Robust Indirect Inference

Marc G. GENTON and Elvezio RONCHETTI

In this article we develop robust indirect inference for a variety of models in a unified framework. We investigate the local robustness properties of indirect inference and derive the influence function of the indirect estimator, as well as the level and power influence functions of indirect tests. These tools are then used to design indirect inference procedures that are stable in the presence of small deviations from the assumed model. Although indirect inference was originally proposed for statistical models whose likelihood is difficult or even impossible to compute and/or to maximize, we use it here as a device to robustify the estimators and tests for models where this is not possible or is difficult with classical techniques such as M estimators. Examples from financial applications, time series, and spatial statistics are used for illustration.

KEY WORDS: Correlated observations; Influence function; Robustness of validity; Robustness of efficiency; Space-time autoregression; Stochastic differential equations.

1. INTRODUCTION

Indirect estimation (Gouriéroux, Monfort, and Renault 1993; Gallant and Tauchen 1996) was proposed as an estimation and inference procedure for models with complex formulations or intractable likelihood functions. Basically, it involves optimizing an auxiliary criterion that does not directly provide a consistent estimator of the parameter of interest. A consistent estimator is then obtained by means of simulation. Indirect inference techniques belong to the class of modern statistical procedures that exploit Monte Carlo methods to derive powerful estimators and tests for complex models. This class includes the bootstrap and Markov Chain Monte Carlo methods, among others.

Suppose that a set of observations $y_1, \ldots, y_n$ has been collected, and assume that these have been generated from a probability model $F(\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ is an unknown parameter. To construct an indirect estimator of $\boldsymbol{\theta}$, it is necessary to be able to draw pseudo-observations from $F(\boldsymbol{\theta})$. Moreover, a simple auxiliary model $\widetilde{F}(\boldsymbol{\pi})$ of $F(\boldsymbol{\theta})$ is needed, where $\boldsymbol{\pi} \in \Pi \subseteq \mathbb{R}^r$ is an unknown auxiliary parameter chosen such that it is easier to estimate than $\boldsymbol{\theta}$. For instance, the auxiliary model could be an approximation of the exact likelihood function, or the exact likelihood function of an approximated model. The auxiliary parameter is then estimated with the n observed data yielding $\hat{\boldsymbol{\pi}}$ (step 1) and with $m = sn$ simulated data ($s \geq 1$) from $F(\boldsymbol{\theta})$ (step 2), yielding $\boldsymbol{\pi}^*$ (step 3), a function of $\boldsymbol{\theta}$ by construction. The indirect estimator (step 4) is then defined as

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

where the matrix Ω can be chosen to maximize efficiency (Gouriéroux and Monfort 1996). It is essential that the pseudo-observations drawn from $F(\boldsymbol{\theta})$ be based on the same random generator seed for all $\boldsymbol{\theta}$ to ensure a successful minimization of (1); otherwise, the objective function would not be deterministic for $\boldsymbol{\theta}$ (see Gouriéroux et al. 1993). For the indirect estimator to be consistent and asymptotically normal, the binding function $h(\boldsymbol{\theta}) = \boldsymbol{\pi}$ needs to be locally injective around the true value of $\boldsymbol{\theta}$ (Gouriéroux et al. 1993). A necessary condition for this is $r \geq p$. In general, the binding function h has no known analytical form. Thus an estimate of $\boldsymbol{\theta}$ cannot be constructed as $h^{-1}(\hat{\boldsymbol{\pi}})$, thereby justifying the use of the aforementioned simulation step. With the particular choice $h(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(T_n)$ for a given estimator T_n of $\boldsymbol{\theta}$, the indirect estimator becomes the target estimator introduced by Cabrera and Fernholz (1999), which has smaller bias and mean squared error than T_n. Indirect inference has been applied successfully to several complex models, examples of which have been given by Gouriéroux and Monfort (1996). A potential drawback of this technique is its dependence on the exact specification of the underlying model $F(\boldsymbol{\theta})$. Robust statistics deals with deviations from exact underlying models and has been developed for a variety of general parametric models (Huber 1981; and Hampel, Ronchetti, Rousseeuw, and Stahel 1986). In this article we investigate the robustness issue of indirect inference from three points of view. This will extend the potential applications of robust statistics to very complex models.

In Section 2 we investigate the theoretical properties of indirect estimators and tests by means of a basic tool in robust statistics, the influence function (Hampel 1974). In particular, we show how the influence function of the indirect estimator is related to the influence function of the auxiliary estimator. The basic (not too surprising) conclusion is that robustness can be integrated in indirect inference using robust auxiliary estimators.

In Section 3.1 we study the behavior of indirect estimators when the underlying model $F(\boldsymbol{\theta})$ does not hold exactly. As a typical (important) example, we consider the estimation of stochastic differential equations, where indirect estimators have been advocated as a bias-reduction technique for estimators derived from a "crude discretization" of the stochastic differential equation. The basic conclusion here is that classical indirect estimators completely lose this property (and their justification) if the model holds only approximately. Because in reality the model is at best an ideal approximation of

Marc G. Genton is Assistant Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27695 (E-mail: *genton@stat.ncsu.edu*). Elvezio Ronchetti is Professor, Department of Econometrics, University of Geneva, CH-1211 Geneva, Switzerland (E-mail: *elvezio.ronchetti@metri.unige.ch*). Part of this research was carried out while the second author was visiting the first at the Massachusetts Institute of Technology, whose hospitality is gratefully acknowledged. The authors thank Dennis Boos and Chris Field for comments on the manuscript, as well as two anonymous referees, the associate editor, and the editor for valuable comments that led to a marked improvement of the article.

Journal of the American Statistical Association
March 2003, Vol. 98, No. 461, Theory and Methods
DOI 10.1198/016214503388619102

the underlying data-generating process, robust estimators are needed. In Section 3.2 we show that robust indirect estimators can be constructed that exhibit a stable and excellent performance in terms of bias and variance in a neighborhood of the ideal model.

In Section 4 we present two applications of robust indirect inference covering autoregressive moving average (ARMA) modeling and spatial statistics. In particular, in Section 4.1 we use the indirect procedure as a device to robustify the inference in ARMA models. Finally, in Section 5 we present some open problems.

2. ROBUST INDIRECT ESTIMATORS AND TESTS

When the likelihood function of the model $F(\boldsymbol{\theta})$ is tractable, a maximum likelihood estimator of $\boldsymbol{\theta}$ can be computed. However, it is well known that the latter is generally not robust in the presence of deviations from the underlying model, and robust methods have been developed for this reason (Huber 1981; Hampel et al. 1986). For instance, when observations are independently distributed, contamination models (or neighborhoods) of the form $(1-\varepsilon)F(\boldsymbol{\theta})+\varepsilon G$, where G is an arbitrary distribution, have been considered. For dependent data, more complex contamination patterns may occur (see Martin and Yohai 1986). The properties of different inferential methods can then be studied under the original assumed model $F(\boldsymbol{\theta})$ and under the contaminated version. Maximum likelihood is asymptotically optimal under the former but often exhibits poor performance under the latter. Therefore, robust estimators and tests have been derived. These exhibit a stable performance in terms of bias and variance on the whole neighborhood (see Hampel et al. 1986); however, for complex situations, such robust inferential methods and their theoretical properties may be intricate, if not impossible to obtain. Indirect estimation may then be a solution, as was first suggested by Genton and de Luna (2000). Gouriéroux et al. (1993) assumed the two estimators $\hat{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}^*$ to be identical, although this does not need to be the case as long as they are consistent. For instance, a robust estimator $\hat{\boldsymbol{\theta}}$ can be obtained with a robust estimator $\hat{\boldsymbol{\pi}}$. On the other hand, $\boldsymbol{\pi}^*$ is evaluated on outlier-free simulated data, and thus is often most conveniently chosen to be efficient under the uncontaminated model. Note, however, that the finite-sample bias is not necessarily negligible, in which case a robust estimator on the simulated data also may be more appropriate (see Sec. 3.2). A schematic illustration of the robust indirect estimation algorithm is given in Figure 1. In this section we study the theoretical properties of robust indirect estimation and further extend this approach to robust indirect tests.

The consistency and asymptotic normality of the robust indirect estimator $\hat{\boldsymbol{\theta}}$ are inherited from the consistency and asymptotic normality of the auxiliary estimator $\hat{\boldsymbol{\pi}}$, for a given matrix Ω and fixed integer s, due to the injectivity of the binding function h. In particular, the asymptotic covariance matrix $V_{\hat{\boldsymbol{\theta}}}$ of the robust indirect estimator $\hat{\boldsymbol{\theta}}$ is given by

$$V_{\hat{\boldsymbol{\theta}}} = BV_{\hat{\boldsymbol{\pi}}}B^T + \frac{1}{s}BV_{\boldsymbol{\pi}^*}B^T, \qquad (2)$$

where $B=(D^T\Omega D)^{-1}D^T\Omega \in \mathbb{R}^{p\times r}$, the rectangular matrix $D\in\mathbb{R}^{r\times p}$ is the Jacobian matrix of the transformation h,

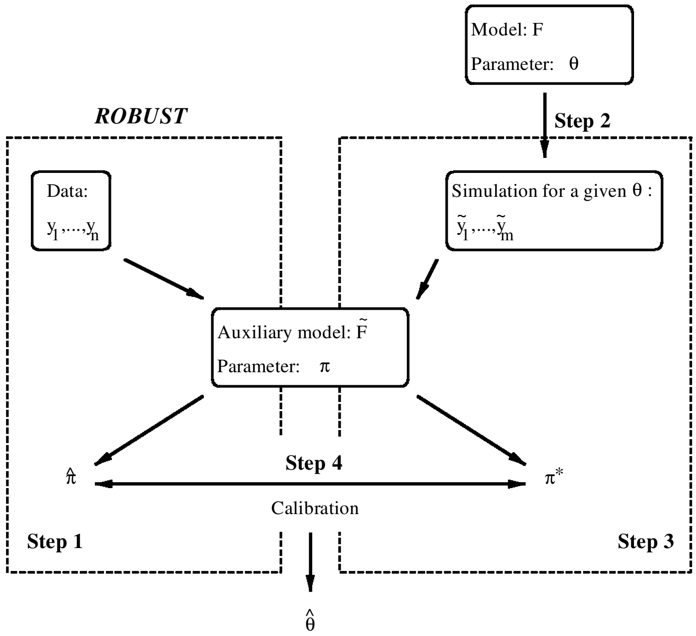

Figure 1. Schematic Illustration of the Robust Indirect Estimation Algorithm. A Robust Estimator $\hat{\boldsymbol{\pi}}$ is Used in Step 1.

and $V_{\hat{\boldsymbol{\pi}}}$ and $V_{\boldsymbol{\pi}^*}$ are the asymptotic covariance matrices of $\hat{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}^*$. In general, $V_{\hat{\boldsymbol{\pi}}}\geq V_{\boldsymbol{\pi}^*}$, because we use a robust estimator on the observed data, thus trading some efficiency for some robustness. Result (2) is a direct consequence of proposition 4.2 of Gouriéroux and Monfort (1996). Note that $\hat{\boldsymbol{\theta}}$ is consistent for any $s\geq 1$, whereas its efficiency is improved when increasing s. For s sufficiently large, we obtain the best efficiency with $\Omega=V_{\hat{\boldsymbol{\pi}}}^{-1}$. The matrix $V_{\hat{\boldsymbol{\theta}}}$ then simplifies to $V_{\hat{\boldsymbol{\theta}}}\cong(D^TV_{\hat{\boldsymbol{\pi}}}^{-1}D)^{-1}$. In practice, $\Omega=V_{\hat{\boldsymbol{\pi}}}^{-1}$ must be expressed as a function of $\boldsymbol{\theta}$ or must be estimated. However, choosing $\Omega=I_r$, the identity matrix in $\mathbb{R}^{r\times r}$, simplifies the computation effort, whereas from practical experience, the loss of efficiency is often limited. A consistent estimator for $V_{\hat{\boldsymbol{\theta}}}$ can be obtained by replacing D by a numerical approximation of $\partial\pi^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^T|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ (see Genton and de Luna 2000), and using a consistent estimator for $V_{\hat{\boldsymbol{\pi}}}$ (see Gouriéroux et al. 1993, app. 2). Note finally that the asymptotic properties have been derived for any value of $r\geq p$ fixed. Increasing r may improve the efficiency, albeit only marginally, after a given level. Thus a heuristic approach to the choice of r is to use an information criterion such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). One may then repeat the procedure for a few values of r to ensure that the estimation is not sensitive to an increase in r. In fact, the efficiency performances of the estimator will in general be similar for a fairly wide range of values of r.

The influence function (Hampel 1974; Hampel et al. 1986) of the indirect estimator has been derived by Genton and de Luna (2000). Denote by P and T the statistical functionals corresponding to the estimators $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\theta}}$ defined previously. Thus P is such that $\hat{\boldsymbol{\pi}}=P(F_n)$ for any n and F_n [or $\hat{\boldsymbol{\pi}}$ tends in probability toward $P(F)$], where F_n is the empirical distribution of the sample and F is the underlying distribution function. In the sequel, Fisher consistency is assumed, that is, $P(F(\boldsymbol{\theta}))=h(\boldsymbol{\theta})$ for all $\boldsymbol{\theta}\in\Theta$ and $T(F(\boldsymbol{\theta}))=\boldsymbol{\theta}$. Let $IF_{\hat{\boldsymbol{\pi}}}$ be the vector influence function of P, and $IF_{\hat{\boldsymbol\theta}}$ be the vector influence function of T. Then

$$IF_{\hat{\boldsymbol\theta}} = B\, IF_{\hat{\boldsymbol\pi}}, \tag{3}$$

where the rectangular matrix B has been defined earlier. The influence function usually can be used to compute the asymptotic variance as $\int IF\, IF^T\, dF$. However, when using (3), only the first part of (2) is obtained, because it does not contain the variability due to the simulations. The following result provides the relation between the self-standardized sensitivities of $\hat{\boldsymbol\theta}$ and $\hat{\boldsymbol\pi}$.

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

$$\|IF_{\hat{\boldsymbol\theta}}\|_{V_{\hat{\boldsymbol\theta}}^{-1}} \le \|IF_{\hat{\boldsymbol\pi}}\|_{V_{\hat{\boldsymbol\pi}}^{-1}}, \tag{4}$$

where $\|\mathbf{x}\|_\Sigma^2 = \mathbf{x}^T\Sigma\mathbf{x}$. Moreover, equality in (4) holds iff $\dim(\boldsymbol\theta) = \dim(\boldsymbol\pi)$ (i.e., $p = r$), and $s \to \infty$. Finally, the lower bound for $\|IF_{\hat{\boldsymbol\theta}}\|_{V_{\hat{\boldsymbol\theta}}^{-1}}$ is $\sqrt{sp/(s+1)}$.

The proof is given in the Appendix. Proposition 1 shows that the robustness of $\hat{\boldsymbol\theta}$ is inherited from the robustness of $\hat{\boldsymbol\pi}$. If $\|IF_{\hat{\boldsymbol\pi}}\|_{V_{\hat{\boldsymbol\pi}}^{-1}}$ is bounded by a constant, say, c then $\|IF_{\hat{\boldsymbol\theta}}\|_{V_{\hat{\boldsymbol\theta}}^{-1}}$ is also bounded by the same constant c, with the lower bound for c given in Proposition 1.

Next, we turn to robust indirect inference by defining bounded-influence tests. The purpose of robust testing procedures is to control the maximal bias on the level and the power of a test that can arise from slight distributional misspecification of a null or an alternative hypothesis. This is generally referred to as robustness of validity and efficiency (Hampel et al. 1986, p. 405). To study the asymptotic local stability of indirect tests, we follow the general approach developed by Heritier and Ronchetti (1994). The key idea is to bound the influence function of the indirect estimator by bounding the influence function of the auxiliary estimator (see Prop. 1) when the underlying null distribution $F(\boldsymbol\theta_0)$ is locally contaminated as

$$F^0_{\varepsilon,n,G} = \left(1 - \frac{\varepsilon}{\sqrt{n}}\right)F(\boldsymbol\theta_0) + \frac{\varepsilon}{\sqrt{n}}G, \tag{5}$$

where G is an arbitrary distribution. Assume that $\boldsymbol\theta$ is partitioned into $\boldsymbol\theta = (\boldsymbol\theta_1^T\ \ \boldsymbol\theta_2^T)^T$, where $\boldsymbol\theta_1$ and $\boldsymbol\theta_2$ have dimensions p_1 and p_2, and consider the null hypothesis $H_0 : \boldsymbol\theta_1 = \boldsymbol\theta_{10}$. Indirect tests based on the Wald statistic or score statistic can be derived. However, we restrict our attention to indirect tests based on the optimal value of the objective function (1) used in the indirect estimation method,

$$\frac{sn}{s+1}\big[(\hat{\boldsymbol\pi} - \boldsymbol\pi^*(\hat{\boldsymbol\theta}_0))^T\Omega(\hat{\boldsymbol\pi} - \boldsymbol\pi^*(\hat{\boldsymbol\theta}_0)) - (\hat{\boldsymbol\pi} - \boldsymbol\pi^*(\hat{\boldsymbol\theta}))^T\Omega(\hat{\boldsymbol\pi} - \boldsymbol\pi^*(\hat{\boldsymbol\theta}))\big], \tag{6}$$

where $\hat{\boldsymbol\theta}_0 = (\boldsymbol\theta_{10}^T\ \ \hat{\boldsymbol\theta}_{20}^T)^T$ is the constrained indirect estimator obtained by minimizing the objective function (1) under $H_0 : \boldsymbol\theta_1 = \boldsymbol\theta_{10}$. Despite the use of simulations, the usual asymptotic equivalence between the indirect tests based on the Wald statistic, score statistic, and (6) remains valid under the null hypothesis. These tests have a $\chi^2_{p_1}$ distribution with p_1 degrees of freedom (see Gouriéroux et al. 1993). If we use a bounded-influence indirect estimator $\hat{\boldsymbol\theta}$ (with bound c on the influence function), then $\hat{\boldsymbol\theta}$ is Fréchet differentiable in the neighborhood (5) (see Ronchetti and Trojani 2001). In this case we can bound the maximal asymptotic bias of the level α of the indirect test over the neighborhood (5) by the inequality

$$\lim_{n\to\infty}\left|\alpha\left(F^0_{\varepsilon,n,G}\right) - \alpha_0\right| \le K\varepsilon^2c^2 + o(\varepsilon^2), \tag{7}$$

where $\alpha_0 = \alpha(F(\boldsymbol\theta_0))$ is the nominal level of the indirect test and the constant K depends only on the model and has been defined by Ronchetti and Trojani (2001). The asymptotic bound (7) can be used to choose the constant c according to the maximal amount of contamination ε expected by the analyst and the maximal bias on the level that he or she can tolerate. Bounds similar to (7) can be derived for the power.

3. APPLICATION TO STOCHASTIC DIFFERENTIAL EQUATIONS

In this section we present an application of robust indirect estimation to an example from finance.

3.1 Robustness Properties of Indirect Estimators

We consider an important field of application of this technique—the estimation of stochastic differential equations from discrete observations. Indirect inference has been suggested for these models as a technique to reduce the bias of the estimators based on a crude discretization of the stochastic differential equation (see Gouriéroux and Monfort 1996). Here we want to study to which extent this property is satisfied when the model does not hold exactly. We derive our results for a particular model—a geometric Brownian motion with drift—because it allows direct analytical computations, but our conclusions hold also for more complex models. It is the model underlying the derivation of the Black–Scholes formula for option pricing and is used widely in finance.

Assume that the price y_t of an asset satisfies the stochastic differential equation (geometric Brownian motion with drift)

$$dy_t = \mu y_t dt + \sigma y_t dW_t, \tag{8}$$

where W_t is a Brownian motion and μ and σ are the drift and volatility parameters. A "crude" discretization of (8) gives the equation

$$\begin{aligned} y_t &= y_{t-1} + \mu y_{t-1} + \sigma y_{t-1}\epsilon_t \\ &= (1+\mu)y_{t-1} + \sigma y_{t-1}\epsilon_t, \end{aligned} \tag{9}$$

where $\{\epsilon_t\}_{t=1,\dots,n}$ are iid standard normal variables. The log-likelihood function is given by

$$\tilde{L}_n = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\,\sigma^2 - \sum_{t=1}^n \log\, y_{t-1} - \frac{1}{2}\sum_{t=1}^n \frac{[y_t - (1+\mu)y_{t-1}]^2}{\sigma^2 y_{t-1}^2}, \tag{10}$$

and the maximum likelihood estimator for (μ, σ^2) is simply

$$\tilde{\mu} = \bar{r}_t - 1 \tag{11}$$

and

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(r_t - \bar{r}_t)^2, \tag{12}$$

where $r_t = y_t/y_{t-1}$.

It is well known (see, e.g., Gouriéroux and Monfort 1996) that the estimators based on the crude discretization are biased. Here indirect inference is used to eliminate this bias. In this case (9) is used as an auxiliary model, (11) and (12) are auxiliary estimators, and the simulation is carried out from a finer discretization of (8). The resulting estimators $(\hat{\mu}_I, \hat{\sigma}_I^2)$ are then unbiased up to order $O(n^{-1})$. In the next illustrative example, we investigate the bias of the auxiliary and indirect inference estimators of the drift when the underlying model does not hold exactly. We still consider (8) as an ideal underlying model, but we take into account that the observed prices are only *approximately* normally distributed.

Illustrative Example. Assume model (9), where the random variables $\{\epsilon_t\}_{t=1,\ldots,n}$ are iid with common distribution $(1-\varepsilon)N(0,1)+\varepsilon N(0,\tau^2)$, $0 \le \varepsilon \le 1$, and $\tau \ge 1$. Let $\tilde{\mu}$ be the auxiliary estimator for μ defined by (11), $\hat{\mu}_I$ the corresponding indirect estimator, and s the number of simulation runs. Then we have

$$\begin{aligned}\text{bias}(\tilde{\mu}, \varepsilon) &= E\tilde{\mu} - \mu \\ &= e^{\mu} - (1+\mu) + \varepsilon e^{\mu}\left(e^{\sigma^2(\tau^2-1)/2} - 1\right)\end{aligned} \tag{13}$$

and

$$\begin{aligned}\text{bias}(\hat{\mu}_I, \varepsilon) &= E\hat{\mu}_I - \mu \\ &= -\frac{1}{n}\frac{1}{2}\left(1+\frac{1}{s}\right)\left(e^{\sigma^2} - 1 + \varepsilon\left(e^{\sigma^2(2\tau^2-1)}\right.\right. \\ &\qquad \left.\left. -2e^{\sigma^2(\tau^2+1)/2} + e^{\sigma^2}\right)\right) \\ &\quad + O(\varepsilon^2) + O(n^{-2}).\end{aligned} \tag{14}$$

The computations are given in the Appendix. This result clearly shows the two components of the bias for both estimators $\tilde{\mu}$ and $\hat{\mu}_I$. For $\tilde{\mu}$, the bias due to the discretization is $e^{\mu} - (1+\mu)$ and the component due to deviations from the underlying model is $\varepsilon e^{\mu}(e^{\sigma^2(\tau^2-1)/2} - 1)$. Both components are positive. The bias of the indirect estimator is negative and of order $O(n^{-1})$. Figures 2 and 3 plot the bias of both estimators as a function of τ and for different values of ε.

Figures 2 and 3 show the bias-reduction effect of the indirect estimator when the model is exact ($\tau = 1$) and when τ is in the interval $[1, 3.4]$. However, when $\tau \ge 3.4$, the bias for the indirect estimator becomes important as well. Finally, when $\tau = 3.7$, the biases of the auxiliary estimator and the indirect estimator are large and comparable, and there is no gain in performing indirect inference. Note that even in this situation, the Kolmogorov distance between the contaminated distribution and the model is less than or equal to ε and the index of tail length (Rosenberger and Gasko 1983, p. 322) of the contaminated distribution is 1.3, compared to 1 for the normal distribution. Therefore, the level of contamination is very low, and it would be difficult to detect such a deviation from the model on a real sample.

To summarize, even if our investigation is limited to the model of the geometric Brownian motion, the evidence seems to indicate that small deviations from the stochastic structure of the model can wipe out the bias improvement obtained with indirect inference. This conclusion will also apply to asymmetric contaminations and more complex models (cf. Ronchetti and Trojani 2001). Therefore, there is a need to robustify estimators and tests obtained by indirect inference; we discuss this issue in the next section.

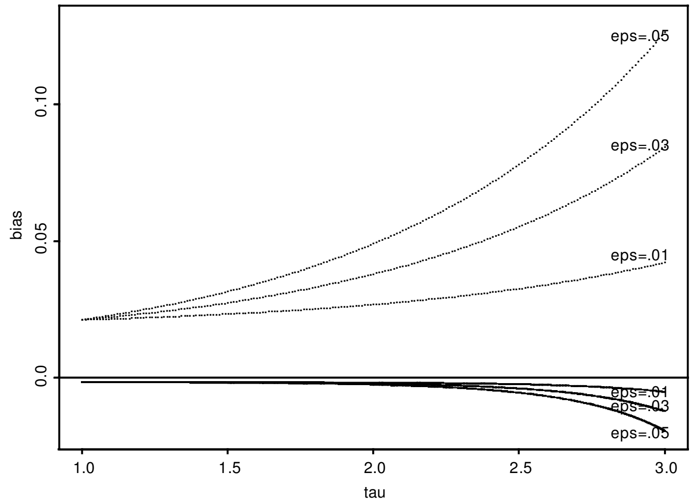

Figure 2. Bias as a Function of τ for the Auxiliary Estimator ($\cdots$) and for the Indirect Inference Estimator (-); $\mu = .2$, $\sigma = .5$, $n = 100$, $s = 10$, τ in [1, 3].

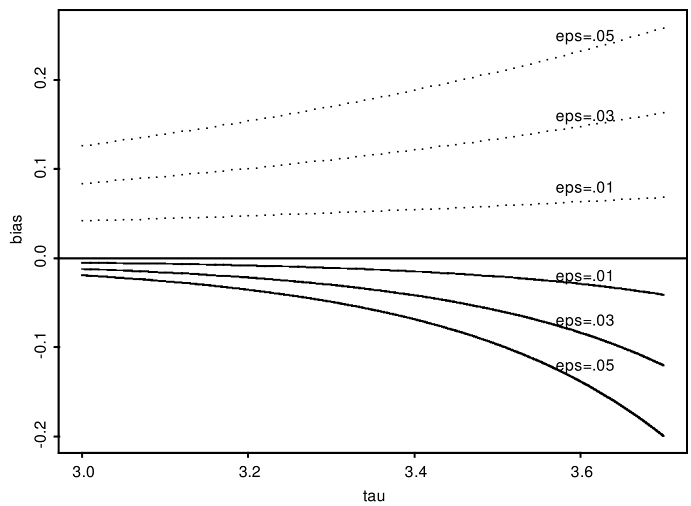

Figure 3. Bias as a Function of τ for the Auxiliary Estimator ($\cdots$) and for the Indirect Inference Estimator (-); $\mu = .2$, $\sigma = .5$, $n = 100$, $s = 10$, τ in [3, 3.7].

3.2 Robust Indirect Estimators for the Geometric Brownian Motion With Drift

We consider again models (8) and (9). The auxiliary estimators $(\tilde{\mu},\tilde{\sigma}^2)$ for (μ,σ^2) given by (11) and (12) have an unbounded influence function, and, according to Proposition 1, the indirect estimators $(\hat{\mu}_I,\hat{\sigma}_I^2)$ will inherit this property and will not be robust. To robustify the procedure, it is natural to replace $(\tilde{\mu},\tilde{\sigma}^2)$ by robust versions $(\tilde{\mu}_R,\tilde{\sigma}_R^2)$ defined as solutions of the equations

$$\sum_{t=1}^{n}\psi_c\left(\frac{r_t-1-\tilde{\mu}_R}{\tilde{\sigma}_R}\right)=0 \tag{15}$$

and

$$\sum_{t=1}^{n}\chi_c\left(\frac{r_t-1-\tilde{\mu}_R}{\tilde{\sigma}_R}\right)=0, \tag{16}$$

where $\psi_c(z)=\min\{c,\max\{-c,z\}\}$ is the Huber function, $\chi_c(z)=\psi_c^2(z)-E_\Phi\psi_c^2$, and c is a tuning constant. Equations (15) and (16) define Huber's proposal 2 (Huber 1981, p. 137) for location-scale estimators on the data r_t-1. We denote by $(\hat{\mu}_{RI},\hat{\sigma}_{RI}^2)$ the indirect estimators based on $(\tilde{\mu}_R,\tilde{\sigma}_R^2)$. In more complex models, such as those of Cox, Ingersoll, and Ross (see Chan, Karloyi, Longstaff, and Sanders 1992), such simple robust estimators are not available. Instead, robust generalized method of moments estimators, as developed by Ronchetti and Trojani (2001), can be used.

Finally, note that model (8) admits an "exact discretization" given by (A.5). This model can be rewritten as

$$l_t=\nu+\sigma\epsilon_t, \tag{17}$$

where $l_t=\log r_t$ and $\nu=\mu-\frac{\sigma^2}{2}$. Therefore, (17) is a location-scale model that leads to the maximum likelihood estimators for (μ,σ^2) given by

$$\hat{\mu}=\bar{l}+\frac{\hat{\sigma}^2}{2} \tag{18}$$

and

$$\hat{\sigma}^2=\frac{1}{n}\sum_{t=1}^{n}(l_t-\bar{l})^2. \tag{19}$$

Similarly, model (17) can be estimated robustly by Huber's proposal 2, leading to the estimators $(\hat{\mu}_R,\hat{\sigma}_R^2)$.

Figures 4–7 show boxplots of 200 simulations for the following estimators of μ and σ^2:

(a) Exact discretization $(\hat{\mu},\hat{\sigma}^2)$
(b) Exact discretization robust $(\hat{\mu}_R,\hat{\sigma}_R^2)$
(c) Crude discretization $(\tilde{\mu},\tilde{\sigma}^2)$
(d) Crude discretization robust $(\tilde{\mu}_R,\tilde{\sigma}_R^2)$
(e) Indirect $(\hat{\mu}_I,\hat{\sigma}_I^2)$
(f) Indirect robust $(\hat{\mu}_{RI},\hat{\sigma}_{RI}^2)$.

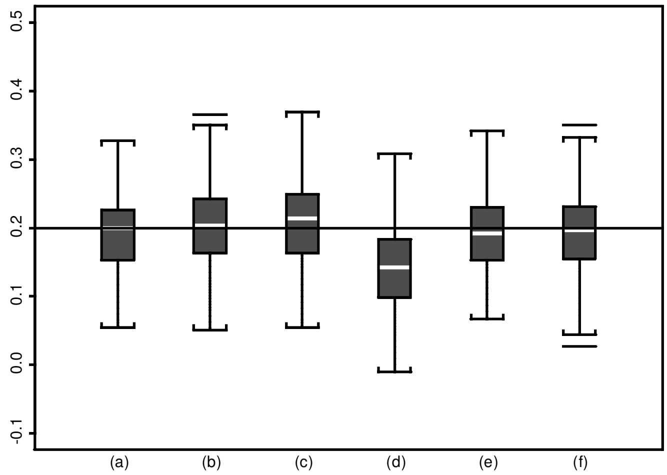

Figure 4. Estimation of μ Without Contamination.

The simulations in Figures 4 and 6 are performed under the normal model $N(0,1)$ for ϵ_t, whereas those in Figures 5 and 7 are performed under 5% additive outlier contamination from the $N(0,5^2)$ distribution. In all cases we generate realizations of the process $\{y_t\}$ with a fine discretization ($\mu=.2$ and $\sigma^2=.25$) and draw samples of size $n=100$. The number of simulations s in the indirect procedure is chosen to be 10. Finally, we choose the tuning constant for the robust estimators $c=1.345$ to achieve 95% efficiency at the normal model.

We now discuss the results. Figure 4 shows that the crude estimators (c) and (d) are biased under the model. As expected, this bias is corrected by their indirect estimators (e) and (f). The biases of the classical estimators (c) and (e) can be read from Figure 2 (at $\tau=1$). A comparison of (a) with (e) and (b) with (f) shows that the indirect estimators behave like the estimators obtained by the "exact discretization." This implies that the former are useful procedures for many models where such an exact discretization is not available. Further, note that the robust estimators are a little more variable than the classical ones; this is to be expected in view of the small loss of efficiency at the normal model. Finally, note that the robust indirect estimator is obtained by robust estimation both on the original and simulated data in the indirect procedure.

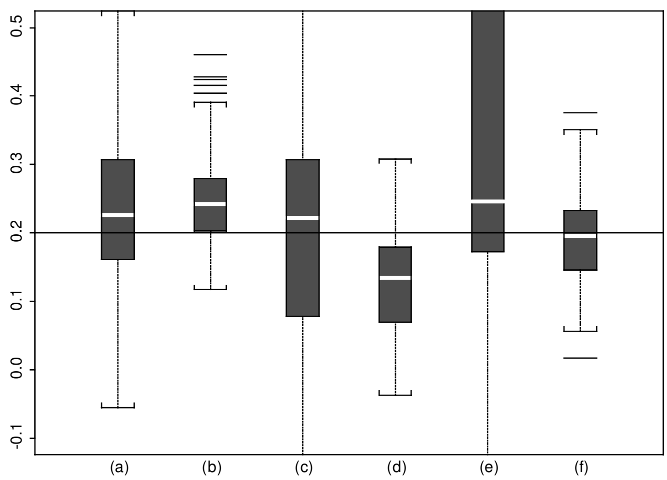

Figure 5. Estimation of μ With 5% Additive Outlier Contamination.

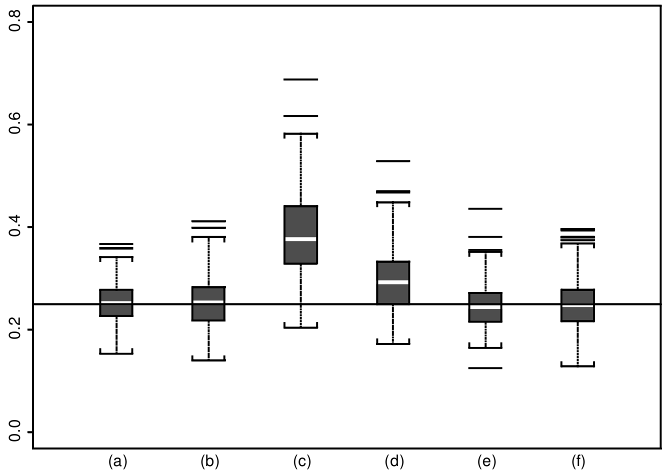

Figure 6. Estimation of σ^2 Without Contamination.

A classical estimation on the simulated data (not shown here) leads to a finite sample–biased (albeit consistent!) indirect estimator.

Figure 6 shows the same situation for the estimation of σ^2. The overall picture is the same except that the gains in bias reduction obtained by indirect estimation are even larger.

Figure 5 shows the different estimators of μ under the contaminated model. The crude estimators (c) and (d) are again biased, and (c) in particular exhibits large variability. In this case the classical indirect estimator (e) cannot correct the bias and reduce the variability, whereas the robust indirect estimator (f) shows the overall best performance. Note that under contamination, the estimators obtained by exact discretization exhibit worse performance in terms of bias and variance than the robust indirect estimator.

Finally, Figure 7 shows the estimator for σ^2 under contamination. Here the effects are much larger than in the case of μ, and the conclusions are the same.

To summarize, the robust indirect estimators for μ and σ^2 exhibit good performance in terms of bias and variance even when the model is not exact and the data are generated by a distribution in a neighborhood of the model. Moreover, indirect inference is a general procedure that can be carried out for general models, for instance, when an exact discretization is not available.

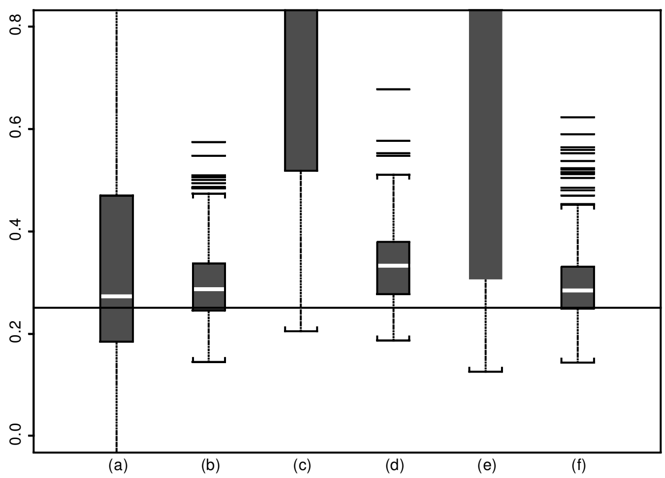

Figure 7. Estimation of σ^2 With 5% Additive Outlier Contamination.

4. APPLICATION TO TIME SERIES AND SPATIAL STATISTICS

4.1 Robust Indirect Inference for Autoregressive Moving Average Models

Robust estimation for time series models has received considerable attention during the last two decades (see Martin and Yohai 1985 for a survey). The main issue is that standard robust procedures become useless for moving average (MA) and mixed ARMA models, because a single outlier at a certain time influences all subsequent terms in the estimation equations. Several ad hoc methods for overcoming this difficulty have been proposed (see, e.g., Bustos and Yohai 1986; Allende and Heiler 1992). Recently, de Luna and Genton (2001) suggested the use of robust indirect estimation in the context of ARMA model estimation. Their proposal compares fairly well with existing procedures, with the main advantage that asymptotic properties (i.e., consistency, asymptotic normality, influence function, breakdown point) can be easily derived. In this section we focus on the robustness properties of p values of indirect tests. For illustrative purposes, we consider a simple MA model of order 1, that is,

$$Y_t = \theta\epsilon_{t-1} + \epsilon_t, \tag{20}$$

where the sequence $\{\epsilon_t\}$ is iid with mean 0 and variance σ^2. The model is identifiable when $|\theta| < 1$. The indirect estimation of this model can be based on the autoregressive auxiliary parameterization (Gouriéroux et al. 1993; de Luna and Genton 2001)

$$Y_t = \pi Y_{t-1} + U_t, \tag{21}$$

where π is such that $\mathrm{E}(U_t Y_{t-1}) = 0$; that is, πY_{t-1} is the best linear predictor of order 1 for Y_t. Higher-order predictors (i.e., $\pi_1 Y_{t-1} + \cdots + \pi_{r-1} Y_{t-r-1}$) may be used, although here for simplicity we focus on (21). Note that U_t is a correlated process with mean 0 and variance σ_U^2.

In this experiment we simulate 200 time series of size $n = 100$ each from the MA(1) model (20) with $\theta = .5$, $\sigma^2 = 1$, and Gaussian innovations $\{\epsilon_t\}$. To test the null hypothesis $H_0 : \theta = .5$, we perform indirect tests based on the optimal value of the objective function as defined in (6) with $s = 30$. We use a classical least squares auxiliary estimator $\hat{\boldsymbol{\pi}} = (\hat{\pi}, \hat{\sigma}_U^2)^T$, as well as a robust version based on a GM estimator (`ar.gm` in S-PLUS, with a Huber ψ-function and default efficiencies `effloc` $= .96$ and `effgm` $= .87$). Figure 8 plots the p values of the indirect tests based on a χ_1^2 distribution under H_0 versus the quantiles of a uniform distribution. Figure 8(a) is obtained using the classical indirect test; 8(b), using the robust indirect test. Both indirect tests, classical and robust, behave similarly, and under the null hypothesis their p values have a distribution close to the uniform. Next we consider the situation of contaminated time series with 5% additive outliers from a $N(0, \tau^2)$ distribution where $\tau = 5$. The classical indirect test now becomes severely affected by

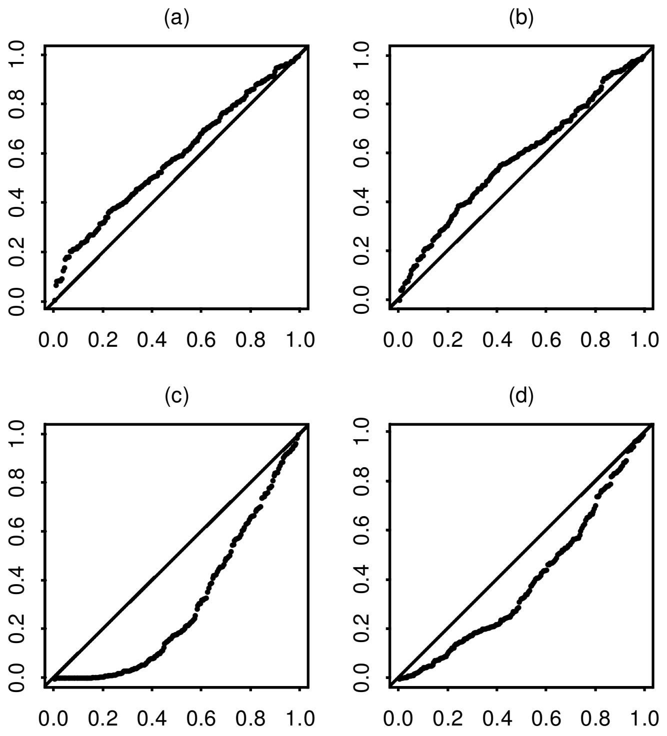

Figure 8. Plots of p Values From Indirect Tests Versus Quantiles of the Uniform Distribution for 200 Simulated Time Series of Size $n = 100$ Each, From the MA(1) Model (20) With $\theta = .5$ and $\sigma^2 = 1$. Without outliers: (a) classical indirect test; (b) robust indirect test. With 5% additive outliers from $N(0, \tau^2)$, where $\tau = 5$: (c) classical indirect test; (d) robust indirect test.

x_2 \ x_1	1	2	3	4	5	6	7	8	9
9	-17.6	-13.3	-11.1	-8.8	-6.6	3.7	-8.1	-5.8	8.6
8	-7.8	-5.6	-0.3	-1.1	-14.2	2.4	-0.3	-3.1	1.2
7	7.9	**21.2**	6.4	0.7	17.9	1.2	5.4	-1.3	-4.1
6	6.7	12.9	7.2	-1.6	-9.3	5.9	13.2	12.4	10.7
5	5.5	2.7	-16.1	-12.8	1.4	8.7	**24.9**	4.2	-0.6
4	6.2	-15.5	-18.3	-6.0	0.2	5.5	11.7	18.9	-1.8
3	-17.0	-13.8	-19.5	-11.3	-5.0	-2.8	0.5	5.7	6.0
2	**26.7**	**40.0**	-1.8	-13.5	-4.3	-0.1	1.2	-9.5	-7.3
1	**26.5**	16.7	-17.0	-11.8	-9.5	-8.3	-6.0	-10.8	-9.5

Figure 9. The Residuals Obtained After Detrending the Reflectance Values on a 9×9 Lattice. Possible outliers are highlighted in bold.

the outliers and consequently often rejects the null hypothesis, as can be seen in Figure 8(c). In 8(d), however, the robust indirect test can cope with the outliers. It is interesting to note that the robust indirect test exhibits an already good performance with a simple low-order auxiliary model.

4.2 Robust Indirect Inference for Simultaneous Autoregression Models

We now turn to the application of our methodology to a real dataset of reflectance values extracted from an extensive aerial survey along the south coast of England. The survey was monitoring pollution levels arising from the pumping of waste material into the English Channel. High pollution levels generate high reflectance values. The original dataset was given by Haining (1990, p. 217) and is located on a 9×9 regular lattice. The pipe carrying the waste material exhausts at several points, and the area that we consider is close to a discharge point (the northwest corner of the lattice). The original dataset shows visual evidence of a trend in the reflectance values from northwest to southeast. Estimation of this trend component provides a measure of the dispersal of pollutants from the source point. Following Haining (1987), we fit by least squares a linear trend to the reflectance values, yielding $17 - 1.248x_1 + 3.759x_2$. Figure 9 shows the corresponding residual values. Several possible outliers are noticeable; these highlighted in bold. This has also been confirmed (Haining 1987) by applying a median polish algorithm to the original data.

We next model the residuals by means of a first-order spatial simultaneous autoregression (SAR). The inclusion of this autocorrelated component in our model may be necessary for two reasons. First, the reflectance value recorded in any pixel of the lattice is obtained by partial averaging of reflectance values in neighboring pixels. This is a general issue with aerial and remotely sensed data and is a characteristic of the recording instrument. Second, pollution in any small area will be affected by local mixing and local diffusion arising from small-scale turbulence. A first-order SAR model, accounting for spatial correlation in the east–west direction with ρ_1 and north–south direction with ρ_2, is defined by

$$\begin{aligned} Z(x_1, x_2) = \rho_1[Z(x_1 - 1, x_2) + Z(x_1 + 1, x_2)] \\ + \rho_2[Z(x_1, x_2 - 1) + Z(x_1, x_2 + 1)] \\ + \epsilon(x_1, x_2), \end{aligned} \qquad (22)$$

where (x_1, x_2) are coordinates on the lattice represented in Figure 9. The process Z is stationary if $|\rho_1| + |\rho_2| < 1/2$. Here ϵ is an iid process with mean 0 and variance σ^2 but is not uncorrelated with the process Z. For this reason, least squares or Yule–Walker estimators for $\boldsymbol{\theta} = (\rho_1\ \rho_2\ \sigma^2)^T$ are not generally consistent for simultaneous models (Whittle 1954). Gaussian maximum likelihood estimation of $\boldsymbol{\theta}$ is consistent (Ali 1979) but difficult to implement, partly because ρ_1 and ρ_2 cannot be simultaneously factored out of the covariance matrix in the likelihood function. Moreover, maximum likelihood is not robust to the presence of outliers in the observations. The use of indirect estimation for consistent and robust estimation of the parameters of SAR models (see de Luna and Genton 2002) avoids the direct estimation of the simultaneous model by considering a unilateral auxiliary model, a so-called

Table 1. Indirect Estimation of the Parameter θ of Model (22) for the Residuals Described in Figure 9: Classical (IYW) and Robust (IYWR) and Unconstrained and Constrained (H_0: $\rho_1 = \rho_2$)

	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\sigma}^2$	$\hat{\rho}_1 = \hat{\rho}_2$	$\hat{\sigma}^2$
IYW	.244	.107	87.5	.168	90.6
IYWR	.306	.100	60.7	.189	66.1

quadrant autoregression, defined by

$$Z(x_1, x_2) = \pi_1 Z(x_1 - 1, x_2) + \pi_2 Z(x_1, x_2 - 1) + \nu(x_1, x_2), \quad (23)$$

where ν is a correlated process with mean 0 and variance τ^2. The parameter $\boldsymbol{\pi} = (\pi_1\ \pi_2\ \tau^2)^T$ can be consistently estimated with Yule–Walker estimators (Tjøstheim 1978; Ha and Newton 1993). Robust Yule–Walker estimators are obtained by means of robust autocovariance estimators (Ma and Genton 2000) as advocated by de Luna and Genton (2002). The former yield indirect estimators of $\boldsymbol{\theta}$ based on Yule–Walker (IYW), whereas the latter yield robust indirect estimators of $\boldsymbol{\theta}$ based on robust Yule–Walker (IYWR).

Table 1 presents the estimates of $\boldsymbol{\theta}$ by indirect estimation (IYW and IYWR) with $s = 9$. The estimates of ρ_1 and ρ_2 are rather similar, but the estimates of the variance σ^2 is larger for the classical indirect procedure than for the robust one.

It is of interest here to test for isotropy of the correlation; that is, investigate the null hypothesis $H_0 : \rho_1 = \rho_2$. This can be easily done with the test (6) based on the optimal value of the objective function of the indirect procedure. The indirect estimates of $\boldsymbol{\theta}$ under H_0 are also presented in Table 1. The p value for the indirect test based on IYW turns out to be .09, but it is .01 for the one based on IYWR. Thus the robust test rejects the null hypothesis of isotropy at the 5% level, whereas the classical test does not find sufficient evidence against H_0. We perform a sensitivity analysis of both indirect tests by moving one observation, say $Z(2, 2)$, from -50 to $+50$ by step 5. The effect on the p values is depicted in Figure 10. The dots represent the p values for the indirect test based on IYW; the pluses, the p values for the indirect test based on IYWR. Even for large outliers, the robust indirect test rejects the null hypothesis; its p value shows stable behavior. In contrast, the p value of the classical indirect test exhibits large changes by moving one single observation.

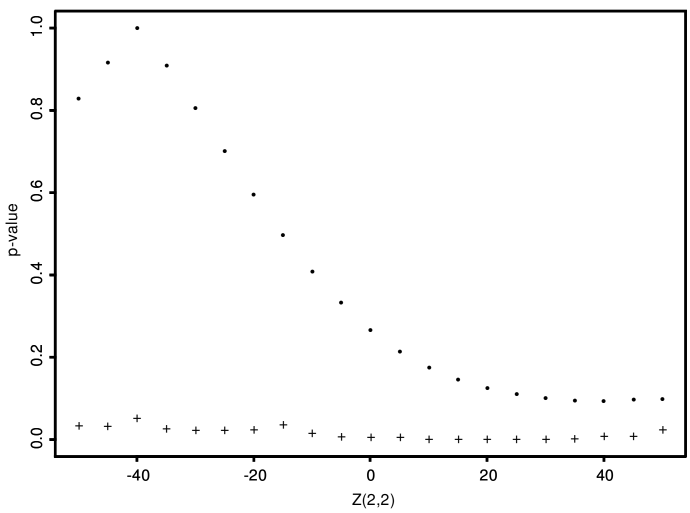

Figure 10. Sensitivity Analysis of the p Value of the Indirect Tests Based on IYW ($\cdots$) and IYWR ($+ + +$) When the Observation Z(2,2) is Moved From -50 to $+50$ by Step 5.

5. CONCLUSION

Indirect inference is a useful general procedure that provides estimators and tests for complex models. The range of its applications includes time series, spatial statistics, and finance. For instance, Gouriéroux and Monfort (1996) discussed applications of indirect inference to latent variable models (e.g., stochastic volatility models) and complex macroeconomics models (e.g., nonlinear simultaneous equations). We have shown that it is necessary to integrate robustness considerations into the procedure to obtain reliable estimators and tests when the model does not hold exactly. Two aspects need to be investigated further. The first aspect is the choice of the auxiliary model. Although the use of robust estimators for the auxiliary parameters seems to reduce the dependence of the final estimator on the auxiliary model (see the example in Sec. 4.1), more precise guidelines on this choice would be welcome. The second point concerns the finite sample properties of the final estimator. Although any consistent estimator for the auxiliary parameter will lead to a consistent indirect estimator, the finite-sample properties of the latter are generally unknown.

APPENDIX: PROOFS AND COMPUTATIONS

Proof of Proposition 1

To prove the first two points, we have

$$\begin{aligned}
\|IF_{\hat{\theta}}\|^2_{V_{\hat{\theta}}^{-1}} &= IF_{\hat{\theta}}^T V_{\hat{\theta}}^{-1} IF_{\hat{\theta}} \\
&= IF_{\hat{\theta}}^T \left(BV_{\hat{\pi}} B^T + \frac{1}{s} BV_{\pi^*} B^T \right)^{-1} IF_{\hat{\theta}} \\
&\le IF_{\hat{\theta}}^T (BV_{\hat{\pi}} B^T)^{-1} IF_{\hat{\theta}} \\
&= IF_{\hat{\pi}}^T [B^T (BV_{\hat{\pi}} B^T)^{-1} B] IF_{\hat{\pi}} \\
&= IF_{\hat{\pi}}^T \big[\Omega D(D^T \Omega D)^{-1} \big((D^T \Omega D)^{-1} D^T \Omega V_{\hat{\pi}} \\
&\qquad \times \Omega D(D^T \Omega D)^{-1}\big)^{-1} (D^T \Omega D)^{-1} D^T \Omega\big] IF_{\hat{\pi}} \\
&= IF_{\hat{\pi}}^T \big[\Omega D\big(D^T \Omega V_{\hat{\pi}} \Omega D\big)^{-1} D^T \Omega\big] IF_{\hat{\pi}} \\
&\le IF_{\hat{\pi}}^T V_{\hat{\pi}}^{-1} IF_{\hat{\pi}} \\
&= \|IF_{\hat{\pi}}\|^2_{V_{\hat{\pi}}^{-1}}.
\end{aligned}$$

The first inequality follows from

$$BV_{\hat{\pi}} B^T + \frac{1}{s} BV_{\pi^*} B^T \ge BV_{\hat{\pi}} B^T.$$

For the second inequality, consider the Choleski decomposition $V_{\hat{\pi}}^{-1} = AA^T$. We need only show that

$$A^T D(D^T AA^T D)^{-1} D^T A \le I_r.$$

Let

$$A^T D = Q \begin{pmatrix} R \\ O \end{pmatrix}$$

be the block QR decomposition of $A^T D$, where Q is orthogonal and R is invertible. Then

$$\begin{aligned}
A^T D(D^T AA^T D)^{-1} D^T A &= Q\begin{pmatrix} R \\ O \end{pmatrix}\left((R^T\ O)Q^T Q\begin{pmatrix} R \\ O \end{pmatrix}\right)^{-1}(R^T\ O)Q^T \\
&= Q\begin{pmatrix} R \\ O \end{pmatrix}(R^T R)^{-1}(R^T\ O)Q^T \\
&= Q\begin{pmatrix} R \\ O \end{pmatrix}R^{-1}R^{-T}(R^T\ O)Q^T \\
&= Q\begin{pmatrix} I \\ O \end{pmatrix}(I\ O)Q^T \\
&= \begin{pmatrix} I_p & O \\ O & O \end{pmatrix} \\
&\le I_r.
\end{aligned}$$

We now prove the lower bound for $\|IF_{\hat{\boldsymbol\theta}}\|_{V_{\hat{\boldsymbol\theta}}^{-1}}$. We have

$$\begin{aligned}
E\left[\|IF_{\hat{\boldsymbol\theta}}\|^2_{V_{\hat{\boldsymbol\theta}}^{-1}}\right] &= E\left[IF_{\hat{\boldsymbol\theta}}^T V_{\hat{\boldsymbol\theta}}^{-1} IF_{\hat{\boldsymbol\theta}}\right] \\
&= \mathrm{tr}\left[V_{\hat{\boldsymbol\theta}}^{-1} E\left[IF_{\hat{\boldsymbol\theta}} IF_{\hat{\boldsymbol\theta}}^T\right]\right] \\
&= \mathrm{tr}\left[\left(BV_{\hat{\boldsymbol\pi}}B^T + \frac{1}{s}BV_{\boldsymbol\pi^*}B^T\right)^{-1} E\left[B\left[IF_{\hat{\boldsymbol\pi}} IF_{\hat{\boldsymbol\pi}}^T\right]B^T\right]\right] \\
&= \mathrm{tr}\left[\left(BV_{\hat{\boldsymbol\pi}}B^T + \frac{1}{s}BV_{\boldsymbol\pi^*}B^T\right)^{-1}(BV_{\hat{\boldsymbol\pi}}^{-1}B^T)\right] \\
&\ge \mathrm{tr}\left[\left(BV_{\hat{\boldsymbol\pi}}B^T + \frac{1}{s}BV_{\hat{\boldsymbol\pi}}B^T\right)^{-1}(BV_{\hat{\boldsymbol\pi}}^{-1}B^T)\right] \\
&= \mathrm{tr}\left[\frac{s}{s+1}I_p\right] \\
&= \frac{sp}{s+1},
\end{aligned}$$

where the inequality follows from $V_{\boldsymbol\pi^*} \le V_{\hat{\boldsymbol\pi}}$. Thus

$$E\left[\|IF_{\hat{\boldsymbol\theta}}\|^2_{V_{\hat{\boldsymbol\theta}}^{-1}}\right] \le E\left[\|IF_{\hat{\boldsymbol\pi}}\|^2_{V_{\hat{\boldsymbol\pi}}^{-1}}\right] \le \sup_x \|IF_{\hat{\boldsymbol\pi}}\|^2_{V_{\hat{\boldsymbol\pi}}^{-1}} = c^2, \quad \text{say,}$$

and thus

$$c \ge \sqrt{\frac{sp}{s+1}}.$$

Computations of the Illustrative Example

The auxiliary estimator $\tilde{\mu}$ can be rewritten as

$$\tilde{\mu} = \bar{r}_t - 1 = b(\mu, \varepsilon) + \frac{1}{\sqrt{n}}A(\mu), \tag{A.1}$$

where $A(\mu) = \frac{1}{\sqrt{n}}\sum_{t=1}^n [r_t - Er_t]$ and

$$\begin{aligned}
b(\mu, \varepsilon) = Er_t - 1 &= E\left(\frac{y_t}{y_{t-1}}\right) - 1 \\
&= E\left[\exp\left(\log\left(\frac{y_t}{y_{t-1}}\right)\right)\right] - 1.
\end{aligned} \tag{A.2}$$

From (8), we have

$$\frac{dy_t}{y_t} = \mu dt + \sigma\, dW_t \tag{A.3}$$

and, by Ito's lemma,

$$d(\log\ y_t) = \left(\mu - \frac{\sigma^2}{2}\right)dt + \sigma\, dW_t. \tag{A.4}$$

This equation provides the "exact" discretization,

$$\log\ y_t = \log\ y_{t-1} + \mu - \frac{\sigma^2}{2} + \sigma\epsilon_t. \tag{A.5}$$

Combining (A.2) and (A.5), we obtain

$$b(\mu, \varepsilon) = E\left[\exp\left(\mu - \frac{\sigma^2}{2} + \sigma\epsilon_t\right)\right] = e^{\mu - \frac{\sigma^2}{2}} M_{\epsilon_t}(\sigma), \tag{A.6}$$

where $M_{\epsilon_t}(\cdot)$ is the moment-generating function of ϵ_t, that is,

$$M_{\epsilon_t}(\lambda) = E[e^{\lambda\epsilon_t}] = (1-\varepsilon)e^{\lambda^2/2} + \varepsilon e^{\lambda^2\tau^2/2}, \tag{A.7}$$

and finally

$$b(\mu, \varepsilon) = e^{\mu}(1 + \varepsilon\ D) - 1, \tag{A.8}$$

where $D = e^{\sigma^2(\tau^2-1)/2} - 1$. Because $EA(\mu) = 0$, (A.1) and (A.8) give the bias of $\tilde{\mu}$ in (13).

To compute the bias of the indirect estimator $\hat{\mu}_I$, we use proposition 4.5 of Gouriéroux and Monfort (1996, p. 77–79). Because (A.1) has the form of (4.23) of Gouriéroux and Monfort (1996, p. 77) with $B(\cdot; \cdot) = 0$, it follows from their (4.26) that

$$\mathrm{bias}(\hat{\mu}_I, \varepsilon) = E\hat{\mu}_I - \mu = \frac{Eb^*}{n} + O(n^{-2}), \tag{A.9}$$

where Eb^* is defined in their proposition 4.5 (p. 79). In our case, $b(\mu, \varepsilon)$ is given by (A.8), $A(\mu)$ is defined in (A.1), $\frac{\partial b}{\partial\mu} = \frac{\partial^2 b}{\partial\mu^2} = b + 1$, $\frac{\partial A}{\partial\mu} = -\sqrt{n}b$, and var A = var r_t. Therefore,

$$Eb^* = -\frac{1}{2}\left(1 + \frac{1}{s}\right)(b+1)^{-2}\mathrm{var}\ r_t. \tag{A.10}$$

Analogously to (A.2) and (A.6), we obtain

$$\begin{aligned}
\mathrm{var}\ r_t = Er_t^2 - (Er_t)^2 &= E\left[\exp\left(2\ \log\left(\frac{y_t}{y_{t-1}}\right)\right)\right] - (b(\mu, \varepsilon) + 1)^2 \\
&= e^{2\mu - \sigma^2} M_{\epsilon_t}(2\sigma) - (b+1)^2 \\
&= e^{2\mu + \sigma^2}\left(1 + \varepsilon\left(e^{2\sigma^2(\tau^2-1)} - 1\right)\right) - (b+1)^2 \\
&= e^{2\mu}\left(e^{\sigma^2} - 1 + \varepsilon\ C\right) + O(\varepsilon^2),
\end{aligned} \tag{A.11}$$

where $C = e^{\sigma^2(2\tau^2-1)} - e^{\sigma^2} - 2e^{\sigma^2(\tau^2-1)/2} + 2$.

Inserting (A.11) and (A.8) in (A.10), we obtain

$$\begin{aligned}
Eb^* &= -\frac{1}{2}\left(1 + \frac{1}{s}\right)(e^{\sigma^2} - 1 + \varepsilon\ C + O(\varepsilon^2))(1 + \varepsilon\ D)^{-2} \\
&= -\frac{1}{2}\left(1 + \frac{1}{s}\right)\left[e^{\sigma^2} - 1 + \varepsilon\left(C - 2D\left(e^{\sigma^2} - 1\right)\right)\right] + O(\varepsilon^2),
\end{aligned}$$

and substituting the expressions for C and D, from (A.9) we finally obtain (14).

[Received April 2001. Revised January 2002.]

REFERENCES

Ali, M. M. (1979), "Analysis of Stationary Spatial-Temporal Processes: Estimation and Prediction," *Biometrika*, 66, 513–518.

Allende, H., and Heiler, S. (1992), "Recursive Generalized M Estimates for Autoregressive Moving-Average Models," *Journal of Time Series Analysis*, 13, 1–18.

Bustos, O. H., and Yohai, V. J. (1986), "Robust Estimates for ARMA Models," *Journal of the American Statistical Association*, 81, 155–168.

Cabrera, J., and Fernholz, L. T. (1999), "Target Estimation for Bias and Mean Square Error Reduction," *The Annals of Statistics*, 27, 1080–1104.

Chan, K. G., Karloyi, G. A., Longstaff, F. A., and Sanders, A. B. (1992), "An Empirical Comparison of Alternative Models of the Short-Term Rate," *Journal of Finance*, 47, 1209–1227.

de Luna, X., and Genton, M. G. (2001), "Robust Simulation-Based Estimation of ARMA Models," *Journal of Computational and Graphical Statistics*, 10, 370–387.

——— (2002), "Simulation-Based Inference for Simultaneous Processes on Regular Lattices," *Statistics and Computing*, 12, 125–134.

Gallant, A. R., and Tauchen, G. (1996), "Which Moments to Match?," *Econometric Theory*, 12, 657–681.

Genton, M. G., and de Luna, X. (2000), "Robust Simulation-Based Estimation," *Statistics and Probability Letters*, 48, 253–259.

Gouriéroux, C., and Monfort, A. (1996), *Simulation-Based Econometric Methods*, Oxford, U.K.: Oxford University Press.

Gouriéroux, C., Monfort, A., and Renault, A. E. (1993), "Indirect Inference," *Journal of Applied Econometrics*, 8 (supplement), S85–S118.

Ha, E., and Newton, H. J. (1993), "The Bias of Estimators of Causal Spatial Autoregressive Processes," *Biometrika*, 80, 242–245.

Haining, R. (1987), "Trend-Surface Models With Regional and Local Scales of Variation With an Application to Aerial Survey Data," *Technometrics*, 29, 461–469.

——— (1990), *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge, U.K.: Cambridge University Press.

Hampel, F. R. (1974), "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association*, 69, 383–393.

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

Heritier, S., and Ronchetti, E. (1994), "Robust Bounded-Influence Tests in General Parametric Models," *Journal of the American Statistical Association*, 89, 897–904.

Huber, P. J. (1981), *Robust Statistics*, New York: Wiley.

Ma, Y., and Genton, M. G. (2000), "Highly Robust Estimation of the Autocovariance Function," *Journal of Time Series Analysis*, 21, 663–684.

Martin, R. D., and Yohai, V. J. (1985), "Robustness in Time Series and Estimating ARMA Models," in *Handbook of Statistics 5*, eds. E. J. Hannan, P. R. Krishnaiah, and M. M. Rao, Amsterdam: Elsevier, pp. 119–155.

——— (1986), "Influence Functionals for Time Series" (with discussion), *The Annals of Statistics*, 14, 781–855.

Ronchetti, E., and Trojani, F. (2001), "Robust Inference With GMM Estimators," *Journal of Econometrics*, 101, 37–69.

Rosenberger, J., and Gasko, M. (1983), "Comparing Location Estimators: Trimmed Means, Medians, and Trimean," in *Understanding Robust and Exploratory Data Analysis*, eds. D. C. Hoaglin, F. Mosteller, and J. W. Tukey, New York: Wiley, pp. 297–338.

Tjøstheim, D. (1978), "Statistical Spatial Series Modeling," *Advances in Applied Probability*, 10, 130–154.

Whittle, P. (1954), "On Stationary Processes in the Plane," *Biometrika*, 41, 434–449.