1989

Confidence Sets Centered at James-Stein Estimators - A Surprise Concerning the Unknown Variance Case

Jiunn T. Hwang

Aman Ullah

ISSN:0318-725X
ISBN:0-7714-1135-9

RESEARCH REPORT 8909

CONFIDENCE SETS CENTERED AT JAMES-STEIN ESTIMATORS - A SURPRISE CONCERNING THE UNKNOWN VARIANCE CASE

by

Jiunn T. Hwang
Department of Mathematics
Cornell University

and

Aman Ullah
Department of Economics
University of Western Ontario

June, 1989

Confidence Sets Centered at James-Stein Estimators
- A Surprise Concerning the Unknown Variance Case

Jiunn T. Hwang
Department of Mathematics
Cornell University

Aman Ullah
Department of Economics
University of Western Ontario

June 1989

Abstract

We compare the confidence set centered at James-Stein point estimator to the usual F-confidence set for the p regression parameters of a linear model. Previous studies usually focused on the known variance case and typically conclude that whatever holds in the known variance case should hold in the unknown variance case when the variance is replaced by its best linear estimator S^2. We are surprised that this is not entirely the picture we observe here. In fact, in many unknown variance cases, the range of the shrinkage factor a for the associate confidence set to have uniformly higher coverage probabilities than its F counterpart can be ten times bigger than $2(p-2)$ (the expected upper bound in the known variance case). This is true especially when the degrees of freedom is small.

Professor Hwang's research is partially supported by NSF with No. DMS-8809016, and that of Professor Ullah's by NSERC.

Key Words: Domination, coverage probability, asymptotic expansion

AMS 1980 Subject Classifications: Primary 62C20; Secondary 62F25.

I. Introduction.

In the past three decades, much progress has been made about the problem of shrinkage estimation, mainly because of the impact of the celebrated work of Stein (1955) and James-Stein (1961). Recent interest in several other areas such as small area estimation has seemed to boost the motivation of constructing shrinkage estimation. There appears to be a need to design procedures that will take advantage of the information from many other strata or "borrow strength" from other populations in the situation that the number of observations in each stratum is small and the number of strata is huge.

The theory of shrinkage estimation appears to have been well studied. However, the other companied problem of constructing associated confidence set has been given less attention relatively, although see Casella and Hwang (1983, 1987), Hwang and Casella (1982, 1984), Hwang and Chen (1986), Shinozaki (1988), and Ullah et al (1988) from the frequentists' view point and Morris (1983) from the empirical Bayes view point. Here we will be taking a frequentist's point of view. For such an approach a review of the area can be found in Robert and Saleh (1989).

Assume a linear model

$$Y = X\beta + \epsilon, \tag{1.1}$$

where ϵ_i are assumed to be i.i.d $N(0,\sigma^2)$ and X is an $N \times p$ matrix having a full rank p. The least squared estimator for β is

$$\hat{\beta} = (X'X)^{-1}X'Y.$$

The positive part James-Stein estimator

$$\hat{\beta}_{JS} = \left[1 - \frac{aS^2/(N-p+2)}{\hat{\beta}'X'X\hat{\beta}}\right]_+ \hat{\beta} \tag{1.2}$$

where

$$S^2 = (y-X\hat{\beta})'(y-X\hat{\beta}),$$

has been shown to have a uniformly smaller risk function than $\hat{\beta}$ under the quadratic loss function $(\beta-\hat{\beta})'(X'X)(\beta-\hat{\beta})$, if and only if $0 < a < 2(p-2)$.

The problem, studied in this paper, is about the construction of confidence sets superior to the F confidence set:

$$\{\beta:(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta)/p \le c^2 \frac{S^2}{N-p}\}, \tag{1.3}$$

where c^2 is chosen to be the α upper quantile of $F_{p,N-p}$ distribution. Therefore the coverage probability of (1.3) is $1-\alpha$. The obvious first stage development will focus on the alternative James-Stein type confidence region:

$$\{\beta:(\hat{\beta}_{JS}-\beta)'(X'X)(\hat{\beta}_{JS}-\beta)/p \le c^2S^2/N-p\}. \tag{1.4}$$

Although several papers including Chen and Hwang (1988), Robert and Casella (1987), and Kim (1987) focus on the comparison of

(1.4) to (1.3), (1.4) has not been proved to dominate (1.3), i.e., having higher coverage probability than (1.3). For the known variance case, with S^2 replaced by σ^2, the domination of (1.4) over (1.3) has been established for a range of a. Furthermore, for the unknown variance case, the evidence in Chen and Hwang (1988) suggests that asymptotically this is true.

However, little has been done in the fixed sample size case, which is our focus in this paper. Based on the calculation of the coverage probability as $|\theta| \to \infty$, one can specify the upper bound of a in (1.2) for the confidence set (1.4) to have coverage probability bigger than $1-\alpha$.

Although the technical argument similar to Hwang and Casella (1982) has been used, the results produced here have been quite a surprise. Based on the experience with the point estimation problem as well as the confidence set problem with known σ^2, it has been expected that the upper bound of a for domination should be about $2(p-2)$. However, the bound for the unknown variance case determined by asymptotic formula can be much bigger when $N-p$ is small. It can sometimes be seven times as big as $2(p-2)$ for $p = 9$ and $N-p = 2$ (in extreme cases, the ratio is even much higher). Numerical calculation of the coverage probabilities based on three dimensional integration indicates that the true upper bound of a for uniform domination is near the bound suggested by the asymptotic formula. For the case $p = 9$ and $N-p = 8$ (a case occurring in practice, say, a two way table with three levels each way and 8 degrees of freedom in the residual), the numerical calculation shows that the upper bound of a for domination is at least 22, a number much larger than

$2(p-2) = 14$. The numerical results also suggest that to gain substantial improvement of coverage probability, a has to be chosen much larger (especially when $N-p$ is small) than what may have been previously expected.

II. Asymptotic Formula.

In comparing (1.3) with (1.4), one can obviously transform the situation into a canonical one. This then reduces to the model $X \sim N(\theta, \sigma^2 I)$ independent of $\frac{S^2}{\sigma^2} \sim \chi^2_n$, $n = N-P$. Also (1.3) and (1.4) reduce to

$$C_{X,S^2} = \{\theta : |X-\theta|^2 \leq pc^2 \frac{S^2}{n}\}, \tag{2.1}$$

and

$$C_{JS,S^2} = \{\theta : |\theta - \left[1 - \frac{aS^2/(n+2)}{|X|^2}\right]_+ X|^2 \leq p \frac{c^2}{n} S^2\}, \tag{2.2}$$

where c^2 is the $(1-\alpha)$ quantile of $F_{p,n}$, the F distribution with degrees of freedom p and n.

To derive the asymptotic coverage probability of C_{JS,S^2} (as $|\theta| \to \infty$) under a simpler notation, we replace $\frac{a}{n+2}$ by a and $c^2 \frac{p}{n}$ by c^2 in the following derivation. Later, we can reverse the substitution to get the asymptotic for (2.2).

Assume without loss of generality that $\sigma^2 = 1$, $X \sim N(\theta, I)$, and $S^2 \sim \chi^2_n$ are independent.

Lemma 2.1. For $c > 0$,

$$P_\theta\left(\left|\left[1 - \frac{aS^2}{|X|^2}\right]_+ X-\theta\right| \le cS\right) \tag{2.3}$$

$$= 1-\alpha+E\ \frac{-aS^2}{2|\theta|^2}(1-\alpha(S)-h(S))(aS^2-2(p-2))+O(|\theta|^{-3})$$

where $1-\alpha = P(|X-\theta| \le cS)$,

$$1-\alpha(s) = P(|X-\theta| \le cS \mid S = s)$$

and

$$h(S) = 1-\alpha(S)-\frac{p^{-1}(cS)^p}{2^{(p-2)/2}\Gamma(\frac{p}{2})}\ e^{-c^2S^2/2}.$$

Proof: We will condition on S and take a derivation similar to the proof of Theorem 3.1 of Hwang and Casella (1984). We can assume without loss of generality that $\sigma^2 = 1$. The key technical difficulty is that S can be an arbitrary large positive number and the Taylor expansion in Hwang and Casella (1984) is only uniformly for S in a compact region. To get around the difficulty, we note that

$$\int_{S\ge|\theta|^{1/2}} S^k f(S)dS = O(|\theta|^{-3}), \tag{2.4}$$

where k is fixed and $f(S)$ is the p.d.f. of S. In fact in (2.4), the order is exponential in $|\theta|$ and is much smaller than $|\theta|^{-3}$. Therefore it is sufficient to consider only $S < |\theta|^{1/2}$. We note also that

$$\left|\left[1 - \frac{aS^2}{|X|^2}\right]_+ X-\theta\right| \le cS \tag{2.5}$$

implies that for all θ

$$|X-\theta| \le (c+\sqrt{a})S.$$

(This can be established by considering two separate regions of X such that $|X|^2 \le aS^2$ and $|X|^2 \ge aS^2$.)

Furthermore we can omit '+' in the derivation, since

$$P\left[\frac{aS^2}{|X|^2} > 1\right] = O(|\theta|^{-3}). \tag{2.6}$$

To establish this, note that (2.6) is bounded above by

$$P(a|\theta| > |X|^2) + P(S^2 > |\theta|) < P(a|\theta| > (|\theta|+Z)^2) + P(S^2 > |\theta|).$$

where Z is $N(0,1)$. The second term of the right hand side is $O(|\theta|^{-3})$ by (2.4). The first term is bounded above, for sufficiently large $|\theta|$, by

$$P(Z > \tfrac{1}{2}|\theta|) = O\left[\frac{1}{|\theta|} e^{-\frac{1}{2}|\theta|^2}\right].$$

Hence (2.6) follows. Therefore it is sufficient to look at only the case

$$\frac{aS^2}{|X|^2} < 1,$$

in which case $\left[1 - \frac{aS^2}{|X|^2}\right]_+ = 1 - \frac{aS^2}{|X|^2}$. Using these keys and following the derivation of Hwang and Casella (1984), we can show that the probability, conditioning on S, that (2.5) holds equals

$$1-\alpha(S) - \frac{aS^2}{2|\theta|^2}[1-\alpha(S)-h(S)](aS^2-2(p-2)) + p(S)O(|\theta|^{-3}),$$

where $p(S)$ is a polynomial of 8 degrees. Taking expectation with respect to S and noting

$$E(1-\alpha(S)) = 1-\alpha \quad \text{and} \quad Ep(S) < \infty,$$

we establish the lemma.

Theorem 2.2. As $|\theta| \to \infty$

$$P(|\left[1 - \frac{aS^2}{|X|^2}\right]_+ X-\theta| \le cS)$$

$$= 1-\alpha - \frac{2ac^p\Gamma(\frac{n+p+2}{2})(\frac{a(n+p+2)}{c^2+1} - 2(p-2))}{p|\theta|^2(c^2+1)^{(n+p+2)/2}\Gamma(\frac{p}{2})\Gamma(\frac{n}{2})} + O(|\theta|^{-3}).$$

Proof: The leading term on the right hand side of (2.3) is

$$\frac{-a}{2|\theta|^2}\,\frac{p^{-1}c^{p}}{2^{(p-2)/2}\Gamma(\frac{p}{2})}\,ES^{p+2}\,e^{-\frac{c^2S^2}{2}}(aS^2-2(p-2)). \qquad (2.7)$$

The expectation of the last expression is

$$aES^{p+4}e^{-\frac{c^2S^2}{2}}-2(p-2)ES^{p+2}e^{-\frac{c^2S^2}{2}}. \qquad (2.8)$$

Using the equation

$$ES^{k}e^{-\frac{c^2S^2}{2}} = 2^{\frac{k}{2}}\,\Gamma(\tfrac{n+k}{2})/((c^2+1)^{\frac{k+n}{2}}\,\Gamma(\tfrac{n}{2})),$$

we show that (2.8) equals

$$\frac{a2^{\frac{p+4}{2}}\,\Gamma(\frac{n+p+4}{2})}{(c^2+1)^{\frac{n+p+4}{2}}\,\Gamma(\frac{n}{2})} - 2(p-2)\,\frac{2^{\frac{p+2}{2}}\,\Gamma(\frac{n+p+2}{2})}{(c^2+1)^{\frac{n+p+2}{2}}\,\Gamma(\frac{n}{2})}$$

$$= \frac{2^{\frac{p+2}{2}}\,\Gamma(\frac{n+p+2}{2})}{(c^2+1)^{\frac{n+p+2}{2}}\,\Gamma(\frac{n}{2})}\left[\frac{a(n+p+2)}{(c^2+1)} - 2(p-2)\right].$$

Consequently (2.7) equals

$$\frac{-2ac^p\,\Gamma(\frac{n+p+2}{2})}{p|\theta|^2(c^2+1)^{(n+p+2)/2}\Gamma(\frac{p}{2})\Gamma(\frac{n}{2})}\left[\frac{a(n+p+2)}{c^2+1}-2(p-2)\right],$$

establishing the theorem.

In order to compare to F confidence set, we replace S^2 by S^2/n. Therefore the James-Stein type confidence set is as given in (2.2) as compared to (2.1), the usual confidence set.

For (2.1) to be a $(1-\alpha)$ confidence set, one can choose c^2 to be the $1-\alpha$ quantile of F distribution with p, n degrees of freedom. The Theorem then implies

<u>Corollary 2.3</u>. The coverage probability of (2.2) is greater than (2.1) for large $|\theta|$ if

$$0 < a < 2(p-2)B, \tag{2.9}$$

where

$$B = \frac{n+2}{n+p+2}\left(\frac{pc^2}{n}+1\right) \tag{2.10}$$

Note that, in point estimation context with respect to the sum of squared error loss, it has been known that the range of a for domination of James-Stein over X is given in (2.9) for B = 1. For the confidence set problem and the known variance case, the similar phenomenon has emerged. The bound 2(p-2) has been suggested by asymptotic calculation and is almost the bound for the James-Stein confidence set to have uniform bigger coverage probability. See Hwang and Casella (1984). Surprisingly, as seen in Tables 1 and 2, the values of B in the Corollary can be quite high and in the extreme case as high as 38. However in many practical cases, say $\alpha = .1$, $p = 19$ and $n = 20$, $B = 1.46$ is still larger than one. As $n \to \infty$, B approach one, followed from (2.10) as well as shown in tables, in agreement with the common belief that the problem reduces to the known variance case where $B = 1$.

We also took another approach based on Edgeworth expansion and obtained an asymptotic formula for the coverage probability of the confidence set (2.2). Similar approach has been taken in Ullah et al (1988). The asymptotic formula appears to be different and can also be used to derive a bound on a so that the asymptotic coverage probability after omitting $O(|\theta|^{-3})$ term is greater than $1-\alpha$. The bound of a thus derived corresponds to

$$B = \frac{G_{0,2}-G_{2,2}}{G_{0,4}-G_{2,4}} \tag{2.11}$$

where

$$G_{\ell,\ell'} = F_{p+\ell,n+\ell'}\left[\frac{pc^2(n+\ell')}{n(p+\ell)}\right],$$

and F_{k_1,k_2} is the cumulative distribution function of F with degrees of freedom k_1 and k_2.

Table 1. Values of B for $\alpha = .1$. Previously, B has been thought to be close and less than one, which is far away from truth when n is small.

p/n	2	4	8	20	40	80	160
3	8.42	2.76	1.61	1.19	1.09	1.04	1.02
5	10.8	3.31	1.8	1.25	1.12	1.06	1.03
7	12.3	3.68	1.94	1.3	1.14	1.07	1.03
9	13.3	3.94	2.04	1.34	1.16	1.07	1.04
19	15.8	4.63	2.34	1.46	1.22	1.11	1.05
20	15.9	4.67	2.35	1.46	1.22	1.11	1.05
40	17.3	5.09	2.56	1.57	1.28	1.14	1.07
80	18.1	5.35	2.7	1.65	1.34	1.18	1.1
160	18.5	5.49	2.78	1.7	1.38	1.21	1.12

Table 2. Value of B for $\alpha = .05$.

p/n	2	4	8	20	40	80	160
3	17	3.96	1.94	1.29	1.13	1.06	1.03
5	21.9	4.81	2.2	1.37	1.17	1.08	1.04
7	25	5.38	2.39	1.43	1.19	1.09	1.05
9	27.1	5.8	2.53	1.47	1.22	1.1	1.05
19	32.3	6.86	2.93	1.63	1.29	1.14	1.07
20	32.6	6.93	2.96	1.64	1.3	1.15	1.07
40	35.5	7.59	3.24	1.77	1.38	1.19	1.1
80	37.2	7.99	3.43	1.87	1.45	1.24	1.13
160	38.1	8.21	3.54	1.94	1.51	1.28	1.16

Although (2.11) looks different from (2.10) and we do not have the analytic proof that they are identical, numerical evaluation of the two bounds leads to the same answers as in Tables 1 and 2. Therefore the two approaches lead to the same bound. This provides an independent confirmation of the correctness of Corollary 2.3.

As reported in Tables 3 and 4, we numerically evaluated the coverage probabilities of the confidence set (2.2). Like in the known variance case, the asymptotic bound is only necessary for uniform domination and not sufficient. In Table 3 we focus on the case that n is moderate, since for large n the results are not very different from the known variance case. For $p = 9$ and $n = 2$, the exact upper bound for a by Table 3 is at least 120. This is larger than 8 times $2(p-2)$ and is closer to the asymptotic bound $13.3 \times 2(p-2) = 186.2$ than $2(p-2) = 14$. For $n = 8$ and the same p, the exact bound is at least 22 about the midpoint of the asymptotic bound $2.08 \cdot 2 \cdot (p-2) = 29.12$ and the $2(p-2) = 14$. However the fact that the exact bound is so much larger than $2(p-2)$ seems to be surprising. Note if one were using the traditional choice $a = p-2$, the largest gain in coverage probabilities may be small especially when n is small. See column one of Tables 3 and 4. The maximum coverage probability for $p = 9$, $n = 2$ and $a = p-2$ is .907, which is strikingly smaller than .999, the probability in the known variance case (i.e., the same situation as here except $n = \infty$).

The coverage probabilities are calculated based on three dimensional integration. Namely by conditioning on S, we write the conditional coverage probabilities as a double integral similar to the first formula after (3.9) in Casella and Hwang (1983). (Of course their $v(\cdot)$ function is a constant function here.) Then we integrate out against S. Since the results seem to be surprising and three dimensional integration seems to have pushed to the limit of numerical integration, we also take another approach of integrating (3.10) of Casella and Hwang (1983). (The $(n+1)!$ in the middle term of their (3.10) is a typographical error and should be corrected as $(n+i)!$.) This approach reduces to a double integration and leads to the numbers agreeing with Tables 3 and 4 up to the third decimal place.

Table 3. Exact coverage probabilities of (2.2). These numbers are greater than .9, the coverage probability of the corresponding usual F confidence sets.

	$p = 9$, $n = 2$		$p = 9$, $n = 8$	
$\lvert\theta\rvert$	$a = p-2 = 7$	$a = 120$	$a = 7$	$a = 22$
0	.907	.953	.961	.989
1	.907	.950	.958	.988
2	.905	.943	.952	.983
4	.903	.921	.933	.947
6	.901	.906	.919	.902
8	.901	.901	.912	.902
10	.901	.901	.908	.905
12	.901	.901	.905	.904
15	.900	.901	.904	.903
20	.900	.901	.902	.902
30	.900	.900	.901	.901
40	.900	.900	.901	.900
50	.900	.900	.900	.900

Table 4. Coverage probabilities of (2.2), greater than the coverage probability 0.9 of the corresponding F sets

	p = 19, n = 4		p = 19, n = 20	
	a = 17	a = 120	a = 17	a = 40
$\lvert\theta\rvert$				
0	.935	.980	.993	1.000
1	.933	.979	.993	.999
2	.930	.975	.991	.999
4	.922	.961	.982	.996
6	.915	.939	.967	.946
8	.910	.915	.950	.904
10	.907	.903	.937	.912
12	.905	.901	.928	.913
15	.904	.905	.919	.911
20	.902	.905	.911	.907
30	.900	.902	.905	.904
40	.900	.901	.903	.902
50	.900	.901	.902	.901

III. Conclusions.

For many problems, especially in decision theory, results hold in the known variance case are proved to hold in the unknown variance case. However, we are surprised that this is not entirely the picture we have seen here. For the known variance case the shrinking constant a, for domination in coverage probability, is less than 2(p-2) but for the unknown variance case, it can be more than 8 times as big as 2(p-2).

It appears that we should choose a much larger shrinkage factor for the unknown variance case to have a substantial gain. This is especially true when the degree of freedom of S^2 is small.

Given that a can be larger for the unknown variance case as reported here, it is surprising that there has been so many unsuccessful attempts for establishing analytical domination results in confidence sets for the unknown variance case.

Acknowledgement

The authors thank Jing-Huei Chen for the numerical computation of the coverage probabilities as reported in Tables 3 and 4.

References

Casella, G. and Hwang, J. T. (1983), Empirical Bayes confidence sets for the mean of a multivariate normal distribution, JASA 78, 688-698.

Casella, G. and Hwang, J. T. (1987), Employing vague prior information in the construction of confidence sets, J. Mult. Anal. 21, 79-104.

Chen, J. and Hwang, J. T. (1988), Improved set estimators for the coefficients of a linear model when the error distribution is spherically symmetric with unknown variances, Can. J. Stat. 16 (2).

Hwang, J. T. and G. Casella (1982), Minimax confidence sets for the mean of a multivariate normal distribution, The Annals of Statistics, 868-881.

Hwang, J. T. and Casella, G. (1984), Improved set estimators for a multivariate normal mean, Stat. Dec. Supplement Issue 1, 3-16.

Hwang, J. T. and Chen, J. (1986), Improved confidence sets for the coefficients of a linear model with spherically symmetric errors, Ann. Stat. 14, 444-460.

James, W. and C. Stein (1961), Estimation with quadratic loss, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 361-379.

Kim, P. T. (1987), Recentered confidence sets for the mean of a multivariate distribution when the scale parameter is unknown, Ph.D. Thesis, Department of Mathematics, University of California at San Diego.

Morris, C. N. (1983), Parametric empirical Bayes inference: Theory and applications, JASA, 47-54.

Robert, C. and Casella, G. (1987), Improved confidence sets in spherically symmetric distributions, Tech. Report #87-51, Department of Statistics, Purdue University.

Robert, C. and Saleh, A. K. Md. E. (1989), Recentered confidence sets: a review, Cornell MSI Technical Report.

Shinozaki, M. (1988), Improved confidence sets for the mean of a multivariate normal distribution, Statistics and Decisions (to appear).

Stein, C. (1955), Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 197-206.

Ullah, A., V. K. Srivastava, R. A. L. Carter and M. S. Srivastava (1988), Unbiased estimation of the MSE matrix of Stein-rule estimators, confidence ellipsoid and hypothesis testing, Research Report, University of Western Ontario, December.