

Bayesian Local Influence Analysis: With an Application to Stochastic Frontiers

January 31, 2018

Abstract

A Bayesian alternative to Zhuo (2018) is presented. The method is of general interest as it presents an explicit formula for the local sensitivity of log marginal likelihood when observations vary by a small amount. The remarkable feature is that the formula is very easy to compute and does not require knowledge of the marginal likelihood which is, invariably, extremely difficult to compute. Similar expressions are derived for posterior moments and other functions of interest, including inefficiency. Methods for examining prior sensitivity in a straightforward way are also presented. The methods are illustrated in the context of a stochastic production frontier.

Keywords: Local influence; Bayesian Analysis; Marginal Likelihood; Posterior moments; Markov Chain Monte Carlo.

JEL Classifications: C11, C13.

Acknowledgments: The author wishes to thank an anonymous referee for useful comments on an earlier version.

1 Introduction

In a recent paper, Zhuo (2018) examined local influence in the context of a stochastic frontier model with normal - half-normal errors. His approach is based on the geometric normal curvature of likelihood displacement. Unfortunately, Zhuo's (2018) cannot be extended easily to examine local sensitivity of certain functions of the parameters or technical inefficiency estimates. In this paper, we derive new results relating to local sensitivity of marginal likelihood and posterior moments and we apply them in the same context, viz. a stochastic frontier model. However, most results are more generally applicable to Bayesian analysis.

Suppose $x_i \in \mathfrak{R}^k$ is a vector of explanatory variables, $\beta \in \mathfrak{R}^k$ is a vector of parameters and the model is:

$$y_i = x_i' \beta + v_i - u_i, \quad i = 1, \dots, n, \quad (1)$$

where

$$v_i \sim \mathcal{N}(0, \sigma_v^2), \quad u_i \sim \mathcal{N}_+(0, \sigma_u^2), \quad i = 1, \dots, n. \quad (2)$$

The two error components are independent and independent from the regressors. The parameter vector is $\theta = [\beta', \sigma_v, \sigma_u]'$.

The likelihood function of the model is given by:

$$\mathcal{L}(\theta; Y) = \left(\frac{2}{\sigma}\right)^n \prod_{i=1}^n \varphi\left(\frac{\varepsilon_i}{\sigma}\right) \prod_{i=1}^n \Phi\left(-\frac{\lambda \varepsilon_i}{\sigma}\right), \quad (3)$$

where $\varepsilon_i = y_i - x_i' \beta$, $\sigma^2 = \sigma_v^2 + \sigma_u^2$, $\lambda = \frac{\sigma_u}{\sigma_v}$ and $Y = \{y, \dots, y_n\}$. Moreover, $\varphi(z) = (2\pi)^{-1/2} e^{-z^2/2}$ and $\Phi(z) = \int_{-\infty}^z \varphi(t) dt$.

Given a prior $p(\theta)$ the posterior is given by Bayes' theorem:

$$p(\theta|Y) \propto \mathcal{L}(\theta; Y)p(\theta). \quad (4)$$

2 Bayesian local sensitivity

The central object of interest in Bayesian analysis is posterior moments. For example, the vector of posterior means is

$$m(Y) = \int \theta p(\theta|Y) d\theta. \quad (5)$$

For model comparison the marginal likelihood is also of considerable interest:

$$\mathcal{M}(Y) = \int \mathcal{L}(\theta; Y)p(\theta) d\theta. \quad (6)$$

The marginal likelihood is very difficult to obtain and often expensive numerical approximations are used (Gelman and Meng, 1998, Han and Carlin, 2001, Kass and Raftery, 1995). Of course, it is simply the integrating constant of the posterior density in (4). It turns out that we can examine the local sensitivity of marginal likelihood without even having to compute it.

Indeed, the expression for local sensitivity can be derived as follows:

$$\frac{\partial \mathcal{M}(Y)}{\partial y_i} = \int \frac{\partial \mathcal{L}(\theta; Y)}{\partial y_i} p(\theta) d\theta = \int \frac{\partial \ln \mathcal{L}(\theta; Y)}{\partial y_i} \mathcal{L}(\theta; Y) p(\theta) d\theta = \int \frac{\partial \ln \mathcal{L}(\theta; Y)}{\partial y_i} \frac{\mathcal{L}(\theta; Y) p(\theta)}{M(Y)} M(Y) d\theta,$$

from which we obtain:

$$\frac{\partial \ln \mathcal{M}(Y)}{\partial y_i} = \int \frac{\partial \ln \mathcal{L}(\theta; Y)}{\partial y_i} \frac{\mathcal{L}(\theta; Y) p(\theta)}{M(Y)} d\theta = \int \frac{\partial \ln \mathcal{L}(\theta; Y)}{\partial y_i} p(\theta|Y) d\theta = \mathcal{E}_{\theta|Y} \left\{ \frac{\partial \ln \mathcal{L}(\theta; Y)}{\partial y_i} \right\}, \quad i = 1, \dots, n. \quad (7)$$

Therefore, the local sensitivity of marginal likelihood to any given observation is a posterior expectation of the derivative of the log likelihood with respect to the observation. This derivative is very often available analytically, and the posterior expectation can be approximated accurately provided a Markov Chain Monte Carlo sample $\{\theta^{(s)}, s = 1, \dots, S\}$ is available:

$$\frac{\partial \ln \mathcal{M}(Y)}{\partial y_i} \simeq S^{-1} \sum_{s=1}^S \frac{\partial \ln \mathcal{L}(\theta^{(s)}; Y)}{\partial y_i}. \quad (8)$$

Let us take now the case of local sensitivity of posterior moments or any function of the parameters, say $f(\theta)$. The posterior expectation, $\mathcal{E}_{\theta|Y} f(\theta)$ is:

$$m(Y) = \int f(\theta) p(\theta|Y) d\theta = \int f(\theta) \frac{\mathcal{L}(\theta; Y) p(\theta)}{\mathcal{M}(Y)} d\theta. \quad (9)$$

Taking derivatives, we have the final expression:

$$\frac{\partial m(Y)}{\partial y_i} = \mathcal{E}_{\theta|Y} f(\theta) \left\{ \frac{\partial \ln \mathcal{L}(\theta; Y)}{\partial y_i} - \frac{\partial \ln \mathcal{M}(Y)}{\partial y_i} \right\}. \quad (10)$$

In this expression, $\frac{\partial \ln \mathcal{M}(Y)}{\partial y_i}$ has been computed in (7), and $\frac{\partial \ln \mathcal{L}(\theta; Y)}{\partial y_i}$ are also available from the intermediate computations in (7).

It should be mentioned that Zhu and Ibrahim (2011) propose techniques organized around the perturbation manifold and use differential geometry to examine simultaneously changes in the data, the prior and the model. Changes in the prior, for example, are modelled using the ϵ -contamination class (Berger, 1990). One disadvantage of their approach is that they need MCMC samples from both the baseline and the perturbed model and some care must be exercised when choosing the types of perturbation to the data and / or the prior. In addition, some analytical work is needed to obtain tangent vectors of the perturbation parameters which are closely associated to Fisher information. For complicated models it seems that the approach in Zhu and Ibrahim (2011) requires both analytical work as well as intensive MCMC which we avoid in our approach. It seems that the types of perturbations they consider, can be examined in the present framework using the techniques we have proposed without additional MCMC: the only requirement is that MCMC samples are available from the baseline model.

3 Application to the stochastic frontier model

From the log of the likelihood of the frontier model in (3) we can obtain the derivatives:

$$\frac{\partial \ln \mathcal{L}(\theta; Y)}{\partial y_i} = -\frac{\varepsilon_i}{\sigma^2} - \frac{\lambda}{\sigma} \frac{\varphi(-\lambda\varepsilon_i/\sigma)}{\Phi(-\lambda\varepsilon_i/\sigma)}, \quad i = 1, \dots, n. \quad (11)$$

To illustrate the concepts, we generate $n = 101$ observations from a frontier model with an intercept and two regressors (generated from standard normal distributions), $\sigma_v = 0.1$, $\sigma_u = 0.5$, $\beta_1 = -2$ and the remaining coefficients set to 0.5. Our prior is:

$$p(\beta, \sigma_v, \sigma_u) \propto \sigma_v^{-1} e^{-0.01/(2\sigma_v^2)} \sigma_u^{-1} e^{-0.01/(2\sigma_u^2)}. \quad (12)$$

The prior for β is flat. For σ_v and σ_u the priors are improper, uninformative relative to the likelihood and take into account the analysis in Fernandez, Osiewalski, and Steel (1997) so that the posterior is proper.

We perform MCMC analysis using a Metropolis algorithm¹ with 60,000 passes the first 10,000 are discarded to mitigate possible start up effects.²

Local sensitivity of log marginal likelihood is presented in Figure 1. Local sensitivity of posterior means of the parameters are presented in Figure 2. Marginal likelihood is, generally, sensitive to omitting observations and particularly observation 64 resulting in a drop of 20 units in log marginal likelihood which is considerable. From Figure 2 we can see that this is because σ_v increases and σ_u decreases considerably as the result of dropping this observation.

In Figure 2, we present also the actual effect on posterior means when an observation is omitted. MCMC is repeated using 60,000 passes with a burn-in phase whose length is 10,000. Local sensitivity tends to overestimate the change in posterior means and “global” sensitivity analysis tends to support the idea that the parameters are fairly robust to omission of observations. Finally, in Figure 3 we provide histograms of sensitivity for posterior means and posterior standard deviations across all parameters so that magnitudes can be appreciated.

4 Local prior sensitivity

Beyond serious doubt, the most frequent question addressed to Bayesians is whether their results are sensitive to the prior. One way to address the question is to change the prior and perform MCMC again to present the new results. In this section we show that this can be avoided. Suppose the prior is indexed by certain hyperparameters α and we denote it by $p_\alpha(\theta)$. As the marginal likelihood is:

$$\mathcal{M}_\alpha(Y) = \int \mathcal{L}(\theta; Y) p_\alpha(\theta) d\theta, \quad (13)$$

¹The Metropolis algorithm is an independence chain whose proposal distributions are uniform for each parameter. The range of the uniform distribution is adjusted during the burn-in phase so that approximately 25% of all candidates are accepted.

²Gauss programs are available on request.

Figure 1: Local sensitivity of log marginal likelihood

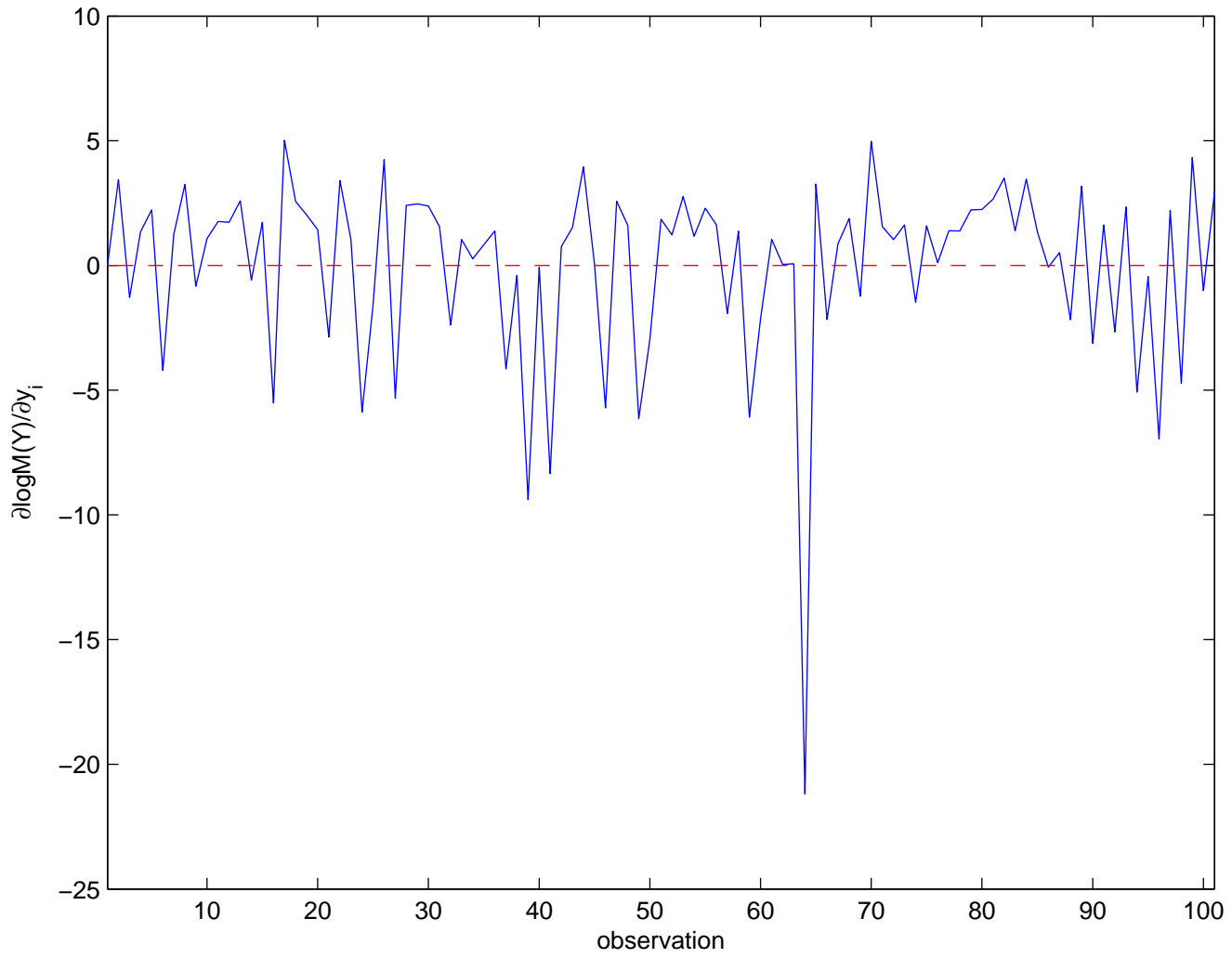


Figure 2: Local sensitivity of posterior means

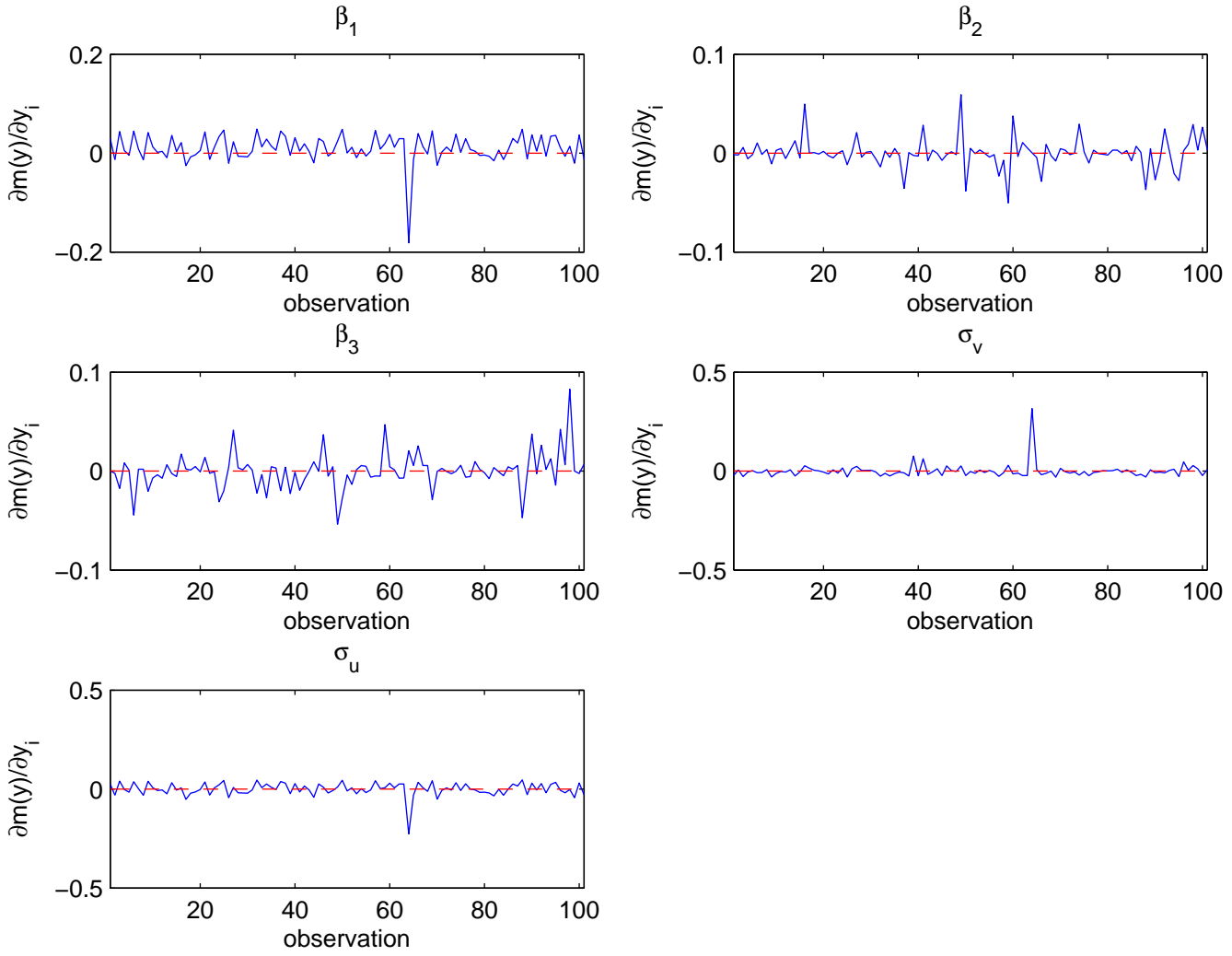
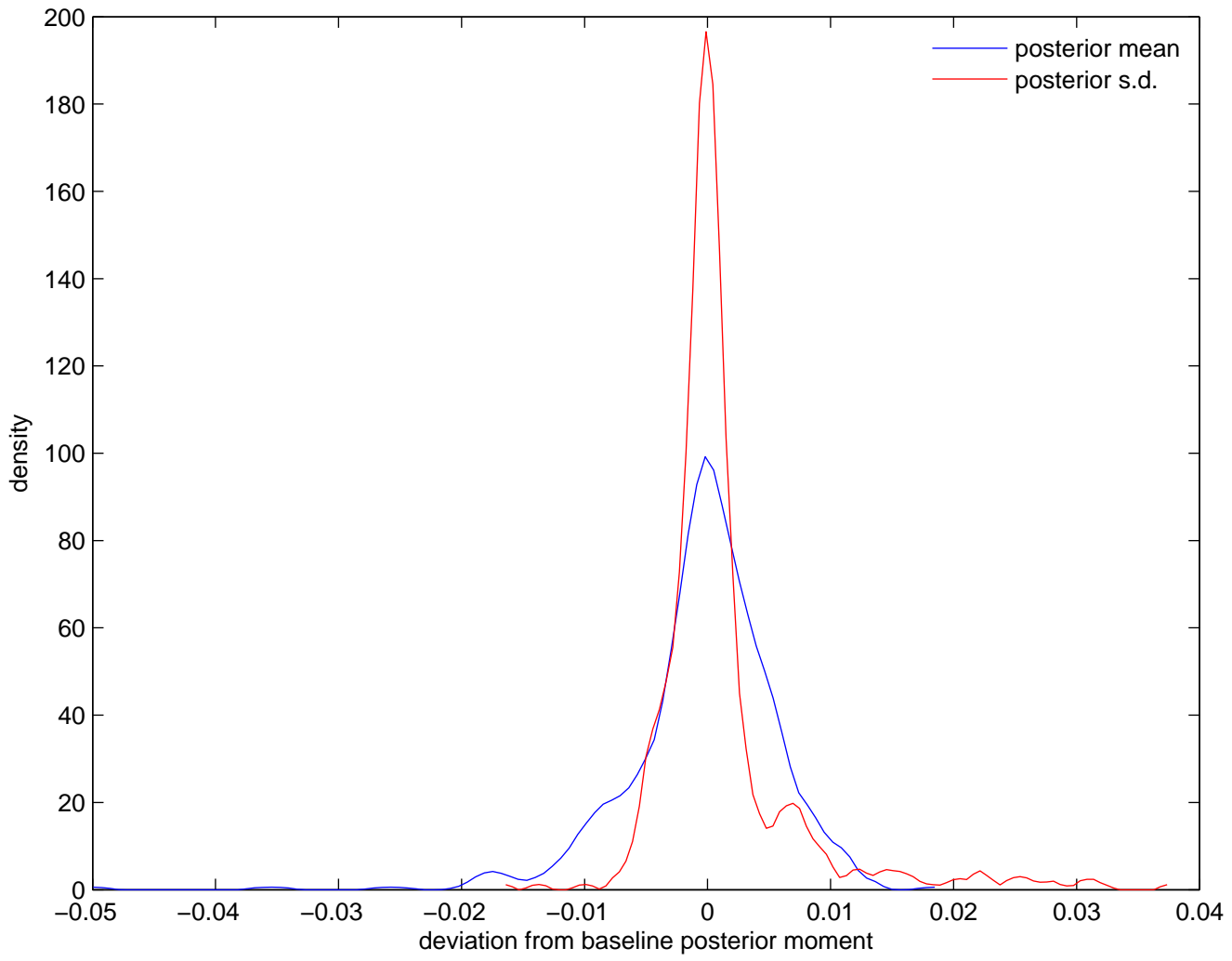


Figure 3: Global sensitivity of posterior means and s.d.



it is easy to show that:

$$\frac{\partial \ln \mathcal{M}_\alpha(Y)}{\partial \alpha} = \mathcal{E}_{\theta|Y} \frac{\partial \ln p_\alpha(\theta)}{\partial \alpha}, \quad (14)$$

which involves only a derivative of the log prior with respect to the hyperparameter(s) and the posterior expectation can be approximated accurately provided a MCMC sample is available. To illustrate in the context of the stochastic frontier model, suppose:

$$p(\sigma_v, \sigma_u) \propto \sigma_v^{-1} e^{-0.01/(2\sigma_v^2)} (\alpha/2)^{-\underline{n}/2} \sigma_u^{-(\underline{n}+1)} e^{-\alpha/(2\sigma_u^2)}. \quad (15)$$

Both priors have the form $p(\sigma) \propto \sigma^{-(\underline{n}+1)} e^{-\alpha/(2\sigma^2)}$. For σ_v we have $\underline{n} = 0$, which makes it improper. For σ_u we set $\underline{n} = 1$ so that the prior is proper, and we examine sensitivity with respect to hyperparameter $\alpha \geq 0$. Since

$$\frac{\partial \ln p_\alpha(\theta)}{\partial \alpha} = \frac{n}{2} - \frac{\alpha}{2} \mathcal{E}_{\theta|Y} \left(\frac{1}{\sigma_u^2} \right), \quad (16)$$

prior sensitivity depends on

$$\frac{\partial \ln \mathcal{M}_\alpha(Y)}{\partial \alpha} = \frac{1}{2} (n - \alpha g), \quad (17)$$

where $g = \mathcal{E}_{\theta|Y} \left(\frac{1}{\sigma_u^2} \right)$ which turns out to be approximately 2 based on our MCMC simulation. Near $\alpha = 0$ the sensitivity is maximal and is negligible for $\alpha \simeq \frac{n}{2}$. The simplicity of prior sensitivity suggests that it should be possible to be used widely in applied econometrics and resolve issues of prior sensitivity in a straightforward manner.

5 Sensitivity of efficiency

Inefficiency can be estimated using:

$$U_i(Y) = \mathcal{E}_{\theta|Y} U_i(\theta, Y), \quad U_i(\theta, Y) = \sigma_* \left\{ \frac{\varphi(\varepsilon_i \lambda / \sigma)}{1 - \Phi(\varepsilon_i \lambda / \sigma)} - \frac{\varepsilon_i \lambda}{\sigma} \right\}, \quad (18)$$

where $\sigma_*^2 = \sigma_v^2 \sigma_u^2 / (\sigma_v^2 + \sigma_u^2)$, see Kumbhakar and Lovell (2000, p. 78). In turn, we have:

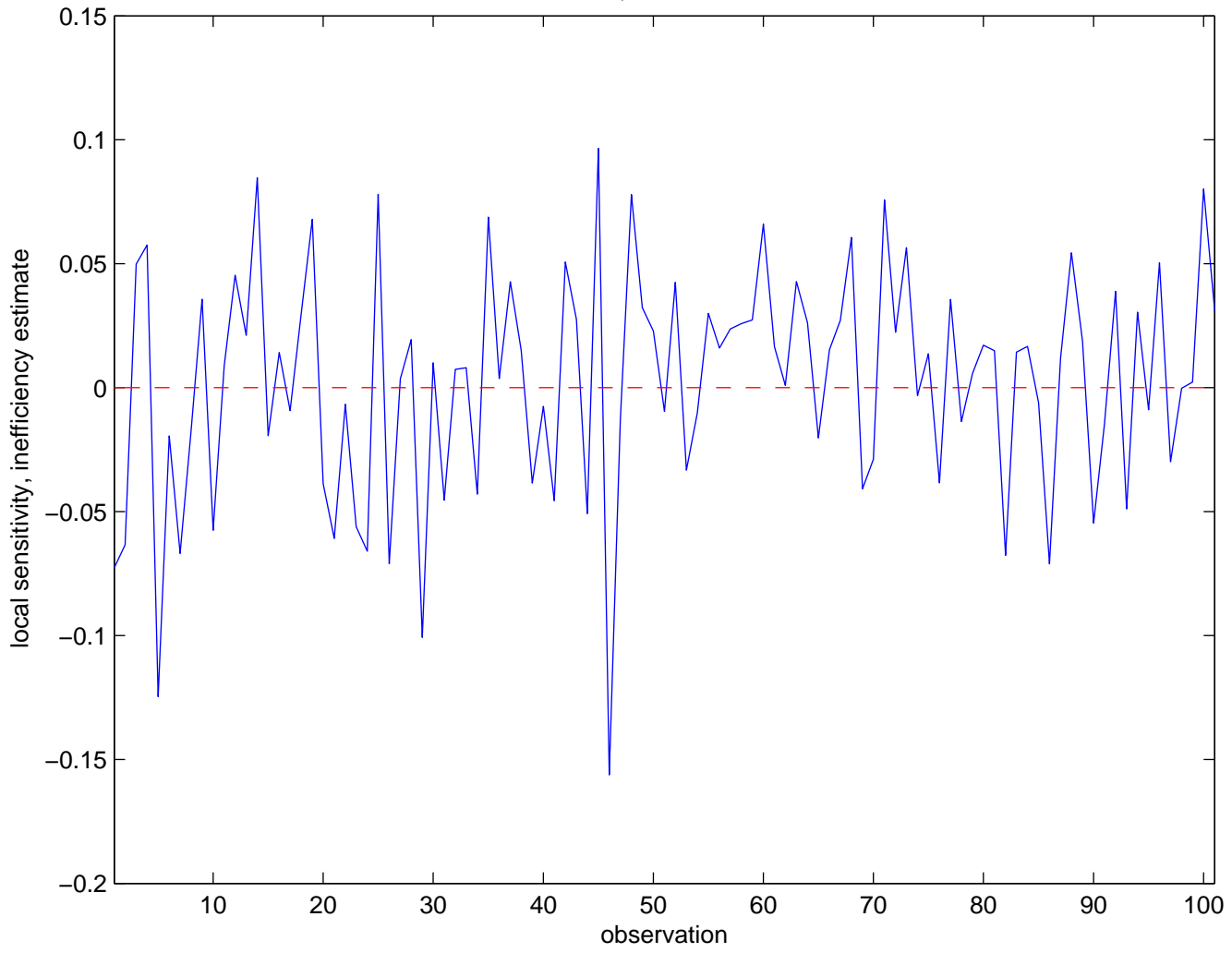
$$\frac{\partial U_i(Y)}{\partial y_i} = \frac{\partial}{\partial y_i} \mathcal{E}_{\theta|Y} U_i(\theta, Y) = \mathcal{E}_{\theta|Y} \frac{\partial U_i(\theta, Y)}{\partial y_i}, \quad (19)$$

where $\frac{\partial U_i(\theta, Y)}{\partial y_i} = -\frac{\lambda \sigma_*}{\sigma} \left\{ \frac{\lambda}{\sigma} \varepsilon_i \Lambda_i(\theta, Y) - \Lambda_i(\theta, Y)^2 + 1 \right\}$, and $\Lambda_i(\theta, Y) = \frac{\varphi(\varepsilon_i \lambda / \sigma)}{\Phi(\varepsilon_i \lambda / \sigma)}$, $i = 1, \dots, n$.

This local sensitivity can be computed easily once MCMC draws are available. The results are presented in Figure 4. Generally, local sensitivity is considerable and can be as large as ten to fifteen percentage points. In such cases, we recommend presentation of posterior densities of efficiency estimates rather than simple sample summaries and, of course, presentation of the local sensitivity measures.

Figure 4: Local sensitivity (%) of efficiency

$$\frac{\partial E_{\theta|Y} U_i(\theta, Y)}{\partial y_i}$$



Concluding remarks

We present several results related to local sensitivity of marginal likelihood, prior hyperparameters and other functions of interest in Bayesian analysis. We illustrate the new concepts in the context of the stochastic frontier model arguing that Zhuo's (2018) local sensitivity measure does not address all issues related to sensitivity in likelihood analysis. Notably, local sensitivity of marginal likelihood (that can be used to obtain Bayes factors for model comparison) does not depend on marginal likelihood estimation and requires only MCMC draws from the posterior. Local sensitivity of efficiency scores is shown to be considerable in some cases and can be as large as 10-15 percentage points. Therefore, presentation of local sensitivity measures along with the scores themselves, is essential.

References

- Berger, J. O. (1990). Robust Bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and Inference*, 25, 303–28.
- Fernandez, C., J. Osiewalski, and M.F.J. Steel (1997). “On the use of panel data in stochastic frontier models with improper priors”. *Journal of Econometrics* 79 (1), 169-193.
- Gelman, A. and Meng, X.-L. (1998). “Simulating normalizing constants: From importance sampling to bridge sampling to path sampling.” *Statistical Science*, 13: 163–185.
- Han, C. and Carlin, B. P. (2001). “MCMC methods for computing Bayes factors: A comparative review.” *Journal of the American Statistical Association*, 96: 1122–1132.
- Kass, R. and Raftery, A. (1995). “Bayes factors.” *Journal of the American Statistical Association*, 90: 773–795.
- Kumbhakar, S.C. and C.A.K. Lovell (2000). “Stochastic Frontier Analysis”. Cambridge, Cambridge University Press.
- Zhu, H. and J. G. Ibrahim (2011). Bayesian influence analysis: a geometric approach. *Biometrika* 98, 307-323.
- Zhuo, S. (2018). “Local influence analysis of stochastic frontier estimation: A case-weights perturbation approach”. *Economics Letters*, forthcoming, <https://doi.org/10.1016/j.econlet.2018.01.008>.