Sankhyā : The Indian Journal of Statistics
2013, Volume 75-B, Part 2, pp. 216-237

A characterization of Bayesian robustness for a normal location parameter

Kamlesh Kumar and Jan R. Magnus
Tilburg University, Tilburg, The Netherlands

Abstract

We consider the Bayesian estimation of a location parameter θ based on one observation x from a univariate normal distribution with mean θ and variance one, together with a prior π. In general, the mean $t(x)$ in the posterior distribution does not satisfy the requirement that $x - t(x)$ vanishes as x approaches ∞ (for example, when π is normal or Laplace), that is, the prior is not robust. In this paper we obtain, under mild regularity conditions on π, a necessary and sufficient (and easy to apply) condition for robustness, and identify classes of robust priors. Special attention is paid to the Subbotin prior because of its role in Bayesian model averaging.

AMS (2000) *subject classification.* Primary 62C10; Secondary 62C20, 62F15, 62P20.
Keywords and phrases. Robustness, posterior moments, powered exponential distribution, model averaging, Subbotin prior.

1 Introduction

We consider the estimation of a location parameter θ based on one observation x from a univariate normal distribution with mean θ and known variance, which (without loss of generality) we set equal to one:

$$x|\theta \sim \mathrm{N}(\theta, 1).$$

Together with a prior $\pi(\theta)$, this leads to a posterior density $p(\theta|x)$ given by

$$p(\theta|x) = \frac{\phi(x-\theta)\pi(\theta)}{\int \phi(x-\theta)\pi(\theta)\, d\theta},$$

where ϕ denotes the standard-normal density. The mean and variance of θ in the posterior distribution are denoted as $t(x)$ and $\sigma^2(x)$, respectively.

If the prior is normal, then so is the posterior. In particular, if the prior distribution of θ is $\mathrm{N}(0, \tau^2)$, then the mean and variance of θ in the posterior distribution are $t(x) = \lambda x$ and $\sigma^2(x) = \lambda$, where $\lambda = \tau^2/(\tau^2 + 1)$ is a constant. The normal prior, although convenient, is often considered inappropriate because the discrepancy between $t(x)$ and x does not vanish when x becomes large, but rather increases linearly without bound. In other words, the normal prior is not discounted when confronted with an observation with which it drastically disagrees. The normal prior is therefore not 'robust' for the normal location problem.

This simple, but disturbing, fact has been known at least since Lindley (1968), more precisely since Mr. Beale's comments following Lindley's paper (pp. 54–55) and Lindley's response (pp. 65–66). In his response Lindley conjectures that if the prior follows a Student rather than a normal distribution, then robustness is achieved.

Lindley's conjecture was proved by Dawid (1973), who considers the 'dual' problem by making the change of variable $y = x - \theta$. This gives

$$t(x) = \frac{\int \theta\phi(x-\theta)\pi(\theta)\,d\theta}{\int \phi(x-\theta)\pi(\theta)\,d\theta} = \frac{\int (x-y)\phi(y)\pi(x-y)\,dy}{\int \phi(y)\pi(x-y)\,dy},$$

and hence, letting $R(x, y) := \left(\pi(x-y) - \pi(x)\right)/\pi(x)$,

$$x - t(x) = \frac{\int y\pi(x-y)\phi(y)\,dy}{\int \pi(x-y)\phi(y)\,dy} = \frac{\int yR(x,y)\phi(y)\,dy}{1 + \int R(x,y)\phi(y)\,dy}.$$

Sufficient for $x - t(x) \to 0$ as $x \to \infty$ is therefore:

$$\lim_{x\to\infty} \int R(x,y)\phi(y)\,dy = 0, \qquad \lim_{x\to\infty} \int yR(x,y)\phi(y)\,dy = 0. \tag{1.1}$$

Dawid (1973) then presents conditions which imply (1.1), and shows that these conditions are satisfied for the Student distribution. His key condition is that $R(x, y) \to 0$ as $x \to \infty$, which he expresses by saying that π is 'almost uniform' for large x. What this means becomes clear when we write $\xi := e^x$, $z := e^{-y}$, and $\pi^*(\xi) := \pi(\log \xi)$. Then,

$$R(x,y) = \frac{\pi^*(\xi z)}{\pi^*(\xi)} - 1,$$

and hence

$$\lim_{x\to\infty} R(x,y) = 0 \iff \lim_{\xi\to\infty} \frac{\pi^*(\xi z)}{\pi^*(\xi)} = 1,$$

so that the prior π is robust if the function π^* is *slowly varying* (de Haan and Ferreira, 2006, Appendix B). One may wonder under which general conditions a slowly varying π^* is also necessary for π to be robust, thus obtaining a characterization of robustness. This, in essence, is what the current paper deals with, except that we will work under the assumption that π is differentiable. This assumption is not, however, an important restriction.

A different approach is based on comparing the mean $t(x)$ and the mode $m(x)$ of the posterior distribution. Meeden and Isaacson (1977) provide conditions such that $t(x) - m(x) \to 0$ as $x \to \infty$. Since

$$\log p(\theta|x) = a(x) + x\theta - \theta^2/2 + \log \pi(\theta),$$

where $a(x)$ does not depend on θ, we obtain

$$\frac{d \log p(\theta|x)}{d\theta} = x - \theta - \omega(\theta), \qquad \omega(\theta) := \frac{-d \log \pi(\theta)}{d\theta}.$$

For large x, the mode $m(x)$ is found from the equation $d \log p(\theta|x)/d\theta = 0$, that is from

$$x - m(x) = \omega(m(x)). \tag{1.2}$$

If $\omega(\theta) \to 0$ as $\theta \to \infty$, then $x - m(x) \to 0$ as $x \to \infty$. In addition, under Meeden and Isaacson's conditions, $t(x) - m(x) \to 0$. Hence, $x - t(x) \to 0$ as $x \to \infty$.

A considerable literature followed, considering in particular the Student and the Laplace (double exponential) priors; see Berger (1994), Carvalho, Polson, and Scott (2010), Choy and Smith (1997), Griffin and Brown (2010), Masreliez (1975), Pericchi and Smith (1992), Polson (1991), and the references contained therein. The theory has mostly been developed in the context of location parameters. The posterior distribution of a scale parameter is more difficult (Andrade and O'Hagan, 2006, 2011), and is not considered in this paper.

Our main purpose is to *characterize* Bayesian robustness, that is, to provide conditions that are both sufficient *and* necessary for $x - t(x)$ to vanish asymptotically. The analysis based on Meeden and Isaacson (1977) suggests that the condition $\omega(\theta) \to 0$ as $\theta \to \infty$ may be sufficient and necessary. We shall prove that this is indeed the case. This simple characterization allows us to identify classes of robust priors. For example, if the prior π follows a Student(k) distribution, then

$$\omega_k(\theta) = \frac{-d \log \pi(\theta)}{d\theta} = \frac{(k+1)\theta}{\theta^2 + k}.$$

We have $\lim_{\theta\to\infty}\omega_k(\theta) = 0$, but $\lim_{\theta\to\infty}\lim_{k\to\infty}\omega_k(\theta) = \infty$, because the sequence of functions $\{\omega_k\}$ $(k = 1, 2, \dots)$ is not uniformly convergent. This explains why the normal prior is not robust, while the Student prior is.

Robust modeling involves choosing distributions with 'appropriate' tail thickness. If the data are generated by $x|\theta \sim \mathrm{N}(\theta, 1)$, then the tail thickness of the prior $\pi(\theta)$ is appropriate when, for large values of x, the posterior distribution of $\theta|x$ has a mean $t(x)$ satisfying $x - t(x) \to 0$ and a variance $\sigma^2(x)$ satisfying $\sigma^2(x) \to 1$, so that the data dominate the prior when the data information is sufficiently strong, and eventually (for $x \to \infty$) the prior information is completely ignored. This is our interpretation of robust modeling, and this is why we are interested in the conditions $x - t(x) \to 0$ and $\sigma^2(x) \to 1$.

The paper is organized as follows. Section 2 relates our statistical problem to econometric modeling and Bayesian model averaging. In Section 3 we formalize robustness, state our assumptions, and obtain two preliminary results. Then, we study the posterior mean $t(x)$, and show that $0 < t(x) < x$ for $x > 0$, that t is increasing on $\mathbb{R}$, and that $t(x) \to \infty$ as $x \to \infty$. Section 4 contains our main result. In Section 5 we study conditions under which the posterior variance $\sigma^2(x)$ approaches one. Next, in Section 6, we restrict the class of robust priors by introducing the concept of 'neutrality'. In Section 7 we attempt to find the 'best' prior within this restricted class of robust and neutral priors. In Section 8 we highlight a specific prior (called the Subbotin prior) which is important because of its role in Bayesian model averaging. Section 9 concludes. A technical appendix contains the proofs of the lemmas and theorems.

2 Relation to econometric model averaging

The problem studied in this paper plays an important role in econometric modeling and Bayesian model averaging, especially in weighted-average least squares (De Luca and Magnus, 2011; Magnus, Powell and Prüfer, 2010), and we sketch this relationship below. We write the econometric model, in its simplest form, as

$$y = X_1\beta_1 + \beta_2 x_2 + \varepsilon, \qquad \varepsilon \sim \mathrm{N}(0, \sigma^2 I_n),$$

where y $(n \times 1)$ is the vector of observations, X_1 $(n \times k_1)$ is a matrix and x_2 $(n \times 1)$ a vector of nonrandom regressors, ε is a random vector of unobservable disturbances, and β_1 and β_2 contain the unknown parameters. We assume that $1 \le k_1 \le n - 2$ and that $(X_1 : x_2)$ has full column-rank. The reason for distinguishing between X_1 and x_2 is that X_1 contains explanatory variables which we want in the model on theoretical or other grounds (irrespective of the found t-ratios of the β_1-parameters), while x_2 contains *one* additional explanatory variable of which we are less certain. The columns of X_1 are called 'focus' regressors and x_2 is the 'auxiliary' regressor. In this simple case there are only two models to consider: restricted (where $\beta_2 = 0$) and unrestricted. We define

$$M_1 := I_n - X_1(X_1'X_1)^{-1}X_1', \qquad q := \frac{1}{\sqrt{x_2'M_1x_2}}(X_1'X_1)^{-1}X_1'x_2.$$

In the restricted model the least-squares estimator of β_1 is $\hat{\beta}_{1r} = (X_1'X_1)^{-1}X_1'y$. In the unrestricted model we have

$$\hat{\beta}_{1u} = \hat{\beta}_{1r} - \sigma\hat{\theta}q, \qquad \hat{\beta}_{2u} = \frac{x_2'M_1y}{x_2'M_1x_2},$$

where

$$\hat{\theta} := \frac{\hat{\beta}_{2u}}{\sigma/\sqrt{x_2'M_1x_2}} \sim \mathrm{N}(\theta, 1), \qquad \theta := \frac{\beta_2}{\sigma/\sqrt{x_2'M_1x_2}}.$$

Now consider the weighted-average least-squares (WALS) estimator of β_1,

$$b_1 := \lambda\hat{\beta}_{1u} + (1-\lambda)\hat{\beta}_{1r} = \hat{\beta}_{1r} - \sigma(\lambda\hat{\theta})q,$$

where $0 \leq \lambda \leq 1$ may depend on $\hat{\theta}$, more precisely on the residuals M_1y. The 'equivalence theorem' (Danilov and Magnus, 2004, Theorem 1; Magnus and Durbin, 1999, Theorem 2) now implies that

$$\mathrm{MSE}(b_1) = \sigma^2\left((X_1'X_1)^{-1} + \mathrm{MSE}(\lambda\hat{\theta})qq'\right). \tag{2.1}$$

The equivalence theorem is in fact much more general, but its essence is well reflected in Equation (2.1): If we can find a λ-function such that $\lambda\hat{\theta}$ is an optimal estimator of θ (in the mean squared error sense), then *the same* λ-function will provide an optimal WALS estimator of β_1. Suppose $t(\hat{\theta})$ is an 'optimal' estimator of θ. Then, since $t(\hat{\theta}) = \lambda(\hat{\theta})\hat{\theta}$, we find $\lambda(\hat{\theta}) = t(\hat{\theta})/\hat{\theta}$. The problem of estimating β_1 in a regression context is thus reduced to estimating θ from a single observation $\hat{\theta} \sim \mathrm{N}(\theta, 1)$, which is the problem studied in this paper.

In a non-Bayesian context the obvious candidate is $t(\hat{\theta}) = \hat{\theta}$ (the 'usual' estimator). Popular is also the 'pretest' estimator,

$$t(\hat{\theta}) = \begin{cases} \hat{\theta} & \text{if } |\hat{\theta}| > c, \\ 0 & \text{if } |\hat{\theta}| \leq c, \end{cases}$$

for some constant $c > 0$. In a Bayesian context, we need a prior π on θ. The prior could be normal, Laplace, or otherwise. The 'usual' estimator is good when $\hat{\theta}$ is large but not when $\hat{\theta}$ is small. The pretest estimator is discontinuous and hence inadmissable. The 'normal' estimator has unbounded risk and $\hat{\theta} - t(\hat{\theta})$ diverges to ∞ as $\hat{\theta} \to \infty$. The Laplace estimator has bounded risk and $\hat{\theta} - t(\hat{\theta})$ converges to a constant, but not to zero. This is undesirable, because we would expect that, for large $\hat{\theta}$, the estimator 'behaves like $\hat{\theta}$' in the sense that $\hat{\theta} - t(\hat{\theta})$ converges to zero (Bayesian robustness). The Subbotin and the reflected Weibull estimators, which are the ones advocated in this paper, possess the property of Bayesian robustness. These estimators depend on a parameter $0 < q < 1$ and curve back to the 45° line $t(\hat{\theta}) = \hat{\theta}$, in contrast to the Laplace estimator.

Bayesian averaging of Bayesian estimators was first proposed by Leamer (1978), and Bayesian averaging of classical estimators by Raftery (1995). A large literature on Bayesian model averaging now exists. An alternative approach, first proposed by Frank and Friedman (1993) and related to the Subbotin prior, is to consider penalized regression by computing

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left((y - X\beta)'(y - X\beta) + \lambda \sum_{j=1}^{k} |\beta_j|^q \right)$$

for some $\lambda > 0$ and $q > 0$. The case $q = 2$ is the familiar ridge regression. Tibshirani (1996) studied the case $q = 1$, which he called the lasso (least absolute shrinkage and selection operator). Hans (2009) and Park and Casella (2008) exploited the fact that the lasso has a Bayesian counterpart, which they called the lasso prior. The lasso prior is in fact the Laplace prior. A different penalty function called SCAD (smoothly clipped absolute deviation) was introduced by Fan and Li (2001). Our proposed (Bayesian) estimators are inspired by the SCAD (non-Bayesian) estimator in that they 'curve back' to the 45° line when x is large, as is well illustrated in Figure 2 of Fan and Li (2001). What they call the 'hard' thresholding function is in fact the pretest estimator, and their lasso and SCAD thresholding functions in a non-Bayesian context are discontinuous counterparts of the Laplace, Subbotin, and reflected Weibull estimators in a Bayesian context. The Subbotin and reflected Weibull priors are then of special importance because they combine transparency in the treatment of ignorance with a trivial computational burden.

3 Assumptions and preliminary lemmas

Throughout we shall impose the following restrictions on the prior density:

Assumption A The prior π is

(A.1) symmetric around zero: $\pi(-\theta) = \pi(\theta)$ for all $\theta > 0$;

(A.2) positive and non-increasing on $(0, \infty)$;

(A.3) differentiable, except possibly at 0; and

(A.4) $\omega(\theta) := -\pi'(\theta)/\pi(\theta)$ has a limit (possibly ∞) as $\theta \to \infty$.

Assumptions (A.1)–(A.3) characterize the prior, allowing a non-differentiable peak at zero. Assumption (A.4) is needed in proving the necessity of the condition in Theorem 4.1 below.

Robustness is formally defined as follows:

Definition A prior π is said to be robust if

$$g(x) := x - t(x) \to 0 \quad \text{as} \quad x \to \infty.$$

If we let, for $j = 0, 1, \dots,$

$$A_j(x) := \int_{-\infty}^{\infty} (x-\theta)^j \phi(x-\theta)\pi(\theta)\, d\theta,$$

then $\mu_j(x) := \mathrm{E}((x-\theta)^j|x) = A_j(x)/A_0(x)$, and hence the mean and variance of θ in the posterior distribution are:

$$t(x) = -\mu_1(x) + x, \qquad \sigma^2(x) = \mu_2(x) - \mu_1^2(x). \tag{3.1}$$

Since $A_0'(x) = -A_1(x)$ and

$$A_j'(x) = jA_{j-1}(x) - A_{j+1}(x) \qquad (j = 1, 2, \dots),$$

we obtain the recursion

$$\mu_j'(x) = j\mu_{j-1}(x) - \mu_{j+1}(x) + \mu_1(x)\mu_j(x) \qquad (j = 1, 2, \dots), \tag{3.2}$$

implying that

$$t(x) = x + \frac{d \log A_0(x)}{dx}, \qquad \sigma^2(x) = 1 + \frac{d^2 \log A_0(x)}{dx^2}$$

in accordance with Pericchi and Smith (1992). We present two preliminary results concerning the functions A_0 and A_1.

Lemma 3.1. *We have*

$$A_0(x) = e^{-x^2/2} \int_0^\infty \left(e^{\theta x} + e^{-\theta x}\right) \phi(\theta)\pi(\theta)\, d\theta,$$

and hence $A_0(x) > 0$ *for all* $x \geq 0$.

Lemma 3.2. *Define* $\eta(x) := (e^x - e^{-x})/x$ *for* $x > 0$ *and* $\eta(0) := 2$*, so that* $\eta(x) \geq 2$ *for all* $x \geq 0$*. Then,*

$$A_1(x) = x\, e^{-x^2/2} \int_0^\infty \theta\eta(\theta x)\omega(\theta)\phi(\theta)\pi(\theta)\, d\theta,$$

and hence $A_1(0) = 0$ *and* $A_1(x) > 0$ *for all* $x > 0$.

In view of the fact that $A_0'(x) = -A_1(x)$ it follows from Lemma 3.2 that A_0 is decreasing on $[0, \infty)$ and hence that $A_0(x) \leq A_0(0) \leq \phi(0)$.

The posterior mean $t(x)$ is an odd function, so that $t(-x) = -t(x)$ and $t(0) = 0$. Assumption A allows us to prove three of its properties.

Lemma 3.3. *The function* t *satisfies the inequality*

$$0 < t(x) < x \qquad (x > 0).$$

Lemma 3.4. *For any* x,

$$t'(x) = \sigma^2(x) > 0,$$

so that t *is increasing on* $\mathbb{R}$*. Furthermore,* $0 < t'(0) < 1$.

Lemma 3.5. *We have*

$$t(x) \to \infty \quad \text{as} \quad x \to \infty.$$

4 Characterization of Bayesian robustness

In the previous section we used only Assumption A to show that t is an increasing function tending to $+\infty$, bounded from above by x. This assumption alone is, however, not sufficient to show that '$t(x)$ is close to x' when x is large, that is, it does not imply Bayesian robustness. Consider, for example, a prior given by the 'powered exponential' family (Box and Tiao, 1973):

$$\pi(\theta) = \frac{qc^{1/q}}{2\Gamma(1/q)}\, e^{-c|\theta|^q} \qquad (c > 0,\, q > 0),$$

special cases of which include the normal density ($q = 2$) and the Laplace density ($q = 1$). The powered exponential family was first proposed by Subbotin (1923), and we shall therefore refer to the important special case $0 < q < 1$ as the Subbotin prior. The powered exponential family satisfies Assumption A and we have, for $\theta > 0$,

$$\omega(\theta) = \frac{-d \log \pi(\theta)}{d\theta} = cq\theta^{q-1}.$$

The discrepancy function $g(x) = x - t(x)$ is unbounded for $q > 1$, approaches a positive constant for $q = 1$, and approaches 0 for $0 < q < 1$ (Choy and Walker, 2003; Sansó and Pericchi, 1992). Bayesian robustness ($g(x) \to 0$) occurs therefore if and only if $0 < q < 1$ (Subbotin prior), that is, if and only if $\omega(\theta) \to 0$ as $\theta \to \infty$.

Our main result shows that the condition $\omega(\theta) \to 0$ is in fact necessary and sufficient, within the general class of priors of Section 3, thus providing a characterization of Bayesian robustness.

Theorem 4.1. *Under Assumption A,*

$$\lim_{\theta \to \infty} \omega(\theta) = 0 \iff \lim_{x \to \infty} g(x) = 0.$$

Our condition can also be phrased, perhaps more intuitively, in terms of the hazard (or failure) rate

$$h(\theta) := \frac{\pi(\theta)}{\int_\theta^\infty \pi(\xi)\, d\xi}.$$

The two conditions $\omega(\theta) \to 0$ as $\theta \to \infty$ and $h(\theta) \to 0$ as $\theta \to \infty$ are equivalent by l'Hôpital's rule and Assumption (A.4). The tail of π can be classified according to the limiting behavior of h, and the condition $h(\theta) \to 0$ means that the tail of π is 'medium-long' or 'long'; see Schuster (1984).

5 The posterior variance

Important is not only whether $x - t(x) \to 0$, but also whether $\sigma^2(x) \to 1$ as $x \to \infty$, because the posterior variance $\sigma^2(x)$ then approaches the 'data' variance (that is the variance of $x|\theta$, which is 1), so that the data dominate the prior when the data information is sufficiently strong, and eventually (for $x \to \infty$) the prior information is completely ignored.

Let us return to the analysis in Meeden and Isaacson (1977), who compare the mean $t(x)$ and the mode $m(x)$ in the posterior distribution. Let

$$\lambda(\theta) := -\log p(\theta|x) = -a(x) - x\theta + \theta^2/2 - \log \pi(\theta),$$

where $a(x)$ does not depend on θ. Then, Meeden and Isaacson provide not only conditions such that $t(x) - m(x) \to 0$, but also such that

$$\sigma^2(x) - \frac{1}{\lambda''(m(x))} \to 0, \tag{5.1}$$

as $x \to \infty$. Letting again $\omega(\theta) = -d\log \pi(\theta)/d\theta$, we find

$$\lambda''(\theta) = 1 + \omega'(\theta).$$

If $\omega'(\theta) \to 0$ as $\theta \to \infty$, then

$$\lim_{x\to\infty} \sigma^2(x) = \lim_{x\to\infty} \frac{1}{1 + \omega'(m(x))} = 1.$$

Thus, while $\omega(\theta) \to 0$ implies that $x - t(x) \to 0$, $\omega'(\theta) \to 0$ implies that $\sigma^2(x) \to 1$.

Additional insight is provided by differentiating (1.2) with respect to x. This gives

$$1 - m'(x) = \omega'(m(x)) \cdot m'(x),$$

which can be written as

$$m'(x) = \frac{1}{1 + \omega'(m(x))}.$$

This, combined with the fact that $\sigma^2(x) = t'(x)$, shows that (5.1) can also be written as $t'(x) - m'(x) \to 0$. If $\omega'(\theta) \to 0$ then $m'(x) \to 1$, and hence, under Meeden and Isaacson's conditions, $\sigma^2(x) = t'(x) \to 1$.

6 Robustness and neutrality

The class of robust priors satisfying Assumption A includes the Student distribution (but not the normal distribution), the reflected generalized gamma distribution

$$\pi(\theta) \propto |\theta|^{-\alpha} e^{-c|\theta|^q} \qquad (c > 0,\ 0 < q < 1,\ 0 \leq \alpha < 1),$$

special cases of which are the reflected Weibull ($\alpha = 1 - q$) and the Subbotin ($\alpha = 0$) distributions, and also the reflected Pareto distribution, among others.

Let us restrict the class of priors further by requiring 'neutrality', a concept which attempts to capture the vague notion of ignorance in an explicit and applicable form. Following Magnus (2002) we define a prior $\pi(\theta)$ to be neutral when the prior median of θ is zero and the prior median of $|\theta|$ is one, that is, when

$$\int_{-\infty}^{-1} \pi(\theta)\, d\theta = \int_{-1}^{0} \pi(\theta)\, d\theta = \int_{0}^{1} \pi(\theta)\, d\theta = \int_{1}^{\infty} \pi(\theta)\, d\theta = \frac{1}{4}.$$

Neutrality places an additional restriction on the prior, thus reducing the number of parameters by one.

Four important robust priors will be considered. First, the Student distribution

$$\pi(\theta) \propto (1+\theta^2/k)^{-(k+1)/2} \qquad (k=1,2,\dots).$$

Neutrality requires that $\int_0^1 \pi(\theta)\, d\theta = 1/4$, which is the case if and only if $k=1$. Hence, the only neutral Student distribution is the Cauchy distribution. Next, the reflected Weibull distribution

$$\pi(\theta) \propto |\theta|^{q-1}\, e^{-c|\theta|^q} \qquad (c>0,\ 0<q<1),$$

with neutrality occurring if and only if $c = \log 2$. Then, the Subbotin distribution

$$\pi(\theta) \propto e^{-c|\theta|^q} \qquad (c>0,\ 0<q<1).$$

Here, neutrality implies a relationship between q and c which can be expressed in terms of the (lower) incomplete gamma function:

$$\int_0^c t^{-1+1/q}\, e^{-t}\, dt = \Gamma(1/q)/2.$$

And finally, the reflected Pareto distribution

$$\pi(\theta) \propto (1+c|\theta|)^{-1/q} \qquad (c>0,\ 0<q<1),$$

with neutrality occurring if and only if $c = 2^{q/(1-q)} - 1$.

To decide which of these priors is 'best' we employ the concept of minimax regret.

7 Minimax regret

Let us write the prior as $\pi(\theta;q)$ to emphasize its dependence on unknown parameters q. In the four robust-neutral cases under consideration, the Cauchy prior has no parameter, while the other three priors have one parameter q. Given a class of priors $\pi(\theta;q)$, the mean in the posterior distribution, that is the estimator of θ, is

$$t(x;q) = \frac{\int \theta\phi(x-\theta)\pi(\theta;q)\,d\theta}{\int \phi(x-\theta)\pi(\theta;q)\,d\theta}.$$

Assuming that $x \sim \mathrm{N}(\theta,1)$, we consider the *risk* of the estimator $t(x;q)$. Under squared error loss, the risk is equal to the mean squared error

$$\mathrm{risk}(\theta;q) = \mathrm{E}_\theta(t(x;q)-\theta)^2.$$

We shall judge the four estimators by comparing their risk functions. However, for two admissible estimators neither dominates the other and some further criterion is needed. The minimax approach sometimes leads to unreasonable or trivial results (Hodges and Lehmann, 1950). On both theoretical and practical grounds we shall adopt the more appealing minimax regret approach, where we minimize maximum regret rather than maximum risk.

The *regret* of the estimator $t(x;q)$ based on the prior $\pi(\theta;q)$ is defined as

$$\mathrm{regret}(\theta;q) = \mathrm{E}_\theta(t(x;q)-\theta)^2 - \frac{\theta^2}{1+\theta^2},$$

where $\theta^2/(1+\theta^2)$ is the lower bound of the risk, according to Theorem A.7 of Magnus (2002). The minimax regret for the class of priors $\pi(\theta;q)$ is then defined as

$$\inf_q \sup_\theta \mathrm{regret}(\theta;q).$$

The minimax results are contained in Table 1. The Laplace prior is a special case of both the reflected Weibull prior and the Subbotin prior by setting $q=1$. It is neutral but not robust, and is included here as a benchmark. The Cauchy prior has no parameters, and hence minimax regret is simply maximum regret. For the remaining three classes of priors, we choose q between 0 and 1, then calculate c to obtain neutrality, and then maximize regret for these values of c and q. Searching over the $(0,1)$ interval, we then select the value of q for which the maximum regret is lowest. This gives the minimax regret in each class of priors.

Table 1: Minimax regret.

Prior	q	c	θ	Regret
Laplace	1.0000	0.6931	4.9319	0.5126
Cauchy	—	—	3.5354	0.6332
Reflected Weibull	0.8877	0.6931	3.5623	0.4546
Subbotin	0.7995	0.9377	3.6912	0.4697
Reflected Pareto	0.0862	0.0676	4.0310	0.4959

The Laplace prior outperforms Cauchy. The reflected Weibull, Subbotin, and reflected Pareto priors outperform Laplace, not only because they are robust while Laplace is not, but also because their minimax regret is lower. The reflected Weibull and Subbotin priors are of particular interest, because they have Laplace as a special case, and appear to be the best choices in the classes of priors considered. Minimax regret is slightly lower for the reflected Weibull prior than for Subbotin, but the difference is small, as exemplified in the figures below.

Combining the prior with the observation $x|\theta \sim \mathrm{N}(\theta, 1)$, we obtain characteristics of the posterior distribution. The deviations $g(x)$ are plotted in Figure 1. We have $t(x) = \lambda x$ with constant $\lambda < 1$ for the normal prior, and hence $g(x) = (1 - \lambda)x$, which diverges to infinity. For the Laplace prior we have $g(x) \to \log 2 \approx 0.6931$, and for the Subbotin(*) and reflected Weibull(*)

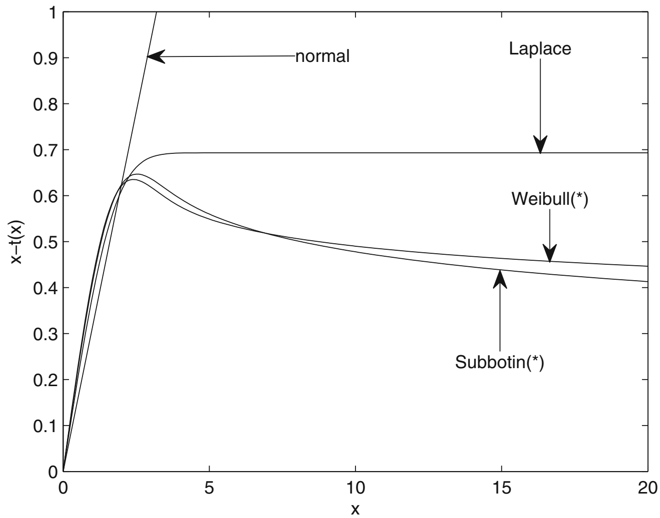

Figure 1: Deviations $g(x) = x - t(x)$.

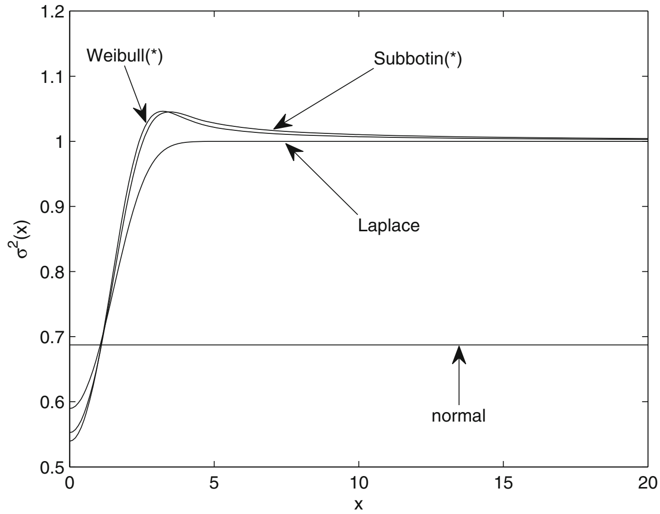

Figure 2: Posterior variances $\sigma^2(x)$.

priors we see that $g(x) \to 0$ in accordance with Theorem 4.1. (The * indicates the minimax member of the class of priors, given in Table 1.)

The posterior variances σ^2 are plotted in Figure 2. For the normal prior, the posterior variance is constant at $\sigma^2 = 1/(2c+1) \approx 0.6873$. For the

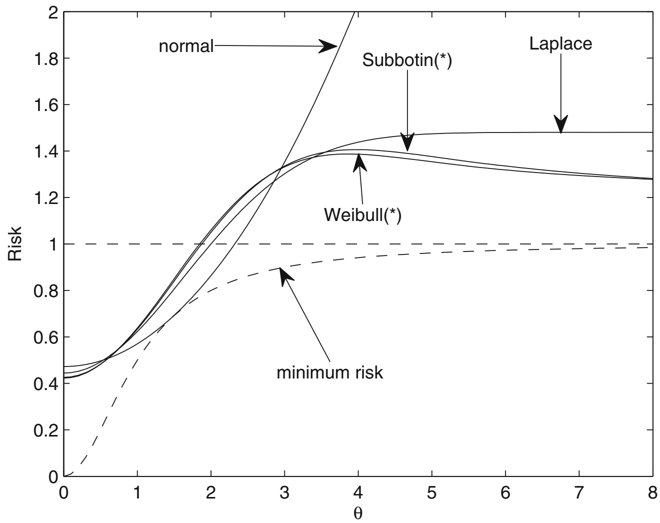

Figure 3: Risk profiles.

Laplace prior the variance increases monotonically to one. For the Subbotin(*) and reflected Weibull(*) priors, the variance also tends to one, but robustness implies that this convergence can not be monotonic (O'Hagan, 1981).

The risk profiles of the four priors are presented in Figure 3, together with the minimum risk $\theta^2/(1+\theta^2)$. The risk associated with the normal prior is unbounded, while the risk associated with the Laplace prior converges to $1 + c^2 = 1.4805$. The risk associated with the Subbotin(*) and reflected Weibull(*) priors converges to one, because of the fact that these priors are robust.

8 The Subbotin prior

While the Weibull distribution is well known, the Subbotin distribution,

$$\pi(\theta) = \frac{qc^{1/q}}{2\Gamma(1/q)}\, e^{-c|\theta|^q} \qquad (c > 0,\ 0 < q < 1),$$

is less well documented. Therefore we present some properties of this distribution below.

All moments of the Subbotin prior exist. Since the density is symmetric around zero, the odd moments vanish, and the even moments are given by

$$\mathrm{E}(\theta^{2m}) = \frac{\Gamma((2m+1)/q)/\Gamma(1/q)}{c^{2m/q}}.$$

In particular, the variance and kurtosis of θ are

$$\mathrm{var}(\theta) = \frac{\Gamma(3/q)/\Gamma(1/q)}{c^{2/q}}, \qquad \mathrm{kur}(\theta) = \frac{\Gamma(5/q)/\Gamma(3/q)}{\Gamma(3/q)/\Gamma(1/q)} - 3.$$

For a neutral Subbotin prior, c is determined by q as given in Table 2 for selected values of q. Under neutrality, both the standard deviation (sd) and kurtosis (kur) of the prior increase when q decreases, because the tail becomes thicker. We also present two quantiles. Neutrality implies that $Q(4/8) = 0$ and $Q(6/8) = 1$. The table gives $Q(5/8)$ and $Q(7/8)$. If the prior is relatively flat between 0 and 1, then $Q(5/8)$ should be close to 0.5. We see that the smaller is q, the less flat is the prior between 0 and 1. Small values of q (like $q = 0.1$) are therefore unappealing from the point of view of ignorance.

Because of the close relationship between mean and mode, the following result is also of interest.

Table 2: Moments and quantiles of some neutral Subbotin priors.

	q	c	sd	kur	$Q(5/8)$	$Q(7/8)$
Normal	2.0	0.2275	1.48	0.00	0.47	1.71
Laplace	1.0	0.6931	2.04	3.00	0.42	2.00
Subbotin	0.9	0.8011	2.19	4.03	0.40	2.06
Subbotin	0.8	0.9369	2.40	5.57	0.39	2.14
Subbotin	0.7	1.1125	2.69	8.06	0.37	2.25
Subbotin	0.6	1.3478	3.14	12.6	0.35	2.38
Subbotin	0.5	1.6783	3.89	22.2	0.33	2.57
Subbotin	0.4	2.1757	5.37	49.0	0.30	2.86
Subbotin	0.3	3.0066	9.21	171	0.25	3.35
Subbotin	0.2	4.6709	27.1	1956	0.20	4.37
Subbotin	0.1	9.6687	691	2,823,513	0.11	8.07

THEOREM 8.1. *If the prior π follows a Subbotin distribution, then the mode in the posterior distribution is given by*

$$m(x) = \begin{cases} x - cqx^{q-1}\Delta_q(x) & \text{if } x > x_q^*, \\ 0 & \text{if } 0 \le x \le x_q^*, \end{cases}$$

where

$$x_q^* = \frac{(2-q)c^{1/(2-q)}}{(2(1-q))^{(1-q)/(2-q)}},$$

and $\Delta_q(x) \to 1$ as $x \to \infty$. For $x < 0$ we have $m(x) = -m(-x)$.

There are precisely two cases for $0 < q < 1$ where Δ_q can be solved explicitly, namely $q = 1/2$ (solving a cubic equation) and $q = 2/3$ (solving a quartic equation).

9 Conclusions

This paper arose from the need to find a suitable prior $\pi(\theta)$ in the problem of estimating the location parameter θ based on one observation x from a univariate normal distribution $x|\theta \sim \mathrm{N}(\theta, 1)$. The Laplace prior has many desirable properties, but it is not robust. Hence we searched for a prior which is close to the Laplace prior and is also robust. This lead us to the Subbotin and reflected Weibull priors. In the process of proving that these priors are indeed robust, we discovered that it is in fact possible to characterize the class of robust priors, that is to obtain a condition which is sufficient *and necessary* (Theorem 4.1). This condition is easy to verify and has an intuitive interpretation in terms of the hazard rate.

Acknowledgement. This paper owes much to John Einmahl who provided the key idea for the proof of main theorem. We are also grateful to Estate Khmaladze and Hannes Leeb for helpful discussions, to Wendun Wang for computational assistance, and to the associate editor and referee for constructive comments.

References

ANDRADE, J.A.A. and O'HAGAN, A. (2006). Bayesian robustness modeling using regularly varying distributions. *Bayesian Anal.*, **1**, 169–188.

ANDRADE, J.A.A. and O'HAGAN, A. (2011). Bayesian robustness modelling of location and scale parameters. *Scand. J. Stat.*, **38**, 691–711.

BERGER, J.O. (1994). An overview of robust Bayesian analysis. *Test*, **3**, 5–58.

BOX, G.E.P. and TIAO, G.C. (1973). *Bayesian inference in statistical analysis.* Addison-Wesley, Reading, Massachusetts.

CARVALHO, C.M., POLSON, N.G. and SCOTT, J.G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, **97**, 465–480.

CHOY, S.T.B. and SMITH, A.F.M. (1997). On robust analysis of a normal location parameter. *J. R. Stat. Soc. Ser. B*, **59**, 463–474.

CHOY, S.T.B. and WALKER, S.G. (2003). The extended exponential power distribution and Bayesian robustness. *Statist. Probab. Lett.*, **65**, 227–232.

DANILOV, D. and MAGNUS, J.R. (2004). On the harm that ignoring pretesting can cause. *J. Econom.*, **122**, 27–46.

DAWID, A.P. (1973). Posterior expectations for large observations. *Biometrika*, **60**, 664–667.

DE HAAN, L. and FERREIRA, A. (2006). *Extreme value theory: An introduction.* Springer, New York.

DE LUCA, G. and MAGNUS, J.R. (2011). Bayesian model averaging and weighted average least squares: Equivariance, stability, and numerical issues. *Stata J.*, **11**, 518–544.

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, **96**, 1348–1360.

FINUCAN, H.M. (1973). The relative position of the two regression curves. *J. R. Stat. Soc. Ser. B*, **35**, 316–322.

FRANK, I.E. and FRIEDMAN, J.H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, **35**, 109–135.

GRIFFIN, J.E. and BROWN, P.J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Anal.*, **5**, 171–188.

HANS, C. (2009). Bayesian lasso regression. *Biometrika*, **96**, 835–845.

HODGES, J.L. and LEHMANN, E.L. (1950). Some problems in minimax point estimation. *Ann. Math. Statist.*, **21**, 182–197.

LEAMER, E.E. (1978). *Specification searches: Ad hoc inference with nonexperimental data.* Wiley, New York.

LINDLEY, D.V. (1968). The choice of variables in multiple regression (with discussion). *J. R. Stat. Soc. Ser. B*, **30**, 31–66.

MAGNUS, J.R. (2002). Estimation of the mean of a univariate normal distribution with known variance. *Econom. J.*, **5**, 225–236.

MAGNUS, J.R. and DURBIN, J. (1999). Estimation of regression coefficients of interest when other regression coefficients are of no interest. *Econometrica*, **67**, 639–643.

MAGNUS, J.R., POWELL, O. and PRÜFER, P. (2010). A comparison of two model averaging techniques with an application to growth empirics. *J. Econom.*, **154**, 139–153.

MASRELIEZ, C.J. (1975). Approximate non-Gaussian filtering with linear state and observation relations. *IEEE Trans. Automat. Control*, **20**, 107–110.

MEEDEN, G. and ISAACSON, D. (1977). Approximate behavior of the posterior distribution for a large observation. *Ann. Statist.*, **5**, 899–908

O'HAGAN, A. (1981). A moment of indecision. *Biometrika*, **68**, 329–330.

PARK, T. and CASELLA, G. (2008). The Bayesian lasso. *J. Amer. Statist. Assoc.*, **103**, 681–686.

PERICCHI, L.R. and SMITH, A.F.M. (1992). Exact and approximate posterior moments for a normal location parameter. *J. R. Stat. Soc. Ser. B*, **54**, 793–804.

POLSON, N.G. (1991). A representation of the posterior mean for a location model. *Biometrika*, **78**, 426–430.

RAFTERY, A.E. (1995). Bayesian model selection in social research. *Sociol. Methodol.*, **25**, 111–163.

SANSÓ, B. and PERICCHI, L.R. (1992). Near ignorance classes of log-concave priors for the location model. *Test*, **1**, 39–46.

SCHUSTER, E.F. (1984). Classification of probability laws by tail behavior. *J. Amer. Statist. Assoc.*, **79**, 936–939.

SUBBOTIN, M.TH. (1923). On the law of frequency of error. *Mat. Sb.*, **31**, 296–301.

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B*, **58**, 267–288.

Appendix: Proofs of lemmas and theorems

Proof of Lemma 3.1. This follows directly from the fact that

$$A_0(x) = \int_{-\infty}^{\infty} \phi(x-\theta)\pi(\theta)\,d\theta = \int_0^{\infty} \left(\phi(x+\theta)+\phi(x-\theta)\right)\pi(\theta)\,d\theta.$$

Proof of Lemma 3.2. Since π is non-increasing on $(0,\infty)$ we have

$$0 < \theta\pi(\theta) \leq \int_0^{\theta} \pi(\xi)\,d\xi \qquad (\theta > 0).$$

The right-hand side converges to 0 for $\theta \downarrow 0$, because π is integrable. Hence,

$$\lim_{\theta\downarrow 0} \theta\pi(\theta) = 0. \tag{A.1}$$

Define the function

$$D(\theta, x) := \phi(x+\theta) - \phi(x-\theta),$$

with partial derivative

$$D'_\theta(\theta, x) = \frac{\partial D(\theta, x)}{\partial \theta} = -(x+\theta)\phi(x+\theta) - (x-\theta)\phi(x-\theta).$$

Integrating by parts gives

$$A_1(x) = -\int_0^\infty D'_\theta(\theta, x)\pi(\theta)\, d\theta = \int_0^\infty D(\theta, x)\pi'(\theta)\, d\theta,$$

where we have used the fact that

$$\begin{aligned}[D(\theta, x)\pi(\theta)]_0^\infty &= \lim_{\theta\to\infty} D(\theta, x)\pi(\theta) - \lim_{\theta\to 0} D(\theta, x)\pi(\theta) \\ &= -\lim_{\theta\to 0} D(\theta, x)\pi(\theta) = 2x\phi(x) \lim_{\theta\to 0} \theta\pi(\theta) = 0,\end{aligned}$$

because of (A.1) and the fact that, for $\theta \to 0$,

$$\frac{D(\theta, x)}{\theta} \to D'_\theta(0, x) = -2x\phi(x).$$

Recalling that $\omega(\theta) = -\pi'(\theta)/\pi(\theta)$, the result follows.

Proof of Lemma 3.3. Let $x > 0$. Lemmas 3.1 and 3.2 then imply that $A_0(x) > 0$ and $A_1(x) > 0$. Since

$$t(x) = -\mu_1(x) + x = -A_1(x)/A_0(x) + x,$$

it follows that $t(x) < x$; see also Finucan (1973).

To prove that $t(x) > 0$ for $x > 0$ we write

$$\begin{aligned}xA_0(x) - A_1(x) &= x\int_0^\infty (\phi(x+\theta) + \phi(x-\theta))\,\pi(\theta)\, d\theta \\ &\quad - \int_0^\infty ((x+\theta)\phi(x+\theta) + (x-\theta)\phi(x-\theta))\,\pi(\theta)\, d\theta \\ &= -\int_0^\infty (\theta\phi(x+\theta) - \theta\phi(x-\theta))\,\pi(\theta)\, d\theta \\ &= x\, e^{-x^2/2}\int_0^\infty \theta^2\eta(\theta x)\phi(\theta)\pi(\theta)\, d\theta > 0. \qquad \text{(A.2)}\end{aligned}$$

Proof of Lemma 3.4. We know from (3.1) that $t'(x) = -\mu_1'(x) + 1$ and $\sigma^2(x) = \mu_2(x) - \mu_1^2(x)$. We know from (3.2) that $\mu_1'(x) = 1 - \mu_2(x) +$

$\mu_1^2(x)$, since $\mu_0(x) = 1$. Hence, $t'(x) = \sigma^2(x)$. The posterior density is non-degenerate for any x, so that $\sigma^2(x) > 0$ for any x, and t is increasing on $\mathbb{R}$. In particular, $t'(0) > 0$.

To prove that $t'(0) < 1$ we note that $(1 - \theta^2)(\pi(\theta) - \pi(1)) \geq 0$. This gives

$$\int_{-\infty}^{\infty} (1 - \theta^2)\phi(\theta)\pi(\theta)\, d\theta > \pi(1) \int_{-\infty}^{\infty} (1 - \theta^2)\phi(\theta)\, d\theta = 0.$$

The inequality is strict, because $(1 - \theta^2)(\pi(\theta) - \pi(1)) = 0$ holds only for all θ if π is constant and hence improper. It follows that $\mu_2(0) < 1$ and hence that $t'(0) = \mu_2(0) < 1$.

PROOF OF LEMMA 3.5. Let $M > 1$. Since

$$\int_0^M e^{\theta x}\phi(\theta)\pi(\theta)\, d\theta \leq e^{Mx} \int_0^M \phi(\theta)\pi(\theta)\, d\theta$$

and

$$\int_M^{\infty} e^{\theta x}\phi(\theta)\pi(\theta)\, d\theta \geq \int_{2M}^{\infty} e^{\theta x}\phi(\theta)\pi(\theta)\, d\theta \geq e^{2Mx} \int_{2M}^{\infty} \phi(\theta)\pi(\theta)\, d\theta,$$

we find, since M is fixed,

$$r(x; M) := \frac{\int_0^M e^{\theta x}\phi(\theta)\pi(\theta)\, d\theta}{\int_M^{\infty} e^{\theta x}\phi(\theta)\pi(\theta)\, d\theta} \to 0 \quad \text{as } x \to \infty. \tag{A.3}$$

In particular, $r(x; M) \leq 1$ for x sufficiently large. Then, using Lemma 3.1, Equation (A.2), and the definition of η in Lemma 3.2, we obtain for x sufficiently large:

$$\begin{aligned} t(x) &= \frac{\int_0^{\infty} \theta \left(e^{\theta x} - e^{-\theta x}\right) \phi(\theta)\pi(\theta)\, d\theta}{\int_0^{\infty} \left(e^{\theta x} + e^{-\theta x}\right) \phi(\theta)\pi(\theta)\, d\theta} \geq \frac{(M/2) \int_M^{\infty} e^{\theta x}\phi(\theta)\pi(\theta)\, d\theta}{2 \int_0^{\infty} e^{\theta x}\phi(\theta)\pi(\theta)\, d\theta} \\ &= \frac{M/4}{1 + r(x; M)} \geq \frac{M}{8}. \end{aligned}$$

Since M can be arbitrarily large, t is unbounded.

PROOF OF THEOREM 4.1. To prove sufficiency, assume that $\omega(\theta) \to 0$ as $\theta \to \infty$. We note, using (A.1), that for every $M > 0$:

$$\begin{aligned} \int_0^M \theta\omega(\theta)\phi(\theta)\pi(\theta)\, d\theta &\leq -\phi(0) \int_0^M \theta\pi'(\theta)\, d\theta \\ &= \phi(0) \left(-M\pi(M) + \int_0^M \pi(\theta)\, d\theta\right) \leq \phi(0)/2. \end{aligned} \tag{A.4}$$

Let $\varepsilon > 0$ be arbitrary. Then, using Lemmas 3.1 and 3.2,

$$\begin{aligned} g(x) = \frac{A_1(x)}{A_0(x)} &= \frac{x\int_0^\infty \theta\eta(\theta x)\omega(\theta)\phi(\theta)\pi(\theta)\,d\theta}{\int_0^\infty \left(e^{\theta x}+e^{-\theta x}\right)\phi(\theta)\pi(\theta)\,d\theta} \\ &\leq \frac{x\int_0^M \theta\eta(\theta x)\omega(\theta)\phi(\theta)\pi(\theta)\,d\theta}{\int_{2M}^\infty e^{\theta x}\phi(\theta)\pi(\theta)\,d\theta} + \frac{\int_M^\infty e^{\theta x}\omega(\theta)\phi(\theta)\pi(\theta)\,d\theta}{\int_M^\infty e^{\theta x}\phi(\theta)\pi(\theta)\,d\theta} \\ &\leq xe^{-2Mx}\eta(Mx)\frac{\int_0^M \theta\omega(\theta)\phi(\theta)\pi(\theta)\,d\theta}{\int_{2M}^\infty \phi(\theta)\pi(\theta)\,d\theta} + \sup_{\theta\geq M}\omega(\theta) \\ &\leq \frac{\phi(0)}{2Me^{Mx}\int_{2M}^\infty \phi(\theta)\pi(\theta)\,d\theta} + \sup_{\theta\geq M}\omega(\theta) < \varepsilon, \end{aligned}$$

using (A.4), where we have chosen $M = M(\varepsilon) > 0$ such that $\sup_{\theta\geq M}\omega(\theta) < \varepsilon/2$, and then $x > x^* = x^*(\varepsilon, M(\varepsilon))$ with

$$e^{Mx^*} = \max\left(1,\ \frac{\phi(0)}{M\varepsilon\int_{2M}^\infty \phi(\theta)\pi(\theta)\,d\theta}\right).$$

This shows that $g(x) \to 0$ as $x \to \infty$.

To prove necessity, assume that it is not true that $\lim_{\theta\to\infty}\omega(\theta) = 0$. Then, $\lim_{\theta\to\infty}\omega(\theta) > 0$ because of Assumption A, in particular (A.4). Hence we have, for some $M > 0$ and $\omega^* > 0$,

$$\inf_{\theta\geq M}\omega(\theta) = \omega^* > 0.$$

This yields

$$\begin{aligned} g(x) &= \frac{\int_0^\infty (e^{\theta x}-e^{-\theta x})\omega(\theta)\phi(\theta)\pi(\theta)\,d\theta}{\int_0^\infty (e^{\theta x}+e^{-\theta x})\phi(\theta)\pi(\theta)\,d\theta} \\ &\geq \omega^*\,\frac{\int_M^\infty (e^{\theta x}-e^{-\theta x})\phi(\theta)\pi(\theta)\,d\theta}{\int_0^\infty (e^{\theta x}+e^{-\theta x})\phi(\theta)\pi(\theta)\,d\theta} \\ &\geq \frac{\omega^*(1-e^{-2Mx})}{2}\cdot\frac{\int_M^\infty e^{\theta x}\phi(\theta)\pi(\theta)\,d\theta}{\int_0^\infty e^{\theta x}\phi(\theta)\pi(\theta)\,d\theta} \\ &= \frac{\omega^*(1-e^{-2Mx})}{2(1+r(x;M))}, \end{aligned} \tag{A.5}$$

where $r(x;M)$ is defined in (A.3). For fixed M, $r(x;M) \to 0$ as $x \to \infty$, according to the proof of Lemma 3.5. Hence, the right-hand side of (A.5) converges to $\omega^*/2$, and $g(x)$ does not converge to zero. This shows that the condition $\omega(\theta) \to 0$ is necessary and concludes the proof.

PROOF OF THEOREM 8.1. For x sufficiently large the mode is given by (1.2):

$$m(x) = x - \omega(m(x)) = x - cq(m(x))^{q-1} = x - cqx^{q-1}\Delta_q(x),$$

where

$$\Delta_q(x) = \left(\frac{m(x)}{x}\right)^{q-1} \to 1.$$

The point x_q^* is determined by the two equations:

$$x\theta - \theta^2/2 - c\theta^q = 0, \qquad x - \theta - cq\theta^{q-1} = 0.$$

KAMLESH KUMAR
CENTER, TILBURG UNIVERSITY
THE NETHERLANDS
E-mail: k.kumar@uvt.nl

JAN R. MAGNUS
DEPARTMENT OF ECONOMETRICS
& OPERATIONS RESEARCH
TILBURG UNIVERSITY, THE NETHERLANDS
E-mail: magnus@uvt.nl

Paper received: 12 January 2012; revised: 15 October 2012.