ELSEVIER

Statistics & Probability Letters 48 (2000) 253–259

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

Robust simulation-based estimation

Marc G. Genton[a,*], Xavier de Luna[b]

[a] Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA
[b] Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT, UK

Received July 1999; received in revised form September 1999

Abstract

The simulation-based inferential method called *indirect inference* was originally proposed for statistical models whose likelihood is difficult or even impossible to compute and/or to maximize. In this paper, indirect estimation is proposed as a device to robustify the estimation for models where this is not possible or difficult with classical techniques such as M-estimators. We derive the influence function of the indirect estimator, and present results about its gross-error sensitivity and asymptotic variance. Two examples from time series are used for illustration. © 2000 Elsevier Science B.V. All rights reserved

Keywords: Asymptotic variance; B-robustness; Gross-error sensitivity; Influence function; M-estimator; Indirect inference; Time series

1. Introduction

Suppose a set of observations $x_1,\dots,x_n$ has been collected, and are assumed to have been generated from a probability model $F(\boldsymbol{\theta})$, where $\boldsymbol{\theta}\in\Theta$ is an unknown parameter. Maximum likelihood is then commonly used to estimate $\boldsymbol{\theta}$. However, the latter is notoriously not robust to the presence of outliers, and robust inferential methods have been developed for that reason. A theory of robustness (Huber, 1981; Hampel et al., 1986) has developed around simple models for contaminated data. For instance, when observations are independently distributed, contamination models are of the form $(1-\varepsilon)F(\boldsymbol{\theta})+\varepsilon\Delta_x$, with ε small and where Δ_x is the measure putting all its mass at x. For dependent data more complex contaminations may occur, see Martin and Yohai (1986). The properties of different inferential methods can be studied under the original assumed model $F(\boldsymbol{\theta})$, and a contaminated version. Maximum likelihood is asymptotically optimal under the former while having often poor performances under the latter. For simple models (typically linear in the parameter $\boldsymbol{\theta}$ and assuming independence), robust inferential methods are readily available (Hampel et al., 1986). However, for more complex situations such robust inferential methods and their theoretical properties may be intricate if not impossible to obtain. Indirect inference (Gouriéroux et al., 1993) may then be a solution as it was first noticed

* Corresponding author.
E-mail addresses: genton@math.mit.edu (M.G. Genton), xavier@stats.ucl.ac.uk (X. de Luna)

PII: S0167-7152(00)00004-3

in de Luna and Genton (1998), where it was used to robustify the estimation of autoregressive and moving average (ARMA) models in a time series context. An indirect estimator of $\boldsymbol{\theta}$ is available when it is possible (and simple) to draw pseudo-observations from $F(\boldsymbol{\theta})$ and if there is a simple auxiliary reparameterization $\tilde{F}(\boldsymbol{\alpha})$ of $F(\boldsymbol{\theta})$, where $\boldsymbol{\alpha}$ is an unknown auxiliary parameter. This auxiliary parameter is chosen such that it is easier to estimate than $\boldsymbol{\theta}$ (the original idea behind the indirect inference proposal), or easier to estimate robustly (the studied concept here). The auxiliary parameter is thus estimated with the observed data yielding $\hat{\boldsymbol{\alpha}}$, and with simulated data (at least as many as were observed) from $F(\boldsymbol{\theta})$, yielding $\boldsymbol{\alpha}^*$, a function of $\boldsymbol{\theta}$ by construction. The indirect estimator is then defined as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)' \Omega (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*), \tag{1}$$

where Ω can be chosen in order to maximize efficiency. Note that the pseudo-observations drawn from $F(\boldsymbol{\theta})$ are based on the same random generator seed for all $\boldsymbol{\theta}$. For the indirect estimator to be consistent and asymptotically normal, the binding function $h(\boldsymbol{\theta}) = \boldsymbol{\alpha}$ needs to be locally injective around the true value of $\boldsymbol{\theta}$ (Gouriéroux et al., 1993). In the latter article, the two estimators $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}^*$ were assumed to be identical, although this does not need to be the case as long as they are consistent, see de Luna and Genton (1998). For instance, a robust estimator $\hat{\boldsymbol{\theta}}$ can be obtained with a robust estimator $\hat{\boldsymbol{\alpha}}$. On the other hand, $\boldsymbol{\alpha}^*$ is evaluated on outlier-free simulated data and thus is most conveniently chosen to be efficient under the uncontaminated model.

Indirect inference was originally introduced and used by econometricians faced with complex dynamic models for which more conventional inferential methods, typically based on a likelihood or on moments, are not applicable because the corresponding estimating criterion is not available neither analytically nor algorithmically; various such examples are discussed in Gouriéroux and Monfort (1996). In the absence of a tractable analytical estimating criterion, (1) is constructed with the help of a data simulator. An intuitive motivation for the minimization of criterion (1) is that it corresponds to looking for $\boldsymbol{\theta}$ such that the distance between $\boldsymbol{\alpha}^*$ and $\hat{\boldsymbol{\alpha}}$ is minimum.

In this paper we study robust indirect estimators in a general framework. For this purpose, we start by defining in Section 2 the influence function, an important analytical tool used to compare robustness properties of estimators. The influence function for the indirect estimator is then derived under general conditions, allowing us to indicate how robust indirect estimation can be obtained. The influence function of some indirect estimators is studied in Section 3 for a moving average model, and in Section 4 for an asymmetric moving average model. Section 5 concludes the paper.

2. Influence function for indirect estimators

The influence function (Hampel, 1974) is a tool used to describe the robustness properties of an estimator. Its importance lies in its appealing heuristic interpretation: it measures the asymptotic bias caused by an infinitesimal contamination of the observations. For a multidimensional statistical functional T at a distribution F, the influence function $\mathrm{IF}(v,T,F)$ at the probability measure v (defining the contamination process) is defined by the limit:

$$\mathrm{IF}(v,T,F) = \lim_{\varepsilon\to 0+} \frac{T(F_\varepsilon) - T(F)}{\varepsilon}, \tag{2}$$

in those v where this limit exists. As $\varepsilon \to 0$, the distribution of the contaminated process $F_\varepsilon \to F$. Hampel's original definition was aimed at the independent and identically distributed case, with $v = \Delta_x$ the measure that puts all its mass at x and $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_x$. Its extension to the time-series setting has first been proposed by Künsch (1984). Martin and Yohai (1986) consider a different contaminated distribution F_ε which they argue is more realistic because it corresponds to usual contaminated models, e.g. additive or replacement outliers.

The general influence function defined by (2) can be derived for the indirect estimator $\hat{\boldsymbol{\theta}}$ from the influence function of the auxiliary estimator $\hat{\boldsymbol{\alpha}}$. They are proportional so that the boundedness of the latter implies the boundedness of the former.

Denoted by A and T the statistical functionals corresponding, respectively, to the estimators $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\theta}}$ defined in the previous section. Thus, A is such that $\hat{\boldsymbol{\alpha}} = A(F_n)$ for any n and F_n (or $\hat{\boldsymbol{\alpha}}$ tends in probability towards $A(F)$), where F_n is the empirical distribution of the sample, and F the generating distribution function (Hampel et al., 1986, p. 82). In the sequel Fisher consistency is assumed, i.e. $A(F(\boldsymbol{\theta})) = h(\boldsymbol{\theta})$, for all $\boldsymbol{\theta} \in \Theta$. The functional T is characterized more precisely in the proof below. Let $\mathrm{IF}_{\mathrm{aux}}(v, A, F)$ be the vector influence function of the auxiliary estimator $\hat{\boldsymbol{\alpha}}$, and $\mathrm{IF}_{\mathrm{ind}}(v, T, F)$ be the vector influence function of the indirect estimator $\hat{\boldsymbol{\theta}}$. We have the following result.

Theorem 1. *Let* $\hat{\boldsymbol{\theta}}$ *be the indirect estimator* (1) *based on* $\hat{\boldsymbol{\alpha}}$ *and* $\boldsymbol{\alpha}^*$, *the latter a consistent estimator of* $h(\boldsymbol{\theta})$ *for all* $\boldsymbol{\theta} \in \Theta$. *Assume further that at* $T(F)$, h *is locally injective and the Jacobian matrix* D *of* h *exists. If the influence function of the auxiliary estimator exists, then the influence function of the indirect estimator is*

$$\mathrm{IF}_{\mathrm{ind}}(v, T, F) = P(T(F))\,\mathrm{IF}_{\mathrm{aux}}(v, A, F), \tag{3}$$

where $P(T(F)) = [D(T(F))'\Omega D(T(F))]^{-1} D(T(F))'\Omega$.

Proof. We need first to characterize the functional T. Because $\boldsymbol{\alpha}^*$ is consistent, i.e. it converges towards $h(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$, either in probability or with probability one, the criterion minimized in (1) converges uniformly in probability towards

$$(A(F) - h(\boldsymbol{\theta}))'\Omega(A(F) - h(\boldsymbol{\theta})). \tag{4}$$

Therefore, $\hat{\boldsymbol{\theta}}$ converges in probability towards the argument of the minimum of (4). This argument is thus $T(F)$. The score corresponding to (4) is equal to zero at $\boldsymbol{\theta} = T(F)$:

$$D(T(F))'\Omega(A(F) - h(T(F))) = \mathbf{0}, \tag{5}$$

where $D(\cdot)$ is the Jacobian matrix which is of full-column rank by the injectivity assumption. Finally, replacing F by F_ε in (5), differentiating with respect to ε and evaluating at $\varepsilon = 0$ yields the theorem's result. □

Therefore, $\mathrm{IF}_{\mathrm{ind}}$ is bounded if $\mathrm{IF}_{\mathrm{aux}}$ is bounded, and in particular, we can replace $\mathrm{IF}_{\mathrm{aux}}$ by either Hampel's (1974) influence function, or Martin and Yohai's (1986) influence function. The former can usually be used to compute the asymptotic variance as

$$V(T, F) = \int \mathrm{IF}(\mathbf{x}, T, F)\mathrm{IF}(\mathbf{x}, T, F)'\,\mathrm{d}F(\mathbf{x}). \tag{6}$$

We have the following relationship between $V(T, F)$'s:

$$V_{\mathrm{ind}}(T, F) = P(T(F))V_{\mathrm{aux}}(A, F)P(T(F))'. \tag{7}$$

Note, however, that $V_{\mathrm{ind}}(T, F)$ does not describe the whole asymptotic variance of the indirect estimator, because it does not contain the variability due to the simulations. The complete expression can be found in de Luna and Genton (1998).

The worst effect of a contamination on the estimator is described by the gross-error sensitivity $\gamma^*(T, F) = \sup_v\{\|\mathrm{IF}(v, T, F)\|\}$, where $\|\cdot\|$ denotes the Euclidean norm. If the gross-error sensitivity is finite, the estimator is said to be B-robust, i.e. robust with respect to the bias. It is often of interest to standardize the gross-error sensitivity (Hampel et al., 1986) in order to make it invariant under one-to-one parameter transformations.

The self-standardized sensitivity is defined in the metric given by the asymptotic variance, i.e. $\gamma_s^*(T,F) = \sup_v \{\mathrm{IF}(v,T,F)' V(T,F)^{-1} \mathrm{IF}(v,T,F)\}^{1/2}$, whereas the information-standardized sensitivity is defined in the local metric given by the Fisher information $J(T(F))$, i.e. $\gamma_i^*(T,F) = \sup_v \{\mathrm{IF}(v,T,F)' J(T(F)) \mathrm{IF}(v,T,F)\}^{1/2}$. He and Simpson (1992) give arguments for such standardizations. Here, we have the following result.

Corollary 1.1. *Under the hypothesis of Theorem* 1, *if* $\hat{\boldsymbol{\alpha}}$ *is B-robust*, *then* $\hat{\boldsymbol{\theta}}$ *is also B-robust. Moreover*, *if* $\dim(\boldsymbol{\alpha}) = \dim(\boldsymbol{\theta})$, *then* $\gamma_{\mathrm{s,ind}}^*(T,F) = \gamma_{\mathrm{s,aux}}^*(A,F)$ *and* $\gamma_{\mathrm{i,ind}}^*(T,F) = \gamma_{\mathrm{i,aux}}^*(A,F)$.

The breakdown point is another important feature of reliability of an estimator (Huber, 1981, 1984; Hampel et al., 1986). It indicates how many data points need to be replaced by arbitrary values to destroy the estimator, i.e. to bring the estimator to the boundaries of the parameter space. Note that this concept can also be applied to a dependent data setting; see, e.g., Lucas (1997) and Ma and Genton (2000) for a time series context, as well as Genton (1998) for a spatial random field situation. The indirect estimator of $\boldsymbol{\theta}$ will inherit the breakdown point of the auxiliary estimator of $\boldsymbol{\alpha}$, because the simulated data are outlier-free. In fact, one can expect that other robustness properties of the auxiliary estimator will be transmitted to the indirect estimator of $\boldsymbol{\theta}$.

3. Example I: first-order moving average model

In de Luna and Genton (1998) it was proposed to use linear predictors as auxiliary parameterization to robustify the inference of autoregressive and moving average models. Here we study the influence function for the simple situation of a moving average model of order one, i.e.

$$X_t = \theta\varepsilon_{t-1} + \varepsilon_t, \tag{8}$$

where the sequence $\{\varepsilon_t\}$ is independently and identically distributed with mean zero and variance σ^2. The model is identifiable when $|\theta| < 1$. The indirect estimation of this model can be based on the auxiliary parameterization (Gouriéroux et al., 1993; de Luna and Genton, 1998)

$$X_t = \alpha X_{t-1} + U_t, \tag{9}$$

where α is such that $E(U_t X_{t-1}) = 0$, i.e. αX_{t-1} is the best linear predictor of order 1 for X_t. Higher order predictors, i.e. $\alpha_1 X_{t-1} + \cdots + \alpha_r X_{t-r}$ may be used, although here for simplicity we focus on the use of (9), further assuming that σ^2 is known.

In order to compute the influence function of the indirect estimator of θ based on (9), we need the binding function, which here is $h(\theta) = \theta/(1+\theta^2)$ and the influence function of α. The latter is not uniquely defined. In the sequel, we follow Martin and Yohai (1986) and work with the influence curve $\mathrm{IC}(\xi)$ obtained by using a degenerate measure for v in Eq. (2), where ξ represents the value of the contamination. For GM-estimators of α, the influence curve $\mathrm{IC}_\alpha(\xi)$ is given in Theorem 5.2 in Martin and Yohai (1986), for additive outliers. We consider Hampel–Krasker–Welsch type of GM-estimators, with Bisquare and Huber ψ-functions, where the tuning constants are respectively 9.36 and 2.52, in order to reach 95% efficiency at the standard normal distribution. As a particular case, the least squares (LS) estimator of α has an explicit influence curve $\mathrm{IC}_\alpha(\xi) = -\alpha(1-\alpha^2)\xi^2$. The matrix P in Theorem 1 is a scalar in our simple example and equals $P(\theta) = (1+\theta^2)^2/(1-\theta^2)$. Therefore, the influence curve for the indirect estimator of θ, based on a least squares estimator of α, is:

$$\mathrm{IC}_\theta(\xi) = \frac{\theta(\theta^4+\theta^2+1)}{\theta^4-1}\xi^2. \tag{10}$$

In Fig. 1, the influence curves $\mathrm{IC}_\theta(\xi)$ of the indirect estimators of θ based on the two GM-estimators with Bisquare (dashed curve) and Huber (dotted curve) ψ-functions, as well as on the least squares estimators

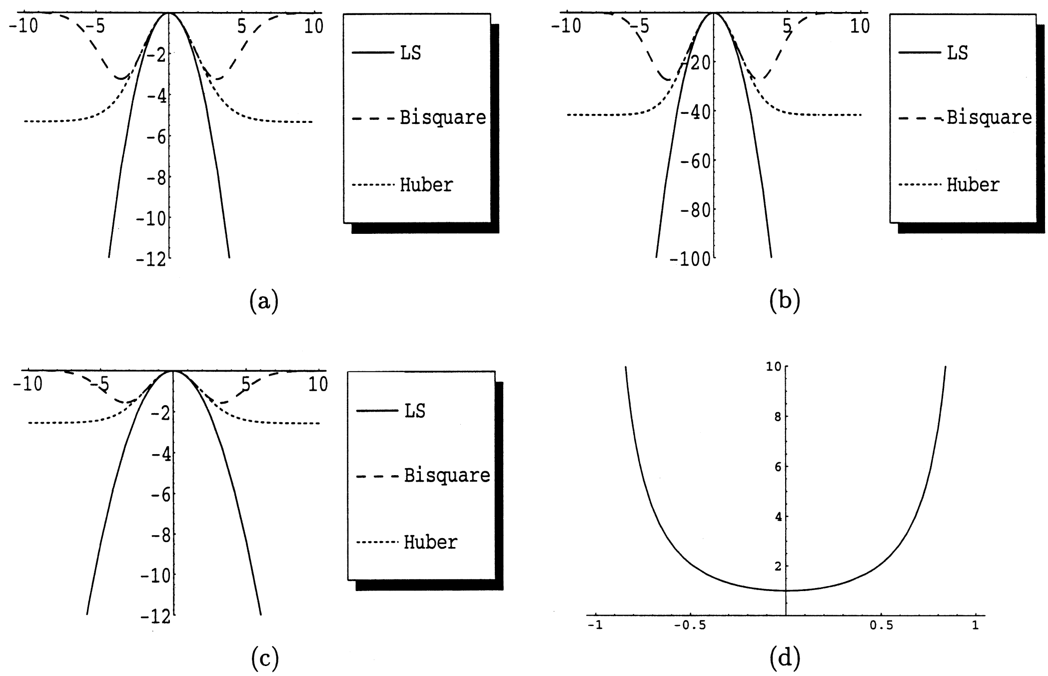

Fig. 1. Influence curves $\mathrm{IC}_\theta(\xi)$ of the indirect estimator for (a) $\theta = 0.5$, (b) $\theta = 0.9$, and influence curve $\mathrm{IC}_\alpha(\xi)$ of the auxiliary estimator for (c) $\alpha = 0.4 = h(0.5)$. The scalar matrix $P(\theta)$ is depicted in (d).

(solid curve), are depicted for $\theta = 0.5$ and 0.9. As expected, the influence curves for the two GM-estimators are bounded, whereas the one for the least squares estimator is not. Figs. 1(a) and (b) may directly be compared to Figs. 3(a) and (b) in Martin and Yohai (1986) where influence curves for direct estimators are displayed. Thus, we note that the least squares influence curves are identical. On the other hand, the indirect estimators of θ have a smaller absolute size in influence curve than the corresponding Martin and Yohai's estimators. Moreover, the latter when based on a Huber-type GM-criterion, are not B-robust while the indirect estimators are. Fig. 1(c) describes the influence curve $\mathrm{IC}_\alpha(\xi)$ of the auxiliary estimator of α, for $\alpha = 0.4 = h(0.5)$. Comparing Figs. 1(c) and (a), we can see that B-robustness is preserved through the indirect inference procedure, although the unstandardized gross-error sensitivity becomes larger. The factor inflating the influence curve of the auxiliary estimator is the scalar matrix $P(\theta)$, depicted in Fig. 1(d).

4. Example II: first-order asymmetric moving average model

In this second example we look at a non-linear moving average model introduced by Wecker (1981). The asymmetric moving average model of order one is of the form

$$X_t = \theta^+ \varepsilon_{t-1}^+ + \theta^- \varepsilon_{t-1}^- + \varepsilon_t, \tag{11}$$

where $\varepsilon_t^+ = \max(0, \varepsilon_t)$, $\varepsilon_t^- = \min(0, \varepsilon_t)$, and the sequence $\{\varepsilon_t\}$ is independently and identically distributed with mean zero and variance σ^2. The model is identifiable for $|\theta^+|, |\theta^-| < 1$ (Wecker, 1981). Note also that when $\theta^+ = \theta^-$ we obtain the linear moving average model (8). Indirect estimation of this model was proposed and studied in Brännäs and de Luna (1998), based on auxiliary parameterizations defined by best linear predictors.

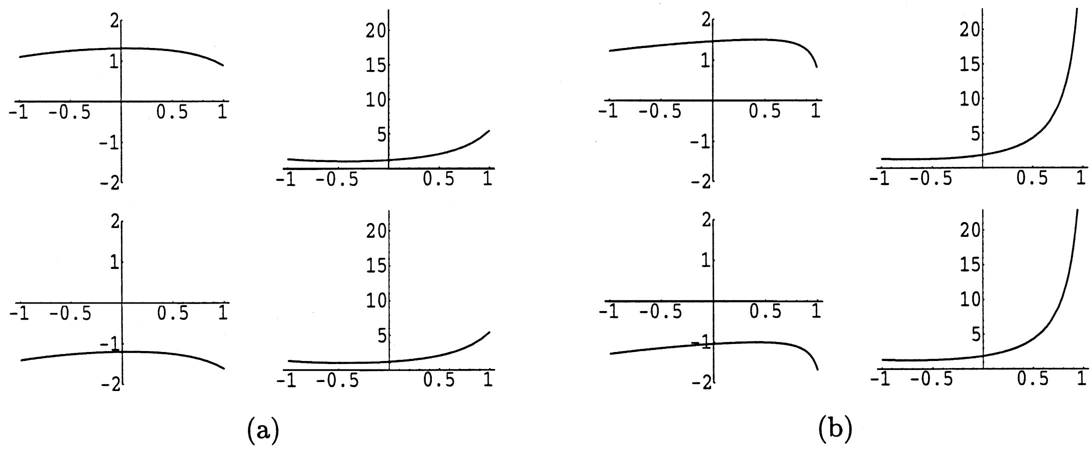

Fig. 2. Four components of the matrix $P(\boldsymbol{\theta})$ for $\theta^+ = 0.5$ (a) and $\theta^+ = 0.9$ (b). On the x-axes we have $-1 < \theta^- < 1$.

On the contrary of the previous example, the stochastic process defined by (11) may have a mean different from zero, and the auxiliary parameterization takes the form

$$(X_t - \alpha_0) = \alpha_1(X_{t-1} - \alpha_0) + U_t, \tag{12}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)'$ is such that $E(U_t(1, X_{t-1})') = \mathbf{0}$, i.e. $\alpha_1(X_{t-1} - \alpha_0)$ is the best linear predictor of order one for $(X_t - \alpha_0)$. Again higher-order linear predictors can be used. Brännäs and de Luna (1998) found this reparameterization to yield inefficient indirect estimators when $\theta^+ - \theta^-$ was far from zero. This is not surprising because then linear predictors are not optimal in a mean squared prediction error sense. On the other hand, a Wald test of the null hypothesis $\theta^+ - \theta^- = 0$ based on the indirect estimator was shown to have good power properties.

For the indirect estimator of $\boldsymbol{\theta} = (\theta^+, \theta^-)'$ based on (12), the binding function is given by (Wecker, 1981)

$$\alpha_0 = \frac{\theta^+ - \theta^-}{\sqrt{2\pi}},$$

$$\alpha_1 = \frac{\theta^+ + \theta^-}{2\gamma},$$

$$\gamma = 1 + \frac{(\theta^+)^2 + (\theta^-)^2}{2} - \frac{1}{2\pi}(\theta^+ - \theta^-)^2,$$

where γ is the variance of $\{X_t\}$. We have set $\sigma^2 = 1$. The Jacobian matrix $D(\boldsymbol{\theta})$ and the matrix $P(\boldsymbol{\theta})$, can be deduced from these formulae. In Fig. 2 the four components of P are plotted for $\theta^+ = 0.5$, 0.9 and $-1 < \theta^- < 1$. We can see that these matrices keep bounded, although, for $\theta^+ = 0.9$ and values of θ^- approaching one, the multiplicative factor may be quite large. This situation is similar to the case of a symmetric moving average model with coefficient close to one, and indeed in Fig. 1(d) we observe a similar phenomenon when this coefficient is approaching one.

Finally, influence functions could be obtained as in Section 3, once $P(\boldsymbol{\theta})$ is known. For instance, robustness is achieved by using GM-estimators for (12).

5. Discussion

In the examples of Sections 3 and 4, the binding function $h(\cdot)$ and its Jacobian could be computed analytically. This, however, does not need to be the case to perform the type of analysis presented. An example where the function $h(\cdot)$ is not possible to retrieve analytically arises when using an indirect estimator for spatial autoregressive models as it was proposed in de Luna and Genton (1999). The link function can be estimated with simulated data, indeed $\hat{h}(\boldsymbol{\theta}) = \boldsymbol{\alpha}^*$ is a consistent estimator of $h(\boldsymbol{\theta})$. The Jacobian can then be approximated numerically, for instance for the one-dimensional parameter case by $\{\hat{h}(\theta + \varepsilon/2) - \hat{h}(\theta - \varepsilon/2)\}/\varepsilon$, with ε small.

The methodology introduced in this article allows us to perform robust inference for models where classical methods such as M-type estimators are not directly implementable. This is achieved by robustifying the indirect estimator of Gouriéroux et al. (1993). Furthermore, previous situations for which indirect inference has proven useful may take advantage of the robust estimator herein proposed. Such instances include inference for spatial processes (de Luna and Genton, 1999), stochastic differential equation models (e.g. models for option pricing and interest rate modeling), latent variable models (e.g. stochastic volatility models), and complex macroeconomics models (non-linear simultaneous equations), see Gouriéroux and Monfort (1996).

Acknowledgements

The authors thank an anonymous referee for valuable comments.

References

Brännäs, K., de Luna, X., 1998. Generalized method of moment and indirect estimation of the ARASMA model. Comput. Statist. 13, 485–494.

de Luna, X., Genton, M.G., 1998. Simulation-based robust estimation of ARMA models. Research Report, 198, Dept. of Statistical Science, University College London.

de Luna, X., Genton, M.G., 1999. Indirect inference for spatio-temporal autoregression models. Proceedings of Spatial-temporal modeling and its application, Leeds, UK, 61–64.

Genton, M.G., 1998. Spatial breakdown point of variogram estimators. Math. Geol. 30, 853–871.

Gouriéroux, C., Monfort, A., Renault, A.E., 1993. Indirect inference. J. Appl. Econometrics 8, 85–118.

Gouriéroux, C., Monfort, A., 1996. Simulation-Based Econometric Methods. Oxford University Press, Oxford.

Hampel, F.R., 1974. The influence curve and its role in robust estimation. J. Amer. Statist. Assoc. 69, 383–393.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics: The Approach based on Influence Functions. Wiley, New York.

He, X., Simpson, D.G., 1992. Robust direction estimation. Ann. Statist. 20, 351–369.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

Huber, P.J., 1984. Finite sample breakdown point of M- and P-estimators. Ann. Statist. 12, 119–126.

Künsch, H., 1984. Infinitesimal robustness for autoregressive processes. Ann. Statist. 12, 843–863.

Lucas, A., 1997. Asymptotic robustness of least median of squares for autoregressions with additive outliers. Comm. Statist. Therory Methods 26, 2363–2380.

Ma, Y., Genton, M.G., 2000. Highly robust estimation of the autocovariance function. J. Time Ser. Anal. (to appear).

Martin, R.D., Yohai, V.J., 1986. Influence functionals for time series (with Discussion). Ann. Statist. 14, 781–855.

Wecker, W.E., 1981. Asymmetric time series. J. Amer. Statist. Assoc. 76, 16–21.