A unified approach to estimation of noncentrality parameters, the multiple correlation coefficient, and mixture models [1]

Tatsuya Kubokawa[a], Éric Marchand[b], William E. Strawderman[c]
a Department of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN e-mail: tatsuya@e.u-tokyo.ac.jp)
b Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, J1K 2R1 (e-mail: eric.marchand@usherbrooke.ca)
c Rutgers University, Department of Statistics and Biostatistics, 501 Hill Center, Busch Campus, Piscataway, N.J., USA, 08855 (e-mail: straw@stat.rutgers.edu)

Summary

We consider a class of mixture models for positive continuous data and the estimation of an underlying parameter θ of the mixing distribution. With a unified approach, we obtain classes of dominating estimators under squared error loss of an unbiased estimator, which include smooth estimators. Applications include estimating noncentrality parameters of chisquare and F-distributions, as well as $\frac{\rho^2}{1-\rho^2}$ where ρ is a multivariate correlation coefficient in a multivariate normal set-up. Finally, the findings are extended to situations where there exists a lower bound constraint on θ.
Keywords and phrases: Estimation; Lower bounded parameter; Mixtures; Multiple correlation coefficient; Noncentrality parameter.
AMS 2010 subject classifications: 62C20, 62C86, 62F10, 62F15, 62F30

1 Introduction

Noncentral chisquare and noncentral Fisher distributions arise in various situations, typically related to quadratic forms in normal linear models and to the distribution of various non-null test statistics in multivariate analysis, such as Hotelling's T^2 test and Pearson's chisquare test for goodness of fit. The statistical inference of noncentral parameters thus becomes an issue and represents interesting and challenging problems, that have been addressed, with both Bayesian and frequentist risk analysis, by several researchers, such as Perlman and Rasmussen (1975), Neff and Strawderman (1976), Chow (1987), Leung and Muirhead (1987), Rukhin (1993), Kubokawa, Robert and Saleh (1993) and Shao and Strawderman (1995).
One motivating issue that arises concerns the fact that benchmark unbiased estimators of such noncentrality parameters take values beyond the range of the parameter space, and that it is desirable to provide smooth or Bayesian improvements. A similar problem arises in estimating either a population multiple correlation coefficient ρ or coefficient of determination ρ^2, and in particular the function $\frac{\rho^2}{1-\rho^2}$ based on the sample coefficient of determination R^2 (see for instance Muirhead, 1985; Muirhead and Leung, 1985; Leung

[1]December 20, 2016

and Muirhead, 1987; Leung, 1994; Lo and Leung, 1996). Analogously, the widely used adjusted R^2 also takes negative values with positive probability, whence the similar problem on how to provide smooth improvements.

To derive positive, and perhaps smooth, estimators improving on the unbiased estimator, Kubokawa, Robert and Saleh (1993) used Stein and chisquare identities. When we apply the same arguments to the estimation of a function of the multiple correlation coefficient, it seems harder to derive improved estimators, because the noncentrality parameter of the conditional distribution depends on the random variables. As well, the issue of estimating such noncentrality parameters under a parametric restriction has, to our knowledge, not been addressed in the literature.

Although it is well known that each of the three models: $W \sim \chi^2_p(\lambda)$, $W \sim \text{Fisher}(m, n, \lambda)$, and $W \sim \frac{R^2}{1-R^2}$ admit similar Gamma or Beta type II mixture representations, with Poisson or Negative Binomial mixing distributions, this has not been exploited in assessing frequentist risk and in searching for improved estimators. In this paper, we provide a unified result for improving on the unbiased estimators above under squared error loss, with or without a lower-bound restriction. This is achieved with Theorems 2.1 and 3.1 by considering indeed a general mixture model (see equation 2.1) with a discrete mixing distribution. The improved estimators given by Kubokawa, Robert and Saleh (1993) for noncentral chisquare and F-distributions can be obtained from Theorem 2.1. Not only are the motivating examples (noncentral chisquare, noncentral F, distribution of $\frac{R^2}{1-R^2}$ in classical multiple regression) covered by our results, but so are many more situations including large classes of Gamma mixtures, type II Beta mixtures, cases with a lower bound restriction $\theta \geq \theta_0$ on θ, and more.

The main findings are presented in Sections 2 and 3, respectively for an unrestricted parameter and for a lower-bound restriction. Classes of dominating estimators of a benchmark unbiased estimator are provided and include smooth estimators taking values on the parameter space. The results are illustrated by examples in Sections 2 to 5. A further application, for estimating the above-mentioned $\frac{\rho^2}{1-\rho^2}$, is expanded upon along with further implications in Section 4.

2 Main Results: the unrestricted parameter case

With the objective of unifying findings applicable to estimating noncentrality parameters of both noncentral chisquare and F distributions, as well as the function $\frac{\rho^2}{1-\rho^2}$ where ρ^2 is a coefficient of determination, we consider mixture distribution models for a random variable W with densities of the form

$$f(w;\theta) = \sum_{k=0}^{\infty} p_\theta(k) f_k(w)\,; \theta \in \Theta \subset \mathbb{R}_+\,, w > 0\,, \tag{2.1}$$

where the p_θ is a mass function on $\mathbb{N}$, and the f_k's are density functions on $\mathbb{R}_+$. We consider situations where there exists an unbiased estimator of θ of the form $c_0W - d_0$, $c_0, d_0 > 0$. With such an estimator possessing the drawback of taking values outside the parameter space, we provide with Theorems 2.1 and 3.1 a unified approach to obtain improvements under squared error loss which include smooth estimators taking values on the parameter space.

We assume that the unbiased estimator has finite risk, which equates to the condition $\mathbb{E}(W^2) < \infty$, as well as the following conditions:

(C0) There exists an unbiased estimator of θ of the form $\hat{\theta}^{UB}(W) = c_0W - d_0$, with $c_0, d_0 > 0$.

(C1) For $h(k) = \mathbb{E}(c_0W - d_0|K = k)$, there exists functions $g_k(\cdot)$ such that

$$\frac{d}{dw} g_k(w) = -(c_0 w - d_0 - h(k)) f_k(w), \text{ for all } w, \tag{2.2}$$

with $\lim_{w\to 0} g_k(w) = \lim_{w\to\infty} g_k(w) = 0$, and with $g_{k+1}(w)/g_k(w) \geq f_{k+1}(w)/f_k(w)$ for $w > 0$ and all k.

(C2) (MLR property) For all $k \in \mathbb{N}$, $f_{k+1}(w)/f_k(w)$ is nondecreasing in $w \in \mathbb{R}_+$.

(C3) $\Theta = [0, \infty)$ or $\Theta = [0, M]$ for some M, and $p_0(0) = 1$.

Condition **(C3)** is satisfied in each one of our motivating examples and by many common discrete distributions with support on $\mathbb{N}$, such as Poisson(θ), NBi($\alpha, p = \frac{1}{1+\theta}$), as well as Poisson and Negative Binomial mixtures. For other cases, and namely cases where there is a lower bound restriction $\theta \geq \theta_0$, we obtain a generalization (Theorem 3.1) requiring a MLR property for the p_θ's. Observe that the above conditions imply that: **(i)** $\sum_{k=0}^{\infty} h(k) p_\theta(k) = \theta$, **(ii)** $h(0) = 0$ taking $\theta = 0$ in previous expression **(i)**, **(iii)** $h(k)$ is nondecreasing in k given the increasing MLR property, and **(iv)** $g_k(w) \geq 0$ for all $k \in \mathbb{N}, w > 0$, since $\frac{d}{dw} g_k(w)$ changes signs from $+$ to $-$ as inferred by (2.2). Finally, integrating both sides of (2.2) on the interval $[0, t]$ yields the solution $g_k(t) = (d_0 + h(k))\, F_k(t) - c_0 \int_0^t w\, f_k(w)\, dw$, F_k being the cdf corresponding to f_k. From this, it follows that $\lim_{w\to 0} g_k(w) = \lim_{w\to\infty} g_k(w) = 0$, so that condition **(C1)** could be reformulated without these limits.

Theorem 2.1 *Consider model (2.1) as a mixture model for W and assume conditions* **(C0)**, **(C1)**, **(C2)**, *and* **(C3)**. *Then, estimators of the form $\hat{\theta}_\psi(W) = c_0W - \psi(W)$ of θ dominate $\hat{\theta}^{UB}(W)$ under squared error loss if the function $\psi(w)$ satisfies the following conditions:*

(a) *$\psi(w)$ is absolutely continuous, nondecreasing in w, and $\lim_{w\to\infty} \psi(w) = d_0$,*

(b) *$\psi(w) \geq \psi_0(w)$, where*

$$\psi_0(w) = c_0 \int_0^w t f_0(t) dt \Big/ \int_0^w f_0(t) dt.$$

Proof. The risk difference between $\hat{\theta}^{UB}$ and $\hat{\theta}_\psi$ is

$$\Delta(\theta) = E[(\hat{\theta}^{UB} - \theta)^2] - E[(\hat{\theta}_\psi - \theta)^2] = 2I_1 - I_2,$$

where $I_1 = E[(c_0W - d_0 - \theta)(\psi(W) - d_0)]$ and $I_2 = E[(\psi(W) - d_0)^2]$.

Noting that $E[c_0W - d_0 - \theta] = 0$, we write I_1 as

$$\begin{aligned} I_1 =& \sum_{k=0}^{\infty} p_\theta(k) \int_0^\infty \{c_0 w - d_0 - h(k)\}\psi(w) f_k(w) dw \\ &+ \sum_{k=0}^{\infty} p_\theta(k)\{h(k) - \theta\} \int_0^\infty \psi(w) f_k(w) dw \\ =& I_{11} + I_{12} \qquad \text{(say)}\,. \end{aligned}$$

Using **(C1)** and integration by parts, we get

$$\begin{aligned} I_{11} =& \sum_{k=0}^{\infty} p_\theta(k) \int_0^\infty \psi(w) \frac{d}{dw}\{-g_k(w)\} dw \\ =& \sum_{k=0}^{\infty} p_\theta(k) \int_0^\infty \psi'(w) g_k(w) dw = \int_0^\infty \psi'(w) g(w;\theta) dw, \end{aligned}$$

with $g(w;\theta) = \sum_{k=0}^{\infty} p_\theta(k)\, g_k(w)$. In evaluating I_{12}, let K be a random variable having mass function p_θ. Observe that $\int_0^\infty \psi(w) f_k(w) dw$ is nondecreasing in k by virtue of the MLR property **(C2)** and condition **(a)** of the Theorem. With h also nondecreasing, the covariance inequality implies that

$$\begin{aligned} I_{12} =& \mathbb{E}\Big[(h(K) - \theta) \int_0^\infty \psi(w) f_K(w) dw\Big] \\ \geq& \, \mathbb{E}[h(K) - \theta]\, \mathbb{E}\Big[\int_0^\infty \psi(w) f_K(w) dw\Big] = 0\,. \end{aligned}$$

For I_2, we re-express it as

$$\begin{aligned} I_2 =& \int_0^\infty (\psi(t) - d_0)^2 f(t;\theta)\, dt \\ =& - \int_0^\infty \{\int_t^\infty \frac{d}{dw} (\psi(w) - d_0)^2\, dw\} f(t;\theta)\, dt \\ =& - 2 \int_0^\infty \{\int_t^\infty \psi'(w)\, (\psi(w) - d_0)\, dw\} f(t;\theta)\, dt \\ =& - 2 \int_0^\infty \psi'(w)\, (\psi(w) - d_0)\, \{\int_0^w f(t;\theta) dt\} dw. \end{aligned}$$

Hence, from the above inequalities, we have

$$\Delta(\theta) = 2I_1 - I_2 \geq 2 \int_0^\infty \Big[g(w;\theta) + (\psi(w) - d_0) \int_0^w f(t;\theta) dt\Big] \psi'(w) dw,$$

which, given that $\psi' \geq 0$ (a.e.), is nonnegative whenever

$$\psi(w) \geq d_0 - \frac{g(w;\theta)}{\int_0^w f(t;\theta)dt}\,, \text{ for all } w > 0, \theta \geq 0. \tag{2.3}$$

We next show that

$$\frac{g(w;\theta)}{\int_0^w f(t;\theta)dt} \geq \frac{g(w;0)}{\int_0^w f(t;0)dt} = \frac{g_0(w)}{\int_0^w f_0(t)dt}\,, \text{ for all } w > 0, \theta \geq 0.$$

This inequality is equivalent to

$$\sum_{k=0}^{\infty} p_\theta(k) \int_0^w g_0(w) f_0(t) \Big\{ \frac{g_k(w)}{g_0(w)} - \frac{f_k(t)}{f_0(t)} \Big\} dt \geq 0,$$

which holds from **(C1)** and **(C2)**. Thus, we get the sufficient condition for dominance:

$$\psi(w) \geq d_0 - \frac{g_0(w)}{\int_0^w f_0(t)dt}, \text{ for all } w > 0.$$

Finally, using **(C1)** again, we get $d_0 \int_0^w f_0(t)dt - g_0(w) = c_0 \int_0^w t f_0(t)dt$, which completes the proof. □

Out of the class in Theorem 2.1, we extract three types of improved and non-negative estimators.

(I) *The estimator* $\hat{\theta}_{\psi_0}$. First, it is noted that the function ψ_0 satisfies conditions **(a)** and **(b)**, and $\hat{\theta}_{\psi_0}$ is thus a smooth estimator dominating the unbiased estimator $\hat{\theta}^{UB}$. Indeed, from **(C1)**, it follows that $\lim_{w\to\infty} \psi_0(w) = c_0 \int_0^\infty t f_0(t)dt = d_0$. As well, we have $\psi(t) = c_0 \mathbb{E}_{\theta=0}[W|W \leq t]$ which is nondecreasing in $t > 0$. From the above representation, we also infer that $\psi_0(w) < c_0 w$ for all $w > 0$ and, thus, $\hat{\theta}_{\psi_0}$ takes positive values only. It is also verified from the steps of the proof of Theorem 2.1 that the estimators $\hat{\theta}_{\psi_0}$ and $\hat{\theta}^{UB}$ have matching risks when $\theta = 0$.

(II) *Positive and truncated linear estimators.* Assume that there exists a positive constant ν such that **(i)** $f_0(t)/t^{\nu-1}$ decreases in t. With T having density proportional to $f_0(t)\mathbb{I}_{[0,w]}(t)$, T^* having density proportional to $t^{\nu-1}\mathbb{I}_{[0,w]}(t)$, and the MLR ordering implied by property **(i)**, it follows that

$$\psi_0(w) = \frac{\int_0^w t f_0(t)dt}{\int_0^w f_0(t)dt} = \mathbb{E}[T] \leq \mathbb{E}[T^*] = \frac{\int_0^w t^\nu dt}{\int_0^w t^{\nu-1}dt} = \frac{\nu}{\nu+1} w. \tag{2.4}$$

Now, let

$$\psi_1(w) = \min \Big\{ c_0 \frac{\nu}{\nu+1} w, d_0 \Big\}.$$

This function satisfies conditions **(a)** and **(b)**. The resulting estimator $\hat{\theta}_{\psi_1}$ is positive and dominates $\hat{\theta}^{UB}$.

(III) *Truncation of* $\hat{\theta}_{UB}$ *onto* $[0,\infty)$. Such a truncation, of course, dominates $\hat{\theta}_{UB}$ without recourse to Theorem 2.1, but it is nevertheless true that such an estimator satisfies the conditions of Theorem 2.1. Indeed, with $\psi_0(w) \leq c_0 w$, the function $\psi_2(w) = \min(c_0 w, d_0)$ satisfies conditions **(a)** and **(b)** of Theorem 2.1. Thus, the resulting estimator $\hat{\theta}_{\psi_2} = \max(\hat{\theta}^{UB}, 0)$ is nonnegative and improves on $\hat{\theta}^{UB}$.

We first provide applications of Theorem 2.1 for noncentral chisquare and F distributions. Another key application, concerning the estimation of a function of a multiple correlation coefficient, is postponed and expanded upon in Section 4. A last class of examples, consisting of further mixtures of Gamma and type II Beta distributions are presented in Section 5.

Example 2.1 (noncentral chisquare distribution) Consider W distributed as noncentral chisquare $\chi^2_p(\theta)$ with p degrees of freedom and noncentrality parameter θ. The distribution admits the mixture representation $W|K = k \sim \chi^2_{p+2k}, K \sim$ Poisson $(\theta/2)$, i.e., as in (2.1) with

$$p_\theta(k) = \frac{(\theta/2)^k}{k!} e^{-\theta/2}, \quad f_k(w) = \frac{2^{-p/2-k}}{\Gamma(p/2+k)} w^{p/2+k-1} e^{-w/2}.$$

Condition **(C3)** holds here. With the estimator $W - p$ (i.e., $c_0 = 1, d_0 = p$) unbiased for θ, it is easy to see that **(C1)** holds with $h(k) = \mathbb{E}(W - p|K = k) = 2k$ and with the choice $g_k(w) = 2w f_k(w)$, i.e.,

$$\frac{d}{dw}\, g_k(w) = -(w - p - 2k) f_k(w)\,.$$

Finally, with the family of densities f_k for W having increasing MLR in W, condition **(C2)** is satisfied. Theorem 2.1 thus applies and provides a class of dominating estimators of $\hat{\theta}^{UB}(W) = W - p$, which includes $W - \psi_0(W)$ with ψ_0 given by

$$\psi_0(w) = \int_0^w t^{p/2} e^{-t/2} dt \Big/ \int_0^w t^{p/2-1} e^{-t/2} dt.$$

As well, (2.4) holds for $\nu \geq p/2$ and the positive truncated linear estimators in **(II)** above given by $W - \psi_1(W)$, with $\psi_1(w) = \min\{\alpha w, p\}$ and $p/(p+2) \leq \alpha < 1$, also dominate $\hat{\theta}^{UB}$.

Example 2.2 (noncentral F distribution) Let $\boldsymbol{X}$ and V be mutually independent random variables such that $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_p)$ and $V/\sigma^2 \sim \chi^2_n\,, n > 2$. Let $W = \boldsymbol{X}^\top \boldsymbol{X}/V$. Then, the distribution of $(n/p)W$ is noncentral F with (p, n) degrees of freedom and noncentrality parameter $\theta = \boldsymbol{\mu}^\top \boldsymbol{\mu}/\sigma^2$, and admits the mixture representation $W|K = k \sim \text{Beta2}(k + p/2, n/2), K \sim$ Poisson $(\theta/2)$, as in (2.1) with

$$p_\theta(k) = \frac{(\theta/2)^k}{k!} e^{-\theta/2}, \quad f_k(w) = \frac{1}{B(p/2+k, n/2)} \frac{w^{p/2+k-1}}{(1+w)^{(n+p)/2+k}}\,.$$

Conditions **(C3)** and **(C0)** hold here with the estimator $(n-2)W-p$ (i.e., $c_0 = n-2, d_0 = p$) unbiased for θ. Condition **(C1)** holds with $h(k) = \mathbb{E}((n-2)W - p|K = k) = 2k$ and $g_k(w) = 2w(1+w)f_k(w)$ as

$$\frac{d}{dw} g_k(w) = -((n-2)w - p - 2k) f_k(w)\,.$$

With the family of densities f_k for W having increasing MLR in W, condition **(C2)** is also satisfied and Theorem 2.1 thus applies and provides a class of dominating estimators of $\hat{\theta}^{UB}(W) = (n-2)W - p$, which includes $(n-2)W - \psi_0(W)$ with ψ_0 given by

$$\psi_0(w) = (n-2) \int_0^w t^{p/2}/(1+t)^{(n+p)/2} dt / \int_0^w t^{p/2-1}/(1+t)^{(n+p)/2} dt.$$

We point that series representation are available via incomplete Beta functions here (see Appendix as well) , and via incomplete Gamma functions in the previous example. Finally, condition (2.4) is satisfied for $\nu \geq p/2$ and the positive truncated linear estimators, given by $(n-2)W - \psi_1(W)$ with $\psi_1(w) = \min\{(n-2)\alpha w, p\}$ and $p/(p+2) \leq \alpha < 1$, also dominate $\hat{\theta}^{UB}$.

Remark 2.1 Condition **(C1)** is satisfied for densities of the form $f_k(t) = c_k(u(t))^k f_0(t)$, with nondecreasing $u(\cdot)$, differentiable $u(\cdot)$ and $f_0(\cdot)$, and with $h(k) = \mathbb{E}(c_0 W - d_0 | K = k) = ak$, for some $a > 0$. More precisely, we claim that the above set-up implies that one can take

$$g_k(w) \;=\; a\, \frac{u(w)}{u'(w)}\, f_k(w)\,, w > 0, \tag{2.5}$$

and that such a choice satisfies the other conditions $\lim_{w\to 0} g_k(w) = \lim_{w\to\infty} g_k(w) = 0$, $g_{k+1}(w)/g_k(w) \geq f_{k+1}(w)/f_k(w)$ for $w > 0$ and all k, as well. To justify this, we will use completeness of the family of densities $\{f_k\}$ (i.e., we are in presence of a natural exponential family with parameter k) for inferring that

$$\mathbb{E}(t(W)|K = k) \;= ak \text{ for all } k \text{ implies } t(w) = c_0 w - d_0 \text{ almost everywhere.}$$

With $\frac{d}{dw} f_k(w) \;=\; f_k(w) \left(k \frac{u'(w)}{u(w)} + \frac{f_0'(w)}{f_0(w)} \right)$, we obtain for g_k in (2.5):

$$\frac{d}{dw} g_k(w) \;=\; f_k(w) \left(h(k) + a \frac{u(w)}{u'(w)} \frac{f_0'(w)}{f_0(w)} \;+\; a \frac{d}{dw} \frac{u(w)}{u'(w)} \right)\,. \tag{2.6}$$

Since $\int_0^\infty g_k(w)\, dw \;=\; 0$, we obtain from the above, setting $T(w) = -a \frac{u(w)}{u'(w)} \frac{f_0'(w)}{f_0(w)} - a \frac{d}{dw} \frac{u(w)}{u'(w)}$, that $\mathbb{E}(h(k) - T(W)|K = k) = 0$ for all k. By completeness, this now implies that $T(w) = c_0 W - d_0$ almost everywhere, and indeed that $\frac{d}{dw} g_k(w) \;=\; f_k(w)\, \{h(k) - (c_0 w - d_0)\}$, which is (2.2). With the above, one readily recovers the $g_k's$ of the previous two examples with $u(w) = w$ and $u(w) = \frac{w}{1+w}$ respectively for the Gamma and Beta type II cases.

3 Improvements under a lower bound restriction

We provide here an extension of the main result of Section 2 to cases with a lower bound parametric restriction $\theta \geq \theta_0 > 0$. A similar finding to Theorem 2.1 is obtained, but the development is more delicate as the distribution of K under θ_0 is, as opposed to the case $\theta = 0$, non-degenerate. We require an additional increasing MLR assumption for the family $\{p_\theta, \theta \geq \theta_0\}$ of mass functions for K. The additional applications are novel and include the estimation of noncentrality parameters under a lower bound restriction for chisquare and F distributions, as well as a lower bounded squared multiple correlation coefficient.

Theorem 3.1 *Consider model (2.1) as a mixture model for W and assume conditions* **(C0)**, **(C1)**, **(C2)**. *Further assume that the family of mass functions p_θ for K has increasing MLR in K for parameter θ. Then, estimators of the form $\hat{\theta}_\psi(W) = c_0 W - \psi(W)$ of θ dominate $\hat{\theta}^{UB}(W)$ for squared error loss, under the parametric restriction $\theta \geq \theta_0$, if the function $\psi(w)$ satisfies the following conditions:*
(a) *$\psi(w)$ is absolutely continuous, nondecreasing in w, and $\lim_{w\to\infty} \psi(w) = d_0$,*
(b) *$\psi(w) \geq \psi_{\theta_0}(w)$, where*

$$\psi_{\theta_0}(w) = c_0 \frac{\int_0^w t f(t;\theta_0) dt}{\int_0^w f(t;\theta_0)\, dt} - \mathbb{E}_w\left(h(L)\right) ,$$

where L is a discrete random variable with mass function proportional to $p_{\theta_0}(\ell)\, F_\ell(w)$, with $F_\ell(w) = \int_0^w f_\ell(t)\, dt$.

Proof. Proceed exactly as in the proof of Theorem 2.1 up to expression (2.3) to show that $\hat{\theta}_\psi$ dominates $\hat{\theta}^{UB}$ whenever

$$\psi(w) \geq d_0 - \frac{g(w;\theta)}{\int_0^w f(t;\theta) dt}, \text{ for all } w > 0, \theta \geq \theta_0. \tag{3.1}$$

Our proof will be complete by establishing that $\psi_{\theta_0}(w)$ is, for all $w > 0$, an upper bound of the r.h.s. of (3.1). This will follow immediately from the identities:

(I) $\frac{g(w;\theta)}{\int_0^w f(t;\theta)\, dt} \geq \frac{g(w;\theta_0)}{\int_0^w f(t;\theta_0)\, dt}$ for all $w > 0$, $\theta \geq \theta_0$;

(II) and $d_0 - \frac{g(w;\theta_0)}{\int_0^w f(t;\theta_0)\, dt} = \psi_{\theta_0}(w)$.

For identity **(I)**, we first have for $w > 0, \theta \geq \theta_0$,

$$\begin{aligned} \frac{g(w;\theta)}{f(w;\theta)} &= \frac{\sum_k \frac{g_k(w)}{f_k(w)} f_k(w)\, p_\theta(k)}{\sum_k f_k(w)\, p_\theta(k)} \\ &\geq \frac{\sum_k \frac{g_k(w)}{f_k(w)} f_k(w)\, p_{\theta_0}(k)}{\sum_k f_k(w)\, p_{\theta_0}(k)} \\ &= \frac{g(w;\theta_0)}{f(w;\theta_0)}\,, \end{aligned} \tag{3.2}$$

given the incresing MLR property of the p_θ's, and thus of the mass functions proportional to $f_k(w)p_\theta(k), k \in \mathbb{N}$, and given that the ratios $\frac{g_k(w)}{f_k(w)}$ are nondecreasing in k by condition (**C2**).
Secondly, we make use of Lemma 5.1, which is stated and proven in the Appendix, and which establishes an MLR property for mixtures such as the densities $f(\cdot;\theta)$. Starting from (3.2), we obtain for $w > 0, \theta \geq \theta_0$:

$$\begin{aligned}
\frac{g(w;\theta)}{g(w;\theta_0)} &\geq \frac{f(w;\theta)}{f(w;\theta_0)} \\
&= \frac{\int_0^w \frac{f(w;\theta)}{f(w;\theta_0)} f(t;\theta_0)dt}{\int_0^w f(t;\theta_0)dt} \\
&\geq \frac{\int_0^w \frac{f(t;\theta)}{f(t;\theta_0)} f(t;\theta_0)dt}{\int_0^w f(t;\theta_0)dt} \\
&= \frac{\int_0^w f(t;\theta)dt}{\int_0^w f(t;\theta_0)dt},
\end{aligned}$$

which is **(I)**.

For identity **(II)**, use condition **(C1)** to first obtain

$$\begin{aligned}
-\sum_k p_{\theta_0} \int_0^w \frac{d}{dt} g_k(t)\,dt &= -\sum_k p_{\theta_0}(k)\, g_k(w) \\
&= \sum_k p_{\theta_0}(k) \int_0^w (c_0 t - d_0) f_k(t)\,dt - \sum_k h(k) p_{\theta_0}(k)\, F_k(w)\,.
\end{aligned}$$

Division by $\int_0^w f(t;\theta_0)dt$ and a shuffling of terms establishes **(II)** and completes the proof of Theorem 3.1. ☐

Among the class of dominating estimators, interest lies in the smooth estimator $\hat{\theta}_{\psi_{\theta_0}}$. Observe that, for $w > 0$,

$$\mathbb{E}_w[h(L)] = \frac{E_{\theta_0}[h(K)F_K(w)]}{E_{\theta_0}[F_K(w)]} \leq E_{\theta_0}[h(K)] = \theta_0\,,$$

where $K \sim p_{\theta_0}$, and where we made use of the covariance inequality with $h(k)$ increasing in k and $F_k(w)$ decreasing in k by virtue of MLR property **(C2)**. Along with the inequality $\frac{\int_0^w t f(t;\theta_0)dt}{\int_0^w f(t;\theta_0)\,dt} \leq w$, we obtain that

$$\hat{\theta}_{\psi_{\theta_0}}(w) = c_0 w - \psi_{\theta_0}(w) \geq c_0 w - (c_0 w - \theta_0) = \theta_0\,,$$

for all $w > 0$. It is thus the case that $\hat{\theta}_{\psi_{\theta_0}}$ takes values solely on the restricted parameter space $\theta \geq \theta_0$.

As shown by the next result, the estimator also satisfies the dominance conditions of Theorem 3.1 for a class of mixture models in (2.1), which include Gamma and Beta type II mixtures.

Proposition 3.1 **(A)** *Under the set-up and conditions of Theorem 3.1, for mixture models (2.1) with densities f_k of the form $f_k(t) \propto (u(t))^k f_0(t)$, with $u(\cdot)$ a nondecreasing function, the estimator $\hat{\theta}_{\psi_{\theta_0}}(W) = c_0 W - \psi_{\theta_0}(W)$ dominates $\hat{\theta}^{UB}(W)$ under squared error loss under the parametric restriction $\theta \geq \theta_0$, and provided that:*
(i) $\frac{g_k(w)}{F_k(w)}$ *is a nondecreasing function of $w \in \mathbb{R}_+$.*

(B) *For $f_k \sim$ Gamma$(a+k, b)$, $a > 1$, and $f_k \sim$ Beta type II$(a+k, b)$, $a, b > 1$, part* **(A)** *applies with condition* **(i)** *verified.*

Proof. **(A)** First, we have by the unbiased property $\lim_{w\to\infty} c_0\, \mathbb{E}_{\theta_0}[T|T \leq w] = c_0\, \mathbb{E}[T] = \theta_0 + d_0$, where $T \sim f(\cdot;\theta_0)$. As well, it is easy to see that $\lim_{w\to\infty} \mathbb{E}_w(h(L)) = \sum_\ell p_{\theta_0}(\ell) h(\ell) = \theta_0$. We thus obtain $\lim_{w\to\infty} \psi_{\theta_0}(w) = d_0$.

There remains to show that ψ_{θ_0} is nondecreasing on $\mathbb{R}_+$. From representation **(II)** contained in the proof of Theorem 3.1, it will suffice to show that

$$\frac{g(w;\theta_0)}{\int_0^w f(t;\theta_0)dt} = \frac{\sum_k g_k(w)\, p_{\theta_0}(k)}{\sum_k F_k(w)\, p_{\theta_0}(k)} \tag{3.3}$$

is nonincreasing in w with $F_k(w) = \int_0^w f_k(t)\, dt$. We now make use of Lemmas 5.2 and 5.4, which are stated and proven in the Appendix. The nonincreasingness in w of (3.3) will follow from Lemma 5.4 by setting $\gamma_k = g_k$ and $\beta_k = F_k$ as long as its conditions **(i)** and **(ii)** are established. Condition **(i)** is assumed here, while condition **(ii)** is a direct consequence of Lemma 5.2 which tells us that, for $w_1 > w_0$, ratios $\frac{F_k(w_1)}{F_k(w_0)}$ are nondecreasing in k under the conditions of our Theorem. The proof of part **(A)** is thus complete.

(B) Both the Gamma and Beta type II f_k densities are of the form $f_k(t) \propto (u(t))^k f_0(t)$ with $u(t) = t, f_0(t) = t^{a-1}e^{-t/b}$; $u(t) = t/(1+t), f_0(t) = t^{a-1}/(1+t)^{a+b}$ respectively. Hence, we can apply part **(A)** and there remains to test condition **(i)**. For the Gamma case, condition **(i)**, which requires that $\frac{g_k(w)}{F_k(w)}$ be nonincreasing in w, is a consequence of Lemma 5.3 since, for the **Gamma case**, we have $g_k(w) = 2w f_k(w)$ as seen in Example 2.1.
For the Beta case, it will suffice to prove that $\frac{g(w)}{F(w)}$ is nonincreasing in w for the Beta type II$(a' = a+k, b)$ case with $a, b > 1$, i.e., for $f(w) \propto \frac{w^{a'-1}}{(1+w)^{a'+b}}, w > 0$. We have $F'(w) = f(w)$, $w(1+w)f'(w) = ((a'-1) - (b+1)w)\, f(w)$ and we use these expressions throughout the proof. For $g(w) = 2w(1+w)f(w)$, as in Example 2.2 we can write

$$\frac{d}{dw}\frac{g(w)}{F(w)} = -\frac{2(1+w)f(w)}{F^2(w)}\, H(w)\,,$$

with

$$H(w) \;=\; w\,f(w) \;+\; \frac{(b-1)w}{1+w}\,F(w) \;-\; \frac{a'}{1+w}\,F(w)\,.$$

We need to show that $H(w) \geq 0$ for all $w \geq 0$. Since $H(0) = 0$, it will suffice to show that H is nondecreasing on $\mathbb{R}_+$. A further differentiation tells us that $\operatorname{sgn} H'(w) \;=\; \operatorname{sgn} T(w)$, with

$$T(w) \;=\; \frac{H'(w)}{(1+w)^2} \;=\; (b+a'-1)\,F(w) \;-\; w(1+w)\,f(w)\,.$$

Yet another differentiation tells us that $T'(w) \;=\; (b-1)\,(1+w)\,f(w) \;>0$. Since $T(0) = 0$ and $\lim_{w\to\infty} T(w) = b + a' - 1 > 0$, we infer that both $T(w)$ and $H'(w)$ are nonnegative for $w \geq 0$, which establishes the result. □

We conclude this section by illustrating how our above finding applies to estimating a lower bounded chisquared noncentrality parameter.

Example 3.1 (lower bounded parameter of a noncentral chisquare distribution) As a follow-up to Example 2.1, consider $W \sim \chi^2_p(\theta)$ with $\theta \geq \theta_0$. For providing improvements on the unbiased estimator $W - p$, Theorem 3.1 applies just as Theorem 2.1 did for Example 2.1, along with the additional MLR property of the Poisson mixing distribution for K. Furthermore, Proposition 4.7 tells us that the function ψ_{θ_0} satisfies the conditions of Theorem 3.1, namely the nondecreasing property. We thus infer that the unbiased estimator $\hat{\theta}^{UB}(W)$ is dominated by $W - \psi_{\theta_0}(W)$, as well as estimators $W - \psi(W)$ with ψ satisfying conditions **(a)** and **(b)** of Theorem 3.1. The function ψ_{θ_0} is given by

$$\psi_{\theta_0}(t) = \mathbb{E}_{\theta_0}(W|W \leq t) - \mathbb{E}_t(h(L))\,,$$

with $h(\ell) = 2\ell$, and where L has mass function proportional to $\frac{e^{-\theta_0/2}(\theta_0/2)^l}{l!}\,F_l(t)$, F_l the cdf of a $\chi^2_{p+2l}(0)$ distribution. Observe, that the first term corresponds to the expectation of a truncated $\chi^2_p(\theta_0)$, while the second term simplifies to

$$\mathbb{E}_t(2L) \;=\; \theta_0 \frac{\mathbb{P}(\chi^2_{p+2}(\theta_0) \leq t)}{\mathbb{P}(\chi^2_p(\theta_0) \leq t)}\,,$$

with the Poisson identity $\mathbb{E}(LF_L(t)) \;=\; \frac{\theta_0}{2}\,\mathbb{E}(F_{L+1}(t))$. Finally, it is easily seen that the trivial improvement $\max(\theta_0, W-p)$ also satisfies the dominance conditions of Theorem 3.1, as well as convex linear combinations of the form $\alpha \max(\theta_0, W-p) + (1-\alpha)\,\hat{\theta}_{\psi_{\theta_0}}(W); \alpha \in (0,1)$.

4 Estimation a Function of the Multiple Correlation Coefficient

4.1 The problem and its motivation

We now expand on the problem of estimating a function of the multiple correlation coefficient. Let Y be a scalar random variable, and let $\boldsymbol{Z} = (Z_1, \ldots, Z_{m-1})^{\mathrm{T}}$ be an $(m-1)$-

dimensional random variable. Suppose that $(Y, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$ has a normal distribution $\mathcal{N}_m(\boldsymbol{\xi}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \boldsymbol{\xi}_2 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

The conditional expectation of Y given $\boldsymbol{Z}$ may be written as

$$\mathbb{E}[Y \mid \boldsymbol{Z}] = \alpha + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z} = \alpha + \beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_{m-1} Z_{m-1},$$

where $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21} = (\beta_1, \ldots, \beta_{m-1})^{\mathrm{T}}$ and $\alpha = \xi_1 - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\xi}_2$. This represents a linear regression model with the dependent variable Y regressed by independent variables $Z_1, \ldots, Z_{m-1}$. The population multiple correlation coefficient is defined by

$$\rho^2 = \boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}/\sigma_{11}, \tag{4.1}$$

which may be rewritten as $\rho^2 = \theta/(1+\theta)$, with

$$\theta = \boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}/\sigma_{11.2} = \rho^2/(1-\rho^2), \tag{4.2}$$

for $\sigma_{11.2} = \sigma_{11} - \boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$.

Let $(Y_1, \boldsymbol{Z}_1^{\mathrm{T}})^{\mathrm{T}}, \ldots, (Y_N, \boldsymbol{Z}_N^{\mathrm{T}})^{\mathrm{T}}$ be a random sample from the m-dimensional normal distribution $\mathcal{N}_m(\boldsymbol{\xi}, \boldsymbol{\Sigma})$. Let $\overline{Y} = N^{-1}\sum_{i=1}^N Y_i$, $\overline{\boldsymbol{Z}} = N^{-1}\sum_{i=1}^N \boldsymbol{Z}_i$, $s_{11} = \sum_{i=1}^N (Y_i - \overline{Y})^2$, $\boldsymbol{s}_{21} = \sum_{i=1}^N (\boldsymbol{Z}_i - \overline{\boldsymbol{Z}})(Y_i - \overline{Y})$ and $\boldsymbol{S}_{22} = \sum_{i=1}^N (\boldsymbol{Z}_i - \overline{\boldsymbol{Z}})(\boldsymbol{Z}_i - \overline{\boldsymbol{Z}})^{\mathrm{T}}$. Based on these statistics, consider the $m \times m$ matrix

$$\boldsymbol{S} = \begin{pmatrix} s_{11} & \boldsymbol{s}_{12} \\ \boldsymbol{s}_{21} & \boldsymbol{S}_{22} \end{pmatrix}.$$

Then $\boldsymbol{S}$ has a Wishart distribution $\mathcal{W}_{m-1}(n, \boldsymbol{\Sigma})$ with degrees of freedom $n = N-1$. Let

$$\begin{aligned} &W = \boldsymbol{s}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{s}_{21}/s_{11.2}, \ s = s_{11.2}/\sigma_{11.2}, \\ &\boldsymbol{U} = \boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{s}_{21}/\sqrt{\sigma_{11.2}}, \ \boldsymbol{V} = \boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{S}_{22}\boldsymbol{\Sigma}_{22}^{-1/2}, \\ &\boldsymbol{\eta} = \boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\sigma}_{21}/\sqrt{\sigma_{11.2}}. \end{aligned}$$

From multivariate normal distribution properties, it follows that s is independent of $(\boldsymbol{U}, \boldsymbol{V})$, $s \sim \chi^2_{n-m+1}$, $\boldsymbol{U} \mid \boldsymbol{V} \sim \mathcal{N}_{m-1}(\boldsymbol{V}\boldsymbol{\eta}, \boldsymbol{V})$ and $\boldsymbol{V} \sim \mathcal{W}_{m-1}(n, \boldsymbol{I})$. Note that

$$\boldsymbol{X} \mid \boldsymbol{V} \sim \mathcal{N}_{m-1}(\boldsymbol{V}^{1/2}\boldsymbol{\eta}, \boldsymbol{I}) \quad \text{for} \quad \boldsymbol{X} = \boldsymbol{V}^{-1/2}\boldsymbol{U}. \tag{4.3}$$

Thus, $W = \boldsymbol{s}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{s}_{21}/s_{11.2} = \|\boldsymbol{X}\|^2/s$.

It is worth commenting on the estimation of θ, some of which was stated in Muirhead (1985). Using the conditional variance $\mathrm{Var}(Y \mid \boldsymbol{Z})$ and unconditional variance $\mathrm{Var}(Y)$, we can express ρ^2 and θ as

$$\begin{aligned} \rho^2 &= \frac{\mathrm{Var}(Y) - \mathrm{Var}(Y \mid \boldsymbol{Z})}{\mathrm{Var}(Y)} = 1 - \frac{\mathrm{Var}(Y \mid \boldsymbol{Z})}{\mathrm{Var}(Y)}, \\ \theta &= \frac{\mathrm{Var}(Y) - \mathrm{Var}(Y \mid \boldsymbol{Z})}{\mathrm{Var}(Y \mid \boldsymbol{Z})} = \frac{\mathrm{Var}(Y)}{\mathrm{Var}(Y \mid \boldsymbol{Z})} - 1, \end{aligned}$$

so that the difference lies in the estimated quantity $\text{Var}(Y \mid \boldsymbol{Z})/\text{Var}(Y)$ and its reciprocal $\text{Var}(Y)/\text{Var}(Y \mid \boldsymbol{Z})$. The R-square statistic, given by $R^2 = W/(1+W)$ is the MLE of ρ^2, and is used to measure fitness of dependent variables regressed on independent variables in the multiple linear regression model. As pointed out in Muirhead (1985) however, the distribution of R^2 is highly skewed on $(0,1)$ and R^2 overestimates ρ^2. Olkin and Pratt (1958) derived the minimum variance unbiased estimator of ρ^2 which is expressible in terms of the hypergeometric function ${}_2F_1$. Transforming to W, whose range is on $(0,\infty)$, yields more symmetric distributions and the estimation of θ is somewhat easier to handle and more tractable. For instance, the unbiased estimator of θ is a linear function of W. When θ can be estimated efficiently, substitution into $\rho^2 = \theta/(1+\theta)$ leads to a reasonable estimate of ρ^2.

4.2 Improved estimators of $\theta = \frac{\rho^2}{1-\rho^2}$

We apply Theorem 2.1 to the estimation of θ. It is known (see Gurland 1968, Muirhead 1982, theorem 5.2.5) that the conditional distribution of $\|\boldsymbol{X}\|^2$ given K is the chisquare distribution with $m-1+2K$ degrees of freedom χ^2_{m-1+2K}, and K has a negative binomial distribution with

$$p_\theta(k) = \frac{\Gamma(n/2+k)}{k!\,\Gamma(n/2)} \frac{\theta^k}{(1+\theta)^{n/2+k}}\,. \tag{4.4}$$

Then, the probability density function of W may be expressed as in (2.1) with the above mixing p_θ's and

$$f_k(w) = \frac{1}{B((m-1)/2+k),(n-m+1)/2)} \frac{w^{(m-1)/2+k-1}}{(1+w)^{n/2+k}}\,. \tag{4.5}$$

Condition **(C0)** is verified with an unbiased estimator of θ given by $\hat{\theta}^{UB} = c_0 W - d_0$, with $c_0 = (n-m-1)/n$, $d_0 = (m-1)/n$. Condition **(C1)** holds here with $h(k) = \mathbb{E}(c_0 W - d_0 | K = k) = 2k/n$ and $g_k(w) = (2/n)w(1+w)f_k(w)$. Alternatively, this follows as a consequence of Remark 2.1. Finally, conditions **(C3)** and **(C2)** are satisfied with $f_{k+1}(w)/f_k(w) \propto (w/(1+w)$ increasing in w. The function $\psi(w)$ in Theorem 2.1 is written as

$$\psi_0(w) = c_0 \frac{\int_0^w t^{(m-1)/2}/(1+t)^{n/2}\mathrm{d}t}{\int_0^w t^{(m-1)/2-1}/(1+t)^{n/2}\mathrm{d}t}\,. \tag{4.6}$$

From Theorem 2.1, we get the following.

Corollary 4.1 *In the multiple regression framework of Section 4.1 or, equivalently, for model (2.1) with the p_θ's and f_k's as defined in (4.4) and (4.5), the estimator $\hat{\theta}_\psi(W) = c_0 W - \psi(W)$ improves on the unbiased estimator $\hat{\theta}^{UB} = c_0 W - d_0 = \frac{(n-m-1)}{n} W - \frac{m-1}{n}$ if the following conditions on $\psi(w)$ hold:*

(a) *$\psi(w)$ is nondecreasing in w and $\lim_{w\to\infty} \psi(w) = d_0$.*

(b) *$\psi(w) \geq \psi_0(w)$.*

As expanded upon in Section 2 following Theorem 2.1, we can provide at least three types of improved estimators from the corollary, which include $\hat{\theta}_{\psi_0}$. An alternative expression for $\psi_0(W)$, which is justified in the Appendix, is as follows: [2]

$$\hat{\theta}_{\psi_0}(w) = c_0 w \sum_{k=0}^{\infty} \frac{k+1}{k+1+(m-1)/2} A_k \Big/ \sum_{k=0}^{\infty} A_k, \tag{4.7}$$

where

$$A_k = \frac{\Gamma(n/2+k)}{\Gamma((m-1)/2+k+1)} \frac{w^k}{(1+w)^k}.$$

Other dominating estimators satisfying the conditions of Corollary 4.1 are provided by $\psi_1(w) = \min\{c_0(m-1)w/(m+1), d_0\}$. This occurs as expression (2.4) holds for $\nu = (m-1)/2$, and we get the dominating truncated linear estimator

$$\begin{aligned} \hat{\theta}_{\psi_1}(W) &= \max\left\{c_0 W - d_0,\ c_0 \frac{2}{m+1} W\right\} \\ &= \frac{1}{n} \max\left\{(n-m-1)W - (m-1), (n-m-1)\frac{2}{m+1} W\right\}. \end{aligned} \tag{4.8}$$

We conclude this section by pointing out that Theorem 3.1 leads to dominating estimators of $\hat{\theta}^{UB}$ under the lower bound constraint $\theta \geq \theta_0$, in other words when $\rho^2 \geq \frac{\theta_0}{1+\theta_0}$.

4.3 Modifying the R^2 and the adjusted R^2

The $R^2 = \frac{W}{1+W}$ and adjusted R^2 (Fisher, 1924) statistics are much used measures of fitness of dependent variables regressed on independent variables in the multiple linear regression model. Efficient estimators $\hat{\theta}$ of $\theta = \frac{\rho^2}{1-\rho^2}$, such as those given in the previous Subsection, lead to interesting estimators $\frac{\hat{\theta}}{1+\hat{\theta}}$ of ρ^2. [3] As an illustration, the estimator

$$\hat{\theta}^{UB*}(W) = \frac{n-m+1}{n} W - \frac{m-1}{n} = \hat{\theta}^{UB}(W) + \frac{2W}{n},$$

is a minor modification of $\hat{\theta}^{UB}$ that leads to the adjusted R^2 given by

$$\overline{R}^2 = \frac{(n-m+1)W - (m-1)}{n + (n-m+1)W - (m-1)} = 1 - \frac{n}{n+1-m}(1-R^2),$$

which is widely used for variable selection in econometrics. This means that the adjusted R-square, which is obtained by adjusting the degrees of freedom, can be derived by substituting the adjusted unbiased estimator of θ. However, it is noted that the unbiased and the adjusted unbiased estimator of θ take negative values with positive probability.

[2] http://dlmf.nist.gov/8.17)

[3] We point out that, working directly with the estimation of ρ^2 under squared error loss, Marchand (2001) obtained affine linear estimators that dominate the adjusted R^2 for $n \geq 6$, as well as estimators that dominate the mle R^2 for $m \geq 8$.

It is indeed such properties that motivate the search for nonnegative improvements and represent a main theme of this paper. In this regard, the positive truncated linear estimator $\hat{\theta}_{\psi_1}$, as well as the estimator $\hat{\theta}_{\psi_1}$, are proposed here as competitors. The former, as given in (4.8), yields the alternative

$$\overline{R}^2_{\psi_1} = \max\Big\{\overline{R}^2,\ \frac{2(n-m+1)W/(m+1)}{n+2(n+1-m)W/(m+1)}\Big\}.$$

It is noted that $\overline{R}^2_{\psi_1}$ is positive almost everywhere, while $\overline{R}^2$ takes negative values with positive probability (for all ρ^2).

Finally, it is interesting to investigate the connection of $\hat{\theta}^{UB*}$ to the Akaike Information Criterion (AIC). Note that $s_{11.2}$ corresponds to $\|\boldsymbol{y}-\hat{\boldsymbol{y}}\|^2$ for the predictor $\hat{\boldsymbol{y}}$ based on the model with $m-1$ regressors. Let $\hat{\sigma}_0^2 = s_{11}/(N-1) = \sum_{i=1}^N (Y_i - \overline{Y})^2/(N-1)$. Since $W = s_{11}/s_{11.2} - 1$, $\hat{\theta}^{UB*}$ is expressed as

$$\hat{\theta}^{UB*} = (N-m)\frac{\hat{\sigma}_0^2}{\|\boldsymbol{y}-\hat{\boldsymbol{y}}\|^2} - 1,$$

which can be approximated as

$$\begin{aligned}-N\log(\hat{\theta}^{UB*}+1) &= N\log\Big(\frac{\|\boldsymbol{y}-\hat{\boldsymbol{y}}\|^2}{N}\Big) - N\log\Big(1-\frac{m}{N}\Big) - N\log\hat{\sigma}_0^2\\ &= N\log\Big(\frac{\|\boldsymbol{y}-\hat{\boldsymbol{y}}\|^2}{N}\Big) + m - N\log\hat{\sigma}_0^2 + O(N^{-1}).\end{aligned}$$

On the other hand, AIC is

$$\mathrm{AIC}(m) = N\log\Big(\frac{\|\boldsymbol{y}-\hat{\boldsymbol{y}}\|^2}{N}\Big) + 2(m+1).$$

Noting that $\hat{\sigma}_0^2$ does not depend on the selected model, it can be seen that $\hat{\theta}^{UB*}$ leads to a smaller penalty than AIC.

5 Concluding Remarks: additional examples

The main theoretical developments (Theorems 2.1, 3.1), which apply for mixture models (2.1), have been applied to Gamma and Beta type Ii f_k's mixed with respect to Poisson and Negative Binomial mixing parameters. But, the results are broader and various other applications can be envisaged.

For instance, consider in (2.1) the Gamma mixture model with a Poisson mixture mixing distribution

$$W|\lambda, K \sim \mathrm{Gamma}(a+K, b)\,,\ K|\lambda,\theta \sim \mathrm{Poisson}(\lambda)\,,\ \lambda|\theta \sim h_\theta\,, \tag{5.1}$$

with $a, b > 0$, $\theta \geq 0$, and h_θ being a mass function or a density on $\mathbb{R}_+$ with increasing MLR and which is degenerate at 0 when $\theta = 0$. The above includes, of course, Poisson mixing (i.e., λ is degenerate at θ) and Negative Binomial mixing when g_θ is a Gamma density. In general, for model (5.1) and denoting $\gamma(\theta) = \mathbb{E}(\lambda|\theta)$, we have that the estimator $\frac{W}{b} - a$ is an unbiased for $\gamma(\theta)$ that takes negative values with probability one (for all θ). Such estimators can be improved by estimators of the form $\frac{W}{b} - \psi(W)$, just as above. To establish this, all that is required is to amend Theorem 2.1 by replacing θ by $\gamma(\theta)$ throughout. Otherwise, the proofs are identical. A similar development applies for the lower bound restriction as in Section 3 given that the family of $p_\theta' s$ have increasing MLR (i.e., as in Lemma 5.1). Finally, similar inferences apply for Beta mixtures with $W|\lambda, K \sim$ Beta type II$(a+K, b)$, as well as for Gamma mixtures $W|\lambda, K \sim Gamma(a, b+K)$.

Acknowledgements

Tatsuya Kubokawa's research is supported in part by Grant-in-Aid for Scientific Research Nos. 15H01943 and 26330036 from the Japan Society for the Promotion of Science, Eric Marchand's research is supported in part by the Natural Sciences and Engineering Research Council of Canada, and William Strawderman's research is partially supported by a grant from the Simons Foundation (#209035).

Appendix

Here are some useful and possibly familiar lemmas, nevertheless accompanied by proofs for sake of completeness. This first result, as well as stronger versions related to variational properties, have been known for some time (e.g., Brown, Johnstone and MacGibbon, 1981). Here is a direct proof.

Lemma 5.1 *Let $\{f_k, k \in \mathbb{K}\}$ be a family of densities for a continuous random variable W with increasing MLR in W for parameter k. Let $\{p_\theta, \theta \in \Theta\}$ be a family of mass functions for a discrete random variable K with increasing MLR in K for parameter θ. Then, the family of mixture densities*

$$f(w;\theta) = \sum_k p_\theta(k)\, f_k(w)$$

has increasing MLR in W for parameter $\theta \in \Theta$.

Proof. We have for $\theta_1 > \theta_0$:

$$\frac{f(w;\theta_1)}{f(w;\theta_0)} = \mathbb{E}_w\Big(\frac{p_{\theta_1}(J)}{p_{\theta_0}(J)}\Big)\,, \tag{5.2}$$

where J is a discrete random variable with mass function proportional to $f_j(w)p_{\theta_0}(j)$, and where the ratio $\frac{p_{\theta_1}(j)}{p_{\theta_0}(j)}$ is nondecreasing in j by assumption. The result follows with this family of mass functions for J having increasing MLR in J for parameter w, which is itself a consequence of the increasing MLR property of the f_k's. □

Lemma 5.2 *Let $f_k, k \in \mathbb{R}_+$ be a family of densities on $\mathbb{R}_+$ of the form $f_k(t) \propto u(t)^k f_0(t)$, with $u(\cdot)$ a nondecreasing function on $\mathbb{R}_+$. Then, for any $w_1 > w_0$, the ratio $\frac{F_k(w_1)}{F_k(w_0)}$ is a nondecreasing function of $k \in \mathbb{R}_+$.*

Proof. We have

$$\frac{F_k(w_0)}{F_k(w_1)} = \mathbb{E}\left(\mathbb{I}_{(0,w_0)}(Z)\right), \text{ with } Z \sim u(z)^k f_0(z)\mathbb{I}_{(0,w_1)}(z)\,.$$

The result follows since this family of densities has an increasing MLR in $u(Z)$, and thus Z, with parameter k, and since $\mathbb{I}_{(0,w_0)}(z)$ is a nonincreasing function of $z \in \mathbb{R}_+$. □

Lemma 5.3 *Let $s_\sigma(x) = \frac{1}{\sigma}\, s_1(\frac{x}{\sigma})$ be a scale family of densities on $\mathbb{R}_+$ for a continuous random variable X with increasing MLR in X. Then, $\frac{x s_1(x)}{S_1(x)}$ is nonincreasing in x, where S_1 is the cdf corresponding to s_1.*

Proof. For $c \in (0,1)$, we have

$$\begin{aligned}
& \frac{cx s_1(cx)}{\int_0^{cx} s_1(t)dt} \geq \frac{x s_1(x)}{\int_0^{x} s_1(t)dt} \\
\iff\ & \frac{s_1(cx)}{\int_0^{x} s_1(ct)dt} \geq \frac{s_1(x)}{\int_0^{x} s_1(t)dt} \\
\iff\ & \int_0^x s_1(x)s_1(t)\{\frac{s_1(cx)}{s_1(x)} - \frac{s_1(ct)}{s_1(t)}\}\,dt\ \geq 0\,,
\end{aligned}$$

since the ratio $\frac{s_1(ct)}{s_1(t)}$ is, by virtue of the MLR property, nondecreasing in $t \in (0,x)$. □

Lemma 5.4 *Let K be a non-negative discrete random variable with mass function p_k. Let $\{\beta_k, k \in \mathbb{K}\}$ and $\{\gamma_k, k \in \mathbb{K}\}$ be functions taking values on $\mathbb{R}_+$ such that:* **(i)** *$\frac{\gamma_k(w)}{\beta_k(w)}$ decreases in w for all k, and* **(ii)** *$\frac{\beta_k(w_1)}{\beta_k(w_0)}$ increases in k for all w_1, w_0 such that $w_1 > w_0$. Then, subject to its existence, $\frac{\mathbb{E}(\gamma_K(w))}{\mathbb{E}(\beta_K(w))}$ is nonincreasing in w.*

Proof. We have for $w_1 > w_0$, making use of assumption **(i)**:

$$\begin{aligned}
\frac{\mathbb{E}\left(\gamma_K(w_1)\right)}{\mathbb{E}\left(\beta_K(w_1)\right)} &= \frac{\sum_k p_k \frac{\gamma_k(w_1)}{\beta_k(w_1)}\beta_k(w_1)}{\sum_k p_k \beta_k(w_1)} \\
&\leq \frac{\sum_k p_k \frac{\gamma_k(w_0)}{\beta_k(w_0)}\beta_k(w_1)}{\sum_k p_k \beta_k(w_1)} \\
&= \mathbb{E}_{w_1}\left(\frac{\gamma_J(w_0)}{\beta_J(w_0)}\right),
\end{aligned}$$

where J has mass function proportional to $p_j\,\beta_j(w_1)$. Observe now that assumption **(ii)** implies that the two-member family of densities proportional to $p_j\,\beta_j(w)$, $w = w_1, w_0$ has increasing MLR in J. Therefore, we have with assumption **(i)** again

$$\frac{\mathbb{E}\left(\gamma_K(w_1)\right)}{\mathbb{E}\left(\beta_K(w_1)\right)} \leq \mathbb{E}_{w_1}\left(\frac{\gamma_J(w_0)}{\beta_J(w_0)}\right) \leq \mathbb{E}_{w_0}\left(\frac{\gamma_J(w_0)}{\beta_J(w_0)}\right) = \frac{\mathbb{E}\left(\gamma_K(w_0)\right)}{\mathbb{E}\left(\beta_K(w_0)\right)}. \qquad \square$$

We conclude with details justifying expression (4.7), which is a series representation for $\hat{\theta}_{\psi_0}$ exploiting familiar techniques for re-expressing incomplete Beta functions.

Proof of (4.7) Starting from $\theta_{\psi_0}(w) = c_0 w - \psi_0(w)$, make the transformation $x = t/w$ in (4.6) to obtain Then,

$$\begin{aligned}\hat{\theta}_{\psi_0}/c_0 =& w - \frac{\int_0^1 (wx)^{(m-1)/2}/(1+wx)^{-n/2}\,\mathrm{d}x}{\int_0^1 (wx)^{(m-1)/2-1}/(1+wx)^{-n/2}\,\mathrm{d}x} \\ =& w\,\frac{\int_0^1 (1-x)x^{(m-1)/2-1}/(1+wx)^{-n/2}\,\mathrm{d}x}{\int_0^1 x^{(m-1)/2-1}/(1+wx)^{-n/2}\,\mathrm{d}x}.\end{aligned}$$

With the transformation $z = 1-x$, we have

$$\hat{\theta}_{\psi_0} = c_0 w \frac{\int_0^1 z(1-z)^{(m-1)/2-1}/(1-\beta z)^{-n/2}\mathrm{d}z}{\int_0^1 (1-z)^{(m-1)/2-1}/(1-\beta z)^{-n/2}\mathrm{d}z}, \tag{5.3}$$

for $\beta = w/(1+w)$. From the negative binomial expansion

$$\frac{1}{(1-z)^{\alpha}} = \sum_{k=0}^{\infty} \frac{\Gamma(\alpha+k)}{k!\,\Gamma(\alpha)} z^k,$$

we obtain for $h = 0, 1$

$$\begin{aligned}\int_0^1 z^h\,(1-z)^{(m-1)/2-1}/(1-\beta z)^{-n/2}\,\mathrm{d}z &= \sum_{k=0}^{\infty} \frac{\Gamma(n/2+k)}{\Gamma(n/2)k!}\beta^k \int_0^1 z^{k+h}(1-z)^{(m-1)/2-1}\,\mathrm{d}z \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(n/2+k)}{\Gamma(n/2)k!} B(k+h+1,(m-1)/2)\,\beta^k.\end{aligned}$$

Along with (5.3), we obtain

$$\hat{\theta}_{\psi_0}(w) = \frac{n-m-1}{n} w \frac{\sum_{k=0}^{\infty} \frac{\Gamma(n/2+k)}{k!}\,B(k+2,(m-1)/2)\,\beta^k}{\sum_{k=0}^{\infty} \frac{\Gamma(n/2+k)}{k!}\,B(k+1,(m-1)/2)\,\beta^k},$$

which leads to expression (4.7). $\square$

References

[1] Brown, L., Johnstone, I., and MacGibbon, B. (1981). Variation diminishing transfromations: a direct approach to total positivity and its statistical applications. *J. Amer. Statist. Assoc.*, **76**, 824-832.

[2] Chow, M.S. (1987). A complete class theorem for estimating a noncentrality parameter. *Ann. Statist.*, **15**, 800-804.

[3] Fisher, R.A. (1924). The influence of rainfall on the yield of wheat at Rothamsted. *Philosophical Transactions of the Royal Society of London*, Ser. B, **213**, 89-124.

[4] Gurland, J. (1968). A relatively simple form of the distribution of the multiple correlation coefficient. *J. Royal Statist. Soc.*, **B, 30**, 276-283.

[5] Kubokawa, T., Robert, C.P. and Saleh, A.K.Md.E. (1993). Estimation of noncentrality parameters. *Canad. J. Statist.*, **21**, 45-57.

[6] Leung, P.L. (1994). Improved estimation of functions of the canonical correlation coefficients. *Commun. Statist. - Theory Methods*, **23**, 831-840.

[7] Leung, P.L. and Muirhead, R.J. (1987). Estimation of parameter matrices and eigenvalues in MANOVA and canonical correlation analysis. *Ann. Statist.*, **15**, 1651-1666.

[8] Lo, M. and Leung, P.L. (1996). Decision theoretic estimation of functions of the canonical correlation coefficients. *Commun. Statist. - Theory Methods*, **25**, 1985-1995.

[9] Marchand, É. (2001). Point Estimation of the Coefficient of Determination. *Statistics and Decisions*, **19**, 137-154.

[10] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*, New York, John Wiley.

[11] Muirhead, R.J. (1985). Estimating a particular function of the multiple-correlation coefficient. *J. Amer. Statist. Assoc.*, **80**, 923-925.

[12] Muirhead, R.J. and Leung, P.L. (1985). Estimating functions of canonical correlation coefficients. *Linear Algebra and its Applications*, **70**, 173-183.

[13] Neff, N. and Strawderman, W.E. (1976). Further remarks on estimating the parameter of a noncentral chi-squared distribution. *Comm. Statist. A*, **5**, 65-76.

[14] Olkin, I., and Pratt, J.W. (1958). Unbiased estimation of certain correlation coefficients. *Ann. Math. Statist.*, **29**, 201-211.

[15] Perlman, M.D. and Rasmussen, V.A. (1975). Some remarks on estimating a noncentrality parameter. *Comm. Statist.*, **4**, 455-468.

[16] Rukhin, A.L. (1993). Estimation of the noncentrality parameter of an F-distribution. *J. Statist. Plann. Inf.*, **35**, 201-311.

[17] Saxena, K.M.L., and Alam, K. (1982). Estimation of the non-centrality parameter of the chisquared distribution. *Ann. Statist.*, **10**, 1012-1016.

[18] Shao, P.Y.-S. and Strawderman. (1995). Improving on the positive part of the UMVUE of a noncentrality parameter of a noncentral chi-square distribution. *J. Multivariate Anal.*, **53**, 52-66.