行政院國家科學委員會補助專題研究計畫 ■ 成 果 報 告
□期中進度報告

分量迴歸的模型設定檢定

計畫類別：個別型計畫
計畫編號：NSC 94－2415－H－004－022－
執行期間：94 年 08 月 01 日至 95 年 07 月 31 日

計畫主持人：林馨怡

成果報告類型(依經費核定清單規定繳交)：精簡報告

本成果報告包括以下應繳交之附件：
出席國際學術會議心得報告及發表之論文各一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢

執行單位：

中 華 民 國 95 年 10 月 31 日

1 Introduction

Different economic or econometric theories usually suggest non-nested models in theoretical and empirical researches. Tests for non-nested hypotheses, henceforth the non-nested tests, are important because researchers are able to choose the true model from non-nested models by the tests. The pioneering works of Cox (1961, 1962), Atkinson (1970) and Pesaran and Deaton (1978) become available to comparing non-nested models. Several papers, such as Davidson and MacKinnon (1981), Fisher and McAleer (1981), Gourieroux, Monfort and Trognon (1983), Mizon and Richard (1986) and Vuong (1989), discuss the theoretical methods for non-nested tests. Many papers apply the non-nested tests in empirical applications; complete surveys can be found in Gourieroux and Monfort (1994) and McAleer (1995). It is often the case that empirical fact displays non-normality behavior, such as models with heavy tail or influential outliers. Most of the existing testing procedures are designed for model with normal distribution and are not robust with respect to misspecification of error distribution.

Aguirre-Torres and Gallant (1983) and Hall (1985) have suggested non-nested tests that incorporate M-estimators and base on classical testing procedure. Although M-estimator is a robust estimator in general, their tests using classical procedure are lack of robustness for model where the error distribution is assumed non-normality. To the best of our knowledge, only Victoria-Feser (1997) constructs a robust non-nested test. She considers a Lagrange multiplier version of the Cox test and extends the optimal bounded influence parametric tests of Heritier and Ronchetti (1994) for testing non-nested hypotheses. Her test limits the influence of small contamination in the data and is robust to model deviations. In order to derive her test statistic, one must specify an explicit density function under the null hypothesis to obtain the log-likelihood function and maximum likelihood (ML) estimators of the model. The test is thus restrictive and strong when applying in practice. In addition, The test statistic involves a very difficult integration problem such as the Cox test and are not easy to compute for applied theorists.

In this paper, we propose a robust testing procedure for the non-nested hypotheses. Several features are as follows. First, the proposed test extends the the rank score test of Gutenbrunner, Jurečková, Koenker and Portnoy (1993); this class of rank test plays an important role especially when the empirical phenomenons are non-normality. Second, the rank test statistic is based on the regression rank process that is computed from parametric linear programming method of quantile regression. Thus, we do not need to specify the complete density function and do not need to estimate ML estimators. Any root-n consistent estimators of non-nested models can be used in the proposed test. Third, unlike the non-nested tests in general use or the test of Victoria-Feser (1997), the proposed test is easy to implement by existing software and the simulation or boostrapping methods are not required. Fourth, we show that under very weak assumptions, the proposed test statistic has asymptotically χ^2 distribution and the test is distribution-free. Fifth, the proposed test can be extended to test one model against several alternative non-nested models. The choice of multiple model selection becomes available. Sixth, local powers of the robust rank test are derived. Finally, Monte Carlo simulations results are provided and shows that the proposed test has good finite sample performances against non-nested hypotheses. Comparing with the J test, our rank score tests is robust when the error term is not standard normal distribution such as $N(0, 4)$, t_2 or the Cauchy distributions. Moreover, our test is robust to the relative number of regressors in the two hypotheses.

This paper is organized as follows. In section 2, the rank score tests for non-nested hypotheses are proposed. Single and multiple alternatives are considered. Local alternatives of our test are discussed in section 3. Some Monte Carlo simulation results are presented in section 4. Section 5 is our conclusion of this paper.

2 Rand Score Tests

2.1 Motivations and Setup

Suppose that we want to choose between two linear models as follows:

$$H_0 : \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}_0,$$

$$H_1 : \mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}_1,$$

where the dependent variable $\mathbf{y}$ is an $n \times 1$ matrix, explanatory variables $\mathbf{X}$ and $\mathbf{Z}$ are $n \times p$ and $n \times q$ matrices, and $\mathbf{e}_0$ and $\mathbf{e}_1$ are error terms, respectively. $\mathbf{X}$ and $\mathbf{Z}$ are two matrices which may contain different variables and the models of H_0 and H_1 are non-nested. To test non-nested hypotheses H_0 and H_1, we consider the following artificial nesting model:

$$\mathbf{y} = (1 - \lambda)\mathbf{X}\boldsymbol{\beta} + \lambda\mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}, \tag{1}$$

where $\lambda = 0$ means the null hypothesis is correct and $\lambda = 1$ means the alternative hypothesis is correct. Under this artificial nesting model, we are able to reconsider the non-nested hypotheses as

$$H_0 : \lambda = 0,$$

$$H_1 : \lambda = 1.$$

We can test the non-nested hypotheses by testing $\lambda = 0$ against $\lambda = 1$ for nesting model (1).

Davidson and MacKinnon (1981) estimated $\boldsymbol{\gamma}$ in (1) by its ML estimator $\hat{\boldsymbol{\gamma}}$ and then estimated λ from model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^* + \lambda\mathbf{Z}\hat{\boldsymbol{\gamma}} + \mathbf{e}, \tag{2}$$

with error density $\mathbf{e}$ independently and identically distributed (i.i.d.) as normal distributions. The J test uses the classical t statistic for $\hat{\lambda}$ to test the non-nested hypotheses $\lambda = 0$. Because the J test is based on the classical testing procedure, the test is not robust to the misspecification of error density. For example, if the error distributions of the non-nested models are non-normality, the J test may lead to incorrect inference.

To overcome the non-robustness problem discussed above, Victoria-Feser (1997) proposes a robust test using the optimal bounded influence parametric tests of Heritier and Ronchetti (1994). The test is to limit the influence of small contamination in the data. Victoria-Feser (1997) considers a Lagrange multiplier version of the Cox test and bounds the level influence function of the test. As one can see that her test bounds the effect of the outlier and is a first paper for robust test for non-nested hypotheses. It is however, in the context of ML method, complete density functions of the models should be specified. In addition, one needs to compute the ML estimator in her test. This makes her test statistic very complicated to compute (see p.722-723 for the computation of her test statistic). Her test is thus restrictive and not operational in practice.

2.2 A Robust Test

In this article, a robust testing procedure for non-nested hypotheses is proposed. The test is based on the regression rank score test of Gutenbrunner, Jurečková, Koenker and Portnoy (1993). They test the parametric hypothesis for quantile regression without estimating the density function that is a nuisance parameters in model of quantile regression. We apply their test to test $\lambda = 0$ in the artificial nesting model (2) for testing non-nested models in H_0 and H_1. When $\lambda = 0$, the restricted model becomes $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, and a regression rank score process $\hat{\mathbf{a}}(t)$ can be obtained from the restricted model. $\hat{\mathbf{a}}(t)$ generates "ranks" of the residuals of the restricted model and is solved from

$$\hat{\mathbf{a}}(t) = \arg\max\left\{y'\mathbf{a}\middle|\mathbf{X}'\mathbf{a} = (1-t)\mathbf{X}'\mathbf{1}_n, \mathbf{a} \in [0,1]^n\right\}, \tag{3}$$

with $\mathbf{1}_n$ an $n \times 1$ vector of ones. It may notice that problem (3) is the dual problem of the objective function of linear quantile regression of Koenker and Bassett (1978) in linear programming algorithm. $\hat{\mathbf{a}}(t)$ can be obtained easily from the existing software since the quantile regression has been available in the standard toolbox of researcher's desk.

Let $\hat{a}_i(t)$ be the i'th element of the rank score process. Consider a score generating function $\varphi(t)$ with bounded variation and integrate $\varphi(t)$ with respect to $\hat{a}_i(t)$ from zero to one with $\hat{a}_i(t)$.

We obtain $\hat{\mathbf{b}}$ with ith element

$$b_i = -\int_0^1 \varphi(t) d\hat{a}_i(t).$$

$\hat{\mathbf{b}}$ is the basic statistic of the rank score test. The underlying idea of the rank score test is to check whether $\hat{\mathbf{b}}$ is sufficiently close to zero. Intuitively, $\hat{\mathbf{b}}$ can be interpreted as the weighted sum of residuals of the regression model in terms of rank, and hence should be small when the restriction is valid. Of course, different score-generating functions $\varphi(\cdot)$ lead different $\hat{\mathbf{b}}$. Three commonly used score functions are Wilcoxon scores, normal scores and sign-median scores. We compare the power performances of our test with different score-generating function in Section 4.

Extending the test of Gutenbrunner, Jurečková, Koenker and Portnoy (1993), we propose a rank score test for Davidson and Mackinnon's artificial nesting model (2). The basic statistic is in the following.

$$\mathbf{S}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = \frac{1}{\sqrt{n}}(\mathbf{Z}\hat{\boldsymbol{\gamma}} - \tilde{\mathbf{Z}})'\hat{\mathbf{b}},$$

where $\hat{\boldsymbol{\gamma}}$ can be any consistent estimator of the restricted model, and $\tilde{\mathbf{Z}}$ is the linear projection of $\mathbf{Z}\hat{\boldsymbol{\gamma}}$ on $\mathbf{X}$:

$$\tilde{\mathbf{Z}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\hat{\boldsymbol{\gamma}}.$$

It follows that

$$\mathbf{S}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = \frac{1}{\sqrt{n}}(\mathbf{M}_X\mathbf{Z}\hat{\boldsymbol{\gamma}})'\hat{\mathbf{b}},$$

where $\mathbf{M}_X = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. The rank score test for non-nested hypotheses is defined as:

$$\mathcal{R} := \mathbf{S}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})'\hat{\mathbf{V}}^{-1}\mathbf{S}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})/\mathbf{A}^2(\varphi),$$

where $\hat{\mathbf{V}} = n^{-1}(\mathbf{M}_X\mathbf{Z}\hat{\boldsymbol{\gamma}})'(\mathbf{M}_X\mathbf{Z}\hat{\boldsymbol{\gamma}})$, and

$$\mathbf{A}^2(\varphi) = \int_0^1 \left(\varphi(t) - \int_0^1 \varphi(t)dt\right)^2 dt,$$

for some score function $\varphi(\cdot)$. The proposed test statistic is only composed of data $\mathbf{X}$ and $\mathbf{Z}$, an estimator $\hat{\boldsymbol{\gamma}}$, and $\hat{\mathbf{b}}$, and is easy to computed. In our test, we do not need to specify the complete density function and the estimating of ML estimator is not required. The proposed test is thus easy to implement. In the following, we show that the limiting distribution of the proposed test is asymptotically chi-square distribution with one degree of freedom.

Let $\mathbb{X} = [\mathbf{X}\ \tilde{\mathbf{Z}}]$ be an $n \times (p+1)$ matrix and $\{x_i, i = 1, \cdots, n\}$ the i-th vector of $\mathbb{X}$. The conditional distribution functions of error term e_i conditional on information set $\mathcal{F}$ are denoted as $F_{e_i|\mathcal{F}}, i = 1, \cdots, n$.

Theorem 2.1. *If* (i) $F_{e_i|\mathcal{F}}, i = 1, \cdots, n$ *are i.i.d. and absolutely continuous with with continuous densities* f_{e_i} *uniformly bounded away from 0 and* ∞. (ii) (a) $x_1 = \mathbf{1}_n$, *with* $\mathbf{1}_n$ *an* $n \times 1$ *vector of ones,* (b) $n^{-1}\mathbb{X}'\mathbb{X} \to \mathbf{D}$, *a positive definite matrix,* (c) $n^{-1}\sum_{i=1}^{n}||x_i||^4 = O(1)$, (d) $\max_{i=1,\cdots,n}||x_i|| = O(n^{1/4}/\log n)$, *and* (iii) $\hat{\mathbf{V}} \to \mathbf{V} := \mathbb{E}_0[(\mathbf{M}_X\mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}})'(\mathbf{M}_X\mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}})]/n$, *a positive definite matrix. Under the null hypothesis,*

$$\mathcal{R} \Rightarrow \chi_1^2.$$

Proof. Under the assumptions (i) and (ii), by the same arguments in Theorems 4.1 and 5.1 of Gutenbrunner, Jurečková, Koenker and Portnoy (1993), we have

$$\mathbf{S}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = \frac{1}{\sqrt{n}}\big(\mathbf{M}_X\mathbf{Z}\hat{\boldsymbol{\gamma}}\big)'\hat{\mathbf{b}} = \frac{1}{\sqrt{n}}\big(\mathbf{M}_X\mathbf{Z}\hat{\boldsymbol{\gamma}}\big)'\mathbf{b} + o_p(1),$$

where $\mathbf{b} = -\int_0^1 \varphi(t)d(\tau - \mathbf{1}_{\{y_i - \mathbf{x}_i'\boldsymbol{\beta}_\tau < 0\}})$. In addition, rewrite

$$\begin{aligned}\frac{1}{\sqrt{n}}\big(\mathbf{M}_X\mathbf{Z}\hat{\boldsymbol{\gamma}}\big)'\mathbf{b} &= \frac{1}{\sqrt{n}}\big(\mathbf{M}_X\mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}}\big)'\mathbf{b} + \frac{1}{\sqrt{n}}\big[\mathbf{M}_X\mathbf{Z}\big(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_{\boldsymbol{\beta}}\big)\big]'\mathbf{b} \\ &= \frac{1}{\sqrt{n}}\big(\mathbf{M}_X\mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}}\big)'\mathbf{b} + \big(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_{\boldsymbol{\beta}}\big)\frac{1}{\sqrt{n}}\big(\mathbf{M}_X\mathbf{Z}\big)'\mathbf{b} \\ &= \frac{1}{\sqrt{n}}\big(\mathbf{M}_X\mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}}\big)'\mathbf{b} + o_p(1),\end{aligned}$$

where the last equality holds because $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_{\boldsymbol{\beta}} = o_p(1)$ and $n^{-1/2}\big(\mathbf{M}_X\mathbf{Z}\big)'\mathbf{b} = O_p(1)$. By central limit theorem and under the null hypothesis,

$$\frac{1}{\sqrt{n}}\big(\mathbf{M}_X\mathbf{Z}\hat{\boldsymbol{\gamma}}\big)'\mathbf{b} \Rightarrow N\big(0, \mathbf{V}\mathbf{A}^2(\varphi)\big).$$

Therefore, under assumption (iii), one has

$$\mathcal{R} \Rightarrow \chi_1^2.$$

□

2.3 Multiple Alternatives

The testing procedure introduced in the aforementioned can be extended to the choice of multiple alternatives. Suppose that there are k different non-nested alternatives as follows:

$$H_1^1 : \mathbf{y} = \mathbf{Z}^1\boldsymbol{\gamma}^1 + \mathbf{e}_1,$$
$$\cdots$$
$$H_1^k : \mathbf{y} = \mathbf{Z}^k\boldsymbol{\gamma}^k + \mathbf{e}_k,$$

where $\mathbf{Z}^1, \cdots, \mathbf{Z}^k$ are $n \times q^1, \cdots, n \times q^k$ matrices, $\boldsymbol{\gamma}^1, \cdots, \boldsymbol{\gamma}^k$ are associated parameters, respectively, and $\mathbf{e}_1, \cdots, \mathbf{e}_k$ represents error terms. To test H_0 against multiple alternatives $H_1^1, \cdots, H_1^k$, we combine these non-nested hypotheses into an artificial nesting model as Davidson and McKinnon (1981) and McAleer (1983):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^* + \mathbf{W}\boldsymbol{\lambda} + \mathbf{e},$$

where $\mathbf{W} = \left(\mathbf{Z}^1\hat{\boldsymbol{\gamma}}^1, \cdots, \mathbf{Z}^k\hat{\boldsymbol{\gamma}}^k\right)$, $\hat{\boldsymbol{\gamma}}^1, \hat{\boldsymbol{\gamma}}^2, \cdots, \hat{\boldsymbol{\gamma}}^k$ are consistent estimators of $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2, \cdots, \boldsymbol{\gamma}^k$, respectively, and $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_k)$ contains k elements, $\mathbf{e} = (e_1, \cdots, e_n)$ are error term in this model. We thus can test the multiple non-nested hypotheses by testing whether all elements of $\boldsymbol{\lambda}$ significantly differ from zero or not.

For the multiple non-nested alternatives case, the rank score test statistic is

$$\mathcal{R}_k := \mathbf{S}_k'\hat{\mathbf{V}}_k^{-1}\mathbf{S}_k/\mathbf{A}^2(\varphi),$$

with

$$\mathbf{S}_k = \frac{1}{\sqrt{n}}\left(\mathbf{W} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\right)'\hat{\mathbf{b}},$$

and $\hat{\mathbf{V}}_k = n^{-1}(\mathbf{M}_X\mathbf{W})'(\mathbf{M}_X\mathbf{W})$. Let $\mathbb{X}^* = [\mathbf{X} \ \ \mathbf{W}]$ an $n \times (p + q^1, \cdots + q^k)$ matrix and $\{x_i^*, i = 1, \cdots, n\}$ the i-th vector of $\mathbb{X}^*$. We have the following theorem. Also let the conditional distribution functions of error terms e_i conditional on information set $\mathcal{F}$ are denoted as $F_{i|\mathcal{F}}, i = 1, \cdots, n$.

Theorem 2.2. *If* (i) $F_{e_i|\mathcal{F}}, i = 1, \cdots, n$ *are i.i.d. and absolutely continuous with continuous densities* f_{e_i} *uniformly bounded away from 0 and* ∞. (ii) (a) $x_1^* = \mathbf{1}_n$, *with* $\mathbf{1}_n$ *an* $n \times 1$ *vector of ones,* (b) $n^{-1}\mathbb{X}^{*\prime}\mathbb{X}^* \to \mathbf{D}^*$, *a positive definite matrix,* (c) $n^{-1}\sum_{i=1}^n ||x_i^*||^4 = O(1)$, (d) $\max_{i=1,\cdots,n} ||x_i^*|| = O(n^{1/4}/\log n)$, *and* (iii) $\hat{\mathbf{V}}_k \to \mathbf{V}_k^* := \mathbb{E}_0[(\mathbf{M}_X\mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}})'(\mathbf{M}_X\mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}})]/n$, *Under the null hypothesis,*

$$\mathcal{R}_k \Rightarrow \chi_k^2,$$

a chi-square distribution with k degree of freedom.

3 Local Powers of the Test

The local power of our test are considered in this section. Pesaran (1982) and Ericsson (1983) have compared the local power of non-nested tests. Consider a local alternative to H_0 as

$$H_{1n} : \lambda_n = \frac{\lambda_0}{\sqrt{n}}.$$

Similar to (2), the resulting artificial nesting model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^* + \lambda_n \mathbf{Z}\hat{\boldsymbol{\gamma}} + \mathbf{e},$$

with $\mathbf{e} = (e_1, \cdots, e_n)$. As sample size n increases, the model converges to the null model.

Under H_{1n},

$$\mathbb{E}_{1n}\left[\mathbf{1}_{\{y_i - \mathbf{x}_i' \boldsymbol{\beta}_\tau < 0\}} | \mathcal{F}\right] = \mathbb{E}_{1n}\left[\mathbf{1}_{\{\epsilon_i < -\frac{\lambda_0}{\sqrt{n}} \mathbf{z}_i' \hat{\boldsymbol{\gamma}}\}} | \mathcal{F}\right] = F_{\epsilon_i|\mathcal{F}}\left(-\frac{\lambda_0}{\sqrt{n}} \mathbf{z}_i \hat{\boldsymbol{\gamma}}\right) = \tau - \frac{\lambda_0}{\sqrt{n}} \mathbf{z}_i' \hat{\boldsymbol{\gamma}} f_{\epsilon_i|\mathcal{F}}(0),$$

and

$$\begin{aligned} \mathbb{E}_{1n}[\mathbf{b}_i | \mathcal{F}] &= \mathbb{E}_{1n}\left[-\int_0^1 \varphi(t) d\left(\tau - \mathbf{1}_{\{y_i - \mathbf{x}_i' \boldsymbol{\beta}_\tau < 0\}}\right) \big| \mathcal{F}\right] \\ &= \int_0^1 \varphi(t) d\left[\tau - \mathbb{E}_{1n}\left(\mathbf{1}_{\{y_i - \mathbf{x}_i' \boldsymbol{\beta}_\tau < 0\}} \big| \mathcal{F}\right)\right] \\ &= \left(\frac{\lambda_0}{\sqrt{n}} \mathbf{z}_i' \hat{\boldsymbol{\gamma}}\right) \int_0^1 \varphi(t) d f_{\epsilon_i|\mathcal{F}}(0) \end{aligned}$$

where $F_{\epsilon_i|\mathcal{F}}$ is the conditional distribution of ϵ_i and $f_{\epsilon_i|\mathcal{F}}$ is the conditional density of error term. If the conditional distribution are i.i.d.,

$$\mathbb{E}_{1n}[\mathbf{b}|\mathcal{F}] = \left(\frac{\lambda_0}{\sqrt{n}} \mathbf{Z}\hat{\boldsymbol{\gamma}}\right) \int_0^1 \varphi(t) d f_{\epsilon|\mathcal{F}}(0).$$

From the proof of Theorem 2.1, it is known that $\mathbf{S}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = n^{-1/2}(\mathbf{M}_X \mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}})' \mathbf{b} + o_p(1)$. It follows that

$$\begin{aligned} \mathbb{E}_{1n}\left[\mathbf{S}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})\right] &= \mathbb{E}_{1n}\left[\frac{1}{\sqrt{n}} (\mathbf{M}_X \mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}})' \, \mathbb{E}_{1n}[\mathbf{b}|\mathcal{F}]\right] \\ &= \left(\frac{\lambda_0}{n}\right) \mathbb{E}\left[(\mathbf{M}_X \mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}})' \mathbf{Z}\hat{\boldsymbol{\gamma}} \int_0^1 \varphi(t) d f_{\epsilon|\mathcal{F}}(0)\right] \\ &= \left(\frac{\lambda_0}{n}\right) \omega(\varphi, f) \, \mathbb{E}_{1n}\left[(\mathbf{M}_X \mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}})' \mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}}\right], \end{aligned}$$

with $\int_0^1 \varphi(t) d f_{e_i|\mathcal{F}}(0) = \int_0^1 \varphi(t) d f_{e|\mathcal{F}}(0) = \omega(\varphi, f)$.

Therefore, under H_{1n}, $\mathbf{S}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ is asymptotically normal distributed with mean

$$\left(\frac{\lambda_0}{n}\right) \omega(\varphi, f) \, \mathbb{E}_{1n}\left[(\mathbf{M}_X \mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}})' \mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}}\right]$$

and variance $\mathbf{V}\mathbf{A}^2(\varphi)$. We obtain the asymptotic distribution of $\mathcal{R}_2$ that is noncentral χ_q^2 distribution with non-centrality parameter

$$\left(\frac{\lambda_0}{n}\right)^2 \left[\frac{\omega^2(\varphi, f)}{\mathbf{A}^2(\varphi)}\right] \mathbb{E}_{1n}\left[(\mathbf{M}_X \mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}})' \mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}}\right]' \mathbf{V}^{-1} \, \mathbb{E}_{1n}\left[(\mathbf{M}_X \mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}})' \mathbf{Z}\boldsymbol{\gamma}_{\boldsymbol{\beta}}\right].$$

4 Monte Carlo Simulations

We use Monte carlo simulations to investigate the finite sample performances and robustness of the proposed tests. The data generating process (DGP) are specified as follows. Given weight $\omega \in [0, 1]$,

$$\mathbf{y} = (1-\omega)\mathbf{X}\boldsymbol{\beta} + \omega\mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}, \tag{4}$$

where $\mathbf{X}, \mathbf{Z}$ are $n \times p, n \times q$ random matrices i.i.d. from $N(0, 1)$ except that $\mathbf{x}_1' = \mathbf{1}_n$, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $p \times 1, q \times 1$ vector of ones. The replication of each simulation is 3000 and we compute the rejection probabilities as the finite sample performances. When $\omega = 0$, the nesting model (4) becomes the null model of H_0; the resulting rejection probability is the finite sample size of our test. When $\omega = 1$, the nesting model becomes the alternative model of H_1 and we can obtain the finite sample power. The nominal level is 5% in this section. Four scenarios are considered: the error terms $\mathbf{e}$ are i.i.d. from standard normal distribution, normal distribution with mean 0 and variance 4, the t distribution with 2 degree of freedom, and the standard cauchy distribution.

To examine the finite sample size, we first consider that the error term is i.i.d. drawn from $N(0, 1)$ and $\omega = 0$. Different score generating functions, sign, Wilcoxon and normal, are considered in the simulation. Table 1 reports the rejection probabilities of the test with sample sizes $n = 50$ and $n = 300$. In Table 1, the finite sample size is over-sized which is common in non-nested J test; see Godfrey and Pesaran (1983) and Gourieroux and Monfort (1994). The finite sample size is more accurate when q is small and is greater than the nominal level when q is large relative p. For example, when $q = 2$, finite sample sizes from $p = 2$ to $p = 6$ are 8.3%, 6.6%, 6.2%, 6.8%, 5.6% which are close to the nominal size. Moreover, when $p = 4$, finite sample sizes from $q = 2$ to $q = 6$ are 6.2%, 7.9%, 9.4%, 11.4%, 12.7% which become greater when q is large relative to p. For different score generating functions, the finite sample performances of our test are similar but the finite sample size is smaller with sign score generating function. In addition, the finite sample size is greater when the sample size $n = 50$ than the finite sample sizes when $n = 300$. This result comes from the asymptotic effect.

To examine the finite sample power, we show finite sample power functions of our test in Figures 1–2 with three different score generating functions. The error term is i.i.d. drawn from standard normal distribution. The sample sizes is 300. Figure 1 is power functions for non-nested models with $p = q = 3$. The horizontal axis is ω and the vertical axis is the rejection probability. When ω deviates from zero, the rejection probabilities is finite sample powers. From Figure 1, we can see that our test has good power performances in all three score generating functions. When ω is about ± 0.25, the finite sample power approximates 1. The tests with Wilcoxon and normal score generating functions have almost the same power function. The test with sign score

Table 1: Finite Sample Sizes, %

	n=50					n=300				
sign										
p	q=2	q=3	q=4	q=5	q=6	q=2	q=3	q=4	q=5	q=6
2	7.0	10.7	12.5	16.1	18.8	8.3	10.7	13.1	14.0	17.9
3	7.2	9.0	10.9	12.5	16.0	6.6	7.7	10.4	12.4	14.6
4	6.3	9.3	10.3	11.0	13.8	6.2	7.9	9.4	11.4	12.7
5	7.3	8.1	9.2	9.7	12.8	6.8	6.8	9.5	11.6	10.8
6	6.4	8.0	8.1	9.2	10.4	5.6	7.9	7.4	8.9	9.9
Wilcoxon										
p	q=2	q=3	q=4	q=5	q=6	q=2	q=3	q=4	q=5	q=6
2	8.6	13.3	17.6	21.2	25.8	8.6	12.4	17.1	19.1	26.2
3	7.8	11.2	13.3	16.3	20.3	7.8	9.6	14.1	15.9	19.8
4	7.7	11.6	13.0	14.8	18.1	7.8	8.5	10.9	14.0	16.5
5	8.3	9.6	11.3	12.5	15.8	7.2	9.1	11.8	13.5	14.8
6	6.5	9.2	9.6	12.6	13.7	7.3	9.3	9.4	10.5	13.0
normal										
p	q=2	q=3	q=4	q=5	q=6	q=2	q=3	q=4	q=5	q=6
2	8.5	13.7	17.5	21.3	27.1	8.7	12.2	17.5	20.3	27.5
3	7.1	10.9	13.4	15.7	20.9	8.2	10.3	13.8	17.3	20.4
4	7.1	11.2	12.5	14.2	17.9	7.7	9.1	12.0	14.0	17.3
5	7.5	8.8	10.7	11.9	15.4	6.9	9.1	11.5	14.1	15.0
6	5.7	8.8	9.1	12.7	12.0	7.3	8.8	9.8	10.8	13.9

generating function has the most accurate finite sample sizes than the test with Wilcoxon and normal score generating functions. The finite sample power of the test with Wilcoxon and normal score generating functions is greater than the one of the test with sign score function. Figure 2 plots the power functions of our test for $p = 5, q = 2$ with Wilcoxon, normal and sign score generating functions. Our test has finite sample sizes which are close to the nominal size in all three score generating functions. When $\omega = \pm 0.2$, the finite sample power approximates 1 and thus the finite sample performances of the proposed test are very good. Similar to the case of $p = q = 3$, finite sample powers of the test with Wilcoxon and normal score generating functions are greater than the ones with sign functions. This result is intuitive since the error term is drawn from the normal distribution.

To examine the robustness of our test, we consider several non-standard scenarios and compare the power performances of our test and the J test. First, Figure 3 plots finite sample power functions of our test and the J test with the error term i.i.d. from $N(0, 1)$. Figures 4–5 plot finite sample power functions of our test and the J test when the error term is i.i.d. drawn from $N(0, 4)$. Both the tests are over-sized but our test has more accurate finite sample size than the one of the J test. Therefore, the over-sized problem is less serious in our test. In addition, powers of the J test is greater than powers of our test. Like Godfrey and Pesaran (1983), we adjust our test and the J test to see whether high power comes form high size. We find that our test has slightly high powers when $\omega \neq 0$. This shows that high power performances of the J test come from high size. Finally, we can see that our test and the J test both perform better when the error term is from $N(0, 1)$ than when the error term is from $N(0, 4)$.

Figures 6, 7, 8 plot finite sample power functions of our test and the J test when the error term is i.i.d. drawn from the t_2 distribution for non-nested models with $p = 2, q = 7$ and $p = q = 3$ and $p = 6, q = 3$, respectively. In Figures 6 and 7, sizes of our test are 8.2% and 8.6% which are close to the nominal size. Sizes of the J test are 17.9% and 20.8% which are much greater than the nominal size. The power of our test approximates 1 when $\omega = \pm 0.3$ and the power of the J test approximates 1 when ω are around 0.6 to ± 0.8. Moreover, when $\omega = \pm 0.1$, the power of our test is greater than the power of the J test. This shows that our test has better power performances in cases when $p = q = 3$ and $p = 6, q = 3$. The power of our test approximates 1 more rapid than the power of the J test. In Figure 8, the $p = 2, q = 7$ case, the J test performs poorly when the number of regressors in the alternative hypothesis is large relative to the one in the null hypothesis. Our test is less sensitive to the relative number of regressors in the two hypotheses. Therefore, our test is more robust when the error distribution in non-normal. In addition, the J test is very sensitive to the relative number of regressors of the two non-nested hypotheses.

Figures 9, 10, 11 plot finite sample power functions of our test and the J test when the error term is i.i.d. drawn from the Cauchy distribution for non-nested models with $p = q = 3$, $p = 6, q = 3$,and $p = 2, q = 7$. In Figures 9 and 10, sizes of our test are 6.4% and 5.6% and sizes of the J test are 28.6% and 29.0%. Our test has correct finite sample size while the finite sample size of the J test are badly distorted when the errors are not assumed to be normal. The powers of our test increase largely when ω deviates from 0 and our test has good power performances under Cauchy distribution. The powers of the J test increase very slowly when ω deviates from 0. When $\omega = \pm 1$, the powers of our test is double of the powers of the J test. In Figure 11, our test also has better power than the J test but the performance of the J test is poor. To sum up, when the DGP is Cauchy distribution, the small sample simulations show that our test is robust in two ways. First, our test is robust to different data generating process, especially the non-standard distribution. Second, our test is robust with respect to the number of the regressors of non-nested models.

5 Conclusion

Robust testing procedures of non-nested tests are desired in theoretical and empirical researches. Unlike the optimal bounded influence parametric test considered by Victoria-Feser (1997), we have suggested another robust test for non-nested tests by extending the regression rank score test of Gutenbrunner, Jurečková, Koenker and Portnoy (1993). We introduce a test statistic $\mathcal{R}$ for two non-nested hypotheses and the statistic $\mathcal{R}_k$ for multiple non-nested alternatives. The limiting distributions of our test statistics in this paper are χ^2 distributions and the limiting distributions of the proposed test under the local alternatives are non-central χ^2 distributions. Monte Carlo simulations show that our test has good finite sample performances against the non-nested alternatives. Our test is robust for testing non-nested hypotheses under non-normal error terms and is robust to the relative number of regressors in the two hypotheses.

References

Aguirre-Torres, V. and R. Gallant (1983). The null and non-null asymptotic distribution of the Cox test for multivariate nonlinear regression: alternatives and a new distribution-free Cox test, *Journal of Econometrics*, **21**, 5–33.

Atkinson, A. (1970). A method for discriminating between models, *Journal of the Royal Statistical Society, Series B*, **32**, 323–353.

Cox, D. (1961). Tests of separate families of hypotheses, in *Fourth Berkeley Symposium on Mathematical Statistical Association*, **1**, 105–123.

——— (1962). Further results on tests of separate families of hypotheses, *Journal of the Royal Statistical Society: series B*, **24**, 406–424.

Davidson, R. and J. MacKinnon (1981). Several tests for model specification in the presence of alternative hypotheses, *Econometrica*, **49**, 781–793.

Ericsson, N. R. (1983). Asymptotic properties of instrumental variables statistics for testing non-nested hypotheses, *Review of Economic Studies*, **50**, 287–304.

Fisher, G. and M. McAleer (1981). Alternative procedures and associated tests of significance for non-nested hypotheses, *Journal of Econometrics*, **16**, 103–119.

Godfrey, L. and M. Pesaran, (1983). Tests of non-nested regression models small sample adjustments and Monte Carlo evidence, *Journal of Econometrics*, **21**, 133–154.

Gouriéroux, C. and A. Monfort (1994). Testing non-nested hypotheses, in: R. Engle and D. Mcfadden, eds., *Handbook of Econometrics, Vol. 4*, Amsterdam: North–Holland.

Gouriéroux, C., A. Monfort and A. Trognon (1983). Testing nested or non-nested hypotheses, *Journal of Econometrics*, **21**, 83–115.

Gutenbrunner, C., J. Jurečkova, R. Koenker and S. Portnoy (1993). Tests of linear hypotheses based on regression rank scores, *Nonparametric Statistics*, **2**, 307–331.

Hall, A. (1985). A simplified method of calculating the distribution free Cox test, *Economic Letters*, **18**, 149–151.

Heritier, S. and E. Rochetti (1994). Robust bounded-influence tests in general parametric models, *Journal of the American Statistical Association*, **89**, 897–904.

Koenker, R. and G. Bassett (1978). Regression quantile, *Econometrica*, **46**, 33–50.

McAleer, M. (1983). Exact tests of a model against non-nested alternatives, *Biometrika*, **70**, 285–288.

——— (1995). The significance of testing empirical non-nested models, *Journal of Econometrics*, **67**, 149–171.

Mizon, G. and J. Richard (1986). The encompassing principle and its application to testing non-nested hypotheses, *Econometrica*, **54**, 657–678.

Pesaran, M. (1982). Comparison of local power of alternative tests of non-nested regression models, *Econometrica*, **50**, 1287–1305.

Pesaran, M. H. and A. S. Deaton (1978). Testing non-nested nonlinear regression models, *Econometrica*, **46**, 677–694.

Victoria-Feser, M.-P. (1997). A robust test for non-nested hypotheses, *Journal of the Royal Statis-*

tical Society, Series B, **59**, 715–727.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, **57**, 307–333.

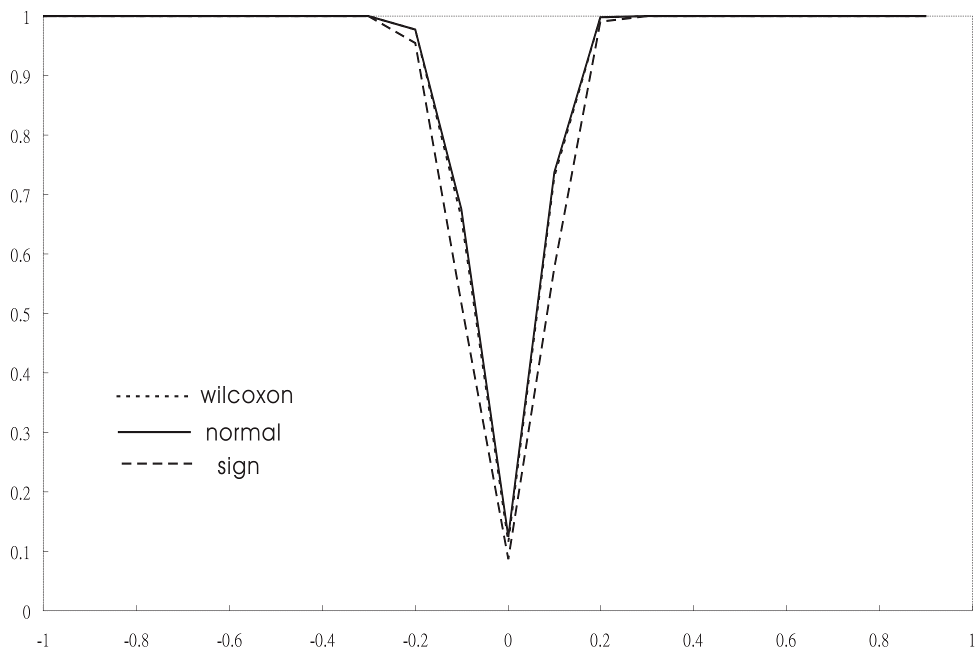

Figure 1: Finite sample powers of $\mathcal{R}$ with $p = 3$ and $q = 3$.

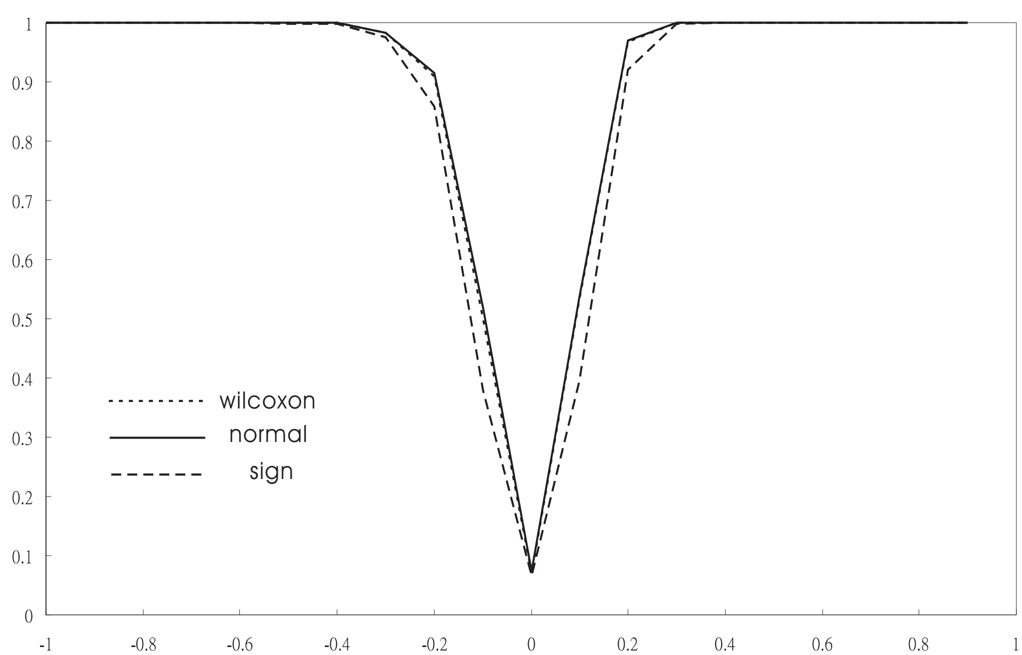

Figure 2: Finite sample powers of $\mathcal{R}$ with $p = 5$ and $q = 2$.

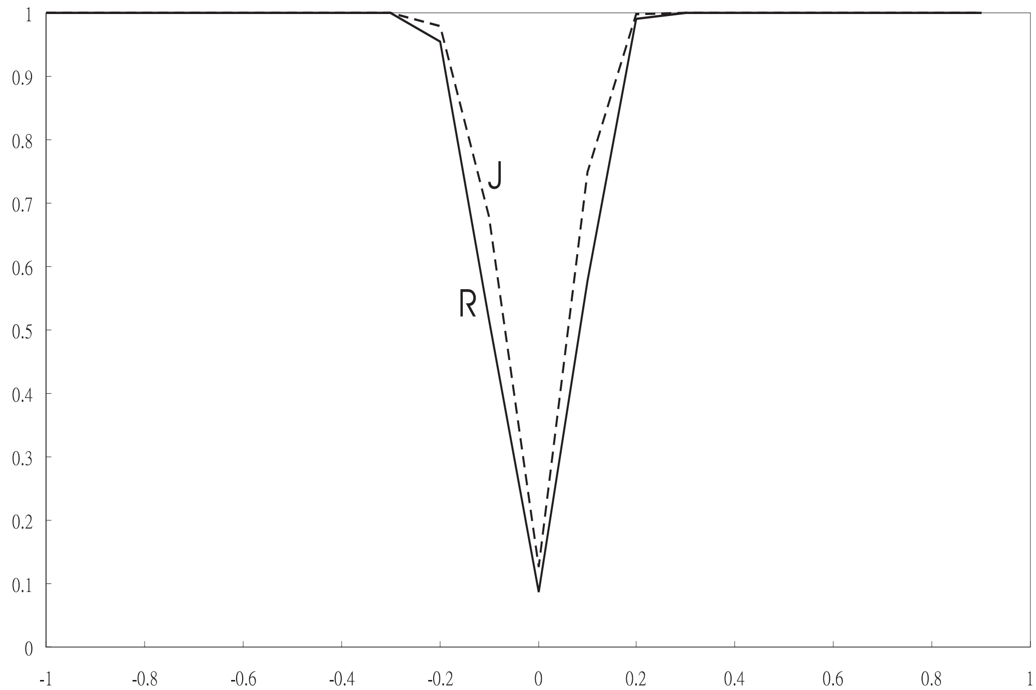

Figure 3: Finite sample powers under $N(0, 1)$ distribution and $p = 3$ and $q = 3$.

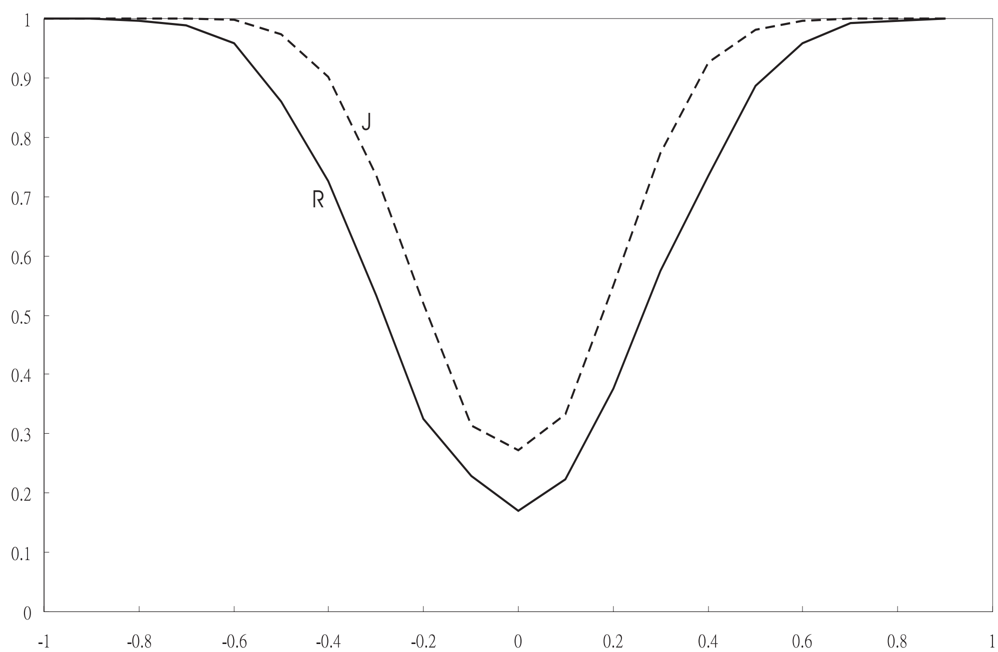

Figure 4: Finite sample powers under $N(0, 4)$ distribution and $p = 3$ and $q = 3$

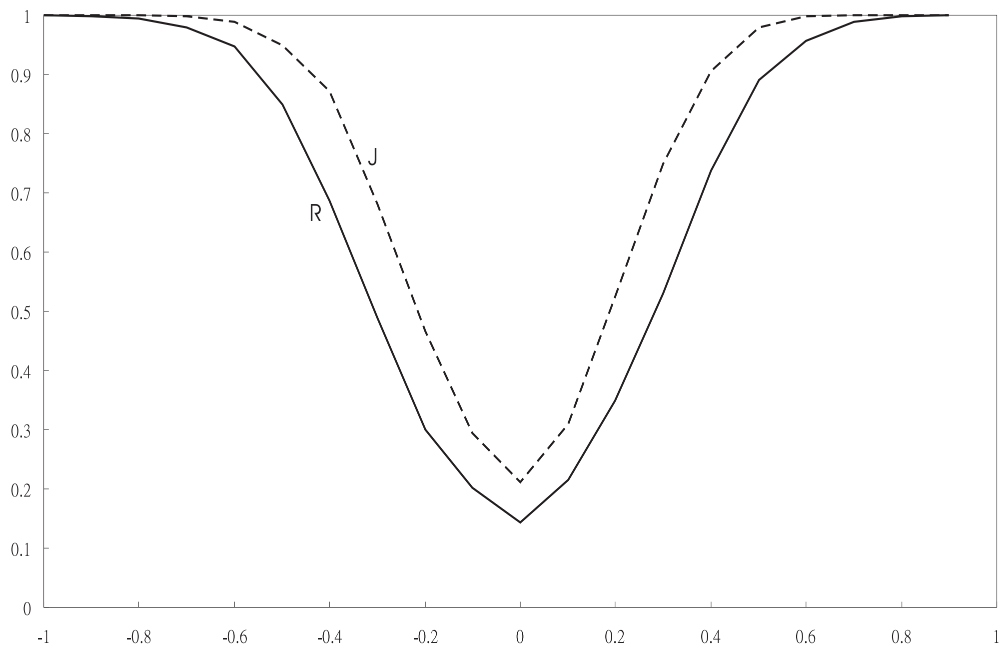

Figure 5: Finite sample powers under $N(0, 4)$ distribution and $p = 6$ and $q = 3$.

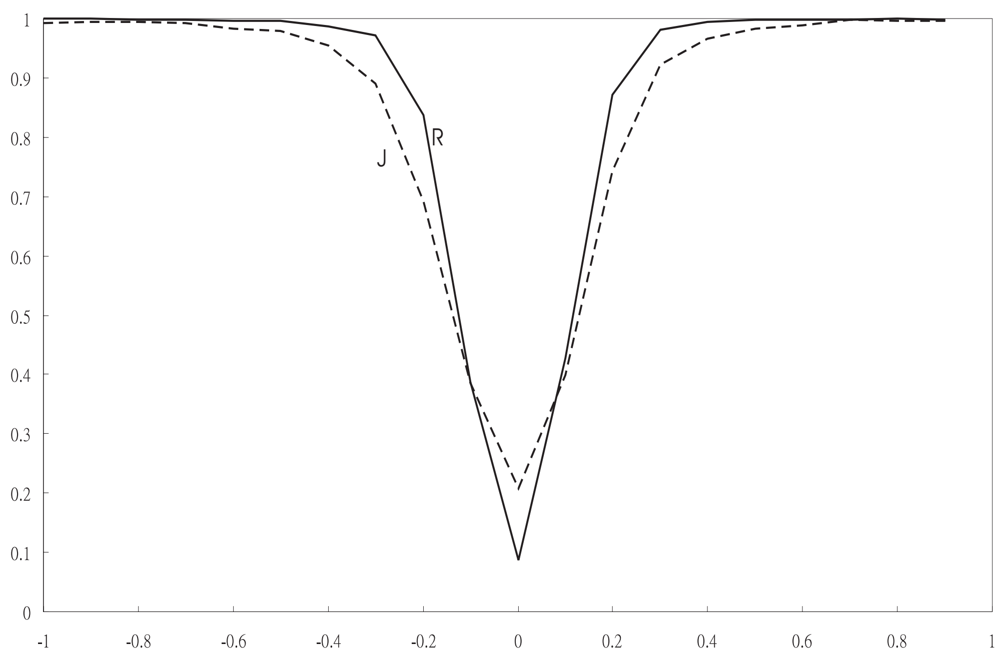

Figure 6: Finite sample powers under t_2 distribution and $p = 3$ and $q = 3$.

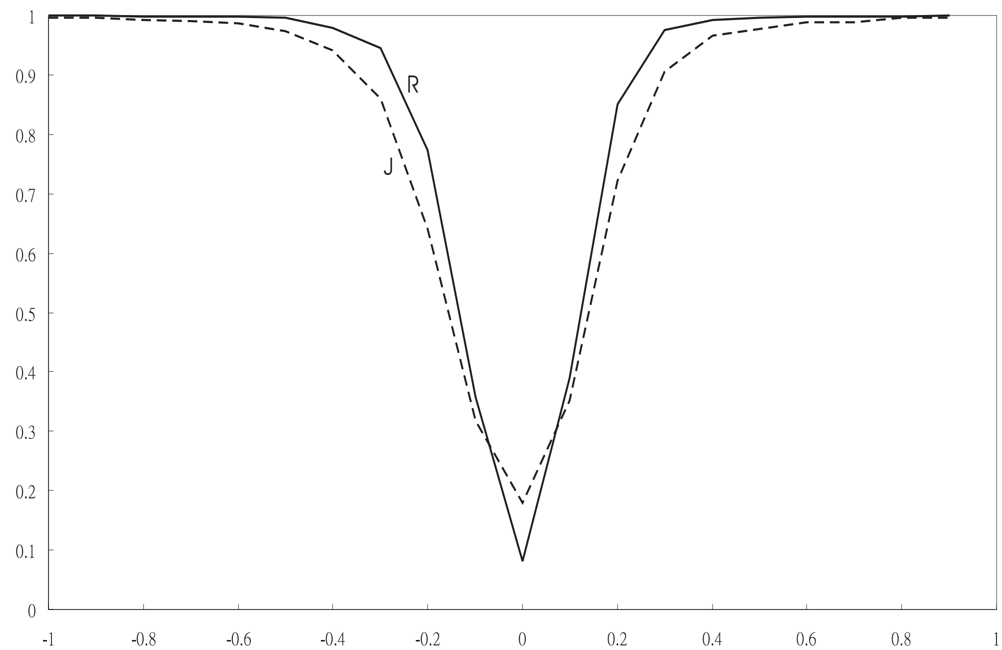

Figure 7: Finite sample powers under t_2 distribution and $p = 6$ and $q = 3$.

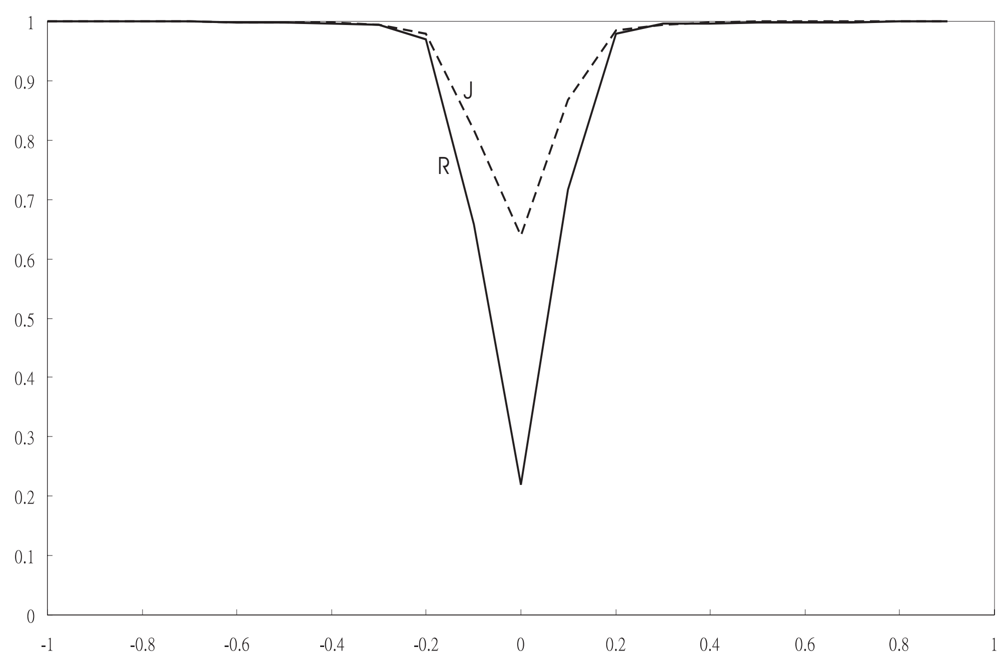

Figure 8: Finite sample powers under t_2 distribution and $p = 2$ and $q = 7$.

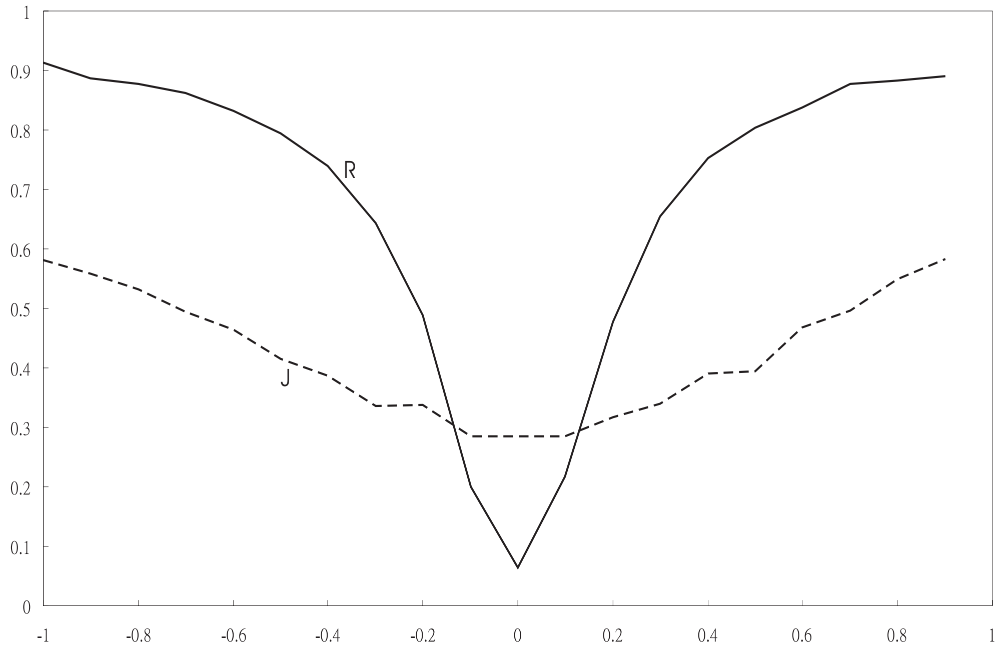

Figure 9: Finite sample powers under Cauchy distribution and $p = 3$ and $q = 3$.

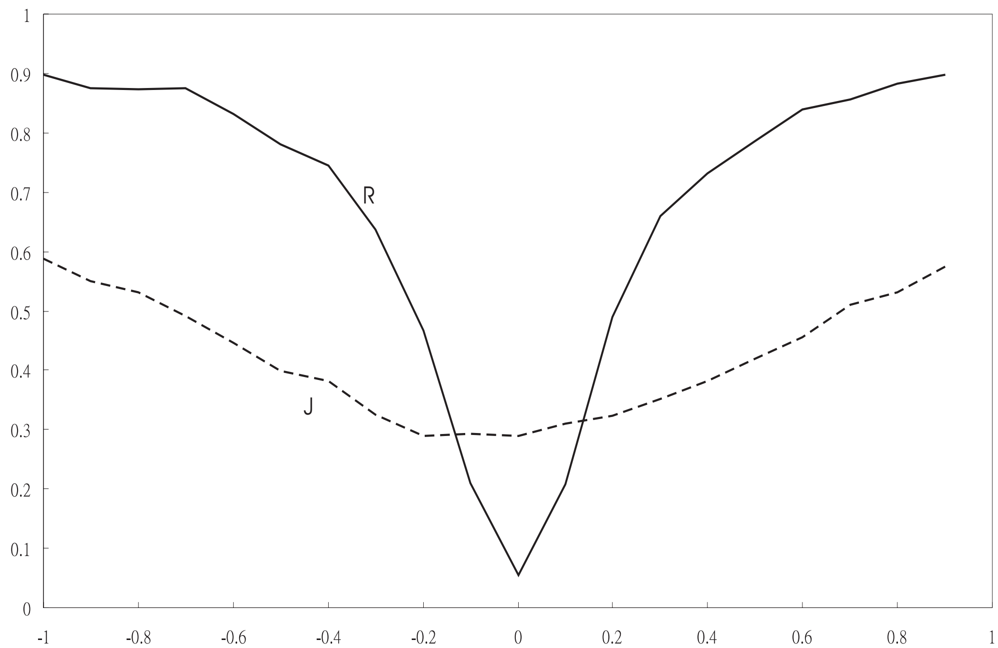

Figure 10: Finite sample powers under Cauchy distribution and $p = 6$ and $q = 3$.

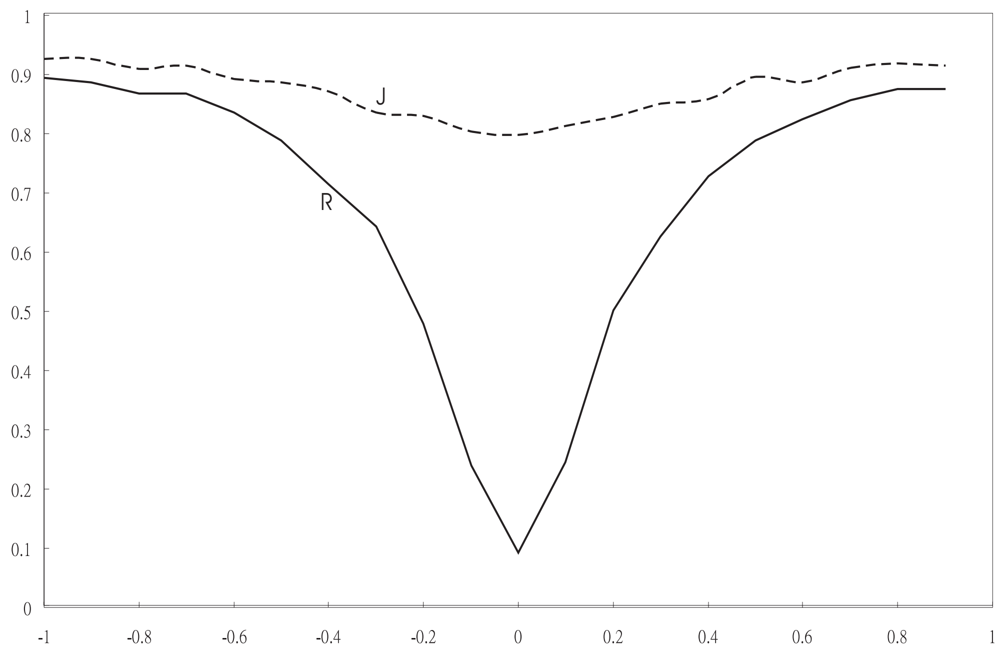

Figure 11: Finite sample powers under Cauchy distribution and $p = 2$ and $q = 7$.