Small Sample Properties of JR Estimators

John Kloke* Joseph McKean†

1. Introduction

This paper is concern with the implementation of the methods described in Kloke, McKean, and Rashid (2009). We present a new sandwich estimator of the variance covariance structure which performs well in simulation studies. We also present an R (R Development Core Team 2010) package, `jrfit` which implements estimation and testing proceedures.

2. Review JR estimation

Rank estimation was first introduced by Jureăková (1971) and Jaeckel (1972). Hettmansperger and McKean (2011) present an overview.

Kloke et al. (2009) showed that R estimation can be extended to cluster correlated data. In this section we briefly review JR estimation and develop some notation.

JR estimation can be applied to cluster correlated data. Assume that there are m clusters and the there are n_k experimental units in the kth cluster ($k = 1, \ldots, m$). Denote the outcome or response variable by y_{ki}. Covariates and treatment assignment information is contained in the design vector $\boldsymbol{x}_{ki}$. A linear model is then formed

$$y_{ki} = \alpha + \boldsymbol{x}_{ki}^T\boldsymbol{\beta} + e_{ki} \text{ for } k = 1, \ldots, m, i = 1, \ldots, n_k \tag{1}$$

where α is the intercept parameter, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and e_{ki} is an error term. We assume that the errors within a block are correlated (i.e. $e_{ki}\&e_{ki'}$) but the errors between blocks are independent (i.e. $e_{ki}\&e_{k'j}$). Further we assume that $e_{ki} \sim F, f$. Now write model (1) in block vector notation as

$$\boldsymbol{y}_k = \alpha\mathbf{1}_{n_k} + \boldsymbol{X}_k\boldsymbol{\beta} + \boldsymbol{e}_k. \tag{2}$$

where $\mathbf{1}_{n_k}$ is an $n_k \times 1$ vector of ones and $\boldsymbol{X}_k = [\boldsymbol{x}_{k1} \ldots \boldsymbol{x}_{kn_k}]^T$ is a $n_k \times p$ design matrix and $\boldsymbol{e}_k = [e_{k1}, \ldots e_{kn_k}]^T$ is a $n_k \times 1$ vector of error terms. Let $N = \sum_{k=1}^m n_k$. Now write as one large model

$$\boldsymbol{y} = \alpha\mathbf{1}_N + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{3}$$

where $\mathbf{1}_N$ is an $N \times 1$ vector of ones and $\boldsymbol{X} = [\boldsymbol{X}_1^T \ldots \boldsymbol{X}_m^T]^T$ is a $N \times p$ design matrix and $\boldsymbol{e} = [\boldsymbol{e}_1^T, \ldots \boldsymbol{e}_m^T]^T$ is a $N \times 1$ vector of error terms. Since there is an intercept in the model, we may assume (WLOG) that $\boldsymbol{X}$ is centered. Then the R estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_\varphi = \text{Argmin}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_\varphi \text{ where } \|\boldsymbol{v}\|_\varphi = \sum_{t=1}^N a(R(v_t))v_t$$

*University of Wisconsin - Madison, Madison, WI 53792
†Western Michigan University, Kalamazoo, MI 49008

is Jaeckel's dispersion function. As shown in Kloke et al. (2009) (c.f. Brunner and Denker 1994), the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ is given by

$$\hat{\boldsymbol{\beta}}_{\varphi} \dot{\sim} N_p \left(\boldsymbol{\beta}, \tau^2 (\boldsymbol{X}^T \boldsymbol{X})^{-1}) \left(\sum_{k=1}^{m} \boldsymbol{X}_k^T \boldsymbol{\Sigma}_{\varphi_k} \boldsymbol{X}_k \right) (\boldsymbol{X}^T \boldsymbol{X})^{-1}) \right)$$

where $\boldsymbol{\Sigma}_k = \text{var}(\varphi(F(\boldsymbol{e}_k)))$ and $F(\boldsymbol{e}_k) = [F(e_{k1}), \ldots, F(e_{kn_k})]^T$. Denote the second term in the variance product by $\boldsymbol{V}_{\varphi} = \left(\sum_{k=1}^{m} \boldsymbol{X}_k^T \boldsymbol{\Sigma}_{\varphi_k} \boldsymbol{X}_k \right)$. In next sections we present estimates of this variance term. The scale parameter τ_{φ} can be esimated with the Koul, Sievers, and McKean (1987) estimate.

2.1 Inference

We present Wald test of the hypothesis

$$H_0 : \boldsymbol{K\beta} = \boldsymbol{0} \text{ versus } H_A : \boldsymbol{K\beta} \neq \boldsymbol{0}. \tag{4}$$

Of course other constants on the right hand side of the hypothesis can tested by subtracting it from the fitted value of the parameter estimate. The Wald test of (4) is based on the statistic

$$T^* = \left(\boldsymbol{K} \hat{\boldsymbol{\beta}} \right)^T \left(\boldsymbol{K} \text{var}(\hat{\boldsymbol{\beta}}) \boldsymbol{K}^T \right) \left(\boldsymbol{K} \hat{(\boldsymbol{\beta})} \right).$$

Kloke et al. (2009) showed that the T^* has an asymptotic $\chi^2(q)$ distribution where where q is the rank of $\boldsymbol{K}$.

3. Estimates of variance of parameter estimates

In this paper we consider two estimates of variance of the parameter estimates. That is we estimate $\boldsymbol{\Sigma}_{\varphi}$ in two ways. The first was consider in the Kloke et al. (2009) paper and the other is a sandwich-type estimator.

If we assume that the errors within a block are exchangeable then the variance of the scores within a block are compound symmetric $\boldsymbol{\Sigma}_{\varphi} = (1 - \rho_{\varphi}) \boldsymbol{I} + \rho_{\varphi} \boldsymbol{J}$ where $\rho_{\varphi} = \text{cov}(\varphi(F(e_{ki}))\varphi(F(e_{kj}))$ for $i \neq j$. As discussed in Kloke et al. (2009) ρ_{φ} can be estimated as

$$\hat{\rho}_{\varphi} = \frac{1}{M-p} \sum_{k=1}^{m} \sum_{i>j} a(R(\hat{e}_{ki})) a(R(\hat{e}_{kj})), \text{ where } M = \sum_{k=1}^{m} \binom{n_k}{2}.$$

An estimate which does not require the fairly strong assumption of exchangeable errors is the sandwich estimator which is given by

$$\frac{m}{m-p} \sum_{k=1}^{m} \boldsymbol{X}_k^T a(R(\hat{\boldsymbol{e}}_k)) a(R(\hat{\boldsymbol{e}}_k))^T \boldsymbol{X}_k.$$

We have included a degrees of freedom correction that seems to work well.

Based on simulation studies (not shown) it was concluded that the denominator degrees of freedome be m when using the sandwich estimator and $N - p - 1 - 1$ where p is the number of parameters in the model and the two additional degrees of freedom are for the intercept and estimation of ρ_{φ}.

4. Simulation Studies

We conducted a series of simulations studies to determine the which critical values should be used in small samples. The main goals, however, were 1. determine if we should be using asymptotic (χ^2, normal) critical values or small sample (F, t) ; 2. compare the two estimates of the variance of the parameter estimates.

The simulations are based on the following model

$$y_{kj} = \alpha + \boldsymbol{x}_{kj}^T\boldsymbol{\beta} + \boldsymbol{w}_{kj}^T\boldsymbol{\Delta} + b_k + e_{kj}.$$

where $\boldsymbol{\Delta}$ is a $k-1\times 1$ vector of effect sizes. For simplicity we assume the reference group is the first group.

The parameter ρ denotes the interclass correlation coefficient; the simulation size is 10000; level of $\alpha = 0.05$ for all tests of hypothesis. In the plots of empirical level we have included horizontal lines at $\alpha \pm 1.96\sqrt{\frac{\alpha(1-\alpha)}{\text{size}}}$. Methods resulting in points consistently within the region represent valid tests.

4.1 RBD

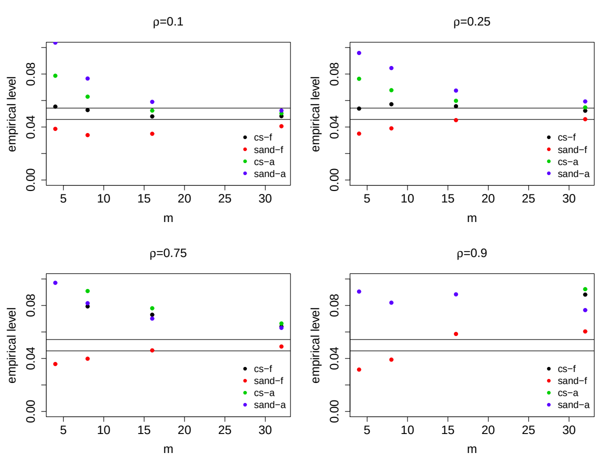
In this section we present a summary of the simulations we conducted which mimic a randomized block design. We simulate both error terms from the normal distribution as well as the covariates. Treatment is balanced within each block.

4.1.1 Empirical Level ($k = 3$ and $p = 1$)

We varied the number of blocks ($m = 4, 8, 16, 32$). The block size was $n = 6$. We considered $k = 3$ treatments and $p = 1$ covariates. We varied the interclass correlation coefficient ($\rho = 0.1, 0.25, 0.75, 0.9$). We test for a treatment effect with the hypothesis $H_0 : \Delta_1 = \Delta_2 = 0$.

The results of the experiement are present in Figure 4.1.1.

Figure 1:



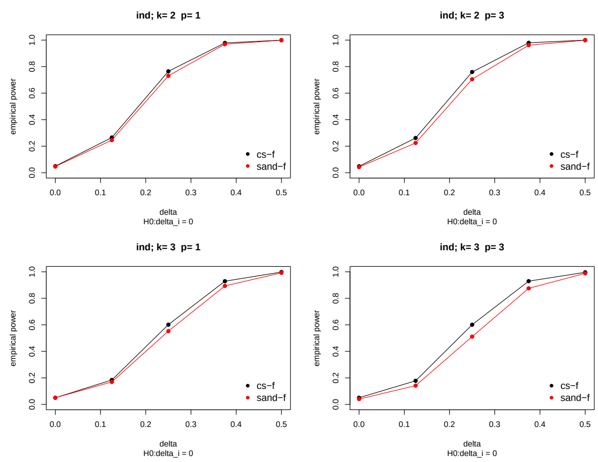
When the correlation is low or moderately low the cs estimates of variance while utilizing F critical values yields valid tests. The sandwich estimate of variance while utilizing F critical values seems to yield slightly concervative tests.

It appears that utilizing the asymptotic χ^2 critical values tends to result in liberal inference and their use is not considered further.

4.2 Power

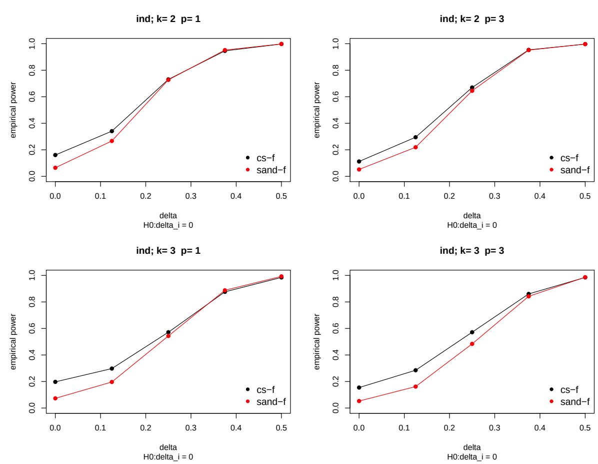
To observe power when utilizing the sandwich estimate we performed an experiement in which we increased the effect size Δ_i for $i = 2, k$. There were 32 blocks in total, 8 each of $n = 6, 12, 18, 24$. We used low ($\rho = 0.1$) and high ($\rho = 0.9$) interclass correlation. The results are presented in Figure 4.2-4.2.

Figure 2: Power curves for low icc ($\rho = 0.1$)



When there is low correlation, using the cs estimate performs better in that it has a slight advantage over the sandwich estimate. When there is high correlation, again, we see that using the CS estimate leads to liberal inference.

Figure 3: Power curves for high icc ($\rho = 0.9$)



4.3 Repeated Measures

In this section we present a summary of the simulations we conducted which mimic a repeated measures design. Here treatment is assigned at random to each block. That is the experimental units is a block. One can think of this as a randomized control trial with measurements taken over time.

The hypotheses we wish to test here are for a profile analysis. The contrast matrices are

$$\boldsymbol{K}_1 = [\mathbf{1} - \boldsymbol{I} - \mathbf{1}\boldsymbol{I}] \text{ and } \boldsymbol{K}_2 = [\mathbf{1}^T - \mathbf{1}^T].$$

$\boldsymbol{K}_1$ tests for parallel profiles and $\boldsymbol{K}_2$ tests for equal median for the treatment. If both the tests $H_0 : \boldsymbol{K}_i\boldsymbol{\beta} = 0$ are not rejected then the test of coincident profiles is also not rejected.

We examined 4 and 8 repeated measures. The number of blocks varied from $m = 12, 25, 50, 75, 100, 150, 200$. There were $k = 2$ treatments where were assigned at random to each of the blocks.

The results are presented in Figures 4.3-4.3.

Figure 4: Empirical Level: $H_0 : K_1\Delta = 0$ (number of blocks = 4)

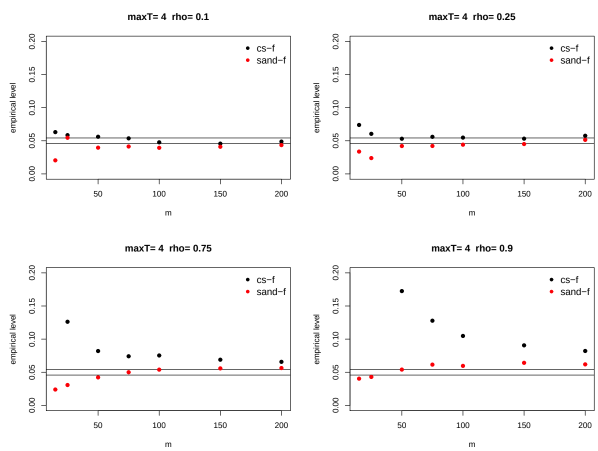
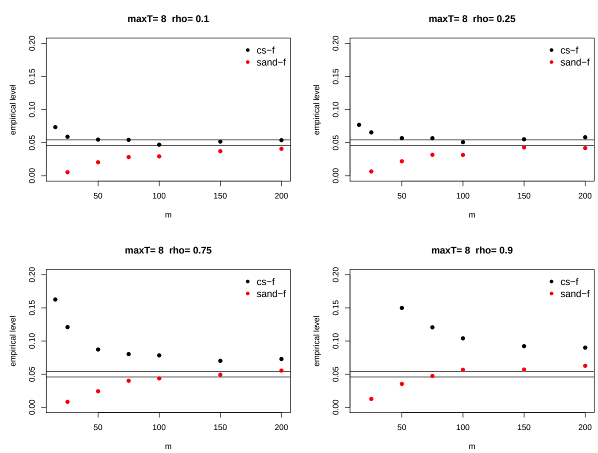


Figure 5: Empirical Level: $H_0 : K_1\Delta = 0$ (number of blocks = 8)



Again, we see the CS estimate can be liberal for small samples, especially when ρ is large.

Figure 6: Empirical Level: $H_0 : K_2\Delta = 0$ (number of blocks = 4)

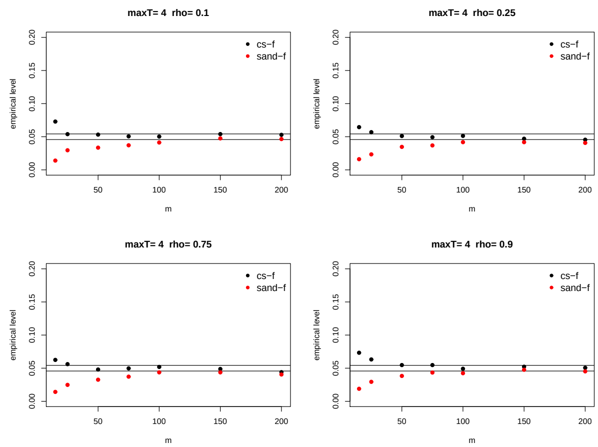
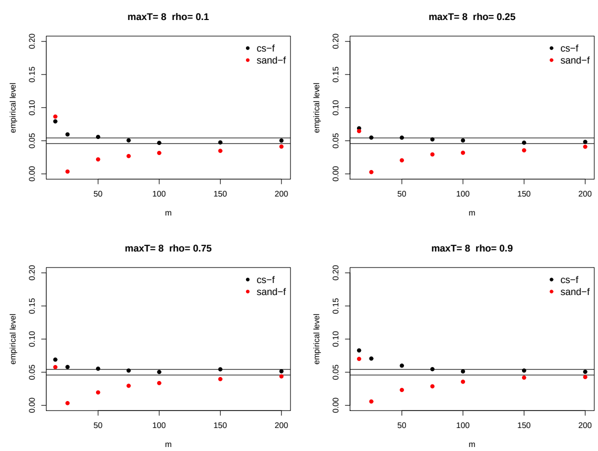


Figure 7: Empirical Level: $H_0 : K_2\Delta = 0$ (number of blocks = 8)



5. Usage of `jrfit`

The R (R Development Core Team 2010) package `jrfit`[1] extends `Rfit` (Kloke and McKean 2012) for the case of cluster correlated data discussed in this paper. The package provides estimation and tests of hypothesis. The methods of covariance estimation discussed in this paper are implemented.

To fix ideas, we present an analysis of a simulated dataset utilizing both the compound symmetry and sandwich estimators discussed in the previous section.

The setup is as follows:

```
> m<-160  # blocks
> n<-4    # observations per block
> p<-1    # baseline covariate
> k<-2    # trtmnt groups
```

We simulated the block effects from a t-distribution with 5 degrees of freedom and the random errors from a t-distribution with 3 degress of freedom. Note the

[1] available at `http://www.biostat.wisc.edu/~kloke/`

assumptions for exchangeable errors is met. Also, the sample size is sufficiently large for one to utilize the sandwich estimator and obtain reliable inference.

Next, we set up the design and simulate a baseline covariate which we assume normally distributed.

```
> trt<-as.factor(rep(sample(1:k,m,replace=TRUE),each=n))
> block<-rep(1:m,each=n)
> x<-rep(rnorm(m),each=n)
```

We set the overall treatment effect to be $\Delta = 0.5$ so that we can form the response as follows.

```
> delta<-0.5
> w<-trt==2
> Z<-model.matrix(~as.factor(block))
> e<-rt(m*n,df=5)
> b<-rt(m,df=3)
> y<-delta*w+Z%*%b+e
```

In this case the covariate is unimportant. First we analyze the data with the compound symmetry assumption.

The three required arguments to `jrfit` are the design matrix, the response vector, and the vector denoting block membership.

```
> library(jrfit)
> X<-cbind(w,x)
> fit<-jrfit(X,y,block,var.type='cs')
> summary(fit)

Coefficients:
   Estimate   Std. Error t-ratio    p.value
w  0.6055401  0.2204266  2.7471278  0.0061823
x -0.1785521  0.1142161 -1.5632832  0.1184834
```

Notice, by default the intercept is not displayed in the output of the `summary` function. If the inference on the intercept is of interest than set the option `int` to `TRUE` in the `jrfit` `summary` function. The cell medians model can also be fit as follows.

```
> library(jrfit)
> W<-model.matrix(~trt-1)
> X<-cbind(W,x)
> fit<-jrfit(X,y,block,var.type='cs')
> summary(fit)

Coefficients:
     Estimate    Std. Error  t-ratio     p.value
trt1 -1.1208e+00  1.7048e-01 -6.5748e+00  1.0165e-10
trt2 -5.1528e-01  1.4972e-01 -3.4417e+00  6.1597e-04
x    -1.7855e-01  1.1422e-01 -1.5633e+00  1.1849e-01
```

Now we present the same analysis utilizing the sandwich estimator. The sandwich estimator is also the default.

```
> X<-cbind(w,x)
> fit<-jrfit(X,y,block,var.type='sandwich')
> summary(fit)

Coefficients:
  Estimate   Std. Error t-ratio    p.value
w  0.6055401  0.2176245  2.7825000  0.0060434
x -0.1785521  0.1238383 -1.4418158  0.1513088

> X<-cbind(W,x)
> fit<-jrfit(X,y,block,var.type='sandwich')
> summary(fit)

Coefficients:
     Estimate    Std. Error  t-ratio     p.value
trt1 -1.1208e+00  1.6947e-01 -6.6139e+00  5.3311e-10
trt2 -5.1528e-01  1.4915e-01 -3.4548e+00  7.0506e-04
x    -1.7855e-01  1.2384e-01 -1.4418e+00  1.5131e-01
```

Not surprizingly, based on what was observed in the simulation studies, the results for both methods are similar in this case.

We have include a function `wald.test.jrfit` computes the test of discussed in Section 2.1. There are two required arguments: the results of a call to the fitting routine `jrfit` and the contrast matrix $\boldsymbol{K}$. The following code demonstrates a test for no treatment effect.

```
> K<-matrix(c(1,-1,0),nrow=1)
> wald.test.jrfit(fit,K)

$statistic
         [,1]
[1,] 7.742992

$p.value
            [,1]
[1,] 0.006041207

$asymptotic
[1] FALSE

$df
[1]   1 160
```

Utilizing the cs estimate.

```
> K<-matrix(c(1,-1,0),nrow=1)
> wald.test.jrfit(jrfit(X,y,block,var.type='cs'),K)

$statistic
         [,1]
[1,] 7.547381
```

```
$p.value
          [,1]
[1,] 0.00617998

$asymptotic
[1] FALSE

$df
[1]   1 636
```

6. Summary

In certain situations, the compound symmetry (cs) estimator may be favorable to the sandwich estimator. That said, we have demonstrated a number of situations for which the cs estimator leads to liberal and it's general use cannot be recommended. The sandwich estimate tends to lead to conservative inference. However, in general, the sandwich estimate seems favorable. Based on the simulations we conducted, we recommend the use of the sandwich estimator. The results of our simulation work on for the gradient test yielded similar results, though we chose to not present.

References

Brunner, E. and Denker, M. (1994), "Rank statistics under dependent observations and applications to factorial designs," *Journal of statistical planning and inference*, 42, 353–378.

Hettmansperger, T. P. and McKean, J. W. (2011), *Robust Nonparametric Statistical Methods, 2nd Edition*, Boca Raton, FL: Chapman-Hall.

Jaeckel, L. A. (1972), "Estimating regression coefficients by minimizing the dispersion of residuals," *The Annals of Mathematical Statistics*, 43, 1449–1458.

Jureăková, J. (1971), "Nonparametric estimate of regression coefficients," *The Annals of Mathematical Statistics*, 42, 1328–1338.

Kloke, J. D. and McKean, J. W. (2012), "Rfit: Rank-based estimation for linear models," *The R Journal*, 4, 57–64.

Kloke, J. D., McKean, J. W., and Rashid, M. (2009), "Rank-Based Estimation and Associated Inferences for Linear Models with Cluster Correlated Errors," *Journal of the American Statistical Association*, 104, 384–390.

Koul, H. L., Sievers, G. L., and McKean, J. W. (1987), "An estimator of the scale parameter for the rank analysis of linear models under general score functions," *Scandinavian Journal of Statistics*, 14, 131–141.

R Development Core Team (2010), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.