Sparse Indirect Inference
Inferenza indiretta sparsa

Stolfi Paola and Bernardi Mauro and Petrella Lea

Abstract In this paper we propose a sparse indirect inference estimator. In order to achieve sparse estimation of the parameters, the Smoothly Clipped Absolute Deviation (SCAD) ℓ_1–penalty of Fan and Li (2001) is added into the indirect inference objective function introduced by Gourieroux et al. (1993). We derive the asymptotic theory and we show that the sparse–Indirect Inference estimator enjoys the oracle properties under mild regularity conditions. The method is applied to estimate the parameters of large dimensional non–Gaussian regression models.
Abstract *In questo lavoro si propone un metodo di stima indiretta sparsa. A tal fine la funzione di penalita SCAD–ℓ_1 di Fan and Li (2001) e introdotta nella funzione obiettivo del metodo di inferenza indiretta di Gourieroux et al. (1993). Sotto usuali condizioni di regolarita vengono inoltre dimostrate la consistenza e la Normalità asintotica unitamente alle proprieta di stimatore ORACLE. Il metodo e illustrato con l'applicazione alla stima di modelli di regressione lineare con distribuzione non–Gaussiana del termine di errore.*

Key words: Indirect inference; sparse regularisation; SCAD penalty, stable non–Gaussian models.

1 Introduction

Indirect inference (II) methods (Gourieroux et al. 1993, Gallant and Tauchen, 1996) are likelihood–free alternatives to maximum likelihood or moment–based estimation methods for parametric inference when a closed–form expression for the density is not available. Throughout the paper we consider the following dynamic model

Stolfi Paola
Istituto per le Applicazioni del Calcolo "Mauro Picone" - CNR, e-mail: p.stolfi@iac.cnr.it.

Bernardi Mauro
Department of Statistical Sciences, University of Padova e-mail: mauro.bernardi@unipd.it

Petrella Lea
MEMOTEF Department, Sapienza University of Rome, e-mail: lea.petrella@uniroma1.it.

Alessandra Petrucci, Rosanna Verde (edited by), *SIS 2017. Statistics and Data Science: new challenges, new generations. 28-30 June 2017 Florence (Italy). Proceedings of the Conference of the Italian Statistical Society*
ISBN (online) 978-88-6453-521-0 (online), CC BY 4.0, 2017 Firenze University Press

$$y_t = r(y_{t-1}, \mathbf{x}_t, u_t, \vartheta) \tag{1}$$
$$u_t = \phi(u_{t-1}, \varepsilon_t, \vartheta), \tag{2}$$

where $\mathbf{x}_t$ are exogenous variables whereas u_t and ε_t are latent variables. We assume that: *(i)* $\mathbf{x}_t$ is an homogeneous Markov process with transition distribution F_0 independent of ε_t and u_t; *(ii)* the process ε_t is a white noise whose distribution G_0 is known, and *(iii)* the process $\{y_t, \mathbf{x}_t\}$ is weekly stationary. We further assume that the joint density function of the observations $\{y_t, \mathbf{x}_t\}_{t=1}^T$ is not known analytically. The II method replaces the maximum likelihood estimator of the parameter ϑ in equations (1)–(2) with a quasi–maximum likelihood estimator which relies on an alternative auxiliary model and then exploits simulations from the original model to correct for inconsistency. Specifically, let $Q_T(y_T, \mathbf{x}_T, \beta)$ the auxiliary criterion function, which depends on the observations $\{y_t, \mathbf{x}_t\}_{t=1}^T$ and on the auxiliary parameter $\beta \in \mathbf{B} \subset \mathbb{R}^q$, such that $\lim_{T\to\infty} Q_T(y_T, \mathbf{x}_T, \beta) = Q_\infty(F_0, G_0, \vartheta_0, \beta)$, a.s., where ϑ_0 is the true parameter of interest, then

$$\hat{\beta}_T = \arg\max_{\beta\in\mathbf{B}} Q_T(y_T, \mathbf{x}_T, \beta). \tag{3}$$

Under the additional assumptions that the limit criterion is continuous in β and has a unique maximum β_0, then the estimator $\hat{\beta}_T$ is a consistent estimator of β_0, that is unknown since it depends on F_0 and ϑ_0 that are unknown. To overcome this problem, the II method simulates, for each value of ϑ, H paths $\tilde{y}_T^h$ for $h = 1, 2, \dots, H$ and computes the QML estimate $\tilde{\beta}_T^h$ for the auxiliary model in equation (3) and subsequently minimises the following objective function

$$\hat{\vartheta}_T = \arg\min_{\vartheta} \left(\hat{\beta}_T - \frac{1}{H}\sum_{h=1}^{H} \tilde{\beta}_T^h\right)' \hat{\Omega}_T \left(\hat{\beta}_T - \frac{1}{H}\sum_{h=1}^{H} \tilde{\beta}_T^h\right), \tag{4}$$

for an appropriately chosen positive–definite square symmetric matrix $\hat{\Omega}_T$. Indirect estimator are consistent and asymptotically Normal under mild regularity conditions, see Gourieroux et al. (1993). The most important condition concerns the binding function that maps the parameter space of the auxiliary model onto the parameter space of the true model

$$b(F, G, \theta) = \arg\max_{\beta\in B} Q_T(F, G, \theta, \beta), \tag{5}$$

must be one–to–one. We further assume that $\frac{\partial b}{\partial \vartheta}(F_0, G_0, \cdot)$ is of full–column rank. In the following Section we introduce the Sparse–II estimator.

2 Sparse indirect inference

In order to achieve sparse estimation of the parameter ϑ we introduce the Smoothly Clipped Absolute Deviation (SCAD) ℓ_1–penalty of Fan and Li (2001) into the II objective function. The SCAD function is a non–convex penalty function with the following form

$$p_{\lambda}(|\gamma|) = \begin{cases} \lambda|\gamma| & \text{if } |\gamma| \leq \lambda \\ \frac{1}{a-1}\left(a\lambda|\gamma| - \frac{\gamma^2}{2}\right) - \frac{\lambda^2}{2(a-1)} & \text{if } \lambda < \gamma \leq a\lambda \\ \frac{\lambda^2(a+1)}{2} & \text{if } a\lambda < |\gamma|, \end{cases} \tag{6}$$

which corresponds to quadratic spline function with knots at λ and $a\lambda$. The SCAD penalty is continuously differentiable on $(-\infty;0) \cup (0;\infty)$ but singular at 0 with its derivatives zero outside the range $[-a\lambda; a\lambda]$. This results in small coefficients being set to zero, a few other coefficients being shrunk towards zero while retaining the large coefficients as they are. The Sparse II estimator minimises the penalised II objective function, as follows

$$\hat{\vartheta}^* = \arg\min_{\vartheta} D^{\star}(\vartheta), \tag{7}$$

where

$$D^{\star}(\vartheta) = \left(\hat{\beta}_T - \frac{1}{H}\sum_{h=1}^{H}\tilde{\beta}_T^h\right)' \hat{\Omega}_T \left(\hat{\beta}_T - \frac{1}{H}\sum_{h=1}^{H}\tilde{\beta}_T^h\right) + n\sum_i p_{\lambda}(|\vartheta_i|), \tag{8}$$

where $\hat{\Omega}_T$ is a positive–definite square symmetric matrix. A similar approach in a different context has been recently proposed by Blasques and Duplinskiy (2015).

3 Asymptotic theory

As shown in Fan and Li (2001), the SCAD estimator, with appropriate choice of the regularisation (tuning) parameter, possesses a sparsity property, i.e., it estimates zero components of the true parameter vector exactly as zero with probability approaching one as sample size increases while still being consistent for the non–zero components. An immediate consequence of the sparsity property of the SCAD estimator is the, so called, oracle property, i.e., the asymptotic distribution of the estimator remains the same whether or not the correct zero restrictions are imposed in the course of the SCAD estimation procedure. More specifically, let $\vartheta_0 = \left(\vartheta_0^1, \vartheta_0^0\right)$ be the true value of the unknown parameter ϑ, where $\vartheta_0^1 \in \mathbb{R}^s$ is the subset of non–zero parameters and $\vartheta_0^0 = 0 \in \mathbb{R}^{k-s}$ and let $A = \left\{i : \vartheta_i \in \vartheta_0^1\right\}$, we consider the following definition of oracle estimator given by Zou (2006).

Definition 1. An oracle estimator $\hat{\vartheta}_{\text{oracle}}$ has the following properties:

(i) consistent variable selection: $\lim_{n\to\infty} \mathbb{P}(A_n = A) = 1$, where $A_n = \left\{ i : \hat{\vartheta}_i \in \hat{\vartheta}^1_{\text{oracle}} \right\}$;

(ii) asymptotic normality: $\sqrt{n}\left(\hat{\vartheta}^1_{\text{oracle}} - \vartheta^1_0\right) \xrightarrow{d} \mathsf{N}(\mathbf{0}, \Sigma)$, as $n \to \infty$, where Σ is the variance covariance matrix of ϑ^1_0.

In the remainder the Section we establish the oracle properties of the penalised SCAD II estimator. To this end, the following set of assumptions are needed:

(i)

$$\xi_T = \sqrt{T}\left(\frac{\partial Q_T}{\partial \beta}(y_T, x_T, \beta_0) - \frac{1}{H}\sum_{h=1}^{H}\frac{\partial Q_T}{\partial \beta}\left(\tilde{y}^h_T, x_T, \beta_0\right)\right), \tag{9}$$

is asymptotically normal with mean zero, and asymptotic variance–covariance matrix given by $W = \lim_{T\to\infty} V(\xi_T)$;

(ii)

$$\lim_{T\to\infty} V\left(\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\tilde{y}^h_T, x_T, \beta_0\right)\right) = I_0, \tag{10}$$

and the limit is independent of the initial values z^h_0, for $h = 1, 2, \ldots, H$;

(iii)

$$\lim_{T\to\infty} \text{Cov}\left(\sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\tilde{y}^h_T, x_T, \beta_0\right), \sqrt{T}\frac{\partial Q_T}{\partial \beta}\left(\tilde{y}^l_T, x_T, \beta_0\right)\right) = K_0, \tag{11}$$

and the limit is independent of z^h_0 and z^l_0 for $h \neq l$;

(iv)

$$\text{plim} - \frac{\partial^2 Q_T}{\partial \beta \partial \beta'}\left(\tilde{y}^h_T, x_T, \beta_0\right) = -\frac{\partial^2 Q_\infty}{\partial \beta \partial \beta'}(F_0, G_0, \vartheta_0, \beta_0), \tag{12}$$

and the limit is independent of z^h_0.

The next Theorem states that the estimator defined in equation (7) satisfies the sparsity property.

Theorem 1. *Given the SCAD penalty function $p_\lambda(\cdot)$, for a sequence of λ_n such that $\lambda_n \to 0$, and $\sqrt{n}\lambda_n \to \infty$, as $n \to \infty$, there exists a local minimiser $\hat{\vartheta}$ of $D^\star(\vartheta)$ in* (7) *with $\|\hat{\vartheta} - \vartheta_0\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. Furthermore, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\hat{\vartheta}^0 = 0\right) = 1. \tag{13}$$

The following theorem establishes the asymptotic normality of the penalised SCAD II estimator; we denote by ϑ^1 the subvector of ϑ that does not contain zero elements and by $\hat{\vartheta}^1$ the corresponding penalised II estimator.

Theorem 2. *Given the SCAD penalty function $p_\lambda(|\vartheta_i|)$, for a sequence $\lambda_n \to 0$ and $\sqrt{n}\lambda_n \to \infty$ as $n \to \infty$, then $\hat{\vartheta}^1$ has the following asymptotic distribution:*

$$\sqrt{n}\left(\hat{\vartheta}^1 - \vartheta_0^1\right) \xrightarrow{d} \mathsf{N}\left(\mathbf{0}, \left(1+\frac{1}{H}\right)\mathbf{W}\right), \tag{14}$$

as $n \to \infty$, where

$$\mathbf{W} = \left(b'(F_0,G_0,\vartheta_0)'\,\Omega b'(F_0,G_0,\vartheta_0)\right)^{-1}\mathbf{W}_1\left(b'(F_0,G_0,\vartheta_0)'\,\Omega b'(F_0,G_0,\vartheta_0)\right)^{-1},$$

and

$$\mathbf{W}_1 = b'(F_0,G_0,\vartheta_0)'\,\Omega\mathbf{J}_0^{-1}(\mathbf{I}_0-\mathbf{K}_0)\,\mathbf{J}_0^{-1}\Omega b'(F_0,G_0,\vartheta_0), \tag{15}$$

where $b'(F_0,G_0,\vartheta_0) = \frac{\partial b(F_0,G_0,\vartheta_0)}{\partial \vartheta^{1'}}$ is the first derivative of the binding function $b(F_0,G_0,\vartheta_0)$.

4 Sparse II algorithm

The objective function of the sparse estimator is the sum of a convex function and a non convex function which complicates the minimisation procedure. Here, we adapt the algorithms proposed by Fan and Li (2001) and Hunter and Li (2005) to our objective function in order to allow a fast procedure for the minimisation problem. To this aim we consider the first order Taylor expansion of the penalty, for $\vartheta_i \approx \vartheta_{i0}$

$$p_\lambda(|\vartheta_i|) \approx p_\lambda(|\vartheta_{i0}|) + \frac{1}{2}\frac{p'_\lambda(|\vartheta_{i0}|)}{|\vartheta_{i0}|}\left(\vartheta_i^2 - \vartheta_{i0}^2\right), \tag{16}$$

where the first derivative of the penalty function has been approximated as follows:

$$\left[p_\lambda(|\vartheta_i|)\right]' = p'_\lambda(|\vartheta_i|)\,\mathrm{sgn}(\vartheta_i) \approx \frac{p'_\lambda(|\vartheta_{i0}|)}{|\vartheta_{i0}|}\vartheta_i, \tag{17}$$

when $\vartheta_i \neq 0$. The objective function $D^\star$ in equation (7) can be locally approximated, except for a constant term by

$$D^\star(\vartheta) \approx \left(\hat{\beta} - \bar{\tilde{\beta}}^h_{\vartheta_0}\right)\hat{\Omega}\left(\hat{\beta} - \bar{\tilde{\beta}}^h_{\vartheta_0}\right) - \frac{1}{H}\sum_{h=1}^{H}\frac{\partial\tilde{\beta}^h_{\vartheta_0}}{\partial\vartheta}\hat{\Omega}\left(\hat{\beta} - \bar{\tilde{\beta}}^h_{\vartheta_0}\right)(\vartheta-\vartheta_0)$$
$$+\frac{1}{2}(\vartheta-\vartheta_0)'\frac{1}{H}\sum_{h=1}^{H}\frac{\partial\tilde{\beta}^h_{\vartheta_0}}{\partial\vartheta}\hat{\Omega}\frac{1}{H}\sum_{h=1}^{H}\frac{\partial\tilde{\beta}^h_{\vartheta_0}}{\partial\vartheta}(\vartheta-\vartheta_0) + \frac{n}{2}\vartheta'\bar{\mathbf{P}}_{\lambda_n}(\vartheta_0)\,\vartheta, \tag{18}$$

where $\bar{\tilde{\beta}}_{\vartheta_0}^h = \frac{1}{H}\sum_{h=1}^H \tilde{\beta}_{\vartheta_0}^h$ and $\bar{\mathbf{P}}_{\lambda_n}(\vartheta) = \mathrm{diag}\left\{\frac{p'_{\lambda_n}(|\vartheta_i|)}{|\vartheta_i|}; \vartheta_i \in \vartheta^1\right\}$. Then the first order condition becomes

$$\begin{aligned}\frac{\partial D^\star(\vartheta)}{\partial\vartheta} &\approx -\frac{1}{H}\sum_{h=1}^H \frac{\partial\tilde{\beta}_{\vartheta_0}^h}{\partial\vartheta}\hat{\Omega}\left(\hat{\beta}-\bar{\tilde{\beta}}_{\vartheta_0}^h\right)\\ &\quad+\frac{1}{H}\sum_{h=1}^H \frac{\partial\tilde{\beta}_{\vartheta_0}^h}{\partial\vartheta}\hat{\Omega}\frac{1}{H}\sum_{h=1}^H \frac{\partial\tilde{\beta}_{\vartheta_0}^h}{\partial\vartheta}(\vartheta-\vartheta_0)+n\bar{\mathbf{P}}_{\lambda_n}(\vartheta_0)\vartheta\\ &= -\frac{1}{H}\sum_{h=1}^H \frac{\partial\tilde{\beta}_{\vartheta_0}^h}{\partial\vartheta}\hat{\Omega}\left(\hat{\beta}-\bar{\tilde{\beta}}_{\vartheta_0}^h\right)+\frac{1}{H}\sum_{h=1}^H \frac{\partial\tilde{\beta}_{\vartheta_0}^h}{\partial\vartheta}\hat{\Omega}\frac{1}{H}\sum_{h=1}^H \frac{\partial\tilde{\beta}_{\vartheta_0}^h}{\partial\vartheta}(\vartheta-\vartheta_0)\\ &\quad+n\bar{\mathbf{P}}_{\lambda_n}(\vartheta_0)(\vartheta-\vartheta_0)+n\bar{\mathbf{P}}_{\lambda_n}(\vartheta_0)\vartheta_0\\ &= 0,\end{aligned} \tag{19}$$

therefore

$$\begin{aligned}\vartheta = \vartheta_0 &- \left[\frac{1}{H}\sum_{h=1}^H \frac{\partial\tilde{\beta}_{\vartheta_0}^h}{\partial\vartheta}\hat{\Omega}\frac{1}{H}\sum_{h=1}^H \frac{\partial\tilde{\beta}_{\vartheta_0}^h}{\partial\vartheta}+n\bar{\mathbf{P}}_{\lambda_n}(\vartheta_0)\right]^{-1}\\ &\times\left[\frac{1}{H}\sum_{h=1}^H \frac{\partial\tilde{\beta}_{\vartheta_0}^h}{\partial\vartheta}\hat{\Omega}\left(\hat{\beta}-\bar{\tilde{\beta}}_{\vartheta_0}^h\right)-n\bar{\mathbf{P}}_{\lambda_n}(\vartheta_0)\vartheta_0\right].\end{aligned} \tag{20}$$

The optimal solution can be found iteratively, as follows

$$\begin{aligned}\vartheta^{(k+1)} = \vartheta^{(k)} &- \left[\frac{1}{H}\sum_{h=1}^H \frac{\partial\tilde{\beta}_{\vartheta^{(k)}}^h}{\partial\vartheta}\hat{\Omega}\frac{1}{H}\sum_{h=1}^H \frac{\partial\tilde{\beta}_{\vartheta^{(k)}}^h}{\partial\vartheta}+n\bar{\mathbf{P}}_{\lambda_n}\left(\vartheta^{(k)}\right)\right]^{-1}\\ &\times\left[\frac{1}{H}\sum_{h=1}^H \frac{\partial\tilde{\beta}_{\vartheta^{(k)}}^h}{\partial\vartheta}\hat{\Omega}\left(\hat{\beta}-\bar{\tilde{\beta}}_{\vartheta_0}^h\right)-n\bar{\mathbf{P}}_{\lambda_n}\left(\vartheta^{(k)}\right)\vartheta^{(k)}\right],\end{aligned} \tag{21}$$

and if $\vartheta_i^{(k+1)} \approx 0$, then $\vartheta_i^{(k+1)}$ is set equal zero. When the algorithm converges the estimator satisfies the following equation

$$\frac{1}{H}\sum_{h=1}^H \frac{\partial\tilde{\beta}_{\vartheta_0}^h}{\partial\vartheta}\hat{\Omega}\left(\hat{\beta}-\bar{\tilde{\beta}}_{\vartheta_0}^h\right)-n\bar{\mathbf{P}}_{\lambda_n}\vartheta_0 = 0, \tag{22}$$

that is the first order condition of the minimisation problem of the Sparse–II estimator.

5 Tuning paramenter selection

The SCAD penalty requires the selection of two tuning parameters (a,λ). The first tuning parameter is fixed at $a=3.7$ as suggested in Fan and Li (2001), while the parameter λ is selected using the cross validation function

$$CV(\lambda)=\sum_{k=1}^{K}\frac{1}{n_k}\left(\hat{\beta}-\frac{1}{H}\sum_{h=1}^{H}\tilde{\beta}^{h}_{\hat{\vartheta}_{\lambda,k}}\right)\hat{\Omega}\left(\hat{\beta}-\frac{1}{H}\sum_{h=1}^{H}\tilde{\beta}^{h}_{\hat{\vartheta}_{\lambda,k}}\right), \tag{23}$$

where $\hat{\vartheta}_{\lambda,k}$ denotes the parameters estimate over the sample $\left(\cup_{i=1}^{K}T_k\right)\setminus T_k$ with λ as tuning parameter. Then the optimal value is chosen as $\lambda^*=\arg\min_\lambda CV(\lambda)$, where again the minimisation is performed over a grid of values for λ.

6 Application

Let $\mathbf{y}=(y_1,y_2,\dots,y_T)'$ be the vector of observations on the scalar response variable Y, $\mathbf{X}=(\mathbf{x}_1',\mathbf{x}_2',\dots,\mathbf{x}_T')'$ is the $(n\times p)$ matrix of observations on the p covariates, i.e., $\mathbf{x}_{j,t}=\left(x_{j,1},x_{j,2},\dots,x_{j,p}\right)$ and consider the following regression model

$$\mathbf{y}=\iota_T\delta+\mathbf{X}\gamma+\varepsilon,\qquad \varepsilon\sim\mathsf{S}_\alpha(0,\sigma), \tag{24}$$

where ι_T is the $T\times 1$ vector of unit elements, $\delta\in\mathbb{R}$ denotes the parameter related to the intercept of the model, $\gamma=(\gamma_1,\gamma_2,\dots,\gamma_p)'$ is the $p\times 1$ vector of regression parameters and $\mathsf{S}_\alpha(0,\sigma)$ denotes the symmetric α–Stable distribution (Samorodnitsky et al. 1994) centred at zero with characteristic exponent $\alpha\in(0,2)$ and scale parameter $\sigma>0$. We further assume that the element of the vector of innovations $\varepsilon=(\varepsilon_1,\varepsilon_2,\dots,\varepsilon_T)$ are independent i.e. $\varepsilon_j \perp\!\!\!\perp \varepsilon_k$, for any $j\neq k$ and they are independent of $\mathbf{x}_l$, for $l=1,2,\dots,p$. Indirect inference for Stable distributions has been previously considered by Lombardi and Veredas (2009). The Sparse–II method requires the definition of the auxiliary model as well as the metric used to compare the synthetic data generated by the method $\tilde{\mathbf{y}}$ with the true data $\mathbf{y}$. As auxiliary distribution we consider the Student–t regression model defined in equation (24), with the only difference that the error term follows a Student–t distribution $\varepsilon\sim\mathsf{T}\left(0,\sigma^2,\nu\right)$. As regards the metric, we consider the L_2 distance between the scores of the auxiliary model evaluated at the true $\mathbf{y}$ and simulated $\tilde{\mathbf{y}}$, i.e., $\|\nabla\left(\hat{\beta},\tilde{y}\right)-\nabla\left(\hat{\beta},y\right)\|_2^2$. In Table 1, we report the empirical inclusion probabilities of the regression parameters obtained over $1,00$ replications of the α–Stable regression model defined in equation (24), for two values of $\alpha=(1.70,1.95)$ with $n=250$. The true parameters are defined in the column (Par.) of Table (24), while the scale parameter of the Stable distribution is held fixed at $\sigma=0.05$. Our simulation results confirm that the sparse Indirect estimator perform well in detecting zeros in linear non–Gaussian regression models.

Par.	True	EIP	EIP	Par.	True	EIP	EIP
		$\alpha = 1.70$	$\alpha = 1.95$			$\alpha = 1.70$	$\alpha = 1.95$
δ	1	0	0	γ_{11}	0	0.6591	0.8919
γ_1	2	0	0	γ_{12}	0	0.7500	0.8378
γ_2	2	0	0	γ_{13}	0	0.8182	0.9459
γ_3	3	0	0	γ_{14}	0	0.7273	0.9189
γ_4	1	0	0	γ_{15}	0	0.7955	0.9730
γ_5	2	0	0	γ_{16}	0	0.7273	0.8378
γ_6	3	0	0	γ_{17}	0	0.7727	0.8919
γ_7	1	0	0	γ_{18}	0	0.8182	0.9189
γ_8	2	0	0	γ_{19}	0	0.8636	0.9459
γ_9	3	0	0	γ_{20}	0	0.8636	1.0000
γ_{10}	0	0	0	γ_{21}	0	0.8636	0.9459

Table 1 Empirical inclusion probabilities (EIP) evaluated over $1,00$ replications for the regression parameters (δ, γ) of the α–Stable regression model defined in equation (24).

7 Conclusion

In this paper we introduce the sparse indirect inference (SII) estimator and we extend the asymptotic theory. Empirical properties of the estimator are evaluated by means of a simulation study where a moderately large linear regression model with non–Gaussian innovations is considered. Our results confirm that the SII estimator performs well in detecting non zero regressor parameters.

References

1. Blasques, F. and Duplinskiy, A. (2015). Penalized Indirect Inference. *Tinbergen Institute*, WorkingPaper, 15-09/III.
2. Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, 96(456):1348–1360, 2001.
3. Gourieroux, C. and Monfort, A. (1996). *Simulation-based econometric methods*. CORE lectures. Oxford University Press, 1996.
4. Gourieroux, C., Monfort, A. and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8(S1):S85–S118, 1993.
5. Hunter, D.R. and Li, R. (2005). Variable selection using mm algorithms. *Ann. Statist.*, 33(4):1617–1642.
6. Lombardi, M. J. and Veredas, D. (2009). Indirect estimation of elliptical stable distributions. *Comput. Statist. Data Anal.*, 6(53):2309–2324.
7. Samorodnitsky, G. and Taqqu, M. S. (1994). Stable non-Gaussian random processes. *Chapman & Hall, New York*, xxii+632.
8. Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.*, 476(101):1418–1429.