# Mathematical Properties of Least Absolute Value Estimation with Serial Correlation

01206310 Shizuoka University Kazuyuki Sekitani*
01205520 Science University of Tokyo Toshiyuki Sueyoshi

We report the paper which includes partially our article [2].

## 1 Introduction

Serial Correlation(SC) is a common feature in regression analysis for time-series data set. The purpose of this article is to explore mathematical properties of Least Absolute Value Estimation (LAVE)-SC, including its formulation, algorithm, multiple solutions, and global optimality, where SC is expressed by a first-order autocorrelated disturbance.

Research rationale supporting the development of LAVE-SC is suggested by the fact that $L_2$-based time series analysis is seriously influenced by the existence of an outlier and/or non-normal error distributions. This research is fully aware of the existence of a few previous studies [3]. For example, Weiss (1990) has investigated LAVE-SC from the perspective of its statistical properties, while Sueyoshi and Sekitani [1] propose a numerical (computer-intensive) approach for LAVE-SC. Unfortunately, the two research efforts do not pay serious attention to analytical features related to the LAVE-SC. Therefore, this article focuses upon the LAVE-SC problem from the analytical view of GP so that we can discuss new mathematical programming properties regarding the estimation technique.

## 2 Formulation

A regression structure and error terms to be examined in this article is expressed by

$$y_t = X_t\beta + \epsilon_t \text{ where } \epsilon_t = \rho\epsilon_{t-1} + \eta_t, {\scriptstyle t=1,\ldots,n} \tag{1}$$

Here, $y_t$ is an observed dependent variable at the $t^{th}$ time period, and $X_t$ represents an $(m \times 1)$ vector of observed independent variables. The vector $\beta$ are parameter coefficients to be measured and $\epsilon_t$ is an error that is further broken down into two error terms: $\epsilon_{t-1}$ and $\eta_t$. Here, $\rho$ is a coefficient for SC and $\eta_t$ is an unobserved, identically distributed error. The condition, $|\rho| < 1$, is usually assumed for the coefficient representing the first-order SC, as well.

Equation (1) can be expressed by

$$\begin{aligned} y_t &= X_t\beta + \rho\epsilon_{t-1} + \eta_t \\ &= \rho y_{t-1} + (X_t - \rho X_{t-1})\beta + \eta_t \end{aligned} \tag{2}$$

To estimate both $\rho$ and $\beta$, we formulate regression (2) as the following LAVE-SC problem:

$$\min \sum_{t=2}^{n} |y_t - \rho y_{t-1} - (X_t - \rho X_{t-1})\beta|. \tag{3}$$

Alternatively,

$$\begin{array}{ll} \min. & \sum_{t=2}^{n} d_t^+ + d_t^- \\ \text{s.t.} & \rho y_{t-1} + (X_t - \rho X_{t-1})\beta + d_t^+ - d_t^- = y_t \\ & d_t^+ \geq 0 \text{ and } d_t^- \geq 0 \quad t = 2, \ldots, n. \end{array} \tag{4}$$

Here, $\eta_t$ is further broken down into two error terms: $d_t^+$ and $d_t^-$ . Each of them represents positive or negative deviations of $\eta_t$, respectively.

Let $g_t(\rho, \beta) = y_t - \rho y_{t-1} - (X_t - \rho X_{t-1})\beta$ and $f(\rho, \beta) = \sum_{t=2}^{n} |g_t(\rho, \beta)|$ then problem (3) can be reformulated as $\min f(\rho, \beta)|$. Functions $g_t(\rho, \beta)$ and $f(\rho, \beta)$ have the following properties.

**Theorem 2.1** *$g_t(\rho, \beta)$ is quasi-convex function or quasi-concave function.*

**Theorem 2.2** *For an arbitrarily fixed $\bar{\rho}$, $f(\bar{\rho}, \beta)$ is a piece-wise linear function of $\beta$. For an arbitrarily fixed $\bar{\beta}$, $f(\rho, \bar{\beta})$ is a piece-wise linear function of $\rho$.*

## 3 Algorithm

Since Theorem 2.1 implies that (4) is a nonlinear programming problem, this study cannot depend upon an ordinal GP algorithm to solve LAVE. However, this study utilizes the following iterative GP procedure to obtain a solution of (4) because of Theorem 2.2.

Step1: Set $k = 1$. Identify a vector of initial $\beta$ estimates. Let $\hat{\beta}_k$ be such a solution. [Ordinary Least Square and LAVE, available in many statistical softwares such as TSP and SAS/ETS, can procedure the initial $\beta$ estimates.]

Step 2: In order to obtain the $\rho$ estimate, the GP model $\min f(\rho, \hat{\beta}_k)$ is solved and let $\hat{\rho}_k$ be SC estimate of $\min f(\rho, \hat{\beta}_k)$.

Step 3: Using $\hat{\rho}_{k+1}$, the GP model $\min f(\hat{\rho}_{k+1}, \beta)$ is solved to obtain the estimates of $\beta$ and let $\hat{\rho}_{k+1}$ be the estimates obtained from $\min f(\hat{\rho}_{k+1}, \beta)$.

Step 4: If the objective function $f(\rho, \beta)$ satisfies the following condition: $f(\hat{\rho}_k, \hat{\beta}_k) \leq f(\hat{\rho}_{k+1}, \hat{\beta}_{k+1})$, then set $(\rho^*, \beta^*) = (\hat{\rho}_k, \hat{\beta}_k)$ and stop this computational process. Otherwise $k = k + 1$ and go to Step1.

This study provides four important perspectives regarding this LAVE-SC algorithm, each of which can be summarized in the following manner.

**(a)** First, many special computer codes have been developed for LAVE in which unique algorithmic features of $\min_\beta f(\hat{\rho}_{k+1}, \beta)$ are fully utilized to reduce its computational effort and time.

**(b)** Second, when solving LAVE $\min_\rho f(\rho, \hat{\beta}_k)$, we can utilize the following simple approach:

$$\hat{\rho}_{k+1} = \text{median}\left[\frac{y_2 - X_2\hat{\beta}_k}{y_1 - X_1\hat{\beta}_k}, \frac{y_3 - X_3\hat{\beta}_k}{y_2 - X_2\hat{\beta}_k}, \ldots, \frac{y_n - X_n\hat{\beta}_k}{y_{n-1} - X_{n-1}\hat{\beta}_k}\right]$$

**(c)** Third, there is a common assumption related to $\min_\rho f(\rho, \hat{\beta}_k)$ and $\min_\beta f(\hat{\rho}_{k+1}, \beta)$; the problem degeneracy does not occur at a vector or optimal parameter estimates. Here, this article describes the degeneracy problem, using $\min_\beta f(\hat{\rho}_{k+1}, \beta)$. The degeneracy may occur when the parameter vector $\beta$ of $\min_\beta f(\hat{\rho}_{k+1}, \beta)$, composing $m$ unknown parameter estimates, is fitted to a data set in which as least $m$ sample observations are on the regression hyperplane. If more than $m$ sample observations are on the regression hyperplane, the degeneracy may occur. Mathematically, this degeneracy can be defined by the following statement: the problem occurs when optimal $d_t^+$ or $d_t^-$ for some $t$ becomes a basic variable and equals zero in $\min_\beta f(\hat{\rho}_{k+1}, \beta)$. In a dual form of $\min_\beta f(\hat{\rho}_{k+1}, \beta)$, the degenerated point is a sample observation with $w_t = 1$ or $-1$ and it is on the regression hyperplane where $w_t$ is the dual variable of the $i^{th}$ constraint of $\min_\beta f(\hat{\rho}_{k+1}, \beta)$.

**(d)** Finally, the above degeneracy corresponds to the "perfect collinearity", which is not a serious problem in a view of statistics because we can solve it by dropping one of perfectly correlated variables. The multicollinearity is exactly speaking "near-collinearity" that several variables closely locate each other in a data space. A difficulty associated with the near-collinearity is that when it occurs, parameter estimates become very large in those magnitudes and/or produce signs opposite to our expectation. To deal with this type of statistical difficulty, we may incorporate additional constraints, expressing prior information on parameter estimations. We put the following two assumptions:

**(A1)** The parameter spaces $(U)$ is closed.

**(A2)** There is some objective value $s$ such that the solution set $S = \{(\rho, \beta) \mid (\rho, \beta) \in U \text{and } f(\rho, \beta) \leq s\}$. Then, we have the following three properties:

**(P1)** When the algorithm does not terminate, it is observed that $f(\hat{\rho}_k, \hat{\beta}_k) > f(\hat{\rho}_{k+1}, \hat{\beta}_{k+1})$.

**(P2)** It the algorithm terminates, then $\min_\rho f(\rho, \beta) = \min_\beta f(\rho^*, \beta^*) = f(\rho^*, \beta^*)$.

**(P3)** Assume that the algorithm generates an infinite sequence $\{(\hat{\rho}_k, \hat{\beta}_k) \mid k = 1, 2, \ldots\}$, then there exists a limit value $f^*$ such that $\lim_{k\to\infty} f(\hat{\rho}_k, \hat{\beta}_k) = f^*$. If $f^* < s$, then the following condition is maintained: $\min_\rho f(\rho, \hat{\beta}) = \min_\beta f(\hat{\rho}, \beta) = f(\hat{\rho}, \hat{\beta})$.

## 4 Numerical Results

In the presentation, we will illustrate difficulty for solving the LAVE-SC problem from simple examples and show how much our GP approach can improve the statistical efficiency by comparing it with a previous algorithmic effort.

## References


[1] Sueyoshi, T. and Sekitani, K., "Goal Programming Regression with Serial Correlation:Policy Implications for Japanese Telecommunications Infrastructure Development", *OMEGA International Journal of Management Science* 26 (1998) 195-205.

[2] Sueyoshi, T. and Sekitani, K., "Mathematical Properties if Least Absolute Value Estimation with Serial Correlation",*Asia-Pacific Journal of Operational Research* 15 (1998) 72-92

[3] Weiss, A.A.,"Least absolute estimation in the presence of serial correlation",*Journal of Econometrics* 44 (1990) 127-158