Adjustments of Rao’s Score Test for Distributional and Local Parametric Misspecifications

Anil K. Bera† Yannis Billias‡ Mann J. Yoon§ Süleyman Taşpinar¶ Osman Doğan∥

December 28, 2016

Abstract
Rao (1948)’s seminal paper introduced a fundamental principle of testing based on the score function and the score test has local optimal properties. When the assumed model is misspecified, it is well known that Rao’s score (RS) test looses its optimality. A model could be misspecified in a variety of ways. In this paper, we consider two kinds of misspecification: distributional and parametric. In the first case, the assumed probability density function differs from the data generating process. Kent (1982) and White (1982) analyzed this case and suggested a modified version of the RS test that involves adjustment of the variance. In our parametric misspecification, the dimension of the parameter space of the assumed model does not match with that of the true one. Using the distribution of the RS test under this situation, Bera and Yoon (1993) developed a modified RS test that is valid under local parametric misspecification. This involves adjustment of both the mean and variance of the standard RS rest. This paper considers the joint presence of distributional and parametric misspecification and develops a modified RS test that is valid under both types of misspecification. Earlier modified tests under misspecification can be obtained as special cases of the proposed test. Two examples are provided to illustrate the usefulness of the suggested test procedure. In a Monte Carlo study, we demonstrate that the modified test statistics have good finite sample properties.

JEL-Classification: C13, C21, C31.
Keywords: Rao’s score tests, Lagrange multiplier tests, robust LM tests, specification testing, locally misspecified models, QMLE, inference.

†Economics Program, University of Illinois, Illinois, United States, email: a-bera@uiuc.edu.
‡Department of Economics, University of Cyprus, Nicosia, Cyprus, email: bilias@ucy.ac.cy
§Department of Economics, California State University at Los Angeles, California, United States, email: myoon@calstatela.edu.
¶Economics Program, Queens College, The City University of New York, United States, email: staspinar@qc.cuny.edu.
∥Metro Istanbul A.S., Project Department, Istanbul, Turkey, email: odogan10@gmail.com.

This research was supported, in part, by a grant of computer time from the City University of New York High Performance Computing Center under NSF Grants CNS-0855217 and CNS-0958379. Please address all correspondence to Süleyman Taşpinar at STaspinar@qc.cuny.edu.
1 Introduction

Since Rao (1948)’s seminal paper introduced a fundamental principle of testing based on the score function, as an alternative to the likelihood ratio (LR) and Wald tests, econometricians have produced a large number of specification tests using Rao’s score (RS) principle, also known as the Lagrange multiplier (LM) test among econometricians. Although the success of the score test in econometrics was phenomenal, the main problem in these specification tests was that they were developed under the assumption that the underlying probability model is correct. When the assumed model is misspecified, it is well known that RS test loses its local optimal properties.

A model could be misspecified in a variety of ways. In this paper, we consider two kinds of misspecifications: distributional and parametric. In the first, the assumed probability density function differs from the data generating process. Kent (1982) and White (1982) analyzed this case and suggested a modified version of the RS test that involves adjustment of the variance of the score function. Although most of the econometric specification tests based on the RS principle are in fact dependent upon distributional assumptions in one way or another, we find only a handful of papers that applied the modified RS test. Koenker (1981)’s robust test for heteroskedasticity can also be put in the framework of White (1982) (see Bera (2000, p. 76-77) for this interpretation). Bera and Mckenzie (1986) investigated the finite sample properties of the modified RS test through simulations. Lucas (1998) developed some inferential procedures on cointegrating ranks based on pseudo-likelihoods using the LR and RS tests. Bera and Bilias (2001) applied White’s approach to obtain adjusted tests for skewness and kurtosis for distributional misspecifications.

In parametric misspecification the dimension of the assumed parameter space does not match with the true one. Bera and Yoon (1993) developed a modified RS test that is valid under local parametric misspecification. This involves adjustment of both the mean and variance of the standard RS test. One of the attractive features of this adjusted test procedure is that it helps to identify specific source(s) of departure(s) from the null hypothesis. Several papers applied this approach to a variety of practical problems: Anselin et al. (1996) and Baltagi and Li (2001), to identify particular source(s) of spatial dependence; Godfrey and Veal (2000) to linear models; Bera et al. (2001) and Baltagi et al. (2002) to test for serial dependence in the presence of random effects, and vice-versa in the context of error component models. Several other potential applications were mentioned in Bera and Yoon (1991) and Yoon (1991), such as testing for duration dependence and heterogeneity. In this paper we consider the presence of distributional and parametric misspecifications simultaneously and develop a modified RS test that is valid under both the misspecifications.

We illustrate the derivation of our test statistics within the context of two specification. For the first specification, we consider an ARCH(1) model, and introduce test statistics for the hypothesis about the constant term of the process. Our results show that the test statistics are sensitive to both types of misspecification. In the second illustration, we illustrate the derivation of test statistics within the context of an error component model to test the presence of random effects and the serial correlation of disturbance term. Our analytical results on test statistics for this model indicate that only distributional misspecification has no effect on the asymptotic distributions of
test statistics. We design a Monte Carlo study based on the error component model to investigate
the finite sample size and power properties of test statistics. Our simulation results demonstrate
that our suggested test statistics perform well and can be useful in a specification search.

The rest of this paper is organized as follows. In Section 2, we provide describe a general
framework under which the QMLE is consistent and has an asymptotic normal distribution. In
Section 3, we introduce some notations and review some basic properties of the RS test under
ideal situation. In Sections 4 and 5, we deal with testing under distributional and parametric
misspecification, respectively. In each case, we provide adjusted test statistic that has central chi-

square distribution under the corresponding misspecification. In Section 6, we formulate the robust
form of the RS test in the simultaneous presence of distributional and parametric misspecification.
In Section 7, we provide two applications to illustrate our procedure. In Section 8, we conduct a
Monte Carlo simulation to investigate the finite sample properties of our suggested test statistics.
Section 9 concludes the paper with some remarks. All technical details and proofs are provided in
an appendix.

2 Model Specification and Assumptions

In this section, we state the usual maximum likelihood regularity condition stated in White (1982)
and Kent (1982) to define the QMLE in a general framework. With respect to data generating
process, we assume the following assumption.

Assumption 1. The random vectors $Y_i$, for $i = 1, \ldots, n$, are independent and identically distributed
(i.i.d) with common joint distribution function $G$ on a measure space $(\Omega, \mathcal{F}, \nu)$, where $\Omega$ is the
sample space, $\mathcal{F}$ is the sigma algebra and $\nu$ is a measure. The distribution function $G$ admits a
measurable Radon-Nikodym density $g = dG/d\nu$.\footnote{For the characterization of $(\Omega, \mathcal{F}, \nu)$, see White (1994) Chapter 2.}

The i.i.d property in Assumption 1 can be easily relaxed by defining some additional regularity
conditions as given in White (1994). Since the true data generating process is usually unknown, we
choose to work with a parametric family of distribution functions $F(y, \theta)$ on $(\Omega, \mathcal{F}, \nu)$, which may
or may not contain $G$, where $\theta$ is a $k \times 1$ vector of parameter. The family $F$ is correctly specified
for $Y$ if it contains $G$, otherwise it is misspecified (White 1994). This family satisfies the conditions
of the following assumption.

Assumption 2. $F(y, \theta)$ has Radon-Nikodym density $f(y, \theta) = dF(y, \theta)/d\nu$. Let $\Theta$ be a compact
subset of $\mathbb{R}^k$. The density function $f(y, \theta)$ is measurable with respect to $y \in \Omega$ and continuous with
respect to $\theta \in \Theta$.

The quasi-log-likelihood function generated by $F$ is defined in the following way

$$L(\theta) = \sum_{i=1}^{N} \log f(y_i, \theta)$$ (2.1)
Then, the QMLE is defined by \( \hat{\theta} = \arg\max_{\theta \in \Theta} L_n(\theta) \). Assumptions 1 and 2 ensures the existence of the QMLE. If \( F \) contains true distribution function, that is, if \( G(y) = F(y, \theta_0) \) for some \( \theta_0 \in \Theta \), then the QMLE is just the MLE of \( \theta_0 \). If \( F \) does not include \( G \), then the QMLE is an estimator of a parameter \( \theta_* \) that minimizes the following Kullback-Leibler Information Criterion (KLIC) (Akaike 1998).

\[
I(g : f, \theta) = \mathbb{E} \left( \log(g(Y_i)/f(Y_i, \theta)) \right) = \int \log g(y) dG(y) - \int \log f(y, \theta) dG(y) \tag{2.2}
\]

Hence, the QMLE minimizes our ignorance about the true structure and therefore White (1982) calls the QMLE the minimum ignorance estimator. To establish the consistency and asymptotic normality for the QMLE, we need additional conditions stated in the following assumptions.

**Assumption 3.** (i) \( \mathbb{E}(\log g(Y_i)) \) exists and \( |\log f(y, \theta)| \leq m(y) \) for \( \theta \in \Theta \), where \( m \) is an integrable function with respect to \( G \). (ii) Identification condition: \( I(g : f, \theta) \) has a unique minimum at \( \theta_* \) in \( \Theta \).

**Assumption 4.** (i) The first order derivatives \( \partial \log f(y, \theta)/\partial \theta_j \), for \( j = 1, \ldots, k \), are measurable with respect to \( y \in \Omega \) for all \( \theta \in \Theta \), and continuously differentiable function of \( \theta \) for each \( y \in \Omega \). (ii) There are integrable functions with respect to \( G \) for all \( y \in \Omega \) and \( \theta \in \Theta \), that dominate \( |\partial^2 \log f(y, \theta)/\partial \theta_i \partial \theta_j| \) and \( |\partial \log f(y, \theta)/\partial \theta_i \cdot \partial \log f(y, \theta)/\partial \theta_j| \) for \( j = 1, \ldots, k \).

**Assumption 5.** \( \theta_* \) is an interior point of \( \Theta \).

By Assumption 3, the KLIC stated in (2.2) is well-defined and \( \theta_* \) is globally identifiable. Assumption 4 ensures that the Liapounov central limit theorem can be applied to the vector of scores to establish the asymptotic normal distribution of the QMLE. By Assumption 4, we can define the following matrices

\[
J(\theta) = -\frac{1}{n} \frac{\partial^2 L_n(\theta)}{\partial \theta \partial \theta'}, \quad K(\theta) = \frac{1}{n} \frac{\partial L_n(\theta)}{\partial \theta} \cdot \frac{\partial L_n(\theta)}{\partial \theta'} \tag{2.3}
\]

\[
\mathcal{J}(\theta) = -\mathbb{E} \left( \frac{1}{n} \frac{\partial^2 L_n(\theta)}{\partial \theta \partial \theta'} \right), \quad \mathcal{K}(\theta) = \mathbb{E} \left( \frac{1}{n} \frac{\partial L_n(\theta)}{\partial \theta} \cdot \frac{\partial L_n(\theta)}{\partial \theta'} \right) \tag{2.4}
\]

Under our stated assumptions, it can be shown that \( \hat{\theta} = \theta_* + o_p(1) \) (White 1982). Furthermore, we have \( \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta)}{\partial \theta} \xrightarrow{d} N(0, \mathcal{K}(\theta_*)) \), \( \mathcal{K}(\hat{\theta}) = \mathcal{K}(\theta_*) + o_p(1) \) and \( \mathcal{J}(\hat{\theta}) = \mathcal{J}(\theta_*) + o_p(1) \). Using these asymptotic results, it can easily be shown that

\[
\sqrt{n} \left( \hat{\theta} - \theta_* \right) \xrightarrow{d} N \left[ 0, \mathcal{J}^{-1}(\theta_*) \mathcal{K}(\theta_*) \mathcal{J}^{-1}(\theta_*) \right] \tag{2.5}
\]

If the model is correctly specified, that is, \( g(y) = f(y, \theta_0) \) for some \( \theta_0 \in \Theta \), then \( I(g : f, \theta) \) attains its unique minimum at \( \theta_* = \theta_0 \). Hence, the QMLE \( \hat{\theta} \) is the consistent estimator of the true parameter vector \( \theta_0 \). In this case, we have the information matrix equivalence stating that \( \mathcal{J}(\theta_0) = \mathcal{K}(\theta_0) \). Thus, the asymptotic covariance of \( \sqrt{n} \left( \hat{\theta} - \theta_0 \right) \) reduces to the inverse of Fisher’s information matrix, that is, \( \sqrt{n} \left( \hat{\theta} - \theta_0 \right) \xrightarrow{d} N \left[ 0, \mathcal{J}^{-1}(\theta_0) \right] \).
Different versions of the sandwich formula in (2.5) are generally attributed to Eicker (1963, 1967) and Huber (1967). However, long before the appearance of these papers, Koopmans et al. (1950) derived the sandwich formula

\[ J^{-1}(\theta_*)K(\theta_*)J^{-1}(\theta_*) \]

while studying the large-sample properties of the maximum-likelihood estimates of the parameters of the system of structural equations. Given the vital role of this formula in the econometrics literature, it would not be out of place to summarize the discussion of Koopmans et al. (1950, p. 134) in their Assumption 3.3.1.4. They assumed joint normality of the disturbances. However, in p.135, they explicitly recognized the possibility that the assumed distribution function "has no necessary connection with the distribution of the observations. Nevertheless, we can use the [assumed distribution] function to define parameters by the same maximizing procedure. In these circumstances", - they write- "we shall call it the quasi-likelihood function, and call the maximizing values of its parameters quasi-maximum-likelihood estimates." Possibly, this is the first appearance of the terms "quasi-likelihood function" and "quasi-maximum-likelihood estimates" in the statistics and econometrics literature.

In Section 3.3.10, Koopmans et al. (1950) studied "Asymptotic sampling variances and covariances of the maximum likelihood estimates." In p.150, they derived the sandwich form of the covariance matrix of the estimates [their expression (3.141)]. However, (p.148) they noted that "Mann and Wald’s analysis shows that the expressions for the asymptotic sampling variances and covariances of the maximum-likelihood estimates are greatly simplified by the normality Assumption 3.3.1.4 regarding the distribution of the disturbances. We shall deal only with that case.” As a result, the sandwich form of the covariance matrix of the estimates is simplified to the inverse of the expected information (or, to be more precise, hessian).

3 Rao’s Score Tests under Ideal Conditions

In this section, we consider Rao’s score tests for simple linear hypotheses when there is no distributional or parametric misspecification. For notational simplicity, we set \( \theta_* = \theta_0 \) even when the information matrix equivalence does not hold. Let us denote \( \theta = \left( \gamma', \psi' \right)' \), where the dimensions of \( \gamma \) and \( \psi \) are, respectively, \( k_\gamma \times 1 \), \( k_\psi \times 1 \) such that \( k_\gamma + k_\psi = k \). Let \( d(\theta) = \frac{1}{n} \frac{\partial L(\theta)}{\partial \theta} \) and \( d_{ab}(\theta) = \frac{1}{n} \frac{\partial^2 L(\theta)}{\partial a \partial b} \), where \( a, b = \gamma, \psi \). Consider the following partition of \( J(\theta) \) and \( d(\theta) \):

\[
J(\theta) = \begin{pmatrix}
J_\gamma(\theta) & J_\psi(\theta) \\
J_{\gamma\psi}(\theta) & J_{\psi\psi}(\theta)
\end{pmatrix}
\]  
\[
d(\theta) = \begin{pmatrix}
d_{\gamma\gamma}(\theta) & d_{\gamma\psi}(\theta) \\
d_{\psi\gamma}(\theta) & d_{\psi\psi}(\theta)
\end{pmatrix}
\]  

Assume that the information matrix \( J(\theta_0) \) is partitioned similarly. Let us now consider testing \( H_0 : \psi_0 = \psi_* \), where \( \psi_* \) is a known parameter value, typically, set to zero. Let \( \hat{\theta} = \left( \hat{\gamma}', \hat{\psi}' \right)' \) be the MLE under \( H_0 \), and we assume that it is readily available. Let \( RS_\psi \) be the RS test statistic for
$H_0$, which is given by

$$RS_\psi = n d_\psi^2(\hat{\theta}) J^{-1}_\psi(\hat{\theta}) d_\psi(\hat{\theta})$$

(3.2)

where $J_{\psi,\gamma}(\theta) = J_\psi(\theta) - J_\psi(\theta) J^{-1}_\gamma(\theta) J_\psi(\theta)$. Given correct specification, $RS_\psi$ is locally optimal and has well known asymptotic distributions under the null and a sequence of local alternatives. This may be summarized as follows:

**Proposition 1.** Under our stated assumptions, the following results hold.

1. Under $H_0$, we have $\sqrt{n} d_\psi(\tilde{\theta}) \overset{d}{\to} N [0, J_{\psi,\gamma}(\theta_0)]$, and hence

$$RS_\psi \overset{d}{\to} \chi^2_{k_\psi}(0).$$

(3.3)

2. Under the sequence of local alternative $H_1 : \psi_0 = \psi_* + \xi/\sqrt{n}$, we have

$$RS_\psi \overset{d}{\to} \chi^2_{k_\psi}(\lambda_1).$$

(3.4)

where $\lambda_1 = \lambda_1(\xi) = \xi' J_{\psi,\gamma}(\theta_0) \xi$ and $\xi \neq 0$.

**Proof.** See Appendix A.

In Proposition 1, we are using $\overset{d}{\to}$ to denote convergence in distribution, and $\chi^2_{k_\psi}(\lambda_1)$ stands for the non-central chi-square distribution with $k_\psi$ degrees of freedom and non-centrality parameter $\lambda_1$. In this ideal situation, the RS principle provides a very convenient approach, and it is, as mentioned above, locally optimal if the assumed probability density function $f(y, \theta_0)$ represents the true data generating process (DGP). For a review of the RS test and a historical account, see Bera and Bilias (2001).

### 4 Testing under Distributional Misspecification

Under the distributional misspecification scenario, we consider hypothesis testing when one does not have confidence in the maintained probability model itself, i.e., when $F$ does not contain $G$. An immediate effect of this is that the information matrix equivalence does not hold. See White (1982) for an illustrative example. The effect on the asymptotic distribution of test statistics can be investigated from the Taylor series expansions of scores. Consider the following Taylor series expansions of $\sqrt{n} d_\psi(\tilde{\theta})$ and $\sqrt{n} d_\gamma(\tilde{\theta})$ around $\theta_0$ when $H_1 : \psi_0 = \psi_* + \xi/\sqrt{n}$ holds:

$$\sqrt{n} d_\psi(\tilde{\theta}) = \sqrt{n} d_\psi(\theta_0) - d_\psi(\theta_0) \xi + d_\psi(\theta_0) \sqrt{n} (\hat{\gamma} - \gamma_0) + o_p(1)$$

(4.1)

$$\sqrt{n} d_\gamma(\tilde{\theta}) = \sqrt{n} d_\gamma(\theta_0) - d_\gamma(\theta_0) \xi + d_\gamma(\theta_0) \sqrt{n} (\hat{\gamma} - \gamma_0) + o_p(1)$$

(4.2)

Solving (4.2) for $\sqrt{n} (\hat{\gamma} - \gamma_0)$ and substituting into (4.1) yields
\[
\sqrt{n}d_{\psi}(\hat{\theta}) = (I_{k_\psi \times k_\psi}, -J_{\psi\gamma}(\theta_0)J^{-1}_{\gamma}(\theta_0)) \left( \frac{\sqrt{n}d_{\psi}(\theta_0)}{\sqrt{n}d_{\gamma}(\theta_0)} \right) + J_{\psi\gamma}(\theta_0)\xi + o_p(1), \tag{4.3}
\]

where \( J_{\psi\gamma}(\theta) = J_{\psi}(\theta) - J_{\psi\gamma}(\theta)J^{-1}_{\gamma}(\theta)J_{\gamma\psi}(\theta) \). Note that we use the asymptotic result \( d(\theta_0) = -J(\theta_0) + o_p(1) \) in deriving \( 4.3 \). Given our assumptions and the asymptotic normality of score functions, the result in \( 4.3 \) implies that

\[
\sqrt{n}d_{\psi}(\hat{\theta}) \overset{d}{\to} N [J_{\psi\gamma}(\theta_0)\xi, B_{\psi\gamma}(\theta_0)] \tag{4.4}
\]

where \( B_{\psi\gamma}(\theta_0) = K_{\psi}(\theta_0) + J_{\psi\gamma}(\theta_0)J^{-1}_{\gamma}(\theta_0)K_{\gamma}(\theta_0)J_{\gamma\psi}(\theta_0) - J_{\psi\gamma}(\theta_0)J^{-1}_{\gamma}(\theta)K_{\gamma\psi}(\theta) - K_{\psi\gamma}(\theta_0)J^{-1}_{\gamma}(\theta)J_{\gamma\psi}(\theta_0) \). Comparing \( 4.4 \) with \( 5.13 \), the distributional misspecification now changes the asymptotic variance of the score vector invalidating the original RS test statistic in \( 3.2 \) based on the asymptotic variance \( J_{\psi\gamma}(\theta_0) \). When there is no distributional misspecification, \( J(\theta_0) = K(\theta_0) \), and therefore \( B_{\psi\gamma}(\theta_0) = J_{\psi}(\theta_0) - J_{\psi\gamma}(\theta_0)J^{-1}_{\gamma}(\theta)J_{\gamma\psi}(\theta_0) \equiv J_{\psi\gamma}(\theta_0) \).

For distributional misspecification, it is very straightforward to obtain a robust form of the RS test. We just have to use the correct variance matrix \( B_{\psi\gamma}(\theta_0) \) in the formation of the test statistic. Let \( RS^*_{\psi}(D) \) denote the robust RS statistic under distributional misspecification for testing \( H_0 : \psi = \psi_0 \). \( RS^*_{\psi}(D) \) can be written as:

\[
RS^*_{\psi}(D) = n \hat{d}_{\psi}(\hat{\theta})B^{-1}_{\psi\gamma}(\hat{\theta})d_{\psi}(\hat{\theta}) \tag{4.5}
\]

where \( B_{\psi\gamma}(\hat{\theta}) = K_{\psi}(\hat{\theta}) + J_{\psi\gamma}(\hat{\theta})J^{-1}_{\gamma}(\hat{\theta})K_{\gamma}(\hat{\theta})J_{\gamma\psi}(\hat{\theta}) - J_{\psi\gamma}(\hat{\theta})J^{-1}_{\gamma}(\hat{\theta})K_{\gamma\psi}(\hat{\theta}) - K_{\psi\gamma}(\hat{\theta})J^{-1}_{\gamma}(\hat{\theta})J_{\gamma\psi}(\hat{\theta}) \).

We summarize the main results of this section in the following proposition.

**Proposition 2.** Under our stated assumptions, the following results hold.

1. Under the distributional misspecification, the results in Proposition 1 are invalid. That is, \( RS_{\psi} \) in \( 3.2 \) does not has an asymptotic chi-square distribution.

2. Under \( H_0 \), we have \( \sqrt{n}d_{\psi}(\hat{\theta}) \overset{d}{\to} N [0, B_{\psi\gamma}(\theta_0)] \), and hence

\[
RS^*_{\psi}(D) \overset{d}{\to} \chi^2_{k_\psi}(0). \tag{4.6}
\]

3. Under local alternatives \( H_1 : \psi_0 = \psi_0 + \xi/\sqrt{n} \), we have

\[
RS^*_{\psi}(D) \overset{d}{\to} \chi^2_{k_\psi}(\lambda_2). \tag{4.7}
\]

where \( \lambda_2 \equiv \lambda_2(\xi) = \xi' J_{\psi\gamma}(\theta_0)B^{-1}_{\psi\gamma}(\theta_0)J_{\psi\gamma}(\theta_0)\xi \) and \( \xi \neq 0 \).

**Proof.** See Appendix A. \( \square \)
White (1982) provides a robust form of the RS test under distributional misspecification for testing the null hypothesis $H_0 : h(\theta_0) = 0$ against the alternative $H_1 : h(\theta_0) \neq 0$, where $h(\cdot)$ is an $k_h \times 1$ vector function. Let the test statistic be denoted by $RS_h^*$. Using our notation, $RS_h^*$ can be written as:

$$RS_h^* = n d_\phi(\hat{\theta}) J^{-1}(\hat{\theta}) H'(\hat{\theta}) \left[ H(\hat{\theta}) B(\hat{\theta}) H'(\hat{\theta}) \right]^{-1} H(\hat{\theta}) J^{-1}(\hat{\theta}) d_\phi(\hat{\theta}),$$

(4.8)

where $d_\phi(\hat{\theta}) = \left( d_{\phi}(\hat{\theta}), d_{\psi}(\hat{\theta}) \right)'$, $B(\theta) = J^{-1}(\theta) K(\theta) J^{-1}(\theta)$, and $H(\theta) = \partial h(\theta)/\partial \theta'$. Under our current setup, we have $h(\theta_0) = \psi_0 - \psi*$ and $H(\theta) = \left[0_{k_\psi \times k_\psi}, I_{k_\psi \times k_\psi} \right]$, and thus we can express $RS_h^*$ as $RS_{\psi}^*$:

$$RS_{\psi}^* = n d_{\psi}(\hat{\theta}) \left[ J_{\psi}(\hat{\theta}) B_{\psi}(\hat{\theta}) J(\hat{\theta}) \right]^{-1} d_{\psi}(\hat{\theta})$$

(4.9)

where $B_{\psi}(\hat{\theta})$ is the $k_\psi \times k_\psi$ block of $B$ that can be written as

$$B(\hat{\theta}) = \begin{pmatrix} B_{\gamma}(\hat{\theta}) & B_{\gamma \psi}(\hat{\theta}) \\ B_{\psi \gamma}(\hat{\theta}) & B_{\psi}(\hat{\theta}) \end{pmatrix} = \begin{pmatrix} J_{\gamma}(\hat{\theta}) & J_{\gamma \psi}(\hat{\theta}) \\ J_{\psi \gamma}(\hat{\theta}) & J_{\psi}(\hat{\theta}) \end{pmatrix}^{-1} \begin{pmatrix} K_{\gamma}(\hat{\theta}) & K_{\gamma \psi}(\hat{\theta}) \\ K_{\psi \gamma}(\hat{\theta}) & K_{\psi}(\hat{\theta}) \end{pmatrix} \begin{pmatrix} J_{\gamma}(\hat{\theta}) & J_{\gamma \psi}(\hat{\theta}) \\ J_{\psi \gamma}(\hat{\theta}) & J_{\psi}(\hat{\theta}) \end{pmatrix}^{-1}$$

(4.10)

Using the partitioned inverse matrix formula and some straightforward algebra, it can be readily seen that $RS_{\psi}^*$ in (4.11) is nothing but our $RS_{\psi}^*(D)$ in (4.5). Note also that, in the special case of $J_{\psi}(\theta) = 0$, the robust formula can be simplified further as:

$$RS_{\psi}^* = RS_{\psi}^* = n d_{\psi}(\hat{\theta}) K_{\psi}(\hat{\theta})^{-1} d_{\psi}(\hat{\theta})$$

(4.11)

That is, the robust form of the RS statistic can be obtained just by using $K_{\psi}(\theta)$ when $J(\theta)$ is block diagonal.

5 Testing under Parametric Misspecification

Suppose we miss out one parameter vector $\phi$ with dimension $k_\phi \times 1$ from our analysis, i.e., our true parameter vector is now represented by $\theta_0 = (\gamma'_0, \psi'_0, \phi'_0)'$, where the dimensions of $\gamma$, $\psi$ and $\phi$ are, respectively, $k_\gamma \times 1$, $k_\psi \times 1$ and $k_\phi \times 1$ such that $k_\gamma + k_\psi + k_\phi = k$. In this section, our focus is on the parametric misspecification arising from the local presence of $\phi_0$ in the alternative model, and therefore we assume that the information matrix equivalence holds. Let $d(\theta) = \frac{1}{n} \frac{\partial L(\theta)}{\partial \theta}$, $d_a(\theta) = \frac{1}{n} \frac{\partial L(\theta)}{\partial a}$ and $d_{ab}(\theta) = \frac{1}{n} \frac{\partial L(\theta)}{\partial a \partial b}$, where $a, b = \gamma, \psi, \phi$. Consider the following partition of $J(\theta)$
and \( d(\theta) \):

\[
J(\theta) = \begin{pmatrix}
J_\gamma(\theta) & J_\psi(\theta) & J_\gamma\phi(\theta) \\
\kappa_\gamma \times k_\gamma & \kappa_\gamma \times k_\phi & \kappa_\gamma \times k_\phi \\
J_\psi\gamma(\theta) & J_\psi(\theta) & J_\psi\phi(\theta) \\
\kappa_\psi \times k_\gamma & \kappa_\psi \times k_\phi & \kappa_\psi \times k_\phi \\
J_\gamma\phi(\theta) & J_\phi(\theta) & J_\phi(\theta) \\
\kappa_\phi \times k_\gamma & \kappa_\phi \times k_\phi & \kappa_\phi \times k_\phi
\end{pmatrix}, \quad d(\theta) = \begin{pmatrix}
d_\gamma(\theta) & d_\psi(\theta) & d_\gamma\phi(\theta) \\
\kappa_\gamma \times k_\gamma & \kappa_\gamma \times k_\phi & \kappa_\gamma \times k_\phi \\
d_\psi(\theta) & d_\psi(\theta) & d_\psi(\theta) \\
\kappa_\psi \times k_\gamma & \kappa_\psi \times k_\phi & \kappa_\psi \times k_\phi \\
d_\phi(\theta) & d_\phi(\theta) & d_\phi(\theta) \\
\kappa_\phi \times k_\gamma & \kappa_\phi \times k_\phi & \kappa_\phi \times k_\phi
\end{pmatrix}.
\] (5.1)

The information matrix \( J(\theta_0) \) is partitioned similarly. Let \( \tilde{\theta} = (\tilde{\gamma}', \tilde{\psi}', \phi_0)' \) be the restricted MLE under \( H_0 : \psi_0 = \psi_* \) and \( H_0 : \phi_0 = \phi_* \). We are interested in the asymptotic distribution of \( \sqrt{n}d_\psi(\tilde{\theta}) \), \( \sqrt{n}d_\phi(\tilde{\theta}) \) and \( \sqrt{n}d_\gamma(\tilde{\theta}) \) under the sequences of local alternatives \( H_1 : \psi_0 = \psi_* + \xi/\sqrt{n} \) and \( H_1 : \phi_0 = \phi_* + \delta/\sqrt{n} \). For this purpose, we consider the following Taylor series expansions of \( \sqrt{n}d_\psi(\tilde{\theta}) \), \( \sqrt{n}d_\phi(\tilde{\theta}) \) and \( \sqrt{n}d_\gamma(\tilde{\theta}) \) around \( \theta_0 \) when \( H_1 : \psi_0 = \psi_* + \xi/\sqrt{n} \) and \( H_1 : \phi_0 = \phi_* + \delta/\sqrt{n} \) hold:

\[ \sqrt{n}d_\psi(\tilde{\theta}) = \sqrt{n}d_\psi(\theta_0) - d_\psi(\theta_0)\xi + d'_\psi(\theta_0)\delta + o_p(1) \] (5.2)

\[ \sqrt{n}d_\phi(\tilde{\theta}) = \sqrt{n}d_\phi(\theta_0) - d_\phi(\theta_0)\xi + d'_\phi(\theta_0)\delta + o_p(1) \] (5.3)

\[ \sqrt{n}d_\gamma(\tilde{\theta}) = \sqrt{n}d_\gamma(\theta_0) - d_\gamma(\theta_0)\xi + d'_\gamma(\theta_0)\delta + o_p(1) \] (5.4)

Solving (5.4) for \( \sqrt{n}(\tilde{\gamma} - \gamma_0) \) and substituting into (5.2) and (5.3) yields

\[ \sqrt{n}d_\psi(\tilde{\theta}) = (I_{k_\psi \times k_\psi} - J_{\psi\gamma}(\theta_0)J_{\gamma}^{-1}(\theta_0)) \left( \frac{\sqrt{n}d_\psi(\theta_0)}{\sqrt{n}d_\gamma(\theta_0)} \right) + J_{\psi\gamma}(\theta_0)\xi + J_{\psi\phi}(\theta_0)\delta + o_p(1), \] (5.5)

\[ \sqrt{n}d_\phi(\tilde{\theta}) = (I_{k_\phi \times k_\phi} - J_{\phi\gamma}(\theta_0)J_{\gamma}^{-1}(\theta_0)) \left( \frac{\sqrt{n}d_\phi(\theta_0)}{\sqrt{n}d_\gamma(\theta_0)} \right) + J_{\phi\gamma}(\theta_0)\delta + J_{\phi\phi}(\theta_0)\xi + o_p(1), \] (5.6)

where \( J_{\psi\gamma}(\theta) = J_{\psi\gamma}(\theta_0) - J_{\psi\gamma}(\theta_0)J_{\gamma}^{-1}(\theta_0)J_{\gamma}(\theta) \) and \( J_{\phi\gamma}(\theta) = J_{\phi\gamma}(\theta_0) - J_{\phi\gamma}(\theta_0)J_{\gamma}^{-1}(\theta_0)J_{\phi}(\theta) \).

Note that we use the asymptotic result \( d(\theta_0) = -J(\theta_0) + o_p(1) \) in deriving (5.5) and (5.6). Under our assumptions and the sequences of local alternatives, \( H_1 : \psi_0 = \psi_* + \xi/\sqrt{n} \) and \( H_1 : \phi_0 = \phi_* + \delta/\sqrt{n} \), the result in (5.5) implies that

\[ \sqrt{n}d_\psi(\tilde{\theta}) \overset{d}{\rightarrow} N \left[ J_{\psi\gamma}(\theta_0)\xi + J_{\psi\phi}(\theta_0)\delta, J_{\psi\gamma}(\theta_0) \right] \] (5.7)

Again, it is interesting to compare (5.7) with (4.4). Unlike the distributional misspecification discussed in the previous section, the parametric misspecification changes the asymptotic mean of the score keeping the asymptotic variance unaffected. That is, the asymptotic mean of the score function \( \sqrt{n}d_\psi(\tilde{\theta}) \) is now contaminated by the local presence of the nuisance parameter \( \phi_0 \).
Obviously, the nonzero mean \( J_{\psi,\gamma}(\theta_0) \delta \) would give rise to the non-centrality of asymptotic null distribution of RS as follows:

\[
RS_{\psi} \xrightarrow{d} \chi_{k_{\psi}}^2(\lambda_3)
\]  

(5.8)

where the non-centrality parameter is \( \lambda_3 = \delta' J_{\psi,\gamma}(\theta_0) J_{\psi,\gamma}^{-1}(\theta_0) J_{\psi,\gamma}(\theta_0) \delta \). The matrix \( J_{\psi,\gamma}(\theta_0) \) can be interpreted as the partial covariance between \( d_{\psi}(\theta_0) \) and \( d_{\phi} \) after eliminating the linear effect of \( d_{\gamma} \) on \( d_{\psi} \) and \( d_{\phi} \) (Anderson [1984] p. 36). Therefore, even asymptotically, the size of the test will not be correct as seen from non-zero \( \lambda_3 \), unless \( \delta \neq 0 \) belongs to the null space of \( J_{\psi,\gamma}(\theta_0) \) and/or \( J_{\psi,\gamma}(\theta_0) \) itself is zero. Following Bera and Yoon (1993), the result in (5.7) can be used to develop a modified RS test that is valid under local parametric misspecification. This involves adjustment of both the mean and variance of the standard RS test. For this purpose, we need to determine the asymptotic distribution of \( \sqrt{n} d_{\psi}(\bar{\theta}) \) under \( H_0 : \psi_0 = \psi_* \) and \( H_1 : \phi_0 = \phi_* + \delta / \sqrt{n} \). The result in (5.6) implies that

\[
\sqrt{n} d_{\psi}(\bar{\theta}) \xrightarrow{d} N [ J_{\psi,\gamma}(\theta_0) \delta, J_{\psi,\gamma}(\theta_0) ]
\]  

(5.9)

Hence, we have \( J_{\psi,\gamma}^{-1}(\theta_0) \sqrt{n} d_{\psi}(\bar{\theta}) \xrightarrow{d} N [ \delta, J_{\psi,\gamma}^{-1}(\theta_0) ] \). Also, under \( H_0 : \psi_0 = \psi_* \) and \( H_1 : \phi_0 = \phi_* + \delta / \sqrt{n} \), the result in (5.7) implies that \( \sqrt{n} d_{\psi}(\bar{\theta}) - J_{\psi,\gamma}(\theta_0) \delta \xrightarrow{d} N [ 0, J_{\psi,\gamma}(\theta_0) ] \). By using these results, we can formulate an adjusted score that has zero mean. Let \( d^*_{\psi}(\theta) \) be the adjusted score. Then, its feasible version is given by

\[
\sqrt{n} d^*_{\psi}(\bar{\theta}) = \sqrt{n} \left[ d_{\psi}(\bar{\theta}) - J_{\psi,\gamma}(\bar{\theta}) J_{\psi,\gamma}^{-1}(\theta_0) d_{\phi}(\bar{\theta}) \right]
\]  

(5.10)

Under \( H_0 : \psi_0 = \psi_* \) and irrespective of whether \( H_0 : \phi_0 = \phi_* \) or \( H_1 : \phi_0 = \phi_* + \delta / \sqrt{n} \) holds, it can be shown that

\[
\sqrt{n} d^*_{\psi}(\bar{\theta}) \xrightarrow{d} N [ 0_{k_{\psi} \times 1}, J_{\psi,\gamma}(\theta_0) - J_{\psi,\gamma}(\theta_0) J_{\psi,\gamma}^{-1}(\theta_0) J_{\psi,\gamma}(\theta_0) ]
\]  

(5.11)

The result in (5.11) can be used to formulate an adjusted RS statistic under parametric misspecification. Let \( RS^*_{\psi}(P) \) be the adjusted RS statistic. In the following proposition, we provide the asymptotic distribution of this statistics along with some other results.

**Proposition 3.** Under our stated assumptions, the following results hold.

1. Under the sequences of local alternatives, \( H_1 : \psi_0 = \psi_* + \xi / \sqrt{n} \) and \( H_1 : \phi_0 = \phi_* + \delta / \sqrt{n} \), we have

\[
RS_{\psi} \xrightarrow{d} \chi_{k_{\psi}}^2(\lambda_4).
\]  

(5.12)

where \( \lambda_4 = \xi' J_{\psi,\gamma}(\theta_0) \xi + \delta' J_{\psi,\gamma}(\theta_0) \xi + \xi' J_{\psi,\gamma}(\theta_0) \delta + \delta' J_{\psi,\gamma}(\theta_0) J_{\psi,\gamma}^{-1}(\theta_0) J_{\psi,\gamma}(\theta_0) \delta \).

\[\text{For details, see the proof of Proposition 3}\]
2. Under the sequences of local alternatives, $H_1 : \psi_0 = \psi_* + \xi/\sqrt{n}$ and $H_0 : \phi_0 = \phi_*$, we have

$$RS_\psi \xrightarrow{d} \chi^2_{k_\psi}(\lambda_5).$$

(5.13)

where $\lambda_5 = \xi' J_{\psi,\gamma}(\theta_0) \xi$.

3. Under $H_0 : \psi_0 = \psi_*$ and irrespective of whether $H_0 : \phi_0 = \phi_*$ or $H_1 : \phi_0 = \phi_* + \delta/\sqrt{n}$ holds, we have

$$RS^*_\psi(P) = n d_\psi^*(\tilde{\theta}) \left[ J_{\psi,\gamma}(\tilde{\theta}) - J_{\psi,\phi_0,\gamma}(\tilde{\theta}) J_{\phi,\gamma}(\tilde{\theta}) J_{\phi,\psi,\gamma}(\tilde{\theta}) \right]^{-1} d_\psi^*(\tilde{\theta}) \xrightarrow{d} \chi^2_{k_\psi}(0).$$

(5.14)

4. Under $H_0 : \psi_0 = \psi_* + \xi/\sqrt{n}$ and irrespective of whether $H_0 : \phi_0 = \phi_*$ or $H_1 : \phi_0 = \phi_* + \delta/\sqrt{n}$ holds, we have

$$RS^*_\psi(P) = n d_\psi^*(\tilde{\theta}) \left[ J_{\psi,\gamma}(\tilde{\theta}) - J_{\psi,\phi_0,\gamma}(\tilde{\theta}) J_{\phi,\gamma}(\tilde{\theta}) J_{\phi,\psi,\gamma}(\tilde{\theta}) \right]^{-1} d_\psi^*(\tilde{\theta}) \xrightarrow{d} \chi^2_{k_\psi}(\lambda_6).$$

(5.15)

where $\lambda_6 = \lambda_6(\xi) = \xi' \left[ J_{\psi,\gamma}(\theta_0) - J_{\psi,\phi_0,\gamma}(\theta_0) J_{\phi,\gamma}(\theta_0) J_{\phi,\psi,\gamma}(\theta_0) \right] \xi$.

**Proof.** See Appendix A. □

Note that $\lambda_5 - \lambda_6 \geq 0$ in Proposition 3. The result in (5.14) is valid both in the presence or absence of the local misspecification $\phi_0 = \phi_* + \delta/\sqrt{n}$, since the asymptotic distribution of $RS^*_\psi(P)$ purges the effect of local departure of $\phi_0$ from $\phi_*$. Therefore, $RS^*_\psi(P)$ will be less powerful than $PS_\psi$ when there is no misspecification. The quantity $\lambda_5 - \lambda_6 = \xi' J_{\psi,\phi_0,\gamma}(\theta_0) J_{\phi,\gamma}(\theta_0) J_{\phi,\psi,\gamma}(\theta_0) \xi$ can be regarded as the "insurance premium" we pay for the validity of $RS^*_\psi(P)$ under local misspecification.

**Remark 1.** Note that the formulation of the adjusted score is based on $J_{\phi,\gamma}^{-1}(\theta_0) \sqrt{n} d_\phi(\tilde{\theta}) \xrightarrow{d} N[\delta, J_{\phi,\gamma}(\theta_0)]$ and $\sqrt{n} d_\phi(\tilde{\theta}) - J_{\psi,\phi_0,\gamma}(\tilde{\theta}) \delta \xrightarrow{d} N[0, J_{\psi,\gamma}(\theta_0)]$, which are based on the results obtained from the Taylor expansions. Bera and Yoon (1993) use a different approach to get an adjusted score. They note that the adjustment requires an estimate of $\delta = \sqrt{n}(\phi_0 - \phi_*)$, and suggest the one-step method-of-scoring method to get the estimate. Let $\tilde{\theta} = (\tilde{\gamma}, \psi_*, \phi_*)'$ be an initial consistent estimator. Then, the one-step method-of-scoring estimator $(\hat{\gamma}', \hat{\psi}', \hat{\phi}')'$ is defined as

$$\left( \begin{array}{c} \hat{\gamma} \\ \hat{\psi} \\ \hat{\phi} \end{array} \right) = \left( \begin{array}{c} \tilde{\gamma} \\ \psi_* \\ \phi_* \end{array} \right) + J_{\gamma}(\tilde{\theta}) J_{\gamma,\phi}(\tilde{\theta})^{-1} \left( \begin{array}{c} d_\gamma(\tilde{\theta}) \\ d_\psi(\tilde{\theta}) \\ d_\phi(\tilde{\theta}) \end{array} \right).$$

(5.16)

This updating in (5.16) can be viewed as an attempt to "correct" the initial estimators $\tilde{\gamma}$ and $\phi_*$ to take account of the local departure of $\phi_0$ from $\phi_*$. Using the fact that $d_\gamma(\tilde{\theta}) = 0$ in (5.16), we get

$$\tilde{\delta} = \sqrt{n}(\tilde{\theta} - \phi_*) = J_{\phi,\gamma}^{-1}(\tilde{\theta}) \sqrt{n} d_\phi(\tilde{\theta})$$

(5.17)

Bera and Yoon (1993) replace $\delta$ in $\sqrt{n} d_\phi(\tilde{\theta}) - J_{\psi,\phi_0,\gamma}(\tilde{\theta}) \delta \xrightarrow{d} N[0, J_{\psi,\gamma}(\theta_0)]$ by its estimate $\tilde{\delta}$ to get...
the adjusted score $\sqrt{n}d^*_\psi(\tilde{\theta})$. They, then, use some well known results to conclude that

$$\sqrt{n}d^*_\psi(\tilde{\theta}) \xrightarrow{d} N\left[0_{k_x \times 1}, \mathcal{J}_{\psi\gamma}(\theta_0) - \mathcal{J}_{\psi\phi}(\theta_0)\mathcal{J}_{\phi\gamma}(\theta_0)^{-1}\mathcal{J}_{\phi\psi}(\theta_0)\right]$$  \hspace{1cm} (5.18)

It should be pointed out that $RS^*_\psi(P)$ in (5.14) is based on $\tilde{\theta} = (\tilde{\gamma}', \tilde{\psi}', \phi_0)'$ circumventing direct estimation of the nuisance parameter $\phi_0$. As shown in Bera and Yoon (1993), $RS^*_\psi(P)$ is asymptotically equivalent to the optimal $C(\alpha)$ test of Neyman (1959). Unlike $RS^*_\psi(P)$, however, the $C(\alpha)$ test requires a $\sqrt{n}$-consistent estimate of the nuisance parameter explicitly.

To provide some more insight of the proposed test in (5.14), we observe that it is based on the so called "effective score" of parameter of interest $\psi_0$:

$\begin{align*}
&d_\psi(\theta_0) = \mathbb{E}(d_\psi(\theta_0)|d_\phi(\theta_0), d_\gamma(\theta_0)) = d_\psi(\theta_0) + C(\theta_0) \times d_\gamma(\theta_0) \\
&- (\mathcal{J}_{\psi\gamma}(\theta_0) - \mathcal{J}_{\psi\phi}(\theta_0)\mathcal{J}_{\phi\gamma}(\theta_0)^{-1}\mathcal{J}_{\phi\psi}(\theta_0)) \times (\mathcal{J}_{\phi\gamma}(\theta_0) - \mathcal{J}_{\phi\psi}(\theta_0)\mathcal{J}_{\psi\gamma}(\theta_0)^{-1}\mathcal{J}_{\psi\phi}(\theta_0))^{-1} d_\phi(\theta_0) \\
&\equiv d_\psi(\theta_0) - \mathcal{J}_{\psi\phi}(\theta_0)\mathcal{J}_{\phi\psi}(\theta_0)d_\phi(\theta_0) + C(\theta_0) \times d_\gamma(\theta_0),
\end{align*}$

(5.19)

where $C(\theta_0)$ depends on the various components of the information matrix $\mathcal{J}(\theta_0)$. Starting from the asymptotic multivariate normal distribution of the score vector, the above residual from regression (or projection) of $d_\psi(\theta_0)$ on $\left(d_\phi(\theta_0), d_\gamma(\theta_0)\right)'$ has an asymptotic normal distribution with mean zero. By construction, the above residual is the part of $d_\psi(\tilde{\theta})$ that is orthogonal to $\left(d_\phi(\theta_0), d_\gamma(\theta_0)\right)'$. The feasible version of the residual is given by $d_\psi(\tilde{\theta}) - J_{\psi\phi}(\tilde{\theta})J_{\phi\gamma}(\tilde{\theta})d_\phi(\tilde{\theta})$, where we use the fact that $d_\gamma(\tilde{\theta}) = 0$. Therefore, it can serve as a testing function for $H_0 : \psi_0 = \psi_*$.

Note also that $RS^*_\psi(P) = RS^\phi_\psi$ when $J_{\psi\phi\gamma}(\theta_0) = 0$. This implies that when the two score vectors $d_\psi$ and $d_\phi$ are asymptotically independent then we do not need any adjustment for testing $H_0 : \psi_0 = \psi_*$ due to the parametric misspecification involving $\phi_0$. Let $RS^*_{\psi\phi}$ be the test statistic for the joint hypothesis $H_0 : \psi_0 = \psi_*$ and $H_0 : \phi_0 = \phi_*$. When $J_{\psi\phi\gamma}(\theta_0) = 0$, the necessary and sufficient condition for the additivity of the RS test is satisfied, i.e., $RS^*_{\psi\phi} = RS^\phi_\psi + RS^\phi_\psi$ (Bera and McKenzie 1987). What is more interesting, however, is the general case where $J_{\psi\phi\gamma}(\theta_0) \neq 0$. We can show after some algebra (see the Appendix) that a modified additivity of the RS test still holds as shown in the following corollary.

**Corollary 1.** Consider $RS^*_{\psi\phi}$, $RS^*_\psi(P)$, $RS^*_\phi$, $RS^*_{\phi\psi}$ and $RS^*_{\psi}$. The following algebraic relationships hold among these test statistics.

$$RS^*_{\psi\phi} = RS^*_\psi(P) + RS^*_{\phi} = RS^*_{\phi}(P) + RS^*_{\psi}$$  \hspace{1cm} (5.20)

**Proof.** See Appendix A \hfill \square

In other words, the two-directional RS test for $\psi$ and $\phi$ can be decomposed into the sum of the uncorrected one-directional test for one type of alternative and the adjusted form for the other alternative. In some cases, one could ease computations considerably using this result, as illustrated in Anselin (2000) and Bera and Sosa-Escudero (2001).
6 Testing under Both Distributional and Parametric Misspecification

We now consider the presence of both distributional and parametric misspecifications. Consider the following partition of $K(\theta)$:

$$
K(\theta) = \begin{pmatrix}
K_\gamma(\theta) & K_\gamma \psi(\theta) & K_\gamma \varphi(\theta) \\
K_{\psi \gamma}(\theta) & K_\psi(\theta) & K_{\psi \varphi}(\theta) \\
K_{\varphi \gamma}(\theta) & K_{\varphi \psi}(\theta) & K_{\varphi \varphi}(\theta)
\end{pmatrix}.
$$

(6.1)

Also, we assume that the outer product matrix $K(\theta_0)$ is partitioned similarly. Again, let $\tilde{\theta} = (\tilde{\gamma}', \tilde{\psi}', \tilde{\varphi}')$ be the restricted MLE under $H_0 : \psi_0 = \psi_*$ and $H_0 : \varphi_0 = \varphi_*$. We are interested in the asymptotic distribution of $\sqrt{n}d_\psi(\tilde{\theta})$, $\sqrt{n}d_\varphi(\tilde{\theta})$ and $\sqrt{n}d_\gamma(\tilde{\theta})$ under the sequences of local alternatives $H_1 : \psi_0 = \psi_* + \xi/\sqrt{n}$ and $H_1 : \varphi_0 = \varphi_* + \delta/\sqrt{n}$. For this purpose, we will start from

$$
\sqrt{n}d_\psi(\tilde{\theta}) = \left( I_{k_\psi \times k_\psi}, -\mathcal{J}_{\psi \gamma}(\theta_0)\mathcal{J}_{\gamma}^{-1}(\theta_0) \right) \delta + \mathcal{J}_{\psi \gamma}(\theta_0)\xi + o_p(1),
$$

(6.2)

$$
\sqrt{n}d_\varphi(\tilde{\theta}) = \left( I_{k_\varphi \times k_\varphi}, -\mathcal{J}_{\varphi \gamma}(\theta_0)\mathcal{J}_{\gamma}^{-1}(\theta_0) \right) \delta + \mathcal{J}_{\varphi \gamma}(\theta_0)\xi + o_p(1),
$$

(6.3)

The asymptotic distribution of $\sqrt{n}d_\psi(\tilde{\theta})$ can be determined from (6.2) by using $\frac{1}{\sqrt{n}} \frac{\partial \ln L(\theta_0)}{\partial \theta} \xrightarrow{d} N[0, K(\theta_0)]$. Thus, it can be shown that

$$
\sqrt{n}d_\psi(\tilde{\theta}) \xrightarrow{d} N[\mathcal{J}_{\psi \gamma}(\theta_0)\xi + \mathcal{J}_{\psi \psi \gamma}(\theta_0)\delta, \mathcal{B}_{\psi \gamma}(\theta_0)]
$$

(6.4)

where

$$
\mathcal{B}_{\psi \gamma}(\theta_0) = \mathcal{K}_{\psi}(\theta_0) + \mathcal{J}_{\psi \gamma}(\theta_0)\mathcal{J}_{\gamma}^{-1}(\theta_0)\mathcal{K}_{\gamma}(\theta_0)\mathcal{J}_{\gamma}^{-1}(\theta_0)\mathcal{J}_{\psi \gamma}(\theta_0)
$$

(6.5)

$$
- \mathcal{K}_{\psi \gamma}(\theta_0)\mathcal{J}_{\gamma}^{-1}(\theta_0)\mathcal{J}_{\psi \gamma}(\theta_0) + \mathcal{J}_{\psi \gamma}(\theta_0)\mathcal{J}_{\gamma}^{-1}(\theta_0)\mathcal{K}_{\gamma}(\theta_0).
$$

Hence, under $H_0 : \psi_0 = \psi_*$ and $H_1 : \varphi_0 = \varphi_* + \delta/\sqrt{n}$, we have

$$
\sqrt{n}d_\psi(\tilde{\theta}) - \mathcal{J}_{\psi \psi \gamma}(\theta_0)\delta \xrightarrow{d} N[0_{k_\psi \times 1}, \mathcal{B}_{\psi \gamma}(\theta_0)]
$$

(6.6)
The result in (6.6) shows the distinct effect of each type of misspecification simultaneously on the asymptotic distribution of the score function. The parametric and distributional misspecifications change the mean and the variance of the score to \( \mathcal{J}_{\psi_\phi \gamma}(\theta_0) \) and \( \mathcal{B}_{\psi_\gamma}(\theta_0) \), respectively. If \( \delta = 0 \) and \( \mathcal{K}(\theta_0) = \mathcal{J}(\theta_0) \), then \( \sqrt{nd_\phi(\tilde{\theta})} \) will have zero mean and the familiar variance \( \mathcal{J}_{\psi_\gamma} \) as shown in Proposition 1. To obtain a modified form of the RS test that is valid under both types of misspecifications, the mean and the variance of the score need to be adjusted accordingly. For this purpose, we need to determine the asymptotic distribution of \( \sqrt{nd_\phi(\tilde{\theta})} \) from (6.3) by using
\[
\frac{1}{\sqrt{n}} \frac{\partial \ln L(\theta_0)}{\partial \theta} \xrightarrow{d} N \left[ 0, \mathcal{K}(\theta_0) \right].
\]
The result in (6.3) implies that
\[
\sqrt{nd_\phi(\tilde{\theta})} \xrightarrow{d} N \left[ \mathcal{J}_{\psi_\gamma}(\theta_0) \delta + \mathcal{J}_{\psi_\phi \gamma}(\theta_0) \xi, \mathcal{B}_{\psi_\gamma}(\theta_0) \right] \tag{6.7}
\]
where
\[
\mathcal{B}_{\psi_\gamma}(\theta_0) = \mathcal{K}_{\psi_\phi}(\theta_0) + \mathcal{J}_{\psi_\phi}(\theta_0) \mathcal{J}_{\gamma}^{-1}(\theta_0) \mathcal{K}_{\gamma}(\theta_0) \mathcal{J}_{\gamma}^{-1}(\theta_0) \mathcal{J}_{\psi_\gamma}(\theta_0)
- \mathcal{K}_{\psi_\phi}(\theta_0) \mathcal{J}_{\gamma}^{-1}(\theta_0) \mathcal{J}_{\psi_\gamma}(\theta_0) - \mathcal{J}_{\psi_\phi}(\theta_0) \mathcal{J}_{\gamma}^{-1}(\theta_0) \mathcal{K}_{\gamma}(\theta_0). \tag{6.8}
\]
Under \( H_0 : \psi_0 = \psi_* \) and \( H_1 : \phi_0 = \phi_0 + \delta / \sqrt{n} \), the result in (6.9) implies that
\[
\mathcal{J}_{\phi_\gamma}^{-1}(\theta_0) \sqrt{nd_\phi(\tilde{\theta})} \xrightarrow{d} N \left[ \delta, \mathcal{J}_{\psi_\gamma}^{-1}(\theta_0) \mathcal{B}_{\psi_\gamma}(\theta_0) \mathcal{J}_{\psi_\gamma}^{-1}(\theta_0) \right] \tag{6.9}
\]
We can now formulate an adjusted score that has mean zero under both parametric and distributional misspecification by using (6.6) and (6.9). Let \( d_\psi^\ast(\theta) \) be the adjusted score. Then, its feasible version is given by
\[
\sqrt{nd_\psi^\ast(\tilde{\theta})} = \sqrt{n} \left[ d_\psi(\tilde{\theta}) - \mathcal{J}_{\psi_\phi \gamma}(\theta_0) d_\phi(\tilde{\theta}) \right] \tag{6.10}
\]
Note that the adjusted score in (6.10) is exactly the same as the adjusted score we get in (5.10) when there is only parametric misspecification. That is, the distributional misspecification does not affect the adjusted score. But, it affects the asymptotic variance of the adjusted score. The asymptotic variance of \( \sqrt{nd_\psi^\ast(\tilde{\theta})} \) is given by
\[
\mathcal{D}_{\psi_\gamma}(\theta_0) = \mathcal{B}_{\psi_\gamma}(\theta_0) + \mathcal{J}_{\psi_\phi \gamma}(\theta_0) \mathcal{J}_{\gamma}^{-1}(\theta_0) \mathcal{B}_{\psi_\gamma}(\theta_0) \mathcal{J}_{\psi_\gamma}(\theta_0)
- \mathcal{J}_{\psi_\phi \gamma}(\theta_0) \mathcal{J}_{\gamma}^{-1}(\theta_0) \mathcal{J}_{\psi_\gamma}(\theta_0) \mathcal{B}_{\psi_\gamma}(\theta_0) - \mathcal{B}_{\psi_\phi \gamma}(\theta_0) \mathcal{J}_{\gamma}^{-1}(\theta_0) \mathcal{J}_{\psi_\gamma}(\theta_0) \tag{6.11}
\]
where
\[
\mathcal{B}_{\psi_\phi \gamma}(\theta_0) = \mathcal{K}_{\psi_\phi}(\theta_0) - \mathcal{J}_{\psi_\gamma}(\theta_0) \mathcal{J}_{\gamma}^{-1}(\theta_0) \mathcal{K}_{\gamma}(\theta_0) - \mathcal{K}_{\psi_\gamma}(\theta_0) \mathcal{J}_{\gamma}^{-1}(\theta_0) \mathcal{J}_{\gamma}(\theta_0)
+ \mathcal{J}_{\psi_\gamma}(\theta_0) \mathcal{J}_{\gamma}^{-1}(\theta_0) \mathcal{K}_{\gamma}(\theta_0) \mathcal{J}_{\gamma}(\theta_0), \tag{6.12}
\]
\(^3\)For details, see the proof of Proposition 4
and $B_{\phi \gamma}(\theta_0)$ is defined similarly. Now a quadratic form can be constructed based on (6.11) to yield an adjusted RS statistic that is valid under both distributional and parametric misspecifications. Let $RS^*_\psi(DP)$ be the resulting test statistic. In the following proposition, we provide the asymptotic distribution results for this statistic.

**Proposition 4.** Under our stated assumptions, the following results hold.

1. Under $H_0 : \psi_0 = \psi_*$ and irrespective of whether $H_0 : \phi_0 = \phi_*$ or $H_1 : \phi_0 = \phi_* + \delta/\sqrt{n}$ holds, we have

$$RS^*_\psi(DP) = n d_\psi^*(\hat{\theta}) D_{\psi,\gamma}^{-1}(\hat{\theta}) d_\psi^*(\hat{\theta}) \xrightarrow{d} \chi^2_{k,\psi}(0),$$

where

$$D_{\psi,\gamma}(\hat{\theta}) = B_{\psi,\gamma}(\hat{\theta}) + J_{\psi,\gamma}(\hat{\theta}) J_{\phi,\gamma}^{-1}(\hat{\theta}) B_{\phi,\gamma}(\hat{\theta}) J_{\phi,\gamma}^{-1}(\hat{\theta}) J_{\psi,\gamma}(\hat{\theta})$$

$$- J_{\psi,\gamma}(\hat{\theta}) J_{\phi,\gamma}^{-1}(\hat{\theta}) B_{\phi,\gamma}(\hat{\theta}) J_{\phi,\gamma}^{-1}(\hat{\theta}) J_{\psi,\gamma}(\hat{\theta}),$$

with

$$B_{\psi,\gamma}(\hat{\theta}) = K_{\psi}(\hat{\theta}) + J_{\psi,\gamma}(\hat{\theta}) J_{\gamma}^{-1}(\hat{\theta}) K_{\gamma}(\hat{\theta}) J_{\gamma}^{-1}(\hat{\theta}) - K_{\psi}(\hat{\theta}) J_{\gamma}^{-1}(\hat{\theta}) J_{\gamma}^{-1}(\hat{\theta}) K_{\gamma}(\hat{\theta}).$$

$$B_{\phi,\gamma}(\hat{\theta}) = K_{\phi,\gamma}(\hat{\theta}) - J_{\psi,\gamma}(\hat{\theta}) J_{\gamma}^{-1}(\hat{\theta}) K_{\phi}(\hat{\theta}) - K_{\psi,\gamma}(\hat{\theta}) J_{\gamma}^{-1}(\hat{\theta}) J_{\gamma}^{-1}(\hat{\theta}) J_{\gamma}^{-1}(\hat{\theta}) K_{\gamma}(\hat{\theta}).$$

2. Under $H_0 : \psi_0 = \psi_* + \xi/\sqrt{n}$ and irrespective of whether $H_0 : \phi_0 = \phi_*$ or $H_1 : \phi_0 = \phi_* + \delta/\sqrt{n}$ holds, we have

$$RS^*_\psi(DP) = n d_\psi^*(\hat{\theta}) D_{\psi,\gamma}^{-1}(\hat{\theta}) d_\psi^*(\hat{\theta}) \xrightarrow{d} \chi^2_{k,\psi}(\lambda_\gamma).$$

where

$$\lambda_\gamma = \lambda(\xi) = \xi \left( J_{\psi,\gamma}(\theta_0) - J_{\phi,\gamma}(\theta_0) J_{\phi,\gamma}^{-1}(\theta_0) J_{\psi,\gamma}(\theta_0) \right)' D_{\psi,\gamma}^{-1}(\theta_0) \times \left( J_{\psi,\gamma}(\theta_0) - J_{\phi,\gamma}(\theta_0) J_{\phi,\gamma}^{-1}(\theta_0) J_{\psi,\gamma}(\theta_0) \right) \xi.$$

**Proof.** See Appendix [A]

Although $RS^*_\psi(DP)$ has has rather lengthy algebraic expression as shown in (6.14), it is actually easy to compute and requires only $\tilde{\theta} = \left( \tilde{\gamma}', \psi_*', \phi_*' \right)'$. Note that when there is no local parametric or distributional misspecification, $RS^*_\psi(DP)$ can simply as shown in the following corollary.
Corollary 2. The $RS^*_\psi(DP)$ defined in (6.14) simplifies in the following ways.

1. When there is no local parametric misspecification, i.e., $\delta = 0$, we have

$$RS^*_\psi(DP) = RS^*_\psi(D)$$

(6.19)

2. When there is no distributional misspecification, i.e. $J(\theta_0) = K(\theta_0)$, we have $B_{\psi^*\gamma}(\tilde{\theta}) = J_{\psi^*\gamma}(\tilde{\theta})$, $B_{\psi^*\gamma}(\tilde{\theta}) = J_{\psi^*\gamma}(\tilde{\theta})$ and $B_{\psi^*\gamma}(\tilde{\theta}) = J_{\psi^*\gamma}(\tilde{\theta})$. Thus,

$$RS^*_\psi(DP) = RS^*_\psi(P)$$

(6.20)

3. Finally, when there is no misspecification, i.e., $\delta = 0$ and $J(\theta_0) = K(\theta_0)$, we have

$$RS^*_\psi(DP) = RS^*_\psi$$

(6.21)

Proof. See Appendix A.

$RS^*_\psi(DP)$, thus, adjusts the standard $RS$ test and $RS_\psi$ and provides a two-way protection against both types of misspecification we considered.

7 Illustrations

In this section, we provide two examples to illustrate the derivation of our test statistics. In the first example, we consider an ARCH(1) model, and introduce test statistics for the hypothesis about the constant term of the process. Our results show that the test statistics are sensitive to both types of misspecification. In the second illustration, we illustrate the derivation of test statistics within the context of an error component model to test the presence of random effects and the serial correlation of disturbance term. Our analytical results on test statistics indicate that distributional misspecification has no effect on the asymptotic distributions of test statistics.

7.1 An ARCH Model

In the first illustration, we consider an extension of an example due to Arnold (1980). Suppose we are interested in testing $H_0 : \psi_0 = \psi^* > 0$ in the following model representing an ARCH(1) process

$$y_t|y_{t-1} \sim I.I.D [0, h_t], \quad h_t = \psi_0 + \phi_0 y_{t-1}^2, \quad \text{for} \quad t = 1, \ldots, T.$$  (7.1)

Here all parameters are scalars and $\gamma_0$ is dropped from the parameter space for simplicity giving $\theta_0 = (\psi_0, \phi_0)'$. The scores and the information matrix evaluated at $\tilde{\theta} = (\psi^*, 0)'$ are given by

$$d_{\psi}(\tilde{\theta}) = \frac{1}{2T\psi^2} \sum_{t=2}^{T} (y_t^2 - \psi^*), \quad d_{\phi}(\tilde{\theta}) = \frac{1}{2T\psi^2} \sum_{i=2}^{T} (y_t^2 y_{t-1}^2 - \psi^* y_{t-1}^2),$$

(7.2)
\[ J(\tilde{\theta}) = \frac{1}{2\psi^2} \begin{pmatrix} 1 & \psi \star \\ \psi & \bar{\mu}_4 \end{pmatrix}, \quad K(\tilde{\theta}) = \frac{1}{4\psi^2} \begin{pmatrix} \bar{\mu}_4 - \psi^2 & \psi \bar{\mu}_4 - \psi^3 \\ \psi \bar{\mu}_4 - \psi^3 & \bar{\mu}_4^2 - \bar{\mu}_4 \psi^2 \end{pmatrix} \] (7.3)

where \( \bar{\mu}_4 = \frac{1}{T} \sum_{t=1}^{T} y_t^4 \), which is an estimate of \( \mu_4 = \mathbb{E}(y_t^4) \). Under the normality assumption, we have \( \mu_4 = 3\psi^2 \), hence the information matrix equivalence holds. Then, the standard RS statistic in (3.2), which completely ignores \( \phi \), and the modified RS test in (6.13) turns out to be the RS test. Finally, we consider the presence of both distributional and parametric misspecifications.

Comparing (7.8) with (7.4), we notice the adjustment in both the mean and variance of the standard RS statistic. The adjusted RS statistic under distributional misspecification is given by

\[ RS_\psi = \frac{1}{2T\psi^2} \left( \sum_{t=2}^{T} (y_t^2 - \psi) \right)^2 \] (7.4)

For the adjusted RS statistic under distributional misspecification, we need to calculate \( B_{\psi-\gamma}(\hat{\theta}) = K_{\psi}(\hat{\theta}) + J_{\psi-\gamma}(\hat{\theta})J_{\psi-\gamma}^{-1}(\hat{\theta})K_{\psi}(\hat{\theta}) - J_{\psi-\gamma}(\hat{\theta})J_{\psi-\gamma}^{-1}(\hat{\theta})K_{\psi}(\hat{\theta}) - K_{\psi-\gamma}(\hat{\theta})J_{\psi-\gamma}^{-1}(\hat{\theta})J_{\psi-\gamma}(\hat{\theta}) \). For this case, we simply have \( B_{\psi-\gamma}(\hat{\theta}) = K_{\psi}(\hat{\theta}) = (\bar{\mu}_4 - \psi^2)/4T\psi^4 \). Thus, the adjusted RS statistic in (4.6) is given by

\[ RS_\psi^D(D) = \frac{1}{T(\bar{\mu}_4 - \psi^2)} \left( \sum_{t=2}^{T} (y_t^2 - \psi) \right)^2 \] (7.5)

Note the variance adjustment in (7.5) for the departure from normality. Turning to parametric misspecification, we introduce the local presence of the parameter \( \phi_0 \) and retain the normality assumption. The adjusted score and its variance are given by

\[ d_\psi^*(\hat{\theta}) = \left[ d_\psi(\hat{\theta}) - J_{\psi-\gamma}(\hat{\theta})J_{\psi-\gamma}^{-1}(\hat{\theta})d_\psi(\hat{\theta}) \right] = \frac{1}{2T\psi^2} \left[ \sum_{t=1}^{T} (y_t^2 - \psi) - \frac{1}{3\psi} \sum_{t=1}^{T} (y_t^2 y_{t-1}^2 - \psi y_{t-1}^2) \right] \] (7.6)

\[ \text{Var} \left( \sqrt{T}d_\psi^*(\hat{\theta}) \right) = \left[ J_{\psi-\gamma}(\hat{\theta}) - J_{\psi-\gamma}(\hat{\theta})J_{\psi-\gamma}^{-1}(\hat{\theta})J_{\psi-\gamma}(\hat{\theta}) \right] = \frac{1}{3T\psi^2} \] (7.7)

Then, the adjusted RS statistic in (5.14) under parametric misspecification is given by

\[ RS_\psi^P(P) = \frac{3}{4T\psi^2} \left( \sum_{t=2}^{T} (y_t^2 - \psi) - \frac{1}{3\psi} \sum_{t=2}^{T} (y_t^2 y_{t-1}^2 - \psi y_{t-1}^2) \right)^2 \] (7.8)

Comparing (7.8) with (7.4), we notice the adjustment in both the mean and variance of the standard RS test. Finally, we consider the presence of both distributional and parametric misspecifications, and the modified RS test in (6.13) turns out to be

\[ RS_\psi^D(P) = \frac{1}{T(\bar{\mu}_4 - \psi^2)(1 - \psi^2)} \left[ \sum_{t=2}^{T} (y_t^2 - \psi) - \psi\bar{\mu}_4^{-1} \sum_{t=2}^{T} (y_t^2 y_{t-1}^2 - \psi y_{t-1}^2) \right]^2 \] (7.9)
Using the test statistics given in (7.4), (7.5), (7.8) and (7.9), we can readily verify the relationships between the adjusted tests and the standard test as given in Corollary 2.

7.2 An Error Component Model

In this second illustration, we introduce our adjusted tests to a one way error component model that combines random individual effects and first order autocorrelation in the disturbance term. This model is introduced by Lillard and Willis (1978) and can be stated as

\[ Y_{it} = X_{it}'\beta_0 + U_{it}, \quad U_{it} = \mu_i + V_{it} \]

\[ V_{it} = \rho_0 V_{i,t-1} + \varepsilon_{it}, \quad |\rho_0| < 1, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \quad (7.10) \]

where \( X_{it} \) is \( k \times 1 \) vector of exogenous variables with matching parameter vector \( \beta_0 \), \( \mu_i \) is the individual random effects, and \( \nu_{it} \) is the disturbance term that has a first order autoregressive process. We assume that \( \mu_i \sim \text{IID}(0, \sigma^2_\mu) \), \( \varepsilon_{it} \sim \text{IID}(0, \sigma^2_\varepsilon) \), and \( V_{i0} \sim \text{IID}(0, \sigma^2_\varepsilon/(1 - \rho^2_0)) \). Further, we assume that \( \mu_i \) and \( V_{it} \) are independent for all \( i \) and \( t \), and they are also independent with \( V_{i0} \). Following Bera et al. (2001), we set \( \theta_0 = (\gamma_0, \psi_0, \phi_0)' = (\sigma^2_\varepsilon, \sigma^2_\mu, \rho_0)' \), since the information matrix is block diagonal between \( \theta_0 \) and \( \beta_0 \). Thus, \( \tilde{\theta} = (\sigma^2_\varepsilon, \psi_s, \phi_s)' \), where \( \psi_s \) and \( \phi_s \) are the hypothesized values of \( \sigma^2_\epsilon \) and \( \rho \), respectively. We are interested in testing \( H_0 : \sigma^2_{0\mu} = 0 \) and \( H_0 : \rho_0 = 0 \), thus \( \psi_s = 0 \) and \( \phi_s = 0 \) in our case. Let \( \tilde{\beta} \) be the OLS estimator from the regression \( Y_{it} = X_{it}'\beta_0 + U_{it} \) and define \( \tilde{U} = (\tilde{U}_{11}, \ldots, \tilde{U}_{1T}, \ldots, \tilde{U}_{NT})' \) as the \( NT \times 1 \) vector of OLS residuals, where \( \tilde{U}_{it} = Y_{it} - X_{it}'\tilde{\beta} \) for \( i = 1, \ldots, N, \quad t = 1, \ldots, T \). The required scores for our test statistics are given in the following (Baltagi and Li 1991, Bera et al., 2001).4

\[
\begin{align*}
d_{\varepsilon}(\tilde{\theta}) &= \frac{\partial L}{\partial \sigma^2_\varepsilon} \bigg|_{\tilde{\theta}} = -\frac{NT}{2\tilde{\sigma}^2_\varepsilon} + \frac{\tilde{U}'\tilde{U}}{2\tilde{\sigma}^2_\varepsilon} \\
d_{\mu}(\tilde{\theta}) &= \frac{\partial L}{\partial \sigma^2_\mu} \bigg|_{\tilde{\theta}} = -\frac{NT}{2\tilde{\sigma}^2_\varepsilon} \left( \frac{\tilde{U}'(I_N \otimes I_T)\tilde{U}}{\tilde{U}'\tilde{U}} \right) \\
d_\rho(\tilde{\theta}) &= \frac{\partial L}{\partial \rho} \bigg|_{\tilde{\theta}} = NT \left( \frac{\tilde{U}'\tilde{U}_{-1}}{\tilde{U}'\tilde{U}} \right) \\
\end{align*}
\]  

(7.11)

where \( \tilde{\sigma}^2_\varepsilon = \tilde{U}'\tilde{U}/NT \), \( I_N \) is the \( N \times N \) identity matrix, \( I_T \) is the \( T \times 1 \) vector of ones, and \( \tilde{U}_{-1} = (0, \tilde{U}_{11}, \tilde{U}_{12}, \ldots, \tilde{U}_{1,T-1}, \ldots, \tilde{U}_{N1}, \tilde{U}_{N2}, \ldots, \tilde{U}_{NT-1})' \) is the \( NT \times 1 \) vector of residuals. Note that the scores with respect to \( \sigma^2_\mu \) and \( \sigma^2_\varepsilon \) are, respectively, denoted as \( d_{\mu} \) and \( d_{\varepsilon} \) in (7.11) to simplify notation. The covariance of scores and the information matrix are evaluated at \( \tilde{\theta} \), and they are

4For this model, we do not use \( 1/NT \) to scale the scores, the Fisher’s information matrix and the outer product matrix. Therefore, the test statistics should be stated without multiplying them by the sample size.
given as

\[
J(\hat{\theta}) = \begin{pmatrix}
J_{\varepsilon\varepsilon}(\hat{\theta}) & J_{\varepsilon\mu}(\hat{\theta}) & J_{\varepsilon\rho}(\hat{\theta}) \\
J_{\mu\varepsilon}(\hat{\theta}) & J_{\mu\mu}(\hat{\theta}) & J_{\mu\rho}(\hat{\theta}) \\
J_{\rho\varepsilon}(\hat{\theta}) & J_{\rho\mu}(\hat{\theta}) & J_{\rho\rho}(\hat{\theta})
\end{pmatrix} = \frac{NT}{2\sigma_2^2} \begin{pmatrix}
1 & 1 & 0 \\
1 & T & (T-1)\sigma_2^2 \\
0 & 2(T-1)\sigma_2^2 & 2(T-1)\sigma_4^2
\end{pmatrix}
\] (7.12)

\[
K(\hat{\theta}) = \begin{pmatrix}
K_{\varepsilon\varepsilon}(\hat{\theta}) & K_{\varepsilon\mu}(\hat{\theta}) & K_{\varepsilon\rho}(\hat{\theta}) \\
K_{\mu\varepsilon}(\hat{\theta}) & K_{\mu\mu}(\hat{\theta}) & K_{\mu\rho}(\hat{\theta}) \\
K_{\rho\varepsilon}(\hat{\theta}) & K_{\rho\mu}(\hat{\theta}) & K_{\rho\rho}(\hat{\theta})
\end{pmatrix} = \frac{NT}{2\sigma_4^2} \begin{pmatrix}
\hat{\mu}_4 - \bar{\sigma}_4^4 & \hat{\mu}_4 - \bar{\sigma}_4^4 & 0 \\
\hat{\mu}_4 - \bar{\sigma}_4^4 & \hat{\mu}_4 + (2T-3)\bar{\sigma}_4^2 & 2(T-1)\bar{\sigma}_4^2 \\
0 & 2(T-1)\bar{\sigma}_4^2 & 2(T-1)\bar{\sigma}_4^4
\end{pmatrix}
\] (7.13)

where \(\hat{\mu}_4\) is the estimate of the fourth moment of \(\varepsilon_{it}\), i.e. \(\mathbb{E}(\varepsilon_{it}^4)\), formulated with \(\hat{\theta}\). When the disturbance terms \(\varepsilon_{it}\)'s are normally distributed, we have \(\mu_4 = 3\sigma_4^2\), and thus, \(J(\theta) = K(\theta)\).

We start with the test statistics for \(H_0 : \sigma_0^2 = 0\). In terms of our notation, we have \(\psi = \sigma_\mu^2\), \(\phi = \rho\) and \(\gamma = \sigma_\varepsilon^2\). The simple one directional test statistic of Proposition 1 is given by

\[
RS_\mu = \frac{NT}{2(T-1)} \times A^2, \quad \text{where} \quad A = 1 - \frac{\tilde{U}'(I_N \otimes I_{T-1})\tilde{U}}{\tilde{U}'\tilde{U}}
\] (7.14)

We need to calculate \(B_{\psi,\gamma}(\hat{\theta})\) for the robust RS statistic in Proposition 2 under distributional misspecification. Simple calculations show that\(^5\)

\[
B_{\psi,\gamma}(\hat{\theta}) = \frac{NT(T-1)}{2\sigma_\varepsilon^4}.
\] (7.15)

Thus, we have

\[
RS_\mu^*(D) = \frac{NT}{2(T-1)} \times A^2
\] (7.16)

Note that \(RS_\mu^*(D)\) equals to \(RS_\mu\). This result indicates that the simple one directional test statistic derived under the normality assumption is already robust to distributional misspecification in the context of this model. That is, we have a case where the distributional misspecification, in the sense of White (1982), under the null hypothesis does not affect the variance of the score statistic.

Next, we determine the adjusted score for the test statistics in Propositions 3 and 4. The adjusted score is given as

\[
d_\psi(\hat{\theta}) = d_\psi(\hat{\theta}) - J_{\psi,\gamma}(\hat{\theta})d_\gamma(\hat{\theta}) = d_\mu(\hat{\theta}) - J_{\mu\varepsilon}(\hat{\theta})d_\varepsilon(\hat{\theta})
\]

\[
= \frac{(NT)^2}{2U'U} \left[ \left( \frac{\tilde{U}'(I_N \otimes J_T)\tilde{U}}{U'U} \right) + 2\frac{\tilde{U}'\tilde{U}_T^{-1}\tilde{U}'U}{U'U} \right] = \frac{(NT)^2}{2U'U} \times (A + 2B)
\] (7.17)

\(^5\)For details, see Appendix B
where $B = \tilde{U}' \tilde{U}^{-1} \tilde{U}'$. For the robust test in Proposition 3, we need to determine

$$J_{\psi, \gamma}(\tilde{\theta}) - J_{\psi, \gamma}(\tilde{\theta})J^{-1}_{\psi, \gamma}(\tilde{\theta})J_{\psi, \gamma}(\tilde{\theta}) = \left[ J_{\mu, \varepsilon}(\tilde{\theta}) - J_{\mu, \varepsilon}(\tilde{\theta})J^{-1}_{\mu, \varepsilon}(\tilde{\theta})J_{\mu, \varepsilon}(\tilde{\theta}) \right]$$

$$= \frac{N(T-1)(T-2)}{2\hat{\varepsilon}^2} = \frac{(NT)^2 N(T-1)(T-2)}{2(\tilde{U}'\tilde{U})^2}$$

(7.18)

Using (7.17) and (7.18), the test statistic of Proposition 3 is then, given by

$$RS^*_\mu(P) = \frac{NT}{2(T-1)(1-2/T)} \times (A + 2B)^2$$

(7.19)

The test statistic (7.28) is exactly the same as the test statistic derived in Bera et al. (2001, p. 7).

Next, we determine the asymptotic variance of the adjusted score under both distributional and parametric misspecification for the test statistic stated in Proposition 4. In Appendix B, we show that

$$D_{\psi, \gamma}(\tilde{\theta}) \equiv D_{\mu, \varepsilon}(\tilde{\theta}) = \frac{(NT)^2 N(T-1)(T-2)}{2(\tilde{U}'\tilde{U})^2},$$

(7.20)

which yields

$$RS^*_\mu(DP) \equiv RS^*_\mu(P) = \frac{NT}{2(T-1)(1-2/T)} \times (A + 2B)^2$$

(7.21)

We conclude that if we wrongly assume normality for the error distribution, the variance of the robust test statistic is not affected. More specifically, $RS^*_\mu(P)$ is also robust to the presence of non-normality in the context of this model.

Next, we consider the test statistics for $H_0 : \rho_0 = 0$. In term of our notation, we have $\psi = \rho$, $\phi = \sigma^2_\mu$ and $\gamma = \sigma^2_\varepsilon$. The simple one directional test statistic of Proposition 1 is now given by

$$RS_\rho = \frac{NT^2}{(T-1)} \times B^2,$$

where $B = \tilde{U}' \tilde{U}^{-1} \tilde{U}'$. (7.22)

To formulate the robust RS statistic in Proposition 2, we calculate $B_{\psi, \gamma} = B_{\rho, \varepsilon} = N(T-1)$, thus we have

$$RS^*_\rho(D) \equiv RS_\rho = \frac{NT^2}{(T-1)} \times B^2$$

(7.23)

\footnote{For details, see Appendix B}
The adjusted score is now given by
\[ d^*_\rho(\tilde{\theta}) = d_\rho(\tilde{\theta}) - J_{\rho \varepsilon}(\tilde{\theta}) J^{-1}_{\mu}(\tilde{\theta}) d_\mu(\tilde{\theta}) = NT \left[ \frac{\tilde{U}' \tilde{U}_{-1}}{\tilde{U}' \tilde{U}} + \frac{1}{T} \left( 1 - \frac{\tilde{U}' (J_N \otimes J_T) \tilde{U}}{\tilde{U}' \tilde{U}} \right) \right] \]
\equiv NT \left( B + \frac{1}{T} A \right) \quad \text{(7.24)}

The asymptotic variance of adjusted score under the parametric misspecification is
\[ \left[ J_{\rho \varepsilon}(\tilde{\theta}) - J_{\rho \mu}(\tilde{\theta}) J^{-1}_{\mu \varepsilon}(\tilde{\theta}) J_{\mu \rho \varepsilon}(\tilde{\theta}) \right] = N (T - 1) \left( 1 - \frac{2}{T} \right) \quad \text{(7.25)} \]

Then, the robust test statistic in Proposition 3 can be stated as
\[ RS^*_\rho(P) = \frac{NT^2 (B + \frac{A}{T})^2}{(T - 1) (1 - \frac{2}{T})} \quad \text{(7.26)} \]

Next, we determine the asymptotic variance of the adjusted score under both distributional and parametric misspecification. In Appendix B, we show that
\[ D_{\rho \varepsilon}(\tilde{\theta}) = N (T - 1) \left( 1 - \frac{2}{T} \right) \quad \text{(7.27)} \]

which implies that
\[ RS^*_\rho(DP) \equiv RS^*_\rho(P) = \frac{NT^2 (B + \frac{A}{T})^2}{(T - 1) (1 - \frac{2}{T})} \quad \text{(7.28)} \]

8 Monte Carlo Simulations

In this section, we conduct some Monte Carlo simulations to investigate the finite sample properties of our proposed tests in the presence of distributional and local parametric misspecifications for the one way error model considered in Section 7.2. Following Bera et al. (2001), we specify the data generating process as
\[ Y_{it} = \beta_{10} + X_{it} \beta_{20} + U_{it}, \quad U_{it} = \mu_i + V_{it} \]
\[ V_{it} = \rho_0 V_{i,t-1} + \varepsilon_{it}, \quad |\rho_0| < 1, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \quad \text{(8.1)} \]

and consider the following test statistics
1. \( RS_\mu \equiv RS^*_\mu(D) \)
2. \( RS^*_\mu(P) \equiv RS^*_\mu(DP) \)
3. \( RS_\rho \equiv RS^*_\rho(D) \)
4. \( RS^*_\rho(P) \equiv RS^*_\rho(DP) \)
5. \( RS_{\mu \rho} \)
6. \( RSO_\mu \)
7. \( RS_{0\mu} \)

The last two test statistics
are one-sided tests suggested first by Honda (1985) and they are defined as

\[
R_{SO\mu} = -\sqrt{\frac{NT}{2(T-1)} \times A},
\]

\[
R_{SO\mu}^* = -\sqrt{\frac{NT}{2(T-1)(1-2/T)} \times (A + 2B)}.
\]

Under the null hypothesis \( H_0 : \sigma^2_\mu = 0 \), \( R_{SO\mu} \) is asymptotically distributed as \( N(0,1) \) when \( H_0 : \rho = 0 \) holds. But, in the presence of serial correlation, this test has an asymptotic standard normal distribution with a non-zero mean of \( \rho \sqrt{2(T-1)/T^2} \) (Bera et al. 2001). The null asymptotic distribution of \( R_{SO\mu}^* \) is a standard normal distribution irrespective of whether \( H_0 : \rho = 0 \) holds or not. The joint test statistic \( R_{SO\mu}^\rho \) for \( H_0 : \sigma^2_\mu = \rho = 0 \) is formulated by Baltagi and Li (1991, 1995), and is defined as

\[
R_{SO\mu}^\rho = \frac{NT^2}{2(T-1)(T-2)} \left[ A^2 + 4AB + 2TB^2 \right]
\]

which has an asymptotic null distribution of \( \chi^2_2 \).

In the data generating process, we set \( \beta_0 = (\beta_{10}, \beta_{20})' = (0.5, 5)' \) and the exogenous variable is generated according to \( X_{it} \sim \text{IID Uniform}[0, 1] \). The individual random effects are generated according to \( \mu_i \sim \text{IID N}(0, 20\sigma) \), where \( \tau \) takes values from \{0, 0.05, 0.10, 0.20, 0.40, 0.60\}. Similarly, we allow \( \rho \) to take values from \{0, 0.05, 0.10, 0.20, 0.40, 0.60\}. We use the following combination of \( N \) and \( T \) for the sample size: \( (N, T) = \{(25, 10), (25, 20), (50, 10)\} \). In a comprehensive Monte Carlo study, Bera et al. (2001) investigate the finite sample size and power properties of these seven tests when there is no distributional misspecification in the data generating process. Here, we want to confirm our analytical results that distributional misspecification has no effect on the performance of test statistics for this model, therefore we specify \( \varepsilon_{it} \sim \text{IID } \chi^2_{\nu} - \nu \), where \( \nu = \{2, 4, 10\} \). Let \( \kappa = \sigma^2_\mu / (\sigma^2_\mu + \sigma^2_\nu) \), where \( \sigma^2_\nu = \sigma^2_{\varepsilon} / (1-\rho^2) \). Our specification allows \( \kappa \) to vary from zero to 0.75. Furthermore, we let \( V_{i0} \sim \text{IID } \chi^2_{\nu} / (1-\rho^2) - \nu / (1-\rho^2) \).

Our design yields 324 combinations and resampling is done 1000 times for each combination. When the nominal size of a test is \( \alpha \), the simulation variability can be assessed by a normal approximation to the binomial distribution. Thus, a 95% confidence interval centered on \( \alpha \) would be given by \( \alpha \pm 1.96(\alpha(1-\alpha)/1000)^{1/2} \). Then, when \( \alpha = 0.05 \), the 95% point-wise interval is \( (0.036, 0.064) \).

Note again that to calculate the seven test statistics under consideration only the residuals from a least squares estimation are needed. Below we will summarize the finite sample size and power properties of the tets statistics. The results presented in tables and figures are based on the nominal size of 0.05. Table 3 presents the empirical sizes of the proposed test statistics. The salient features of size properties are summarized in the following list.

1. Generally, the empirical sizes of all tests lie inside the 95% confidence interval.
Table 1: Empirical Size

<table>
<thead>
<tr>
<th>ν</th>
<th>(N, T)</th>
<th>RS_μ</th>
<th>RS_μ⁺</th>
<th>RS_ρ</th>
<th>RS_ρ⁺</th>
<th>RSO_μ</th>
<th>RSO_μ⁺</th>
<th>RS_μρ</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>(25,10)</td>
<td>0.038</td>
<td>0.034</td>
<td>0.050</td>
<td>0.050</td>
<td>0.038</td>
<td>0.042</td>
<td>0.036</td>
</tr>
<tr>
<td>2</td>
<td>(25,20)</td>
<td>0.040</td>
<td>0.046</td>
<td>0.044</td>
<td>0.041</td>
<td>0.055</td>
<td>0.056</td>
<td>0.040</td>
</tr>
<tr>
<td>2</td>
<td>(50,10)</td>
<td>0.044</td>
<td>0.041</td>
<td>0.051</td>
<td>0.047</td>
<td>0.035</td>
<td>0.042</td>
<td>0.046</td>
</tr>
<tr>
<td>4</td>
<td>(25,10)</td>
<td>0.042</td>
<td>0.043</td>
<td>0.053</td>
<td>0.048</td>
<td>0.051</td>
<td>0.050</td>
<td>0.045</td>
</tr>
<tr>
<td>4</td>
<td>(25,20)</td>
<td>0.036</td>
<td>0.030</td>
<td>0.041</td>
<td>0.044</td>
<td>0.046</td>
<td>0.046</td>
<td>0.037</td>
</tr>
<tr>
<td>4</td>
<td>(50,10)</td>
<td>0.050</td>
<td>0.048</td>
<td>0.052</td>
<td>0.054</td>
<td>0.054</td>
<td>0.047</td>
<td>0.042</td>
</tr>
<tr>
<td>10</td>
<td>(25,10)</td>
<td>0.041</td>
<td>0.038</td>
<td>0.047</td>
<td>0.045</td>
<td>0.049</td>
<td>0.037</td>
<td>0.040</td>
</tr>
<tr>
<td>10</td>
<td>(25,20)</td>
<td>0.041</td>
<td>0.039</td>
<td>0.035</td>
<td>0.034</td>
<td>0.051</td>
<td>0.043</td>
<td>0.033</td>
</tr>
<tr>
<td>10</td>
<td>(50,10)</td>
<td>0.049</td>
<td>0.056</td>
<td>0.057</td>
<td>0.055</td>
<td>0.054</td>
<td>0.055</td>
<td>0.051</td>
</tr>
</tbody>
</table>

2. $RS_μ⁺$ is generally more undersized compared to $RS_μ$. As the variation of $\varepsilon_{it}$ increases, the empirical sizes of both tests improve. Also, the empirical sizes of $RS_μ$ and $RS_μ⁺$ improve as $N$ gets larger relative to $T$.

3. $RS_ρ$ and $RS_ρ⁺$ have similar empirical sizes. As $T$ gets larger relative to $N$, the empirical sizes of $RS_ρ$ and $RS_ρ⁺$ decrease.

4. The one-sided tests $RSO_μ$ and $RSO_μ⁺$ are generally more over sized compared to their two-sided counterparts.

We also present the empirical size properties of $RSO_μ$, $RSO_μ⁺$, $RSO_ρ$ and $RSO_ρ⁺$ in Q–Q plots for the case of $\nu = 10$, $N = 25$ and $T = 10$ in Figure 1. To save space figures for the other eight combinations of $\nu$ and $(N,T)$ are not included. We also do not present the figures for the joint and one-sided tests, since they resemble those reported for the other tests.

1. The empirical distributions of the test statistics diverge from that of the $\chi^2_1$ at the right tail.

2. For $RS_μ$ the points are below the $45^\circ$ line about up to 99th percentile leading to the lower size of the test than 0.05. It then changes to above the $45^\circ$ line beyond the 99th percentile leading to higher size of the test than 0.01.

3. For $RS_μ⁺$ there is a similar pattern. However, beyond the 99th percentile the points are much closer to the $45^\circ$ line compared to those for $RS_μ$. Therefore, its size is closer to 0.01 beyond the 99th percentile as opposed to $RS_μ$.

4. For $RS_ρ$ and $RS_ρ⁺$, there is a similar degree of departure. But, the departures are in the opposite direction leading to lower sizes of the tests.

Next we will analyse the power properties of the test statistics. To save space we only present the results for the case of $\nu = 10$, $N = 25$ and $T = 10$ in Table 2 and Figures 2 and 3. We will start with the tests for the presence of random effects.
Table 2: Estimated rejection probabilities of different tests. Sample size: \( \nu = 10, N = 25 \) and \( T = 10 \)

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>( \rho )</th>
<th>( RS_{\mu} )</th>
<th>( RS_{\mu}^* )</th>
<th>( RS_{\rho} )</th>
<th>( RS_{\rho}^* )</th>
<th>( RSO_\mu )</th>
<th>( RSO_\mu^* )</th>
<th>( RS_{\mu\rho} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>0.041</td>
<td>0.038</td>
<td>0.047</td>
<td>0.045</td>
<td>0.049</td>
<td>0.037</td>
<td>0.040</td>
</tr>
<tr>
<td>0.00</td>
<td>0.05</td>
<td>0.062</td>
<td>0.057</td>
<td>0.093</td>
<td>0.093</td>
<td>0.084</td>
<td>0.049</td>
<td>0.089</td>
</tr>
<tr>
<td>0.00</td>
<td>0.10</td>
<td>0.127</td>
<td>0.073</td>
<td>0.286</td>
<td>0.242</td>
<td>0.186</td>
<td>0.071</td>
<td>0.243</td>
</tr>
<tr>
<td>0.00</td>
<td>0.20</td>
<td>0.326</td>
<td>0.138</td>
<td>0.833</td>
<td>0.706</td>
<td>0.396</td>
<td>0.125</td>
<td>0.764</td>
</tr>
<tr>
<td>0.00</td>
<td>0.40</td>
<td>0.860</td>
<td>0.382</td>
<td>1.000</td>
<td>0.999</td>
<td>0.905</td>
<td>0.418</td>
<td>1.000</td>
</tr>
<tr>
<td>0.00</td>
<td>0.60</td>
<td>0.995</td>
<td>0.859</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.884</td>
<td>1.000</td>
</tr>
<tr>
<td>0.05</td>
<td>0.00</td>
<td>0.331</td>
<td>0.283</td>
<td>0.130</td>
<td>0.051</td>
<td>0.424</td>
<td>0.376</td>
<td>0.280</td>
</tr>
<tr>
<td>0.05</td>
<td>0.05</td>
<td>0.380</td>
<td>0.282</td>
<td>0.280</td>
<td>0.077</td>
<td>0.483</td>
<td>0.353</td>
<td>0.349</td>
</tr>
<tr>
<td>0.05</td>
<td>0.10</td>
<td>0.531</td>
<td>0.317</td>
<td>0.525</td>
<td>0.223</td>
<td>0.618</td>
<td>0.410</td>
<td>0.569</td>
</tr>
<tr>
<td>0.05</td>
<td>0.20</td>
<td>0.703</td>
<td>0.389</td>
<td>0.940</td>
<td>0.693</td>
<td>0.767</td>
<td>0.461</td>
<td>0.907</td>
</tr>
<tr>
<td>0.05</td>
<td>0.40</td>
<td>0.934</td>
<td>0.583</td>
<td>1.000</td>
<td>0.994</td>
<td>0.954</td>
<td>0.640</td>
<td>1.000</td>
</tr>
<tr>
<td>0.05</td>
<td>0.60</td>
<td>0.997</td>
<td>0.891</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
<td>0.912</td>
<td>1.000</td>
</tr>
<tr>
<td>0.10</td>
<td>0.00</td>
<td>0.681</td>
<td>0.619</td>
<td>0.265</td>
<td>0.040</td>
<td>0.769</td>
<td>0.702</td>
<td>0.613</td>
</tr>
<tr>
<td>0.10</td>
<td>0.05</td>
<td>0.760</td>
<td>0.627</td>
<td>0.501</td>
<td>0.080</td>
<td>0.818</td>
<td>0.715</td>
<td>0.709</td>
</tr>
<tr>
<td>0.10</td>
<td>0.10</td>
<td>0.769</td>
<td>0.591</td>
<td>0.689</td>
<td>0.194</td>
<td>0.826</td>
<td>0.669</td>
<td>0.770</td>
</tr>
<tr>
<td>0.10</td>
<td>0.20</td>
<td>0.887</td>
<td>0.639</td>
<td>0.977</td>
<td>0.674</td>
<td>0.924</td>
<td>0.702</td>
<td>0.974</td>
</tr>
<tr>
<td>0.10</td>
<td>0.40</td>
<td>0.975</td>
<td>0.727</td>
<td>1.000</td>
<td>0.993</td>
<td>0.985</td>
<td>0.758</td>
<td>1.000</td>
</tr>
<tr>
<td>0.10</td>
<td>0.60</td>
<td>0.998</td>
<td>0.937</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.948</td>
<td>1.000</td>
</tr>
<tr>
<td>0.20</td>
<td>0.00</td>
<td>0.942</td>
<td>0.926</td>
<td>0.632</td>
<td>0.030</td>
<td>0.964</td>
<td>0.953</td>
<td>0.916</td>
</tr>
<tr>
<td>0.20</td>
<td>0.05</td>
<td>0.968</td>
<td>0.921</td>
<td>0.787</td>
<td>0.054</td>
<td>0.983</td>
<td>0.948</td>
<td>0.947</td>
</tr>
<tr>
<td>0.20</td>
<td>0.10</td>
<td>0.979</td>
<td>0.929</td>
<td>0.916</td>
<td>0.146</td>
<td>0.986</td>
<td>0.949</td>
<td>0.977</td>
</tr>
<tr>
<td>0.20</td>
<td>0.20</td>
<td>0.989</td>
<td>0.924</td>
<td>0.997</td>
<td>0.559</td>
<td>0.992</td>
<td>0.944</td>
<td>0.997</td>
</tr>
<tr>
<td>0.20</td>
<td>0.40</td>
<td>0.998</td>
<td>0.917</td>
<td>1.000</td>
<td>0.988</td>
<td>0.999</td>
<td>0.942</td>
<td>1.000</td>
</tr>
<tr>
<td>0.20</td>
<td>0.60</td>
<td>1.000</td>
<td>0.971</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.975</td>
<td>1.000</td>
</tr>
<tr>
<td>0.40</td>
<td>0.00</td>
<td>0.998</td>
<td>0.999</td>
<td>0.938</td>
<td>0.010</td>
<td>0.999</td>
<td>0.999</td>
<td>0.998</td>
</tr>
<tr>
<td>0.40</td>
<td>0.05</td>
<td>1.000</td>
<td>1.000</td>
<td>0.979</td>
<td>0.032</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.40</td>
<td>0.10</td>
<td>0.999</td>
<td>0.995</td>
<td>0.990</td>
<td>0.093</td>
<td>0.999</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>0.40</td>
<td>0.20</td>
<td>1.000</td>
<td>0.998</td>
<td>1.000</td>
<td>0.460</td>
<td>1.000</td>
<td>0.999</td>
<td>1.000</td>
</tr>
<tr>
<td>0.40</td>
<td>0.40</td>
<td>1.000</td>
<td>0.994</td>
<td>1.000</td>
<td>0.979</td>
<td>1.000</td>
<td>0.995</td>
<td>1.000</td>
</tr>
<tr>
<td>0.40</td>
<td>0.60</td>
<td>1.000</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.997</td>
<td>1.000</td>
</tr>
<tr>
<td>0.60</td>
<td>0.00</td>
<td>1.000</td>
<td>1.000</td>
<td>0.993</td>
<td>0.006</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.60</td>
<td>0.05</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
<td>0.024</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.60</td>
<td>0.10</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
<td>0.063</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.60</td>
<td>0.20</td>
<td>1.000</td>
<td>0.999</td>
<td>1.000</td>
<td>0.298</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.60</td>
<td>0.40</td>
<td>1.000</td>
<td>0.998</td>
<td>1.000</td>
<td>0.939</td>
<td>1.000</td>
<td>0.998</td>
<td>1.000</td>
</tr>
<tr>
<td>0.60</td>
<td>0.60</td>
<td>1.000</td>
<td>0.999</td>
<td>1.000</td>
<td>0.966</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>
1. The marginal tests $RS_\mu$ and $RSO_\mu$ are the optimal tests when $\rho = 0$. They have more power compared to their robust counterparts as expected. For example, when $\tau = 0.1$ and $\rho = 0$, the rejection rate of $RS_\mu$ is 0.681, whereas $RS_\mu^*$ has 0.619. This can also be confirmed from Figure 2(a). However, the discrepancy in power between $RS_\mu$ and $RS_\mu^*$ (under no parametric misspecification) disappears as the value of $\tau$ increases. When $\tau = 0.4$ both tests have power equal to about 1.

2. In case of a local parametric misspecification in $\rho$, we see that $RS_\mu^*$ and $RSO_\mu^*$ perform well. For example, when $\tau = 0$ and $\rho = 0.1$ the rejection rates of $RS_\mu^*$ and $RSO_\mu^*$ are 0.073 and
As the value of $\rho$ increases, the false rejection probabilities of $RS_{\mu}$ and $RSO_{\mu}$ increase.

3. As expected, the higher the magnitude of parametric misspecification in $\rho$ the performance of the adjustment methodology worsens. Figure 2(b) confirms this finding. However, the false rejection rate of $RS^*_\rho$ is still significantly lower than that of $RS_{\mu}$ even for larger values of $\rho$.

4. In Table 2 and Figures 2(c) and 2(d) we see that for a given $\tau > 0$ the rejection probabilities of $RS_{\mu}$ and $RS^*_\mu$ increase as the value of $\rho$ increases. As the value of $\tau$ increases, the discrepancy in rejection probabilities as a function of the value of $\rho$ decreases and power converges to 1.

Next, we summarize the findings on the the tests for the presence of serial correlation.

1. The marginal test $RS_{\rho}$ is the optimal tests when $\tau = 0$. It has more power compared to its adjusted counterpart as expected. For example, when $\rho = 0.2$ and $\tau = 0$, the rejection rates of $RS_{\rho}$ is 0.833 , whereas $RS^*_\rho$ has 0.726. This can also be confirmed from Figure 3(a). However, the discrepancy in power between $RS_{\rho}$ and $RS^*_\rho$ (under no parametric misspecification) disappears as the value of $\rho$ increases. When $\rho = 0.4$ both tests have power equal to about 1.

2. In case of a local parametric misspecification in $\tau$, we see that $RS^*_\rho$ performs well. For example, when $\rho = 0$ and $\tau = 0.05$ the rejection rate of $RS^*_\rho$ is 0.051, whereas $RS_{\rho}$ has 0.130. As the value of $\tau$ increases, the false rejection probabilities of $RS_{\rho}$ increase.

3. Surprisingly, the magnitude of parametric misspecification in $\tau$ does not seem to worsen the performance of the adjustment methodology significantly. Figure 3(b) confirms this finding. However, we see that for very large values of $\tau$ the false rejection rate of $RS^*_\rho$ becomes lower than the level of the test.

4. Table 2 and Figures 3(c) and 3(d) we see that for a given $\rho > 0$ the rejection probability of $RS_{\rho}$ increases as the value of $\tau$ increases. On the other hand, the rejection probability of $RS^*_\rho$ decreases as the value of $\tau$ increases.

5. As the value of $\rho$ increases, for $RS_{\rho}$ the discrepancy in rejection probabilities as a function of the value of $\tau$ decreases and power converges to 1. For $RS^*_\rho$, the highest discrepancy in rejection probabilities as a function of the value of $\tau$ happens when the value of $\rho$ is about 0.2.

6. Comparing Figures 3(c) and 3(d) we observe that the power of $RS_{\rho}$ is strongly affected by the presence of random effects, while there is virtually no significant effect on the power of $RS^*_\rho$.

Finally, we discuss briefly the performance of the joint statistic $RS_{\mu\rho}$.

1. Note that the joint test statistic is optimal when $\sigma^2_\mu > 0$ and $\rho \neq 0$. As we can see from Table 2 in this case $RS_{\mu\rho}$ generally has the highest power.
2. When the misspecification is one-directional, say $\sigma^2 > 0$ and $\rho = 0$, $RS_{\mu \rho}$ and $RS_\mu$ will have the same non-centrality parameter. Since $RS_{\mu \rho}$ and $RS_\mu$ has the $\chi^2_2$ and $\chi^2_1$ distributions, respectively, there will be a loss of power in using $RS_{\mu \rho}$. For example, when $\tau = 0.10$ and $\rho = 0$, the powers for $RS_\mu$ and $RS_{\mu \rho}$ are 0.681 and 0.613, respectively.

3. Although $RS_{\mu \rho}$ has overall good power, it cannot help to identify the exact source of misspecification when there is only a one-directional misspecification.

Overall, the patterns revealed for size and power properties in our simulations are highly similar.
to those obtained in Baltagi and Li (2001) for a simulation study based on a DGP with no distributional misspecification. These observations confirm our analytical results that the asymptotic distributions of test statistics are not affected in the presence of distributional misspecification.

9 Conclusions and Discussions

In this section, we summarize the main points of paper and discuss some specifications in which the adjusted tests can be useful for the detection of exact source(s) of misspecification. We have
generalized the adjusted Rao’s score tests suggested in White (1982) and Bera and Yoon (1993) to take account of the simultaneous presence of distributional and local parametric misspecifications. Although White’s adjustment has not yet been much used in the econometric literature, we believe this procedure has high potential.

Kent (1982) shows that the standard Rao’s score statistic is asymptotically equivalent to a linear combination of central chi-square distributed random variables in the presence of distributional misspecification. This approach can provide a unified framework that can be useful to summarize our results. To set our framework, we begin by giving a standard definition from linear algebra. Let $M$ be an $n \times n$ matrix. Then, the spectral decomposition of $A$ is defined by

$$M = \sum_{j=1}^{m} \eta_j E_j$$

where $\eta_1, \ldots, \eta_m$ are distinct eigenvalues of $M$, and $E_j$’s are non-negative $n \times n$ matrices satisfying $E_jE_i = 0$, for $j \neq i$, and $E_j^2 = E_j$ for $j = 1, \ldots, m$. Further, let $S$ and $V$ be two $n \times n$ symmetric real matrices. If $V$ is positive definite then the product $SV$ has a spectral decomposition (Baldessari 1967). The following lemma would be useful to show the asymptotic results of quadratic forms under various scenarios.

**Lemma 1.** Let $Y$ be a $n \times 1$ random vector and assume that $Y \sim N(\mu, \Omega)$, where $\Omega$ is a non-singular matrix. Consider the quadratic form $Q = Y' \Sigma Y$, where $\Sigma$ is a symmetric non-singular $n \times n$ matrix. Consider the spectral decomposition $\Omega \Sigma = \sum_{j=1}^{p} \eta_j E_j$, where $\eta_1, \ldots, \eta_p$ are distinct non-zero real eigenvalues of $\Omega \Sigma$, and $E_j$’s are real non-negative definite matrices satisfying $E_j^2 = E_j$ and $E_jE_i = 0$ for $j \neq i$. Then

$$Q \sim \sum_{i=1}^{p} \eta_i V_i$$

where $V_i$’s are independent random variables such that $V_i \sim \chi^2_{r_i}(\delta_i)$, $r_i$ is the algebraic multiplicity of $\eta_i$, and $\delta_i = \mu' \Omega^{-1} E_i \mu$ for $i = 1, \ldots, p$.

**Proof.** See Appendix A

**Remark 2.** In the special case where $B = A^{-1}$ and $\mu = 0$, we have $AB = I_m$ and $\delta_i = 0$ for each $i$. Then, we have the standard result on the quadratic form, $Q \sim \chi^2_m$.

We use Lemma 1 to summarize the asymptotic null distributions of tests statistics under misspecification. We start with the standard test statistic stated in Proposition 1

$$RS_\psi = n d_\psi'(\tilde{\theta})J_{\psi, \gamma}^{-1}(\tilde{\theta})d_\psi(\tilde{\theta})$$

With respect to this statistics, we have the following conclusion.

1. (i) The asymptotic null distribution of $RS_\psi$ is a central chi-square distribution when there is no misspecification.

(ii) Consider the following spectral decomposition, $B_{\psi, \gamma}(\theta_0)J_{\psi, \gamma}^{-1}(\theta_0) = \sum_{j=1}^{p} \mu_j(\theta_0)E_j(\theta_0)$, where $\mu_1(\theta_0), \ldots, \mu_p(\theta_0)$ are distinct eigenvalues of $B_{\psi, \gamma}(\theta_0)J_{\psi, \gamma}^{-1}(\theta_0)$, and $E_j(\theta_0)$’s are non-negative $k_\psi \times k_\psi$ matrices satisfying $E_j(\theta_0)E_i(\theta_0) = 0$, for $j \neq i$, and $E_j^2(\theta_0) = E_j(\theta_0)$ for
Our stated results for $RS_{\psi}$ are as follows.

1. The asymptotic null distribution of $RS_{\psi}$ is a central chi-square distribution when there is no distributional misspecification. Under the null hypothesis and parametric misspecification, our results indicate that

$$RS_{\psi} \sim \chi^2_{\tilde{k}_{\psi}} (\lambda_3(\theta_0))$$

where $\lambda_3(\theta_0) = \delta' \mathcal{J}_{\psi,\gamma}(\theta_0) \mathcal{J}_{\psi,\gamma}^{-1}(\theta_0) \mathcal{J}_{\psi,\gamma}(\theta_0) \delta$.

2. The asymptotic null distribution of $RS_{\psi}(D)$ is a central chi-square distribution when there is no parametric misspecification. Under the null hypothesis and parametric misspecification, our results indicate that

$$RS_{\psi}(D) \sim \chi^2_{\tilde{k}_{\psi}} (\lambda_8(\theta_0))$$

where and $\lambda_8(\theta_0) = \delta' \mathcal{J}_{\psi,\gamma}(\theta_0) \mathcal{B}_{\psi,\gamma}^{-1}(\theta_0) \mathcal{J}_{\psi,\gamma}(\theta_0) \delta$.

Our stated results for $RS_{\psi}(P)$ can be summarized in the following.

3. The asymptotic null distribution of $RS_{\psi}(P)$ is a central chi-square distribution when there is no distributional misspecification. Consider the following spectral decomposition, $\mathcal{B}_{\psi,\gamma}(\theta_0) \left[ \mathcal{J}_{\psi,\gamma}(\theta_0) - \mathcal{J}_{\psi,\gamma}(\theta_0) \mathcal{J}_{\psi,\gamma}^{-1}(\theta_0) \mathcal{J}_{\psi,\gamma}(\theta_0) \right]^{-1} = \sum_{j=1}^{g} \rho_j(\theta_0) \mathcal{O}_j(\theta_0)$, where $\rho_1(\theta_0), \ldots, \rho_g(\theta_0)$ are distinct eigenvalues of $\mathcal{B}_{\psi,\gamma}(\theta_0) \left[ \mathcal{J}_{\psi,\gamma}(\theta_0) - \mathcal{J}_{\psi,\gamma}(\theta_0) \mathcal{J}_{\psi,\gamma}^{-1}(\theta_0) \mathcal{J}_{\psi,\gamma}(\theta_0) \right]^{-1}$, and $\mathcal{O}_j(\theta_0)$'s are non-negative $k_{\psi} \times k_{\psi}$ matrices satisfying $\mathcal{O}_j(\theta_0) \mathcal{O}_i(\theta_0) = 0$, for $j \neq i$, and $\mathcal{O}_j^2(\theta_0) = \mathcal{O}_j(\theta_0)$ for $j = 1, \ldots, p$. Then, under the null hypothesis and distributional misspecification, we have

$$RS_{\psi}(P) \sim \sum_{i=1}^{p} \rho_i Z_i$$

where $Z_i$'s are independent random variables satisfying $Z_i \sim \chi^2_{\tilde{b}_i}$ and $b_i$ is the algebraic multiplicity of $\rho_i(\theta_0)$ for $i = 1, \ldots, g$. 

\[ j = 1, \ldots, p. \]
Finally, we have the following conclusion for $RS_{\psi}(DP)$.

4. The adjusted statistic $RS_{\psi}(DP)$ is robust to both types of misspecification and its null asymptotic distribution is a central chi-square distribution.

The modified score test for local parametric misspecification has already found its way to many useful applications. For example, among others, see Anselin et al. (1996), Baltagi and Li (2001), Bera and Bilias (2001), and Baltagi et al. (2002). Therefore, the modified test proposed here could be successfully applied to various testing problems in econometrics. The illustrations we provided here highlights the inter-relationship among various adjustments needed for different kinds of misspecification. Recently, models that simultaneously encompass various effects are suggested in the literature. Our adjusted test statistic can be formulated for these models to reveal the source(s) of misspecification. Recently, models that simultaneously encompass various effects are suggested in the literature. Our adjusted test statistic can be formulated for these models to reveal the source(s) of misspecification. Recently, models that simultaneously encompass various effects are suggested in the literature. Our adjusted test statistic can be formulated for these models to reveal the source(s) of misspecification.

Another application can be based on the generalized specification suggested by Lee and Yu (2012b) for static spatial panel data model with fixed or random effects. This specification nests our error component model in (7.10), the spatial panel data models considered in Anselin (1988), Kapoor et al. (2007), Anselin et al. (2008), Baltagi et al. (2007) and Baltagi et al. (2013). The adjusted statistic $RS_{\psi}(DP)$ is robust to both types of misspecification and its null asymptotic distribution is a central chi-square distribution.

The modified score test for local parametric misspecification has already found its way to many useful applications. For example, among others, see Anselin et al. (1996), Baltagi and Li (2001), Bera and Bilias (2001), and Baltagi et al. (2002). Therefore, the modified test proposed here could be successfully applied to various testing problems in econometrics. The illustrations we provided here highlights the inter-relationship among various adjustments needed for different kinds of misspecification. Recently, models that simultaneously encompass various effects are suggested in the literature. Our adjusted test statistic can be formulated for these models to reveal the source(s) of misspecification. Recently, models that simultaneously encompass various effects are suggested in the literature. Our adjusted test statistic can be formulated for these models to reveal the source(s) of misspecification. Recently, models that simultaneously encompass various effects are suggested in the literature. Our adjusted test statistic can be formulated for these models to reveal the source(s) of misspecification.
individual effects $\mu$ are considered as fixed parameters, then the spatial structure defined through $\mu = \lambda_{30} W_3 \mu + (I_n + \delta_{30} M_3)c$ would be irrelevant. If these effects are random, then the spatial effects in $\mu$ can be considered as permanent spillover effects (Baltagi et al. 2013). Under the joint null hypothesis $H_0 : \lambda_{10} = \lambda_{30} = \delta_{20} = \delta_{30} = \rho_0 = 0$, the model again reduces to a conventional one-way model, which can be estimated by an OLS type estimator. Our adjusted test, which only requires this OLS estimator, can be formulated to test the presence of various effects to reveal the correct specification in applied studies.

In both (9.1) and (9.2), our adjusted tests would be robust in the sense that they would have central chi-square distributions in the presence of both types of misspecification. These tests requires the estimation of the corresponding model in its simplest form, and thus the practitioners can easily conduct their specification search without going through any complex estimation.
Appendix

A  Main Proofs

Proof of Proposition 2 These results are well-known in the literature. For example, see Davidson and MacKinnon (1987) and Saikkonen (1989).

Proof of Proposition 3 Based on the conventional Taylor series expansions and the asymptotic normality of scores, we showed that

\[
\sqrt{n}d_\psi(\tilde{\theta}) \xrightarrow{d} N \left[ J_{\psi,\gamma}(\theta_0)\xi, B_{\psi,\gamma}(\theta_0) \right] \tag{A.1}
\]

where \( B_{\psi,\gamma}(\theta_0) = K_\psi(\theta_0) + J_{\psi,\gamma}(\theta_0)J_{\gamma,\gamma}^{-1}(\theta_0)K_\gamma(\theta_0)J_{\gamma,\gamma}^{-1}(\theta_0)J_{\gamma,\gamma}(\theta_0) - J_{\psi,\gamma}(\theta_0)J_{\gamma,\gamma}^{-1}(\theta_0)K_\gamma(\theta_0) - K_{\psi,\gamma}(\theta_0)J_{\gamma,\gamma}^{-1}(\theta_0)J_{\gamma,\gamma}(\theta_0) \). The \( RS_\psi \) in \((3.2)\) is not weighted with \( B_{\psi,\gamma}(\theta_0) \) and hence it will not has an asymptotic chi-square distribution. The last two results directly follows from \((A.1)\).

Proof of Proposition 3 The first two results directly follows from the following asymptotic distribution result:

\[
\sqrt{n}d_\psi(\tilde{\theta}) \xrightarrow{d} N \left[ J_{\psi,\gamma}(\theta_0)\xi + J_{\psi,\phi}(\theta_0)\delta, J_{\psi,\gamma}(\theta_0) \right], \tag{A.2}
\]

where \( J_{\psi,\phi}(\theta) = J_{\psi,\gamma}(\theta) - J_{\psi,\gamma}(\theta)J_{\gamma,\gamma}^{-1}(\theta)J_{\gamma,\phi}(\theta) \) and \( J_{\phi,\psi,\gamma}(\theta) = J_{\phi,\psi}(\theta) - J_{\phi,\gamma}(\theta)J_{\gamma,\gamma}^{-1}(\theta)J_{\gamma,\gamma}(\theta) \).

In the rest of this proof, we will show the last two results. For this purpose, we will consider the following equations as a system:

\[
\sqrt{n}d_\psi(\tilde{\theta}) = \left( I_{k_\psi \times k_\psi} - J_{\psi,\gamma}(\theta_0)J_{\gamma,\gamma}^{-1}(\theta_0) \right) \left( \sqrt{n}d_\psi(\theta_0) \right) + J_{\psi,\gamma}(\theta_0)\xi + J_{\psi,\phi}(\theta_0)\delta + o_p(1), \tag{A.3}
\]

\[
\sqrt{n}d_\phi(\tilde{\theta}) = \left( I_{k_\phi \times k_\phi} - J_{\phi,\gamma}(\theta_0)J_{\gamma,\gamma}^{-1}(\theta_0) \right) \left( \sqrt{n}d_\phi(\theta_0) \right) + J_{\phi,\gamma}(\theta_0)\delta + J_{\phi,\psi,\gamma}(\theta_0)\xi + o_p(1), \tag{A.4}
\]

The combined system is given by

\[
\begin{pmatrix}
\sqrt{n}d_\psi(\tilde{\theta}) \\
\sqrt{n}d_\phi(\tilde{\theta})
\end{pmatrix} = \begin{pmatrix}
I_{k_\psi \times k_\psi} & 0_{k_\psi \times k_\phi} \\
0_{k_\phi \times k_\psi} & I_{k_\phi \times k_\phi}
\end{pmatrix} \begin{pmatrix}
\sqrt{n}d_\psi(\theta_0) \\
\sqrt{n}d_\phi(\theta_0)
\end{pmatrix} + \begin{pmatrix}
J_{\psi,\gamma}(\theta_0)\xi + J_{\psi,\phi}(\theta_0)\delta \\
J_{\phi,\gamma}(\theta_0)\delta + J_{\phi,\psi,\gamma}(\theta_0)\xi
\end{pmatrix} + o_p(1) \tag{A.5}
\]
The joint asymptotic distribution of $\sqrt{n}d_{\psi}(\tilde{\theta})$ and $\sqrt{n}d_{\phi}(\tilde{\theta})$ can now be determined from (A.5) by using $\frac{1}{\sqrt{n}} \frac{\partial \ln L(\theta_0)}{\partial \theta} \xrightarrow{d} N[0, K(\theta_0)]$. Thus, we have

$$
\begin{pmatrix}
\sqrt{n}d_{\psi}(\tilde{\theta}) \\
\sqrt{n}d_{\phi}(\tilde{\theta})
\end{pmatrix} \xrightarrow{d} N\left[
\begin{pmatrix}
J_{\psi,\gamma}(\theta_0) + J_{\psi,\phi}(\theta_0)\delta \\
J_{\phi,\gamma}(\theta_0) + J_{\phi,\phi}(\theta_0)\xi
\end{pmatrix},
\begin{pmatrix}
J_{\psi,\gamma}(\theta_0) & J_{\psi,\phi}(\theta_0) \\
J_{\phi,\gamma}(\theta_0) & J_{\phi,\phi}(\theta_0)
\end{pmatrix}
\right]
$$

(A.6)

Note that the adjusted score can be written as

$$
\sqrt{n}d_{\psi}^*(\tilde{\theta}) = \left(I_{k_{\psi} \times k_{\psi}} - J_{\psi,\phi}(\theta_0)J_{\phi,\phi}^{-1}(\theta_0)\right) \left(\sqrt{n}d_{\psi}(\tilde{\theta})/\sqrt{n}d_{\phi}(\tilde{\theta})\right) + o_p(1)
$$

(A.7)

Using (A.7), under $H_0 : \psi_0 = \psi_*$ and $H_1 : \phi_0 = \phi_* + \delta/\sqrt{n}$, we have

$$
\sqrt{n}d_{\psi}^*(\tilde{\theta}) \xrightarrow{d} N\left[0_{k_{\psi} \times 1}, J_{\psi,\gamma}(\theta_0) - J_{\psi,\phi}(\theta_0)J_{\phi,\phi}^{-1}(\theta_0)J_{\phi,\phi}(\theta_0)\right]
$$

(A.8)

Hence, $RS_{\psi}^*(P) \xrightarrow{d} \chi^2_{k_{\psi}}(0)$, which proves the third part of proposition. Using (A.6) and (A.7), under $H_0 : \psi_0 = \psi_* + \xi/\sqrt{n}$ and $H_0 : \phi_0 = \phi_*$, we have

$$
\sqrt{n}d_{\psi}^*(\tilde{\theta}) \xrightarrow{d} N\left[\left(J_{\psi,\gamma}(\theta_0) - J_{\psi,\phi}(\theta_0)J_{\phi,\phi}^{-1}(\theta_0)J_{\phi,\phi}(\theta_0)\right)\xi, J_{\psi,\gamma}(\theta_0) - J_{\psi,\phi}(\theta_0)J_{\phi,\phi}^{-1}(\theta_0)J_{\phi,\phi}(\theta_0)\right]
$$

(A.9)

Thus, we have

$$
RS_{\psi}^*(DP) \xrightarrow{d} \chi^2_{k_{\psi}}(\lambda_7).
$$

(A.10)

where

$$
\lambda_7 \equiv \lambda_7(\xi) = \xi' \left(J_{\psi,\gamma}(\theta_0) - J_{\psi,\phi}(\theta_0)J_{\phi,\phi}^{-1}(\theta_0)J_{\phi,\phi}(\theta_0)\right)' D_{\psi,\gamma}^{-1}(\tilde{\theta}) \times \left(J_{\psi,\gamma}(\theta_0) - J_{\psi,\phi}(\theta_0)J_{\phi,\phi}^{-1}(\theta_0)J_{\phi,\phi}(\theta_0)\right) \xi.
$$

(A.11)
Similarly, it can be shown that

\[
J(\theta) = \begin{pmatrix}
J_\gamma(\theta) & J_\psi(\theta) & J_\phi(\theta) \\
J_\gamma(\theta) & J_\psi(\theta) & J_\phi(\theta) \\
J_\gamma(\theta) & J_\psi(\theta) & J_\phi(\theta)
\end{pmatrix}^{-1}
\]

\[
\equiv \begin{pmatrix}
J_{11}(\theta) & J_{12}(\theta) \\
J_{21}(\theta) & J_{22}(\theta)
\end{pmatrix}
\]

\[
\equiv \begin{pmatrix}
J_{11}(\theta) & J_{12}(\theta) \\
J_{21}(\theta) & J_{22}(\theta)
\end{pmatrix}
\]

\[
\equiv \begin{pmatrix}
J_{11}(\theta) & J_{12}(\theta) \\
J_{21}(\theta) & J_{22}(\theta)
\end{pmatrix}
\]

Proof of Corollary 1. Consider the following partition

\[
J(\theta) = \begin{pmatrix}
J_\gamma(\theta) & J_\psi(\theta) & J_\phi(\theta) \\
J_\gamma(\theta) & J_\psi(\theta) & J_\phi(\theta) \\
J_\gamma(\theta) & J_\psi(\theta) & J_\phi(\theta)
\end{pmatrix}^{-1}
\]

\[
\equiv \begin{pmatrix}
J_{11}(\theta) & J_{12}(\theta) \\
J_{21}(\theta) & J_{22}(\theta)
\end{pmatrix}
\]

Using the results on the inversion of partition matrices, we get \(J_{22}(\theta) = (J_{22}(\theta) - J_{21}(\theta)J_{11}^{-1}(\theta)J_{12}(\theta))^{-1}\), which can be further written as

\[
J_{22}(\theta) = \begin{pmatrix}
J_{\psi,\gamma}(\theta) & J_{\psi,\phi}(\theta) \\
J_{\phi,\gamma}(\theta) & J_{\phi,\phi}(\theta)
\end{pmatrix}^{-1}
\]

\[
= \begin{pmatrix}
J_{\psi,\gamma}(\theta) & J_{\psi,\phi}(\theta) \\
J_{\phi,\gamma}(\theta) & J_{\phi,\phi}(\theta)
\end{pmatrix}^{-1}
\]

\[
\equiv \begin{pmatrix}
J_{11}(\theta) & J_{12}(\theta) \\
J_{21}(\theta) & J_{22}(\theta)
\end{pmatrix}
\]

where \(J_{\psi,\phi}(\theta) = J_{\psi,\gamma}(\theta) - J_{\phi,\gamma}(\theta)J_{\phi,\gamma}^{-1}(\theta)J_{\phi,\phi}(\theta)\). The \(RS_{\psi,\phi}\) statistic for the joint null hypothesis, \(H_0 : \psi = \psi_*\) and \(\phi = \phi_*\), is given by

\[
RS_{\psi,\phi} = n \begin{pmatrix}
d_\psi(\tilde{\theta}) \\
d_\phi(\tilde{\theta})
\end{pmatrix} J_{22}(\tilde{\theta}) \begin{pmatrix}
d_\psi(\tilde{\theta}) \\
d_\phi(\tilde{\theta})
\end{pmatrix}
\]

Using (A.12), the \(RS_{\psi,\phi}\) statistic in (A.14) can be written as

\[
RS_{\psi,\phi} = n \left( d_\psi(\tilde{\theta}) - J_{\psi,\gamma}(\tilde{\theta})J_{\phi,\gamma}^{-1}(\theta)d_\phi(\tilde{\theta}) \right) + n d_\phi(\tilde{\theta})J_{\phi,\gamma}^{-1}(\theta)d_\phi(\tilde{\theta})
\]

\[
= RS_{\psi,\phi}^*(P) + RS_{\phi}
\]

Similarly, it can be shown that \(RS_{\psi,\phi} = RS_{\phi}^*(P) + RS_{\psi}\).

Proof of Proposition 4. The proof requires the asymptotic distribution of the adjusted score

\[
\sqrt{n}d_{\psi,\phi}^*(\tilde{\theta}) = \sqrt{n} \left[ d_\psi(\tilde{\theta}) - J_{\psi,\gamma}(\tilde{\theta})J_{\phi,\gamma}^{-1}(\theta)d_\phi(\tilde{\theta}) \right].
\]

For this purpose, we will consider the combined
The joint asymptotic distribution of \( \sqrt{n}d_{\psi}(\hat{\theta}) \) and \( \sqrt{n}d_{\phi}(\hat{\theta}) \) can now be determined from (A.16) by using \( \frac{1}{\sqrt{n}} \frac{\partial \ln L(\theta_0)}{\partial \theta} \rightarrow N \left[ 0, \mathcal{K}(\theta_0) \right] \). Thus, we have

\[
\begin{pmatrix}
\sqrt{n}d_{\psi}(\hat{\theta}) \\
\sqrt{n}d_{\phi}(\hat{\theta})
\end{pmatrix}
\xrightarrow{d}
N \left[ \begin{pmatrix}
J_{\psi,\gamma}(\theta_0) \xi + J_{\psi,\phi}(\theta_0) \delta \\
J_{\phi,\gamma}(\theta_0) \delta + J_{\phi,\psi}(\theta_0) \xi
\end{pmatrix}, \begin{pmatrix}
B_{\psi,\gamma}(\theta_0) & B_{\psi,\phi}(\theta_0) \\
B_{\phi,\psi}(\theta_0) & B_{\phi,\gamma}(\theta_0)
\end{pmatrix} \right]
\]

(A.17)

where

\[
B_{\psi,\gamma}(\theta_0) = \mathcal{K}_{\psi}(\theta_0) + J_{\psi,\gamma}(\theta_0)J_{\gamma,\gamma}(\theta_0) - \mathcal{K}_{\psi,\gamma}(\theta_0)J_{\gamma,\gamma}(\theta_0) - \mathcal{K}_{\psi,\gamma}(\theta_0)J_{\gamma,\gamma}(\theta_0)
\]

(A.18)

\[
B_{\psi,\phi}(\theta_0) = \mathcal{K}_{\psi}(\theta_0) - J_{\psi,\gamma}(\theta_0)J_{\gamma,\gamma}(\theta_0) - \mathcal{K}_{\psi,\gamma}(\theta_0)J_{\gamma,\gamma}(\theta_0) + J_{\psi,\gamma}(\theta_0)J_{\gamma,\gamma}(\theta_0)J_{\gamma,\gamma}(\theta_0)
\]

(A.19)

\[
B_{\phi,\gamma}(\theta_0) \text{ and } B_{\phi,\psi}(\theta_0) \text{ are defined similarly. Under our assumptions, we have}
\]

\[
\sqrt{n}d_{\psi}^*(\tilde{\theta}) = \left( I_{k_{\psi} \times k_{\psi}} - J_{\psi,\phi,\gamma}(\theta_0)J_{\phi,\gamma}(\theta_0) \right) \left( \sqrt{n}d_{\psi}(\hat{\theta}) \right) + o_p(1)
\]

(A.20)

Using the asymptotic normality of score functions, under \( H_0 : \psi_0 = \psi_* \) and \( H_1 : \phi_0 = \phi_* + \delta/\sqrt{n} \), we have

\[
\sqrt{n}d_{\psi}^*(\tilde{\theta}) = N \left[ 0_{k_{\psi} \times 1}, D_{\psi,\gamma}(\theta_0) \right],
\]

(A.21)

where

\[
D_{\psi,\gamma}(\theta_0) = B_{\psi,\gamma}(\theta_0) + J_{\psi,\phi,\gamma}(\theta_0)J_{\phi,\gamma}(\theta_0)B_{\phi,\gamma}(\theta_0)J_{\phi,\gamma}(\theta_0) - J_{\psi,\phi,\gamma}(\theta_0)J_{\phi,\gamma}(\theta_0)B_{\phi,\psi}(\theta_0) - B_{\psi,\gamma}(\theta_0)J_{\phi,\gamma}(\theta_0)J_{\phi,\psi}(\theta_0)
\]

(A.22)

The first result of proposition directly follows from (A.21). Now, consider the asymptotic distribution of \( \sqrt{n}d_{\psi}^*(\tilde{\theta}) \) under \( H_1 : \psi_0 = \psi_* + \xi/\sqrt{n} \) and \( H_0 : \phi_0 = \phi_* \). Using (A.17) and (A.20), we
have
\[
\sqrt{n}d^*_\psi(\tilde{\theta}) = N \left[ \left( J_{\psi^{-1}}(\theta_0) - J_{\psi^{-1}}(\theta_0)J_{\phi^{-1}}(\theta_0)J_{\phi^{-1}}(\theta_0) \right) \xi, D_{\psi^{-1}}(\theta_0) \right],
\] (A.23)
which implies the last result of the proposition.

Proof of Corollary 2. These results on the simplification of \( RS^*_\psi(DP) \) are obvious, and therefore their proofs are omitted.

Proof of Lemma 1. This lemma is a special version of theorems stated in Baldessari (1967) and Tan (1977), therefore its proof is omitted.

B Asymptotic Variance of Adjusted Scores For Error Component Model

In this section, we provide details on the calculation of \( D_{\psi^{-1}} \). First, we consider \( H_0 : \sigma^2 = 0 \) so that \( \psi = \sigma^2 \) and \( \phi = \rho \). All terms in this section are evaluated at \( \tilde{\theta} \). For this hypothesis, we need to determine the following items.

\[
D_{\psi^{-1}}(\tilde{\theta}) = B_{\psi^{-1}}(\tilde{\theta}) + J_{\psi^{-1}}(\tilde{\theta})J_{\phi^{-1}}(\tilde{\theta})B_{\phi^{-1}}(\tilde{\theta})J_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta}) - J_{\psi^{-1}}(\tilde{\theta})J_{\phi^{-1}}(\tilde{\theta})B_{\phi^{-1}}(\tilde{\theta})
\]

- \( B_{\psi^{-1}}(\tilde{\theta}) = K_{\psi^{-1}}(\tilde{\theta}) + J_{\psi^{-1}}(\tilde{\theta})J_{\phi^{-1}}(\tilde{\theta})K_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta}) - J_{\psi^{-1}}(\tilde{\theta})J_{\phi^{-1}}(\tilde{\theta})K_{\phi^{-1}}(\tilde{\theta}) \) (B.1)

where

\[
B_{\psi^{-1}}(\tilde{\theta}) = K_{\psi^{-1}}(\tilde{\theta}) + J_{\psi^{-1}}(\tilde{\theta})J_{\phi^{-1}}(\tilde{\theta})K_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta}) - J_{\psi^{-1}}(\tilde{\theta})J_{\phi^{-1}}(\tilde{\theta})K_{\phi^{-1}}(\tilde{\theta}) \)

- \( B_{\phi^{-1}}(\tilde{\theta}) = K_{\phi^{-1}}(\tilde{\theta}) - J_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta})K_{\psi^{-1}}(\tilde{\theta}) - K_{\psi^{-1}}(\tilde{\theta})J_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta}) \) (B.2)

\[
B_{\phi^{-1}}(\tilde{\theta}) = K_{\phi^{-1}}(\tilde{\theta}) - J_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta})K_{\psi^{-1}}(\tilde{\theta}) - K_{\psi^{-1}}(\tilde{\theta})J_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta}) + J_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta})K_{\psi^{-1}}(\tilde{\theta}) \) (B.3)

and

\[
B_{\phi^{-1}}(\tilde{\theta}) = K_{\phi^{-1}}(\tilde{\theta}) + J_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta})K_{\psi^{-1}}(\tilde{\theta})J_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta}) - J_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta})K_{\psi^{-1}}(\tilde{\theta}) \)

- \( B_{\phi^{-1}}(\tilde{\theta}) = K_{\phi^{-1}}(\tilde{\theta}) + J_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta})K_{\psi^{-1}}(\tilde{\theta})J_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta}) - J_{\phi^{-1}}(\tilde{\theta})J_{\psi^{-1}}(\tilde{\theta})K_{\psi^{-1}}(\tilde{\theta}) \) (B.4)
With our notation, the variance of the score under only distributional misspecification is $B_{\psi, \gamma} = B_{\mu, \epsilon}$, which is given by

$$B_{\mu, \epsilon}(\hat{\theta}) = K_{\mu}(\hat{\theta}) + J_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})K_{\epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})J_{\epsilon, \mu}(\hat{\theta}) - J_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})K_{\epsilon, \mu}(\hat{\theta})$$

$$- K_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})J_{\epsilon, \mu}(\hat{\theta}) = \frac{NT(T - 1)}{2\sigma_{\epsilon}^4} \tag{B.5}$$

Next, let us compute the other two quantities $B_{\psi, \gamma} = B_{\mu, \epsilon}$ and $B_{\phi, \gamma} = B_{\mu, \epsilon}$:

$$B_{\mu, \epsilon}(\hat{\theta}) = K_{\mu}(\hat{\theta}) + J_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})K_{\epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})J_{\epsilon, \mu}(\hat{\theta})$$

$$+ J_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})K_{\epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})J_{\epsilon, \mu}(\hat{\theta})$$

$$= \frac{N(T - 1)}{\sigma_{\epsilon}^2} - \left( \frac{NT}{2\sigma_{\epsilon}^4} \right)^{-1} - 0 - K_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta}) \cdot 0 + J_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})K_{\epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta}) \cdot 0$$

$$= \frac{N(T - 1)}{\sigma_{\epsilon}^2} = J_{\mu, \epsilon}(\hat{\theta}) = J_{\mu, \epsilon}(\hat{\theta}), \tag{B.6}$$

$$B_{\mu, \epsilon}(\hat{\theta}) = K_{\mu}(\hat{\theta}) + J_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})K_{\epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})J_{\epsilon, \mu}(\hat{\theta})$$

$$- J_{\mu, \epsilon}(\hat{\theta})K_{\epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})J_{\epsilon, \mu}(\hat{\theta}) - K_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})J_{\epsilon, \mu}(\hat{\theta})$$

$$= N(T - 1) + 0 - J_{\epsilon}^{-1}(\hat{\theta})K_{\epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta}) \cdot 0 - 0 - J_{\epsilon}^{-1}(\hat{\theta}) \cdot 0 - 0 - J_{\epsilon}^{-1}(\hat{\theta}) \cdot 0$$

$$= N(T - 1) = J_{\mu, \epsilon}(\hat{\theta}). \tag{B.7}$$

In the following, we use that $B_{\mu, \epsilon}(\hat{\theta}) = J_{\mu, \epsilon}(\hat{\theta}) = N(T - 1)$ and $J_{\mu, \epsilon}(\hat{\theta}) = J'_{\mu, \epsilon}(\hat{\theta})$ and $B_{\mu, \epsilon}(\hat{\theta}) = B'_{\mu, \epsilon}(\hat{\theta})$. Hence,

$$D_{\psi, \gamma}(\hat{\theta}) = D_{\mu, \epsilon}(\hat{\theta}) = B_{\mu, \epsilon}(\hat{\theta}) + J_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})B_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})J_{\epsilon, \mu}(\hat{\theta}) - J_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})B_{\mu, \epsilon}(\hat{\theta})$$

$$- B_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})J_{\epsilon, \mu}(\hat{\theta})$$

$$= \frac{NT(T - 1)}{2\sigma_{\epsilon}^2} + \frac{N(T - 1)}{\sigma_{\epsilon}^2} \cdot \frac{NT(T - 1)}{N(T - 1)} - 2 \frac{N(T - 1)}{\sigma_{\epsilon}^2} \cdot \frac{1}{N(T - 1)} - \frac{N(T - 1)}{\sigma_{\epsilon}^2}$$

$$= \frac{N(T - 1)(T - 2)}{2\sigma_{\epsilon}^2} \tag{B.8}$$

Next, we consider the test statistics for $H_0 : \rho = 0$. In term of our notation, we have $\psi = \rho$, $\phi = \sigma_{\mu}^2$ and $\gamma = \sigma_{\epsilon}^2$. The variance of original (not-adjusted) Rao’s score under only distributional misspecification is:

$$B_{\psi, \gamma}(\hat{\theta}) = B_{\mu, \epsilon}(\hat{\theta}) = K_{\mu}(\hat{\theta}) + J_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})K_{\epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})J_{\epsilon, \mu}(\hat{\theta}) - J_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})K_{\epsilon, \mu}(\hat{\theta})$$

$$- K_{\mu, \epsilon}(\hat{\theta})J_{\epsilon}^{-1}(\hat{\theta})J_{\epsilon, \mu}(\hat{\theta}) = N(T - 1). \tag{B.9}$$
The asymptotic variance of the adjusted score under only parametric misspecification is given by

\[
\left[ J_{\rho \varepsilon}(\hat{\theta}) - J_{\rho \mu \varepsilon}(\hat{\theta})J_{\mu \varepsilon}^{-1}(\hat{\theta})J_{\mu \rho \varepsilon}(\hat{\theta}) \right] = N(T - 1) - \frac{N(T - 1)}{\hat{\sigma}_\varepsilon^2} \left( \frac{NT(T - 1)}{2\hat{\sigma}_\varepsilon^4} \right)^{-1} \frac{N(T - 1)}{\hat{\sigma}_\varepsilon^2} \\
= N(T - 1) - \frac{2N(T - 1)}{T} = N(T - 1) \left( 1 - \frac{2}{T} \right). \quad (B.10)
\]

Finally, the asymptotic variance of the adjusted score under both parametric and distributional misspecification is

\[
D_{\rho \varepsilon}(\hat{\theta}) = B_{\rho \varepsilon}(\hat{\theta}) + J_{\rho \mu \varepsilon}(\hat{\theta})J_{\mu \varepsilon}^{-1}(\hat{\theta})B_{\mu \varepsilon}(\hat{\theta})J_{\mu \varepsilon}^{-1}(\hat{\theta})J_{\mu \rho \varepsilon}(\hat{\theta}) - J_{\rho \mu \varepsilon}(\hat{\theta})J_{\mu \varepsilon}^{-1}(\hat{\theta})B_{\mu \rho \varepsilon}(\hat{\theta}) \\
= N(T - 1) + \left( \frac{N(T - 1)}{\hat{\sigma}_\varepsilon^2} \right)^2 \left( \frac{NT(T - 1)}{2\hat{\sigma}_\varepsilon^4} \right)^{-1} \frac{N(T - 1)}{\hat{\sigma}_\varepsilon^2} \\
- 2 \left( \frac{N(T - 1)}{\hat{\sigma}_\varepsilon^2} \right) \left( \frac{NT(T - 1)}{2\hat{\sigma}_\varepsilon^4} \right)^{-1} \frac{N(T - 1)}{\hat{\sigma}_\varepsilon^2} \\
= N(T - 1) + \frac{2N(T - 1)}{T} - 2 \frac{2N(T - 1)}{T} \\
= N(T - 1) \left( 1 - \frac{2}{T} \right). \quad (B.11)
\]
References


Bera, Anil K. and Yannis Bilias (2001). “Rao’s score, Neyman’s C(α) and Silvey’s LM tests: an essay on historical developments and some new results”. In: Journal of Statistical Planning and Inference 97.1, pp. 9 –44.


