Linkage Between WLAD and GLAD and its Applications for Autoregressive Analysis

A.V. Panyukov
Department of Computational Mathematics and Informatics
South Ural National Research University
Chelyabinsk, Russia
e-mail: paniukovav@susu.ac.ru

I.A. Tetin
Department of Computational Mathematics and Informatics
South Ural National Research University
Chelyabinsk, Russia
e-mail: Ilya.Tetin@susu.ru

Ya.A. Mezal
Department of Computational Mathematics and Informatics
South Ural National Research University
Chelyabinsk, Russia
e-mail: yaser_ali_84@yahoo.com

Abstract[1]

The most commonly used method to determine the coefficients of the regression equation is the ordinary least squares method (OLS). Usage of OLS assumes low correlation of explanatory variables, independence and normal distribution of measurement errors. It is known that even minor violations of these assumptions drastically reduce the efficiency of estimates. Procedure of OLS-estimation is unstable in the presence of large measurement errors, the estimators become inconsistent. Computation of autoregressive estimates is very complicated due to poor conditionality of equations system representing the necessary conditions to find the minimum sum of squares. An alternative to OLS is Least Absolute Deviations (LAD) method, which gives robust estimates even under violation of OLS assumptions. The paper considers two types of LAD: Weighted LAD and Generalized LAD. It is stated that interdependence of this methods allows to reduce the problem of computing GLAD-estimates to the iterative computation of WLAD-estimates. The latter are calculated by solving the corresponding linear programming problem.

1. Introduction

Consider the evaluation of the coefficients of the linear equation autoregression equation

$$x_t = \sum_{j=1}^{m} a_j x_{t-j} + \varepsilon_t, \quad t = 1, 2, \ldots, n, \tag{1}$$

here $y_1, y_2, \ldots, y_n$ are the values of the state variable, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are random errors, $a_1, a_2, a_3 \ldots, a_m$ are unknown coefficients.

Ordinary Least Squares is the parametric method and it is the most commonly used method for estimation the coefficients of regression equation. To use OLS we need some strict assumptions. They include independence and normal distribution of errors and determinacy of explanatory variables [1-3]. Even minor violations of stated assumptions dramatically lower the efficiency of estimators. Note the instability of OLS estimation process in case of presence of large measurements errors. In this case, estimated coefficients become inconsistent. Finding estimates of autoregressive equation becomes substantially complicated due to poor conditionality of the system of equations representing necessary conditions for minimization of sum of squared deviations.

Least Absolute Deviations method is an alternative method to OLS to obtain robust errors in case of violation of OLS assumptions [4]. We present two types of LAD: Weighted LAD and Generalized LAD. In the paper, we find the interrelation of these methods, and this fact allowed us to reduce the problem of determining GALD-estimates to an iterative procedure with WALD-estimators. The latter are calculated by solving the corresponding linear programming problem. A sufficient condition imposed on the loss function is found. It ensures the stability of GLAD-estimates of autoregressive models in terms of outliers.

2. Weighted Least Absolute Deviations Method (WLAD)

One can get the WLAD estimations of coefficients by solving the problem

Proceedings of the 4th International Conference "Information Technologies for Intelligent Decision Making Support", May 17 - 19, Ufa, Russia, 2016

$$\left(a_1^*, a_2^*, a_3^* \ldots, a_m^*\right) = \arg \min_{(a_1,a_2,a_3\ldots,a_m)\in \mathbb{R}^m} \sum_{t=1}^{n} p_t \left| x_t - \sum_{j=1}^{m} a_j x_{t-j} \right|, \quad (2)$$

In which $p_t \geq 0, t = 1,2,\ldots,n$ are some pre-defined factors. This task is the problem of a convex piecewise-linear optimization. The introduction of additional variables is reduced problem (2) to a linear programming problem

$$\min_{\substack{(a_1,a_2,a_3\ldots,a_m)\in \mathbb{R}^m \\ (u_1,u_2,u_3,\ldots,u_n)\in \mathbb{R}^n}} \left\{ \begin{array}{c} \sum_{t=1}^{n} p_t u_t : -u_t \leq x_t - \sum_{j=1}^{m} a_j x_{t-j} \leq u_t, \\ u_t \geq 0, t = 1,2,\ldots,n \end{array} \right\}. \quad (3)$$

This task has the canonical form of $n+m+1$ variables and $3n$ constraints inequalities, including the conditions of non-negativity of the variables u_j, $j = 1,2,\ldots,n.$

Weighted least absolute deviations (WLAD) can be used in the following cases. First, if there is reason to believe that the error variance is functionally dependent on one or more of the proportionality factors [2]. The problem here is the same as for the weighted OLS. The search procedure is ambiguous and weighting factors usually leads to multiple solutions. As a result, it is not clear how to use the weighting.

Second, as shown in [2], the LAD-estimation of autoregression coefficients are not stable (not consistent) in the case of large errors. Usage of some functions from previous values of $y_{t-1}, y_{t-2}, \ldots, y_{t-m}$ as weight coefficients p_t proposed in [2]. Estimates in this case are consistent.

The main difficulty in usage of GLAD is the absence of general formal rules of selection of weight coefficients. Therefore, this approach requires additional research.

3. Generalized Least Absolute Deviations Method (GLAD)

In [4] for a stable estimation of autoregression coefficients of the equation proposed by generalized least absolute deviations (GLMM), consisting in solving problems

$$\left(a_1^*, a_2^*, a_3^* \ldots, a_m^*\right) = \arg \min_{(a_1,a_2,a_3\ldots,a_m)\in \mathbb{R}^m} \sum_{t=1}^{n} \rho\left(\left| x_t - \sum_{j=1}^{m} a_j x_{t-j} \right| \right), \quad (4)$$

where ρ(*) - a concave monotone increasing twice continuously differentiable function such that ρ(0) = 0. It is following from [4] that should hold

Theorem 1. *All local minima of GLMM estimation for autoregression equation coefficients belong to the set*

$$U = \left\{ \begin{array}{l} \left(a_1^{(\mathbf{k})}, a_2^{(\mathbf{k})}, a_3^{(\mathbf{k})} \ldots, a_m^{(\mathbf{k})}\right) : x_t = \sum_{j=1}^{m} a_j^{(\mathbf{k})} x_{t-j}, \\ t \in \mathbf{k} = \{k_1, k_2, \ldots, k_m : 1 \leq k_1 < k_2 < \ldots < k_m \leq n\} \end{array} \right\}. \quad (5)$$

Set U consists of solutions of system of n algebraic equations with m unknowns. It is obvious that the number of systems equals to $\mathbf{C}_n^m$. Therefore, solution of problem (4) may be reduced to the choice of the best $\mathbf{C}_n^m$ solutions of algebraic equation systems. This approach can be used for $m \leq 3$. To compute GLAD-estimates for higher order dimension problems the interrelation between WLAD and GLAD-estimates have to be used from the theorem stated in [4].

Theorem 2. *Let U – be the set of local extremums of the problem (4), then:*

(1) for each collection of weights $\{p_t \geq 0\}_{t=1}^{n}$

$$\arg \min_{(a_1,a_2,a_3\ldots,a_m)\in \mathbb{R}^m} \sum_{t=1}^{n} p_t \left| x_t - \sum_{j=1}^{m} a_j x_{t-j} \right| \in U; \quad (6)$$

(2) for each $\left(a_1^*, a_2^*, a_3^* \ldots, a_m^*\right) \in U$ there is a collection of weights $\{p_t \geq 0\}_{t=1}^{n}$ such as

$$\left(a_1^*, a_2^*, a_3^* \ldots, a_m^*\right) \in \arg \min_{(a_1,a_2,a_3\ldots,a_m)\in \mathbb{R}^m} \sum_{t=1}^{n} p_t \left| x_t - \sum_{j=1}^{m} a_j x_{t-j} \right|. \quad (7)$$

Theorems 1 and 2 let us, from the one hand, to reduce the problem (4) to the solution of the sequence of linear programming tasks and from the other hand, give us the way to compute the weight coefficients for the problem (2).

4. Algorithm for computing GLAD-estimates

Direct solution of problem (4) is based on the usage of theorem 1 and involves finding all node points and choosing one of them as a solution that ensures the minimum of the objective function.

The brute force algorithm requires the solution of $\mathbf{C}_n^m$ systems of linear equations of order m. For large values of n and m this leads to a significant computational complexity. An alternative approach is based on reduction of this problem to the sequence linear programming problems (3). Consider possible algorithms based on this approach.

Algorithm GLAD-estimate

Input: *number of measures n;*

values $\left\| \{y_t\}_{t=0}^{n} \right\|$ *of the dependent variable*; *function* ρ(*).

Output: *estimation of coefficients* $\{a_j\}_{j=1}^{m}$ *of autoregressive equation*

Step 1. *For all* $t = 1,2,\ldots,n$ *define* $p_t := 1$;

$k := 0$;

$$\begin{pmatrix} \left(a_1^{(k)}, a_2^{(k)}, a_3^{(k)} \ldots, a_m^{(k)}\right) \\ \left(u_1^{(k)}, u_2^{(k)}, u_3^{(k)}, \ldots, u_n^{(k)}\right) \end{pmatrix} :=$$

$$\arg \min_{\substack{(a_1,a_2,a_3\ldots,a_m)\in \mathbb{R}^m \\ (u_1,u_2,u_3,\ldots,u_n)\in \mathbb{R}^n}} \left\{ \begin{array}{l} \sum_{t=1}^{n} p_t u_t : -u_t \le x_t - \sum_{j=1}^{m} a_j x_{t-j} \le u_t, \\ u_t \ge 0, t = 1, 2, \ldots, n \end{array} \right\}.$$

Step 2. *For all* $t = 1, 2, \ldots, n$ *define* $p_t := \rho'\left(u_t^{(k)}\right)$;

k:=k+1;

$$\begin{pmatrix} \left(a_1^{(k)}, a_2^{(k)}, a_3^{(k)} \ldots, a_m^{(k)}\right) \\ \left(u_1^{(k)}, u_2^{(k)}, u_3^{(k)}, \ldots, u_n^{(k)}\right) \end{pmatrix} :=$$

$$\arg \min_{\substack{(a_1,a_2,a_3\ldots,a_m)\in \mathbb{R}^m \\ (u_1,u_2,u_3,\ldots,u_n)\in \mathbb{R}^n}} \left\{ \begin{array}{c} \sum_{t=1}^{n} p_t u_t : -u_t \le x_t - \sum_{j=1}^{m} a_j x_{t-j} \le u_t, \\ u_t \ge 0, t = 1, 2, \ldots, n \end{array} \right\}.$$

Step 3. *If* $\left(a_1^{(k)}, a_2^{(k)}, a_3^{(k)} \ldots, a_m^{(k)}\right) \ne \left(a_1^{(k-1)}, a_2^{(k-1)}, a_3^{(k-1)} \ldots, a_m^{(k-1)}\right)$, *go to* **step 2**.

Step 4. *Stop. Target values are* $\left(a_1^{(k)}, a_2^{(k)}, a_3^{(k)} \ldots, a_m^{(k)}\right)$.

Performance justification of this algorithm leads us to the following theorem.

Theorem 3. *If the loss function ρ(∗) is convex upward monotonically increasing twice continuously differentiable function on positive half-axis, such as ρ′(0)=M<∞, then the sequence* $\left(a_1^{(k)}, a_2^{(k)}, a_3^{(k)} \ldots, a_m^{(k)}\right)$, *built by* **GLAD-estimate** *algorithm converges to the global extremum of the problem* (4).

Proof. The requirements imposed on the function ρ(*) implies that at any point $u^{(k)}$ an approximation (which is the majorant) for

$$\rho(u) : \nu^{\left(u^{(k)}\right)}(u) = \rho(u^{(k)}) - \rho'(u^{(k)}) \cdot u^{(k)} + \rho'(u^{(k)}) \cdot u$$

is defined, i.e.

$$\rho(u_k) = \nu(u_k) \text{ and } (\forall u \ne u_k)\left(\rho(u) < \nu_{u^{(k)}}(u)\right).$$

Therefore, in accordance with the algorithm

$$\sum_{t=1}^{n} \rho\left(\left|x_t - \sum_{j=1}^{m} a_j^{(k)} x_{t-j}\right|\right) = \sum_{t=1}^{n} \begin{pmatrix} \rho\left(\left|x_t - \sum_{j=1}^{m} a_j^{(k)} x_{t-j}\right|\right) - \\ p_t \cdot \left|x_t - \sum_{j=1}^{m} a_j^{(k)} x_{t-j}\right| + \\ p_t \cdot \left|x_t - \sum_{j=1}^{m} a_j^{(k)} x_{t-j}\right| \end{pmatrix} \ge$$

$$\sum_{t=1}^{n} \left(\rho\left(\left|x_t - \sum_{j=1}^{m} a_j^{(k)} x_{t-j}\right|\right) - p_t \cdot \left|x_t - \sum_{j=1}^{m} a_j^{(k)} x_{t-j}\right|\right) +$$

$$\min_{(a_1,a_2,a_3\ldots,a_m)\in \mathbb{R}^m} \sum_{t=1}^{n} \left(p_t \cdot \left|x_t - \sum_{j=1}^{m} a_j x_{t-j}\right|\right) =$$

$$\sum_{t=1}^{n} \left(\rho\left(\left|x_t - \sum_{j=1}^{m} a_j^{(k)} x_{t-j}\right|\right) - p_t \cdot \left|x_t - \sum_{j=1}^{m} a_j^{(k)} x_{t-j}\right|\right) +$$

$$\sum_{t=1}^{n} \left(p_t \cdot \left|x_t - \sum_{j=1}^{m} a_j^{(k+1)} x_{t-j}\right|\right) =$$

$$\sum_{t=1}^{n} \nu^{\left(u^{(k)}\right)}\left(\left|x_t - \sum_{j=1}^{m} a_j^{(k+1)} x_{t-j}\right|\right) \ge \sum_{t=1}^{n} \rho\left(\left|x_t - \sum_{j=1}^{m} a_j^{(k)} x_{t-j}\right|\right).$$

Therefore

$$\sum_{t=1}^{n} \rho\left(\left|x_t - \sum_{j=1}^{m} a_j^{(k)} x_{t-j}\right|\right) =$$

$$\sum_{t=1}^{n} \nu^{\left(u^{(k)}\right)}\left(\left|x_t - \sum_{j=1}^{m} a_j^{(k+1)} x_{t-j}\right|\right) \ge \sum_{t=1}^{n} \rho\left(\left|x_t - \sum_{j=1}^{m} a_j^{(k+1)} x_{t-j}\right|\right),$$

moreover, equality is attained only if $\rho\left(u_t^{(k)}\right) = \rho\left(u_t^{(k+1)}\right)$ for all $t = 1, 2, \cdots n$ and for all $k = 1, 2, \cdots, m$. That's why, the sequence

$$\left\{ \sum_{t=1}^{n} \rho\left(\left|x_t - \sum_{j=1}^{m} a_j^{(k)} x_{t-j}\right|\right) \right\}_{k=0,1,\ldots}$$

is monotonically decreasing and bounded below by zero, hence it has a unique limit point. An existence of limit point in the sequence $\left\{\left(a_1^{(k)}, a_2^{(k)}, a_3^{(k)} \ldots, a_m^{(k)}\right)\right\}_{k=1,2,\ldots}$ follows from continuity and monotonicity of functions $\rho(*)$.

The limit point $\left(a_1^*, a_2^*, a_3^*, \ldots, a_m^*\right)$, built by the algorithm is the global minimum, because for any set $\left(a_1, a_2, a_3, \ldots, a_m\right)$ and any $t = 1, 2, \ldots, n$ we have the following sequence of statements

$$\rho\left(\left|x_t - \sum_{j=1}^{m} a_j^* x_{t-j}\right|\right) =$$

$$\nu^{\left(u^*\right)}\left(\left|x_t - \sum_{j=1}^{m} a_j^* x_{t-j}\right|\right) \le \nu^{\left(u^*\right)}\left(\left|x_t - \sum_{j=1}^{m} a_j x_{t-j}\right|\right) \Leftrightarrow$$

$$\rho'\left(\left|x_t - \sum_{j=1}^{m} a_j^* x_{t-j}\right|\right) \cdot \left(\left|x_t - \sum_{j=1}^{m} a_j^* x_{t-j}\right|\right) \le$$

$$\rho'\left(\left|x_t - \sum_{j=1}^{m} a_j^* x_{t-j}\right|\right) \cdot \left(\left|x_t - \sum_{j=1}^{m} a_j x_{t-j}\right|\right) \Leftrightarrow$$

$$\left(\left|x_t - \sum_{j=1}^{m} a_j^* x_{t-j}\right|\right) \le \left(\left|x_t - \sum_{j=1}^{m} a_j x_{t-j}\right|\right) \Rightarrow$$

$$\rho\left(\left|x_t - \sum_{j=1}^{m} a_j^* x_{t-j}\right|\right) \le \rho\left(\left|x_t - \sum_{j=1}^{m} a_j x_{t-j}\right|\right).$$

Theorem 3 is proved.

The advantage of the proposed algorithm compared to brute force algorithm is a quite high speed of convergence for the effective use of methods of linear programming. Indeed, the linear programming task in step 2 for iteration k is different from the corresponding problem in step k-1 only by the coefficients of the objective function which allows the initial basic solution at the current iteration to use the optimum basic solution of the previous iteration.

5. Application features of the algorithm GLAD-estimate

For the implementation of the algorithm GLAD-estimate a function $\rho(*)$ is required that meets the conditions of theorems 2 and 3. Examples of such functions are

$$\arctan(|x|),\quad \frac{|x|}{|x|+1},\quad 1-\exp(-|x|),$$

$$\ln(|x|+1),\quad \sqrt{|x|+1}.$$

Another feature for computing the estimates of autoregressive equation of high order is the great sensitivity of the algorithm to rounding errors. To resolve this problem one may use the error-free implementation of the basic arithmetic operations over the rationals [6-9] and the use of parallel algorithms.

6. Conclusion

Established linkage between Weighted Least Absolute Deviation and Generalized Least Absolute Deviation methods allowed us to reduce the problem of determining GLAD-estimates to an iterative procedure of WLAD-estimates. The latter are calculated by solving the corresponding linear programming problem. The sufficient condition imposed on the loss function has found. It ensures the stability of GLAD-estimates of coefficients in autoregressive models in terms of outliers.

References

1. Pan J., Wang H., Qiwei Y. Weighted Least Absolute Deviations Estimation for ARMA Models with Infinite Variance // Econometric Theory, 23, 2007, 852-879. DOI: 10.1017/S0266466607070363
2. Powell J. L. Least Absolute Deviations Estimation for the Censored Regression Model // Journal of Econometrics 25 (1984) 303-325.
3. Dielman T. E. Least absolute value regression: recent contributions. // Journal of Statistical Computation and Simulation. – Vol. 75. No. 4, April 2003, 263-286
4. Martin R.D., Yohai V.J. Influence functionals for time series//Annals of Statistics, 1986. V.14. №4. P.781-818.
5. Panyukov A.V., Tyrsin A.N. Vzaimosvyaz' vzveshennogo i obobshchennogo metodov naimen'shih modulej // Izvestiya Chelyabinskogo nauchnogo centra UrO RAN. 2007. № 1. S. 6-11. http://elibrary.ru/item.asp?id=9572542& [Panyukov A.V., Tyrsin A.N. The Relationship Between Weighted and Generalized Least Absolute Deviation methods // Proceedings of the Chelyabinsk scientific center of UB RAS 2007. № 1. Pp. 6-11. http://elibrary.ru/item.asp?id=9572542&]
6. Panyukov A.V., Tyrsin A.N. Stable Parametric Identification of Vibratory Diagnostics Objects // Journal of Vibroengineering. 2008. T. 10. № 2. C. 142-146.
7. Panyukov A.V., Gorbik V.V. Exact and Guaranteed Accuracy Solutions of Linear Programming Problems by Distributed Computer Systems with MPI // Tambov University REPORTS: A Theoretical and Applied Scientific Journal. Series: Natural and Technical Sciences. – Volume 15, Issue 4, 2010. – P. 1392-1404.
8. Panyukov A.V., Gorbik V.V. Using Massively Parallel Computations for Absolutely Precise Solution of the Linear Programming Problems // Automation and Remote Control. 2012. T. 73. № 2. C. 276-290. DOI: 10.1134/S0005117912020063
9. Panyukov A.V., Golodov V.A. Parallel Algorithms of Integer Arithmetic in Radix Notations for Heterogeneous Computation Systems with Massive Parallelism // Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software (Bulletin SUSU MMCS), 2015, vol. 8, no. 2, pp. 117-126