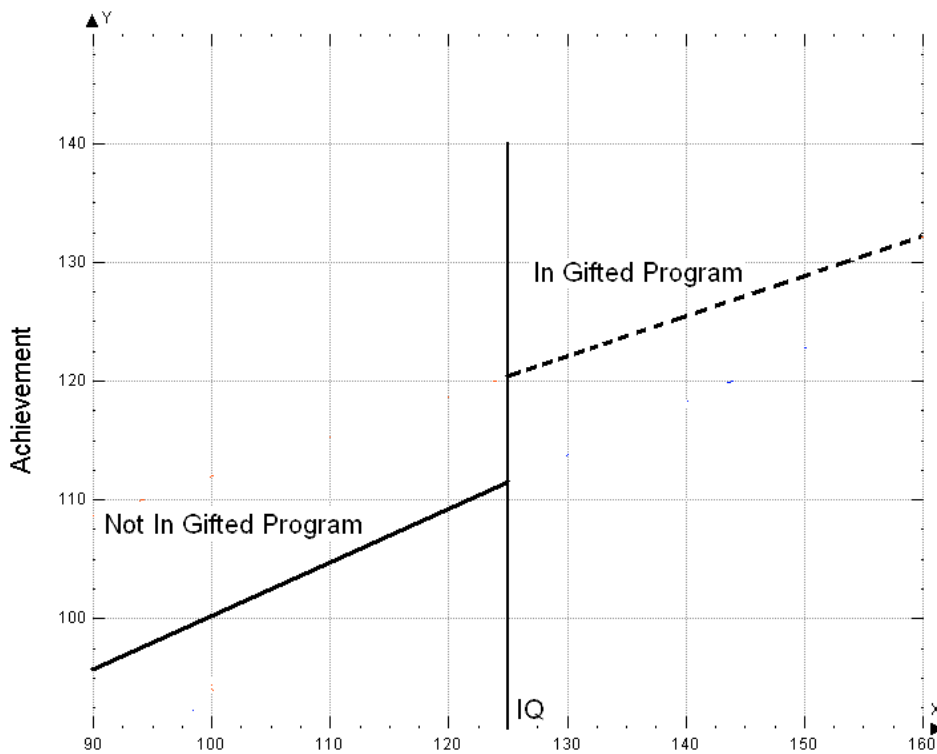


Figure 2. Regression lines for two groups using Braden & Bryant data, $N = 88$.

Another Example.

Seaver & Quarton (1976) used regression discontinuity to analyze the effects of a student being on the Dean's honor list. Two groups were identified and tracked over three academic terms. The first group consisted of students who achieve a grade-point average that allowed them to be on the Dean's list. The other group consisted of students who did not achieve grades to qualify them to be on the Dean's list. Seaver & Quarton (1976) found that membership in the Dean's list had a positive effect on grade point average in subsequent terms. They concluded that early membership on the Dean's list helps maintain the quality of academic work.

For purposes of this paper, Seaver & Quarton's (1976) study was partially replicated.

Similar data were obtained on 1845 psychology undergraduate students at this author's university (California State University, Northridge) for the academic year of 2005-2006. The number of students that made the Dean's list was 632. The number of students excluded was 1213. To qualify for the Dean's List, a student had to carry 12 semester units or more and earns a grade point average of 3.40 or higher.

Using SPSS regression analysis subprogram and formulas for the Chow Test, the results were similar to those found by Seaver & Quarton (1976). The numbers are given below.

$$n = 632; \quad m = 1213 \quad p = 2$$

$$n + m = 1845 \quad \mathbf{e}'_1\mathbf{e}_1 = 715.615 \quad \mathbf{e}'_2\mathbf{e}_2 = 173.817 \quad \mathbf{e}'\mathbf{e} = 901.796$$

$$F = \frac{\left[\frac{\mathbf{e}'\mathbf{e} - \mathbf{e}'_1\mathbf{e}_1 - \mathbf{e}'_2\mathbf{e}_2}{p} \right]}{\left[\frac{\mathbf{e}'_1\mathbf{e}_1 + \mathbf{e}'_2\mathbf{e}_2}{(n+m-2p)} \right]} = \frac{(901.796 - 715.615 - 173.817) / 2}{(715.615 + 173.817) / (1845 - 4)}$$

$$= \frac{12.364 / 2}{892.770 / 1841} = \frac{6.182}{.485} = 12.748$$

Critical F value ($\alpha = .01, df = 2, 1841$) = 4.60.

Since $F = 12.748 > 4.60$, the hypothesis of equality is rejected. This tells us that being on the Dean's List yield different (positive) results than not being on the Dean's list. Figure 3 shows the two regression lines.

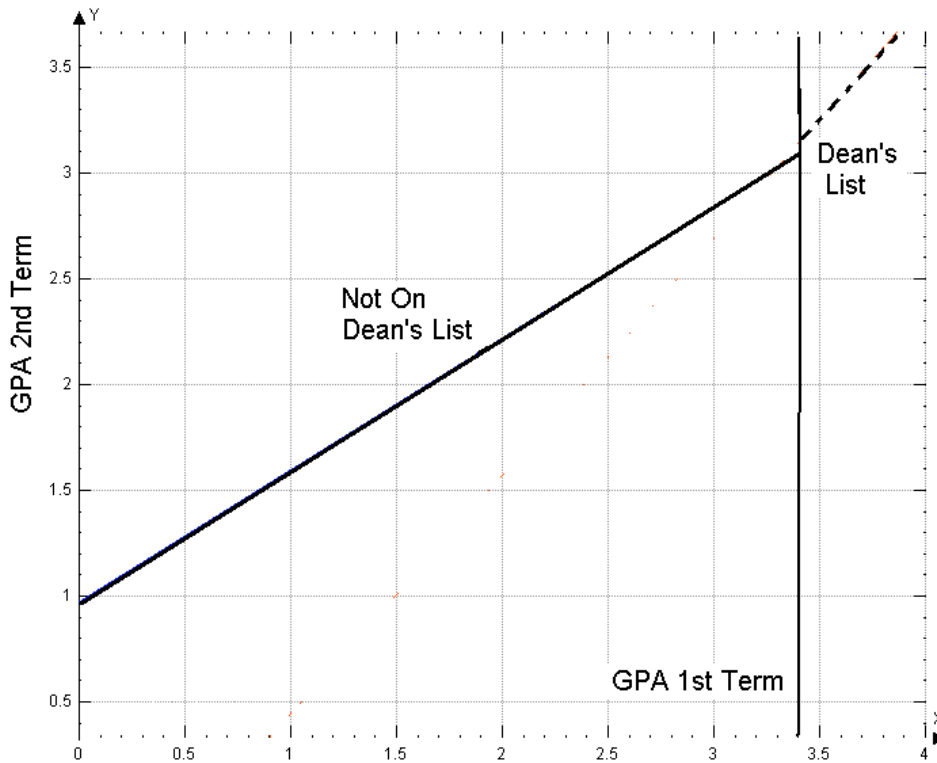


Figure 3. Regression lines for two groups using Dean's List Data $N = 1845$.

Comparison to Analysis of Covariance

The three data sets were also analyzed using the analysis of covariance (ANCOVA). This was done to compare the results obtained from using the Chow Test. The analysis of covariance agreed with the conclusion drawn from using the Chow Test. In the first example of the Braden & Bryant data the ANCOVA did not find a statistically significant difference between those in the gifted program and those not in the gifted program. In the second example using the Braden & Bryant data, there was a statistically significant difference. Tables 1 and 2 shows the ANCOVA summary tables for both examples. As stated previously, the Braden & Bryant data presented in this paper are not the actual data used in their article.

In the third example, using grade-point average collected from archival data at California State University, Northridge, the Chow Test also agreed with the conclusion drawn from the analysis of covariance. Both tests found a difference between those on the Dean's list and those who were not on the Dean's list in terms of later academic performance. Table 3 presents the ANCOVA summary table.

Table 1. ANCOVA SUMMARY TABLE - Braden & Bryant Example 1

Dependent Variable: achieve

Source	Sum of Squares	df	Mean Square	F	Sig.
IQ-Covariate	240.442	1	240.442	3.419	.070
Group	20.091	1	20.091	.286	.595
Error	4009.091	57	70.335		
Total	5115.933	59			

Table 2. ANCOVA SUMMARY TABLE- Braden & Bryant Example 2

Dependent Variable: achieve

Source	Sum of Squares	df	Mean Square	F	Sig.
IQ-covariate	738.484	1	738.484	9.867	.002
Group	489.506	1	489.506	6.540	.012
Error	6361.865	85	74.845		
Total	13806.318	87			

Table 3. ANCOVA SUMMARY TABLE - Seaver & Quarton Example

Dependent Variable: TeGpa2

Source	Sum of Squares	df	Mean Square	F	Sig.
Gpa1-Covariate	261.812	1	261.812	538.808	.000
Group	6.751	1	6.751	13.893	.000
Error	895.045	1842	.486		
Total	1534.997	1844			

Discussion

This paper introduced the Chow test and its possibility in analyzing data from a regression discontinuity study. The formulas used in the Chow Test are not difficult to compute. Obtaining the values for the Chow Test is straightforward and can be obtained by using the regression subprogram of canned statistical packages such as SPSS. The Chow Test gives the researcher using this type of research design an alternative method of analysis from the traditional simple regression or analysis of covariance approaches.

The Chow Test was compared to the analysis of covariance (ANCOVA) and the results from both analysis agreed in terms of the conclusion that can be drawn about the data. However, using the ANCOVA through a canned program such as SPSS provides a more precise measurement of the probability of a Type 1 error. The ANCOVA also involves only one computer run while the Chow Test requires three. However, for those who are unfamiliar with ANCOVA or have issues with the method (see Elashoff, 1969), would benefit from using the Chow Test for studies that use the regression discontinuity design. Potentially, the Chow Test could analyze relationships that are complex and difficult to accomplish in ANCOVA.

All possibilities of the Chow Test were not explored. The regression examples used here only involved linear regressions with one criterion variable and one predictor variable. Some possible variations that need exploration would involve the use of multiple predictor variables such those found in multiple regression. Other possibilities would involve nonlinear regressions such as quadratics and exponential relationships. Assumption issues such as homoscedasticity were addressed in Toyoda. (1974) and Schmidt & Sickles (1977). Future research would need to be done to determine the effects of sample size, statistical power and error rates.

The Chow Test was shown to be useful in the situations demonstrated in this paper. With knowledge of the Chow Test, perhaps researchers in the social sciences will be more willing to develop a greater number of suitable studies using the regression discontinuity design.

References

- Braden, J. P. & Bryant, T. J. (1990). Regression discontinuity designs: Applications for school psychology. *School Psychology Review*, 19(2), 232-239.
- Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regression. *Econometrica*, 26(3), 591-605.
- Christensen, L. B. (2006). *Experimental methodology*, 10th Ed. Allyn & Bacon.

- Cook, T.D. & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago: Rand McNally.
- Elashoff, J. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal*, 6, 383-401.
- Fisher, F. M. (1970). Tests of equality between two sets of coefficients in two linear regression: An expository note. *Econometrica*. 38(2), 361-366.
- Gormley, W.T. Jr., Gayer, T. & Phillips, D. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology*, 41(6), 872-884.
- Graziano, A. M. & Raulin, M. L. (1993). *Research methods: A process of inquiry*. 2nd Ed. New York: Harper Collins Publishers.
- Johnston, J. (1972). *Econometric methods*. 2nd, Ed. New York: McGraw-Hill.
- Johnston, J. & Dinardo, J. (1996). *Econometric methods*. New York: McGraw-Hill.
- Kerlinger, F. N. & Lee, H.B. (2000). *Foundations of behavioral research*, 4th Ed. Fort Worth, TX: Harcourt.
- McGuigan, F. J. (1997). *Experimental psychology: Methods of research*. 7th Ed. Upper Saddle River, NJ: Prentice Hall.
- Schmidt, P. & Sickles, R. (1977). Some further evidence on the use of the Chow Test under heteroskedasticity. *Econometrica*. 45(5), 1293-1298.
- Seaver, W. B. & Quarton, R. J. (1976). Regression discontinuity analysis of dean's list effects. *Journal of Educational Psychology*, 68(4), 459-465.
- Shadish, W. P., Cook, T.D. & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin Company.
- Toyoda, T. (1974). Use of the Chow Test under heteroskedasticity. *Econometrica*. 42, 601-608.
- Trochim, W. M. K. (1984). *Research design for program evaluation: The regression-discontinuity approach*. Newbury Park, CA: Sage.
- Zuckerman, I. H.; Lee, E. & Wutoh, A. K. (2006). Application of regression-discontinuity analysis in pharmaceutical health services research. *Health Services Research*, 41(2), 550-563.