Journal of Modern Applied Statistical Methods

Volume 8 | Issue 2 Article 5

11-1-2009

Level Robust Methods Based on the Least Squares Regression Estimator

Marie Ng
University of Hong Kong, marieng@uw.edu

Rand R. Wilcox
University of Southern California, rwilcox@usc.edu

Journal of Modern Applied Statistical Methods
November 2009, Vol. 8, No. 2, 384-395

1538 – 9472/09/$95.00

Level Robust Methods Based on the Least Squares Regression Estimator

Marie Ng
University of Hong Kong

Rand R. Wilcox
University of Southern California

Heteroscedastic consistent covariance matrix (HCCM) estimators provide ways for testing hypotheses about regression coefficients under heteroscedasticity. Recent studies have found that methods combining the HCCM-based test statistic with the wild bootstrap consistently perform better than non-bootstrap HCCM-based methods (Davidson & Flachaire, 2008; Flachaire, 2005; Godfrey, 2006). This finding is more closely examined by considering a broader range of situations which were not included in any of the previous studies. In addition, the latest version of HCCM, HC5 (Cribari-Neto, et al., 2007), is evaluated.

Key words: Heteroscedasticity, level robust methods, bootstrap.

Introduction

Consider the standard simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i, i=1, ..., n, \quad (1)$$

where β_0 and β_1 are unknown parameters and ε_i is the error term. When testing the hypothesis,

$$H_0 : \beta_1 = 0 \quad (2)$$

the following assumptions are typically made:

1. $E(\varepsilon_i) = 0$.
2. $Var(\varepsilon_i) = \sigma^2$ (Homoscedasticity).
3. ε_i's are independent of X.
4. ε_i's are independent and identically distributed (i.i.d).

This article is concerned with testing (2) when assumption 2 is violated.

Let $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ be the least squares estimate of $\beta = (\beta_0, \beta_1)$. When there is homoscedasticity (i.e., assumption 2 holds), an estimate of the squared standard error of $\hat{\beta}$ is $Var(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1}$, where $\hat{\sigma}^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta})/(n - 2)$ is the usual estimate of the assumed common variance; X is the design matrix containing an $n \times 1$ unit vector in the first column and X_{i1}'s in the second column.

Marie Ng is an Assistant Professor in the Faculty of Education. Email: marieng@uw.edu. Rand R. Wilcox is a Professor of Psychology. Email: rwilcox@usc.edu.

However, when heteroscedasticity occurs, the squared standard error based on such an estimator is no longer accurate (White, 1980). The result is that the usual test of (2) is not asymptotically correct. Specifically, using the classic *t*-test when assumptions are violated can result in poor control over the probability of a Type I error. One possible remedy is to test the assumption that the error term is homoscedastic before proceeding with the *t*-test. However, it is unclear when the power of such a test is adequate to detect heteroscedasticity.

One alternative is to use a robust method that performs reasonably well under homoscedasticity and at the same time is robust to heteroscedasticity and non-normality. Many methods have been proposed for dealing with heteroscedasticity. For example, a variance stabilizing transformation may be applied to the dependent variable (Weisberg, 1980) or a weighted regression with each observation weighted by the inverse of the standard deviation of the error term may be performed (Greene, 2003).

Although these methods provide efficient and unbiased estimates of the coefficients and standard error, they assume that heteroscedasticity has a known form. When heteroscedasticity is of an unknown form, the best approach to date, when testing (2), is to use a test statistic (e.g., quasi-t test) based on a heteroscedastic consistent covariance matrix (HCCM) estimator. Several versions of HCCM have been developed that provide a consistent and an unbiased estimate of the variance of coefficients even under heteroscedasticity (White, 1980; MacKinnon & White, 1985).

Among all the HCCM estimators, HC4 was found to perform fairly well with small samples (Long & Ervin, 2000; Cribari-Neto, 2004). Recently, Cribari-Neto, et al. (2007) introduced a new version of HCCM (HC5) arguing that HC5 is better than HC4, particularly at handling high leverage points in X. However, in their simulations, they only focused on models with $\varepsilon \sim N(0, 1)$. Moreover, only a limited number of distributions of X and patterns of heteroscedasticity were considered. In this study, the performances of HC5-based and HC4-based quasi-t statistics were compared by looking at a broader range of situations.

Methods combining an HCCM-based test statistic with the wild bootstrap method perform considerably better than non-bootstrap asymptotic approaches (Davidson & Flachaire, 2008; Flachaire, 2005). In a recent study, Godfrey (2006) compared several non-bootstrap and wild bootstrap HCCM-based methods for testing multiple coefficients ($H_0\colon \beta_1 = \ldots = \beta_p = 0$). It was found that when testing at the $\alpha = 0.05$ level, the wild bootstrap methods generally provided better control over the probability of a Type I error than the non-bootstrap asymptotic methods. However, in the studies mentioned, the wild bootstrap and non-bootstrap methods were evaluated in a limited set of simulation scenarios.

In Godfrey's study, data were drawn from a data set in Greene (2003) and only two heteroscedastic conditions were considered. Here, more extensive simulations were performed to investigate the performance of the various bootstrap and non-bootstrap HCCM-based methods. More patterns of heteroscedasticity were considered, as well as more types of distributions for both X and ε. Small sample performance of one non-bootstrap and two wild bootstrap versions of HC5-based and HC4-based quasi-t methods were evaluated.

Finally, two variations of the wild bootstrap method were compared when generating bootstrap samples. One approach makes use of the lattice distribution. Another approach makes use of a standardized continuous uniform distribution: Uniform(-1, 1). The former approach has been widely considered (Liu, 1988; Davidson & Flachaire, 2000; Godfrey, 2006) and was found to work well in various multiple regression situations. Of interest is how these two approaches compare when testing (2) in simple regression models. Situations were identified where wild bootstrap methods were unsatisfactory.

Methodology

HC5-Based Quasi-T Test (HC5-T)

The HC5 quasi-t statistic is based on the standard error estimator HC5, which is given by

$$\ddot{V} = (X'X)^{-1} X'\text{diag}\left[\frac{r_i^2}{\sqrt{(1-h_{ii})^{\alpha_i}}}\right] X(X'X)^{-1}, \tag{3}$$

where r_i, i = 1, ..., n are the usual residuals, X is the design matrix,

$$\begin{aligned}\alpha_i &= \min\left\{\frac{h_{ii}}{\bar{h}}, \max\left\{4, \frac{kh_{max}}{\bar{h}}\right\}\right\} \\ &= \min\left\{\frac{nh_{ii}}{\sum_{i=1}^{n} h_{ii}}, \max\left\{4, \frac{nkh_{max}}{\sum_{i=1}^{n} h_{ii}}\right\}\right\}\end{aligned} \tag{4}$$

and

$$h_{ii} = x_i \left(X'X\right)^{-1} x_i' \tag{5}$$

where x_i is the i^{th} row of X, $h_{max} = \max\{h_{11}, \ldots, h_{nn}\}$ and k is set at 0.7 as suggested by Cribari-Neto, et al. (2007, 2008). The motivation behind HC5 is that when high leverage observations are present in X, the standard error of the coefficients are often underestimated. HC5 attempts to correct such a bias by taking into account the maximal leverage.

For testing (2), the quasi-t test statistic is,

$$T = \hat{\beta}_1 - 0 / \sqrt{\ddot{V}_{22}} \tag{6}$$

where $\ddot{V}_{22}$ is the 2^{nd} entry along the diagonal of $\ddot{V}$. Reject (2) if $|T| \geq t_{1-\alpha/2}$ where $t_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the Student's t-distribution with $n-2$ degrees of freedom.

HC4-Based Quasi-T Test (HC4-T)

The HC4 quasi-t statistics is similar to HC5-T, except the standard error is estimated using HC4, which is given by

$$\tilde{V} = \left(X'X\right)^{-1} X'\mathrm{diag}\left[\frac{r_i^2}{\left(1-h_{ii}\right)^{\delta_i}}\right] X\left(X'X\right)^{-1} \quad (7)$$

where

$$\delta_i = \min\{4, \frac{h_{ii}}{\bar{h}}\} = \min\{4, \frac{nh_{ii}}{\sum_{i=1}^{n} h_{ii}}\} \quad (8)$$

HC5-Based Wild Bootstrap Quasi-T Test (HC5WB-D and HC5WB-C)

The test statistic for testing (2) is computed using the following steps:

1. Compute the HC5 quasi-t test statistics (T) given by (6).
2. Construct a bootstrap sample $Y_i^* = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + a_i r_i$, i = 1, ..., n, where a_i is typically generated in one of two ways. The first generates a_i from a two-point (lattice) distribution:

   $$a_i = \begin{cases} -1, & \text{with probability } 0.5 \\ 1, & \text{with probability } 0.5 \end{cases}$$

   The other method uses

   $$a_i = \sqrt{12}(U_i - 0.5)$$

   where U_i is generated from a uniform distribution on the unit interval. We denote the method based on the first approach HC5WB-D and the method based on the latter approach HC5WB-C.

3. Compute the quasi-t test statistics (T^*) based on this bootstrap sample, yielding

   $$T^* = \frac{\hat{\beta}_1^* - 0}{\sqrt{\ddot{V}_{22}^*}} \quad (9)$$

4. Repeat Steps 2 - 3 *B* times yielding T_b^*, b = 1, ..., *B*. In the current study, *B* = 599.
5. Finally, a *p*-value is computed:

   $$p = \frac{\#\{|T_b^*| \geq |T|\}}{B} \quad (10)$$

6. Reject H_0 if $p \leq \alpha$.

HC4-Based Wild Bootstrap Quasi-T Test (HC4WB-D and HC4WB-C)

The procedure for testing (2) is the same as that of HC5WB-D and HC5W-C except that HC4 is used to estimate the standard error.

Simulation Design

Data are generated from the model:

$$Y_i = X_i\beta_1 + \tau(X_i)\varepsilon_i \quad (11)$$

where τ is a function of X_i used to model heteroscedasticity. Data are generated from a *g*-and-*h* distribution. Let Z be a random variable generated from a standard normal distribution,

$$X = \left(\frac{\exp(gZ) - 1}{g}\right)\exp(hZ^2/2) \quad (12)$$

has a *g*-and-*h* distribution. When *g* = 0, this last equation is taken to be $X = Z\exp(hZ^2/2)$. When *g* = 0 and *h* = 0, X has a standard normal distribution. Skewness and heavy-tailedness of the *g*-and-*h* distributions are determined by the values of *g* and *h*, respectively. As the value of *g* increases, the distribution becomes more skewed. As the value of *h* increases, the distribution becomes more heavy-tailed. Four types of distributions are considered for X: standard normal (*g* = 0, *h* = 0), asymmetric light-tailed (*g* = 0.5, *h* = 0), symmetric heavy-tailed (*g* = 0, *h* = 0.5) and asymmetric heavy-tailed (*g* = 0.5, *h* = 0.5). The error term (ε_i) is also randomly generated based on one of these four *g*-and-*h* distributions. When *g*-and-*h* distributions are asymmetric (*g* = 0.5), the mean is not zero. Therefore, ε_i's generated from these distributions are re-centered to have a mean of zero.

Five choices for $\tau(X_i)$ are considered: $\tau(X_i) = 1$, $\tau(X_i) = \sqrt{|X_i|}$, $\tau(X_i) = 1 + \frac{2}{|X_i| + 1}$, and $\tau(X_i) = |X_i + 1|$. These functions are denoted as variance patterns (VP), VP1, VP2, VP3, VP4 and VP5, respectively. Homoscedasticity is represented by $\tau(X_i) = 1$.

Moreover, $\tau(X_i) = \sqrt{|X_i|}$, $\tau(X_i) = 1 + \frac{2}{|X_i| + 1}$, and $\tau(X_i) = |X_i + 1|$ represent a particular pattern of variability in Y_i based upon the value of X_i. All possible pairs of X_i and ε_i distributions are considered, resulting in a total of 16 sets of distributions. All five variance patterns are used for each set of distributions. Hence, a total of 80 simulated conditions are considered. The estimated probability of a Type I error is based on 1,000 replications with a sample size of n = 20 and when testing at α = 0.05 and α = 0.01. According to Robey and Barcikowski (1992), 1,000 replications are sufficient from a power point of view. If the hypothesis that the actual Type I error rate is 0.05 is tested, and power should be 0.9 when testing at the 0.05 level and the true α value differs from 0.05 by 0.025, then 976 replications are required. The actual Type I error probability is estimated with $\hat{\alpha}$, the proportion of *p*-values less than or equal to 0.05 and 0.01.

Results

First, when testing at both α = 0.05 and α = 0.01, the performances of HC5-T and HC4-T are extremely similar in terms of control over the probability of a Type I error (See Tables 1 and 2). When testing at α = 0.05, the average Type I error rate was 0.038 (SD = 0.022) for HC5-T and 0.040 (SD = 0.022) for HC4-T. When testing at α = 0.01, the average Type I error rate was 0.015 (SD = 0.013) for HC5-T and 0.016 (SD = 0.013) for HC4-T.

Theoretically, when leverage points are likely to occur (i.e. when X is generated from a distribution with h = 0.5), HC5-T should perform better than HC4-T; however, as shown in Table 1, this is not the case. On the other hand, when leverage points are relatively unlikely (i.e., when X is generated from a distribution with h = 0), HC5-T and HC4-T should yield the same outcomes. As indicated by the results of this study, when X is normally distributed (g = 0 and h = 0), the actual Type I error rates resulting from the two methods are identical. However, when X has a skewed light-tailed distribution (g = 0.5 and h = 0), HC5-T and HC4-T do not always yield the same results. Focus was placed on a few situations where HC4-T is unsatisfactory, and we considered the extent it improves as the sample size increases. We considered sample sizes of 30, 50 and 100. As shown in Table 3, control over the probability of a Type I error does not improve markedly with increased sample sizes.

Second, with respect to the non-bootstrap and bootstrap methods, results suggest that the bootstrap methods are not necessarily superior to the non-bootstrap ones. As shown in Figures 1 and 4, when testing at α = 0.05, under VP 1 and 4, the bootstrap methods outperform the non-bootstrap methods. Specifically, the non-bootstrap methods tended to be too conservative under those conditions. Nonetheless, under VP 3 and 5 (see Figures 3 and 5), the non-bootstrap methods, in general, performed better than the bootstrap methods. In particular, the actual Type I error rates yielded by the bootstrap methods in those situations tended to be noticeably higher than the nominal level. In one situation, the actual Type I error rate was as high as 0.196. When testing at α = 0.01, HC5WB-C and HC4WB-C offered the best performance in general; however, situations were found where non-bootstrap methods outperform bootstrap methods.

Finally, regarding the use of the continuous uniform distribution versus the lattice distribution for generating bootstrap samples, results suggest that the former has slight practical advantages. When testing at α = 0.05, the average Type I error rates yielded by the two approaches are 0.059 for HC5WB-C and HC4WB-C and 0.060 for HC5WB-D and HC4WBD. When testing at α = 0.01, the average Type I error rates are 0.015 for HC5WB-C and HC4WB-C and 0.021 for HC5WB-D and HC4WB-D. Overall, the actual Type I error rates yielded by HC5WB-C and HC4WB-C appear to deviate from the nominal level in fewer cases.

Conclusion

This study expanded on extant simulations by considering ranges of non-normality and heteroscedasticity that had not been considered previously. The performance of the latest HCCM estimator (HC5) was also closely considered. The non-bootstrap HC5-based and HC4-based quasi-t methods (HC5-T and HC4-T) were compared, as well as their wild bootstrap counterparts (HC5WB-D, HC5WB-C, HC4WB-D and HC4WB-C). Furthermore, two wild bootstrap sampling schemes were evaluated - one based on the lattice distribution; the other based on the continuous standardized uniform distribution.

As opposed to the findings of Cribari-Neto, et al. (2007), results here suggest that HC5 does not offer striking advantages over HC4.

Both HC5-T and HC4-T perform similarly across all the situations considered. In many cases, HC5-T appears more conservative than HC4-T. One concern is that, for the situations at hand, setting k = 0.7 when calculating HC5 may not be ideal; thus, whether changing the value of k might improve the performance of HC5-T was examined. As suggested by Cribari-Neto, et al., values of k between 0.6 and 0.8 generally yielded desirable results, for this reason k = 0.6 and k = 0.8 were considered. However, as indicated in Tables 4 and 5, regardless of the value of k, no noticeable difference was identified between the methods.

Moreover, contrary to both Davidson and Flachaire (2008) and Godfrey's (2006) findings, when testing the hypothesis $H_0: \beta_1 = 0$ in a simple regression model, the wild bootstrap methods (HC5WB-D, HC5WB-C, HC4WB-D and HC4WB-C) do not always outperform the non-bootstrap methods (HC5-T and HC4-T). By considering a wider range of situations, specific circumstances where the non-bootstrap methods outperform the wild bootstrap methods are able to be identified and vice versa. In particular, the non-bootstrap and wild bootstrap approaches are each sensitive to different patterns of heteroscedasticity.

For example, the wild bootstrap methods generally performed better than the non-bootstrap methods under VP 1 and 4 whereas the non-bootstrap methods generally performed better than the wild bootstrap methods under VP 3 and 5. Situations also exist (1988), Davidson and Flachaire (2008) and Godfrey (2006). The actual Type I error rates resulting from the methods HC5WB-C and HC4WB-C were generally less variable compared to those resulting from HC5WB-D and HC4WB-D. In many cases, the performances between the two approaches are similar, but in certain situations such as in VP3, HC5WB-C and HC4WB-C notably outperformed HC5WB-D and HC4WB-D.

References

Cribari-Neto, F. (2004). Asymptotic inference under heteroscedasticity of unknown form. *Computational Statistics & Data Analysis*, *45*, 215-233.

Cribari-Neto, F., Souza, T. C., & Vasconcellos, A. L. P. (2007). Inference under heteroskedasticity and leveraged data. *Communication in Statistics -Theory and Methods*, *36*, 1877-1888.

Cribari-Neto, F., Souza, T. C., & Vasconcellos, A. L. P. (2008). Errata: Inference under heteroskedasticity and leveraged data. *Communication in Statistics -Theory and Methods*, *36*, 1877-1888. *Communication in Statistics -Theory and Methods*, *37*, 3329-3330.

Davidson, R., & Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics, 146,* 162-169.

Flachaire, E. (2005). Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap. *Computational Statistics & Data Analysis*, *49*, 361-376.

Greene, W. H. (2003). *Econometric analysis* (*5th Ed.*). New Jersey: Prentice Hall.

Godfrey, L. G. (2006). Tests for regression models with heteroscedasticity of unknown form. *Computational Statistics & Data Analysis*, *50*, 2715-2733.

Liu, R. Y. (1988). Bootstrap procedures under some non i.i.d. models. *The Annuals of Statistics*, *16*, 1696-1708.

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, *54*, 217-224.

MacKinnon, J. G., & White, H. (1985). Some heteroscedasticity consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, *29*, 305-325.

Robey, R. R. & Barcikowski, R. S. (1992). Type I error and the number of iterations in Monte Carlo studies of robustness. *British Journal of Mathematical and Statistical Psychology*, *45*, 283–288.

Weisberg, S. (1980). *Applied Linear Regression*. NY: Wiley.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test of heteroscedasticity. *Econometrica*, *48*, 817-838.

Table 1: Actual Type I Error Rates when Testing at α = 0.05

X g	X h	e g	e h	VC	HC5WB-C	HC5WB-D	HC5-T	HC4WB-C	HC4WB-D	HC4-T
0	0	0	0	1	0.051	0.036	0.030	0.050	0.035	0.030
				2	0.077	0.059	0.068	0.075	0.061	0.068
				3	0.052	0.036	0.048	0.053	0.038	0.048
				4	0.058	0.040	0.038	0.055	0.042	0.038
				5	0.085	0.064	0.072	0.086	0.065	0.072
0	0	0	0.5	1	0.048	0.044	0.022	0.052	0.046	0.022
				2	0.053	0.048	0.032	0.055	0.050	0.032
				3	0.055	0.054	0.036	0.054	0.051	0.036
				4	0.058	0.046	0.022	0.053	0.046	0.022
				5	0.053	0.054	0.036	0.055	0.051	0.036
0	0	0.5	0	1	0.059	0.047	0.041	0.055	0.045	0.041
				2	0.066	0.063	0.046	0.063	0.059	0.046
				3	0.058	0.050	0.054	0.058	0.047	0.054
				4	0.065	0.054	0.038	0.066	0.056	0.038
				5	0.093	0.074	0.072	0.091	0.070	0.072
0	0	0.5	0.5	1	0.036	0.028	0.017	0.036	0.030	0.017
				2	0.044	0.037	0.024	0.045	0.040	0.024
				3	0.057	0.053	0.036	0.054	0.056	0.036
				4	0.048	0.045	0.018	0.051	0.046	0.018
				5	0.157	0.152	0.118	0.165	0.154	0.118
0	0.5	0	0	1	0.053	0.049	0.028	0.060	0.050	0.033
				2	0.059	0.050	0.043	0.063	0.056	0.049
				3	0.051	0.055	0.039	0.056	0.053	0.043
				4	0.048	0.035	0.018	0.042	0.030	0.020
				5	0.060	0.051	0.045	0.059	0.050	0.052
0	0.5	0	0.5	1	0.044	0.042	0.008	0.045	0.041	0.009
				2	0.055	0.063	0.024	0.050	0.063	0.028
				3	0.043	0.054	0.023	0.042	0.048	0.027
				4	0.036	0.038	0.007	0.032	0.035	0.008
				5	0.043	0.068	0.029	0.044	0.067	0.031
0	0.5	0.5	0	1	0.058	0.044	0.031	0.051	0.048	0.033
				2	0.070	0.054	0.053	0.067	0.058	0.056
				3	0.054	0.052	0.047	0.056	0.051	0.050
				4	0.050	0.041	0.023	0.050	0.039	0.024
				5	0.055	0.052	0.041	0.056	0.055	0.045
0	0.5	0.5	0.5	1	0.039	0.043	0.013	0.042	0.037	0.013
				2	0.048	0.055	0.026	0.043	0.053	0.030
				3	0.049	0.062	0.030	0.045	0.063	0.037
				4	0.023	0.042	0.006	0.024	0.046	0.006
				5	0.071	0.090	0.045	0.078	0.086	0.054
0.5	0	0	0	1	0.067	0.061	0.049	0.068	0.055	0.050
				2	0.070	0.057	0.055	0.068	0.057	0.060
				3	0.061	0.064	0.057	0.064	0.064	0.058
				4	0.061	0.048	0.038	0.066	0.047	0.038
				5	0.075	0.095	0.066	0.083	0.088	0.069

Table 1: Actual Type I Error Rates when Testing at α = 0.05 (continued)

X g	X h	e g	e h	VC	HC5WB-C	HC5WB-D	HC5-T	HC4WB-C	HC4WB-D	HC4-T
0.5	0	0	0.5	1	0.052	0.047	0.023	0.056	0.045	0.023
				2	0.056	0.059	0.029	0.053	0.055	0.031
				3	0.057	0.071	0.034	0.056	0.071	0.035
				4	0.041	0.036	0.020	0.043	0.041	0.020
				5	0.065	0.089	0.037	0.067	0.092	0.038
0.5	0	0.5	0	1	0.053	0.048	0.040	0.058	0.049	0.041
				2	0.073	0.055	0.072	0.081	0.064	0.073
				3	0.078	0.073	0.062	0.072	0.074	0.064
				4	0.046	0.038	0.027	0.040	0.040	0.027
				5	0.107	0.113	0.087	0.108	0.111	0.087
0.5	0	0.5	0.5	1	0.044	0.044	0.019	0.046	0.047	0.019
				2	0.065	0.062	0.050	0.068	0.065	0.051
				3	0.059	0.081	0.055	0.070	0.083	0.055
				4	0.048	0.046	0.019	0.046	0.048	0.019
				5	0.168	0.190	0.120	0.168	0.196	0.124
0.5	0.5	0	0	1	0.080	0.056	0.034	0.076	0.056	0.041
				2	0.062	0.065	0.040	0.064	0.067	0.047
				3	0.064	0.080	0.047	0.063	0.072	0.051
				4	0.050	0.042	0.017	0.047	0.038	0.019
				5	0.069	0.089	0.044	0.073	0.092	0.057
0.5	0.5	0	0.5	1	0.035	0.048	0.013	0.035	0.044	0.013
				2	0.038	0.057	0.017	0.036	0.059	0.018
				3	0.042	0.077	0.028	0.041	0.079	0.034
				4	0.036	0.036	0.007	0.028	0.033	0.008
				5	0.082	0.122	0.053	0.080	0.118	0.058
0.5	0.5	0.5	0	1	0.058	0.041	0.026	0.058	0.040	0.029
				2	0.061	0.057	0.043	0.061	0.055	0.054
				3	0.048	0.062	0.036	0.050	0.066	0.043
				4	0.045	0.038	0.016	0.049	0.035	0.016
				5	0.059	0.083	0.035	0.057	0.078	0.041
0.5	0.5	0.5	0.5	1	0.036	0.039	0.010	0.038	0.041	0.012
				2	0.057	0.057	0.021	0.055	0.059	0.031
				3	0.062	0.094	0.046	0.063	0.094	0.050
				4	0.030	0.041	0.007	0.036	0.036	0.008
				5	0.084	0.116	0.058	0.086	0.117	0.065
				Max	0.168	0.190	0.120	0.168	0.196	0.124
				Min	0.023	0.028	0.006	0.024	0.030	0.006
				Average	0.059	0.060	0.038	0.059	0.060	0.040
				SD	0.022	0.026	0.022	0.023	0.027	0.022

Table 2: Actual Type I Error Rates when Testing at α = 0.01

X g	X h	e g	e h	VC	HC5WB-C	HC5WB-D	HC5-T	HC4WB-C	HC4WB-D	HC4-T
0	0	0	0	1	0.016	0.005	0.013	0.017	0.011	0.013
				2	0.016	0.009	0.016	0.014	0.009	0.016
				3	0.018	0.010	0.017	0.017	0.009	0.017
				4	0.016	0.010	0.008	0.017	0.011	0.008
				5	0.024	0.018	0.024	0.025	0.021	0.024
0	0	0	0.5	1	0.013	0.010	0.008	0.012	0.007	0.008
				2	0.013	0.018	0.010	0.013	0.015	0.010
				3	0.014	0.025	0.012	0.011	0.019	0.012
				4	0.006	0.004	0.001	0.008	0.005	0.001
				5	0.013	0.024	0.009	0.013	0.019	0.009
0	0	0.5	0	1	0.016	0.016	0.010	0.019	0.016	0.010
				2	0.022	0.011	0.023	0.021	0.011	0.023
				3	0.021	0.010	0.020	0.017	0.008	0.020
				4	0.012	0.011	0.011	0.013	0.010	0.011
				5	0.026	0.019	0.025	0.026	0.018	0.025
0	0	0.5	0.5	1	0.006	0.009	0.004	0.007	0.009	0.004
				2	0.015	0.010	0.010	0.013	0.010	0.010
				3	0.014	0.012	0.014	0.015	0.010	0.014
				4	0.008	0.010	0.001	0.006	0.008	0.001
				5	0.060	0.063	0.047	0.054	0.071	0.047
0	0.5	0	0	1	0.011	0.010	0.006	0.010	0.010	0.006
				2	0.010	0.007	0.015	0.013	0.008	0.018
				3	0.012	0.017	0.014	0.016	0.014	0.015
				4	0.005	0.002	0.003	0.006	0.004	0.004
				5	0.017	0.022	0.021	0.018	0.030	0.023
0	0.5	0	0.5	1	0.005	0.013	0.004	0.004	0.009	0.004
				2	0.005	0.019	0.009	0.008	0.022	0.011
				3	0.006	0.028	0.006	0.006	0.028	0.007
				4	0.007	0.007	0.004	0.006	0.006	0.004
				5	0.009	0.021	0.009	0.007	0.020	0.012
0	0.5	0.5	0	1	0.009	0.005	0.009	0.012	0.007	0.010
				2	0.016	0.020	0.020	0.018	0.016	0.023
				3	0.014	0.022	0.023	0.017	0.022	0.024
				4	0.005	0.007	0.006	0.006	0.006	0.006
				5	0.009	0.016	0.013	0.008	0.015	0.015
0	0.5	0.5	0.5	1	0	0.011	0	0.001	0.006	0
				2	0.009	0.018	0.010	0.007	0.016	0.012
				3	0.016	0.027	0.020	0.011	0.026	0.024
				4	0.004	0.012	0.001	0.004	0.011	0.001
				5	0.015	0.036	0.021	0.018	0.033	0.024
0.5	0	0	0	1	0.011	0.008	0.009	0.007	0.007	0.010
				2	0.019	0.021	0.023	0.021	0.021	0.027
				3	0.024	0.027	0.028	0.025	0.023	0.029
				4	0.015	0.008	0.009	0.012	0.009	0.009
				5	0.023	0.028	0.030	0.021	0.029	0.030

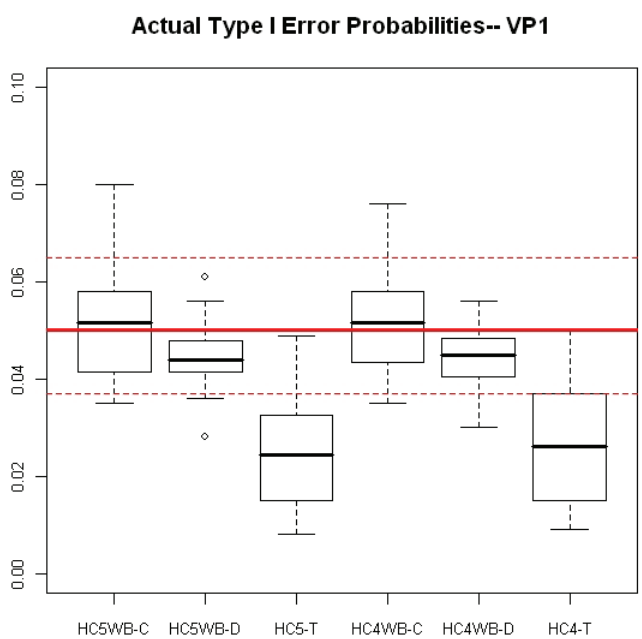
Table 2: Actual Type I Error Rates when Testing at α = 0.01 (continued)

X g	X h	e g	e h	VC	HC5WB-C	HC5WB-D	HC5-T	HC4WB-C	HC4WB-D	HC4-T
0.5	0	0	0.5	1	0.008	0.007	0.005	0.007	0.004	0.005
				2	0.015	0.027	0.013	0.012	0.024	0.013
				3	0.013	0.031	0.009	0.014	0.030	0.009
				4	0.010	0.011	0.006	0.013	0.012	0.006
				5	0.021	0.057	0.015	0.023	0.057	0.015
0.5	0	0.5	0	1	0.010	0.007	0.005	0.010	0.010	0.005
				2	0.017	0.011	0.022	0.017	0.008	0.023
				3	0.026	0.034	0.040	0.029	0.031	0.040
				4	0.005	0.009	0.005	0.003	0.005	0.005
				5	0.028	0.049	0.030	0.023	0.05	0.031
0.5	0	0.5	0.5	1	0.010	0.011	0.004	0.008	0.011	0.004
				2	0.026	0.028	0.019	0.028	0.029	0.020
				3	0.032	0.039	0.032	0.033	0.041	0.034
				4	0.008	0.014	0.005	0.010	0.013	0.005
				5	0.078	0.114	0.072	0.076	0.111	0.079
0.5	0.5	0	0	1	0.008	0.005	0.011	0.011	0.005	0.012
				2	0.019	0.016	0.022	0.021	0.020	0.025
				3	0.007	0.036	0.013	0.006	0.037	0.013
				4	0.004	0.006	0.003	0.005	0.004	0.003
				5	0.012	0.034	0.014	0.012	0.032	0.021
0.5	0.5	0	0.5	1	0.004	0.011	0.002	0.006	0.011	0.002
				2	0.009	0.031	0.008	0.010	0.029	0.013
				3	0.006	0.029	0.008	0.008	0.027	0.010
				4	0.004	0.005	0	0.007	0.004	0.001
				5	0.074	0.114	0.075	0.081	0.116	0.076
0.5	0.5	0.5	0	1	0.003	0.003	0.006	0.005	0.004	0.006
				2	0.012	0.016	0.021	0.015	0.015	0.026
				3	0.015	0.027	0.022	0.016	0.030	0.026
				4	0.004	0.003	0	0.001	0.006	0
				5	0.017	0.036	0.020	0.014	0.038	0.024
0.5	0.5	0.5	0.5	1	0.010	0.011	0.004	0.010	0.014	0.004
				2	0.010	0.023	0.015	0.012	0.020	0.017
				3	0.014	0.045	0.021	0.013	0.047	0.029
				4	0.008	0.014	0.002	0.005	0.011	0.004
				5	0.025	0.059	0.024	0.027	0.060	0.031
				Max	0.078	0.114	0.075	0.081	0.116	0.079
				Min	0	0.002	0	0.001	0.004	0
				Average	0.015	0.021	0.015	0.015	0.021	0.016
				SD	0.013	0.020	0.013	0.013	0.020	0.014

Table 3: Actual Type I Error Rates when Testing at α = 0.05 with Sample Sizes 30, 50 and 100 for HC4-T

X		e		VP	n = 30	n = 50	n = 100
g	h	g	h				
0	0.5	0	0.5	1	0.022	0.021	0.018
0.5	0.5	0.5	0.5	1	0.012	0.019	0.020
0	0.5	0	0.5	4	0.014	0.007	0.011
0	0.5	0.5	0.5	4	0.011	0.009	0.023
0	0	0.5	0.5	5	0.118	0.123	0.143
0.5	0	0.5	0	5	0.093	0.070	0.078
0.5	0	0.5	0.5	5	0.190	0.181	0.174

Figure 1: Actual Type I Error Rates for VP1 when Testing at α = 0.05



The solid horizontal line indicates α = 0.05, the dashed lines indicate the upper and lower confidence limits for α, (0.037, 0.065).

Figure 2: Actual Type I Error Rates Under VP2

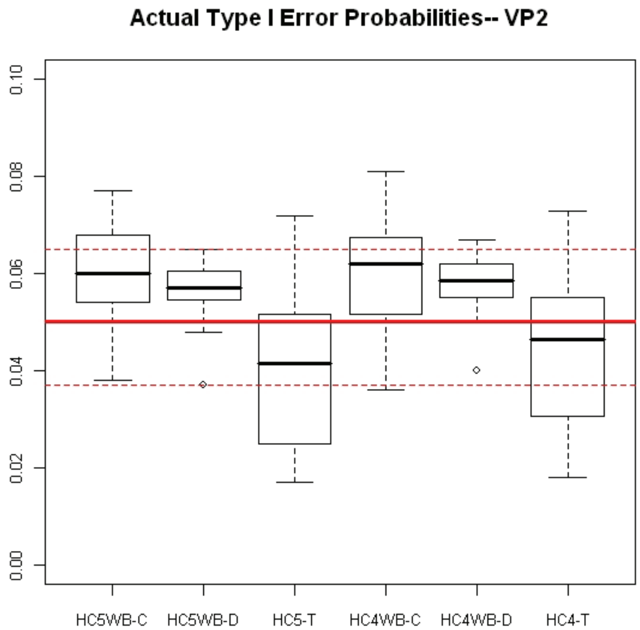


Figure 3: Actual Type I Error Rates Under VP3

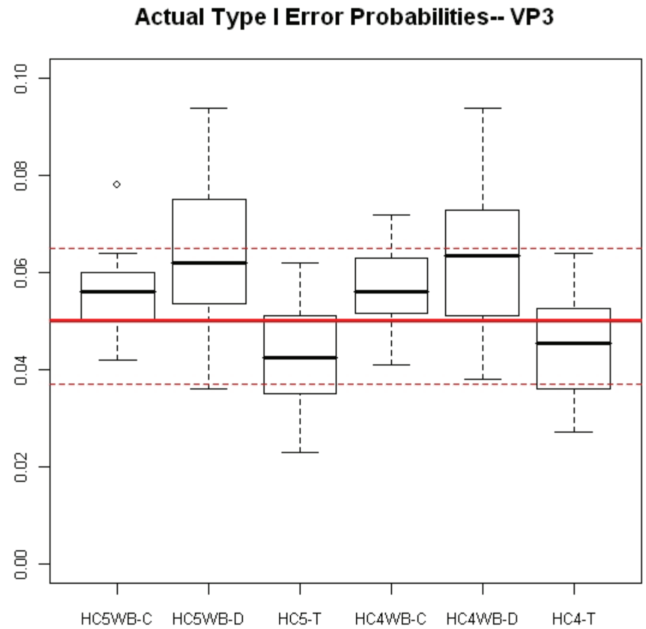


Figure 4: Actual Type I error rates under VP4

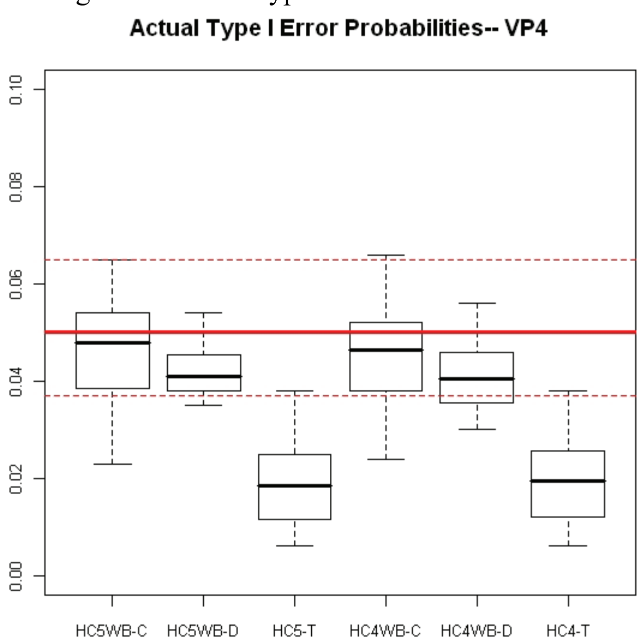


Figure 5: Actual Type I error rates under VP5

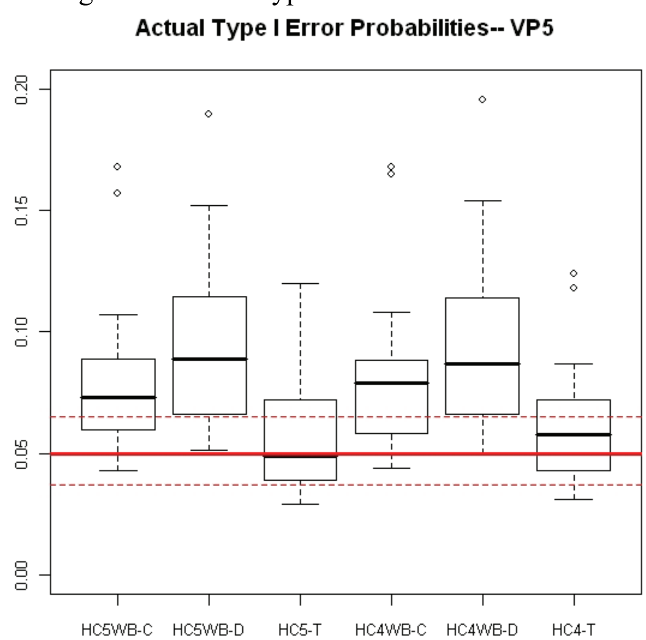


Table 4: Actual Type I Error Rates when Testing at $\alpha = 0.05$, $k = 0.6$ for HC5-T

X		e		VP	HC5-T	HC4-T
g	h	g	h			
0	0.5	0.5	0.5	4	0.013	0.013
0.5	0.5	0	0.5	4	0.015	0.015
0	0	0.5	0.5	5	0.106	0.106
0.5	0	0.5	0.5	5	0.135	0.136

Table 5: Actual Type I Error Rates when Testing at $\alpha = 0.05$, $k = 0.8$ for HC5-T

X		e		VP	HC5-T	HC4-T
g	h	g	h			
0	0.5	0.5	0.5	4	0.005	0.006
0.5	0.5	0	0.5	4	0.007	0.007
0	0	0.5	0.5	5	0.106	0.106
0.5	0	0.5	0.5	5	0.128	0.131