V-Robust Design for AR(1) Processes

by

Xiaozhou Wang

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Statistics

Department of Mathematical and Statistical Sciences
University of Alberta

Abstract

Two types of robust designs, robust design against a misspecified response function and robust design against autocorrelation, are reviewed in the first chapter of this dissertation. Among robust designs against autocorrelation, this dissertation focuses on V-robust designs. Chapter 2 reviews and extends the work done by Wiens and Zhou (1997) on the construction of V-robust design for MA(1) processes. On the basis of their work, *most* V-robust design for AR(1) processes is introduced in Chapter 3, with its application on a toy sales example. Compared to *most* V-robust design for MA(1), *most* V-robust design for AR(1) performs better if the number of observations is small and there is some bias in fitted model.

Acknowledgements

My first and deepest gratitude goes to my supervisor, Professor Douglas Wiens for his continuous support and guidance throughout my study. Without his immense knowledge, and patience in correcting my writing, this thesis would not be completed.

In addition to the great supervision from Professor Wiens, I am also thankful to all professors who taught me a lot in the classes. I appreciate everything I learned from this department since I came here as a newbie who barely knew Statistics.

My sincere gratitude goes to my great-grandmother and grandfather who both passed away this year, for their greatest love and support. My sincere gratitude goes to my family, for their love and encouragement that make me who I am today. I am also grateful to my girlfriend, Yao Xie, for her understanding and support throughout the last 7 years.

I would like to thank all my friends who helped me and brought happiness into my study in University of Alberta. Special thanks goes to Wenrui Ye for his help in measure theory.

Last but not least, I would like to thank Professor Edit Gombay, Professor Ivan Mizera and Professor Chris Frei for serving on my dissertation examination committee.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Classical Optimal Design

An experimenter anticipates investigating the effect of some independent vector variable $\mathbf{x}$ on a dependent random variable Y, by making a number of observations of that variable Y corresponding to the vector $\mathbf{x}$. It is a challenge for the experimenter to choose the values of the variable $\mathbf{x}$ to observe, which will result in an optimal design that may provide maximum information at minimum cost. Also, it is important to construct such a design because limited time or cost may not allow large number of observations or repeated experiments in real life. Therefore, an optimal design is desirable for the experimenter.

A simple model is introduced in order to defining the optimal design mathematically. Suppose that there is an exact linear relation between the response variable Y and the explanatory variable $\mathbf{x}$, that is,

$$y = \mathbf{x}^T\theta + \varepsilon, \tag{1.1}$$

where $\mathbf{x} = (1, x)^T$, $\theta = (\theta_0, \theta_1)^T$, and the error terms ε_i are i.i.d with zero mean and variance σ_ε^2.

The experimenter wishes to observe n values of the response variable y corresponding to predictor variables $\mathbf{x}$. A 'design measure', denoted as ξ, will show the experimenter which values of $\mathbf{x}$ to select, with how many repetitions, from a design space S. If the design places n_i out of n observations at the point $\mathbf{x}_i \in S$, then $\xi(\mathbf{x}_i) = \frac{n_i}{n}$. The experimenter is confident about the fitted regression model,

$$E(y_i|\mathbf{x}_i) = \mathbf{x}_i^T\theta. \tag{1.2}$$

The estimate of θ is

$$\hat{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}, \tag{1.3}$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T$, and $\mathbf{Y} = (y_1, y_2, \ldots, y_n)^T$.

And it is unbiased with covariance matrix

$$COV(\hat{\theta}) = \sigma_\varepsilon^2(\mathbf{X}^T\mathbf{X})^{-1}.$$

Since the covariance matrix depends on the data through $(\mathbf{X}^T\mathbf{X})^{-1}$, an optimal design ξ may determine the most efficient estimate. In this example,

$$\begin{aligned} COV(\hat{\theta}) &= \sigma_\varepsilon^2(\mathbf{X}^T\mathbf{X})^{-1} \\ &= \frac{\sigma_\varepsilon^2}{n\sum x_i^2 - n^2\bar{x}^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}. \end{aligned}$$

Thus, if $trace(COV(\hat{\theta}))$ is to be minimized, we have

$$Var(\hat{\theta}_0) = \sigma_\varepsilon^2(1 + \frac{\bar{x}^2}{S_x^2}), \quad Var(\hat{\theta}_1) = \frac{\sigma_\varepsilon^2}{S_x^2},$$

where, $S_x^2 = \frac{\sum (x_i - \bar{x})^2}{n}$.

Assume $S = [-1, 1]$. Both variances are minimized (so is the trace) by putting half of the x's at each of ± 1. That is, the optimal design has

$$\xi(-1) = \xi(1) = \frac{1}{2}.$$

This optimality criterion that minimizes $tr(COV(\hat{\theta}))$ is recognized as A-optimality.

Also, if the determinant of covariance matrix

$$det(COV(\hat{\theta})) = \frac{\sigma_\varepsilon^4}{n \sum x_i^2 - n^2 \bar{x}^2}$$

is to be minimized, which is known as D-optimality, the design is again putting half of the x's at each of ± 1.

There are other kinds of optimality criteria, such as E-optimality and G-optimality. Further and detailed discussion can be found in Kiefer (1959), Chernoff (1953), Ehrenfeld (1955), etc.

1.2 Robust Design Against a Misspecified Response Function

The development of this section follows that in Wiens (*Robustness of Design* - an upcoming Handbook chapter). The classical optimal design assume the proposed model is exactly correct, which is not the case mostly. Then the 'best' design for even a slightly wrong model can be much more than slightly sub-optimal. As the above example shows, the optimal design may put all mass at the extreme points ± 1, and there is no other information about the points within the design space. It is impossible to detect any curvature using only the boundary points, and the fitted model can be far away from the true one. Hence, it is necessary to study more appropriate designs, robust designs, which can be applied for this situation.

Suppose the true mean response is $E(y|x) = \phi_0 + \phi_1 x + \phi_2 x^2$ instead of (1.1), with uncorrelated, equal varied errors. We still fit the regression model and the least squares estimate is (1.3).

Define

$$\tau_k = \sum \frac{x_i^k}{n},$$

and assume that $\tau_1 = \tau_3 = 0$ (for instance, the design is symmetric).

Then it can be found that, under the true quadratic model,

$$E(\hat{\theta}) = \begin{pmatrix} \phi_0 + \tau_2 \phi_2 \\ \phi_1 \end{pmatrix},$$

and

$$COV(\hat{\theta}) = \frac{\sigma_\varepsilon^2}{n} \begin{pmatrix} 1 & 0 \\ 0 & \phi_2^{-1} \end{pmatrix},$$

so that each prediction $\hat{y}(x) = \theta_0 + \theta_1 x$ has *mse*:

$$MSE\left[\hat{y}(x)\right] = E\left[\{\hat{y}(x) - E[y(x)]\}^2\right] = \frac{\sigma_\varepsilon^2}{n}\left(1 + \frac{x^2}{\tau_2}\right) + (\phi_2(\tau_2 - x^2))^2.$$

A common measure of performance is the IMSE, integrated mean squared error of the predictors, which in this example is

$$IMSE = \int_{-1}^{1} MSE[\hat{y}(x)]dx$$

$$= \left\{\frac{2\sigma_\varepsilon^2}{n}\left(1 + \frac{1}{3\tau_2}\right)\right\} + \left\{2\phi_2^2\left(\left(\tau_2 - \frac{1}{3}\right)^2 + \frac{4}{45}\right)\right\}. \qquad (1.4)$$

The first term is the integrated variance, and is minimized by the design with half of the observations at each of $x = \pm 1$. The second term is the integrated bias which can be minimized if $\tau_2 = \frac{1}{3}$. More detailed discussion can be found in Box and Draper (1959), a seminal work in this area. And the conclusion drawn by them is that " .....the optimal design in typical situations in which both variance and bias occur is very nearly the same as would be obtained if *variance were ignored completely* and the experiment designed so as to *minimize bias alone*. "

To obtain *IMSE* for more general cases, it is necessary to introduce some notation first.

Let the fitted model be

$$E[Y(\mathbf{x})] = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}, \tag{1.5}$$

where $\mathbf{z}(\mathbf{x})$ is a p-vector with each element being a function of q-vector $\mathbf{x} = (x_1, \ldots, x_q)$, and $\mathbf{x}$ is chosen from a design space $\mathcal{S}$. A precise model is also defined as

$$E[Y(\mathbf{x})] = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta} + f(\mathbf{x}), \tag{1.6}$$

for some unknown function $f(x)$. But now the parameter $\boldsymbol{\theta}$ is not identifiable, since one might equally well write

$$E[Y(\mathbf{x})] = \mathbf{z}^T(\mathbf{x})(\boldsymbol{\theta} + \boldsymbol{\phi}) + (f(\mathbf{x}) - \mathbf{z}^T(\mathbf{x})\boldsymbol{\phi}), \tag{1.7}$$

for arbitrary $\boldsymbol{\phi}$.

To avoid this problem, we can define $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta} = \arg\, min_\eta \int_{\mathcal{S}} \left(E\left[Y(\mathbf{x})\right] - \mathbf{z}^T(\mathbf{x})\eta\right)^2 d\mathbf{x}. \tag{1.8}$$

Taking derivative of (1.8),

$$2\int_{\mathcal{S}} \mathbf{z}(\mathbf{x}) \left(E(Y(\mathbf{x}) - \mathbf{z}^T(\mathbf{x}\boldsymbol{\theta}))\right) d\mathbf{x} = 0 \tag{1.9}$$

leading to the orthogonality requirement

$$\int_{\mathcal{S}} \mathbf{z}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = 0. \tag{1.10}$$

Recall that ξ is a design measure, and $\xi_i = \xi(\mathbf{x}_i) = n_i/n$ if n_i of the n

observations are to be made at $\mathbf{x}_i$.

Define

$$\mathbf{B} = \sum \frac{n_i}{n} \mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i) = \sum \xi_i \mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i) = E_\xi[\mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})], \tag{1.11}$$

and

$$\mathbf{b}(f, \xi) = \sum \frac{n_i}{n} \mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i) = \sum \xi_i \mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i) = E_\xi[\mathbf{z}(\mathbf{x}) f(\mathbf{x})]. \tag{1.12}$$

Then,

$$\begin{aligned} COV(\hat{\boldsymbol{\theta}}) &= \sigma_\varepsilon^2 (\mathbf{Z}^T\mathbf{Z})^{-1} \\ &= \sigma_\varepsilon^2 (n \sum \frac{n_i}{n} \mathbf{z}(\mathbf{x}_i)\mathbf{z}^T(\mathbf{x}_i))^{-1} \\ &= \frac{\sigma_\varepsilon^2}{n} \mathbf{B}^{-1}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{bias} &= E\left((\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Y}\right) - \boldsymbol{\theta} \\ &= E\left((\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T(\mathbf{Z}\boldsymbol{\theta} + \mathbf{f} + \boldsymbol{\varepsilon})\right) - \boldsymbol{\theta} \\ &= n(\mathbf{Z}^T\mathbf{Z})^{-1} \sum \frac{n_i}{n} \mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i) \\ &= \mathbf{B}^{-1}\mathbf{b}(f, \xi). \end{aligned}$$

Therefore, the Mean Squared Error matrix of $\hat{\boldsymbol{\theta}}$ can be written as

$$\begin{aligned} MSE &= E\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^T\right] \\ &= COV\left(\hat{\boldsymbol{\theta}}\right)+\mathbf{bias}\cdot\mathbf{bias}^T \\ &= \frac{\sigma_\varepsilon^2}{n}\mathbf{B}^{-1}+\mathbf{B}^{-1}\mathbf{b}(f,\xi)\mathbf{b}^T(f,\xi)\mathbf{B}^{-1}. \end{aligned}$$

If we define $\mathbf{A}$ as

$$\mathbf{A}=\int_{\mathcal{S}}\mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})d\mathbf{x},$$

then finally IMSE can be found to be

$$\begin{aligned} IMSE &= \int_{\mathcal{S}} MSE\left[\hat{Y}(\mathbf{x})\right]d\mathbf{x} \\ &= \frac{\sigma_\varepsilon^2}{n}\text{trace}\left(\mathbf{A}\mathbf{B}^{-1}\right)+\mathbf{b}^T(f,\xi)\mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1}\mathbf{b}(f,\xi)+\int_{\mathcal{S}}f^2(\mathbf{x})d\mathbf{x}. \end{aligned}$$

In addition, the following condition on the function $f(\mathbf{x})$ was used by Huber (1975)

$$\int_{\mathcal{S}}f^2(\mathbf{x})d\mathbf{x}\leq\frac{\tau^2}{n}, \tag{1.13}$$

for some given constant τ^2; this class was generalized by Wiens (1992).

To achieve robustness, we maximize IMSE over all f satisfying (1.10) and (1.13), and then find a design minimizing this maximum loss. Namely, we construct a *minimax* design.

1.3 Robust Design Against Autocorrelation

In addition to seeking robust designs against a misspecified response function, it is also valuable to obtain robustness against a misspecified error structure. For instance, experimental observations may be autocorrelated because they are gathered serially. This type of problem has been investigated by Bickel and Herzberg (1979), where they developed an asymptotic theory for studying the effect of dependence of the observations for linear regression model. Specifically it is assumed that for N observations, the correlation function is given by

$$\rho_N(t) = \rho_1(Nt),$$

where, $\rho_1(t) \to 0$ as $t \to \infty$. They proved the continuous uniform design is asymptotically optimal for estimating location, and asymptotically optimal for estimating slope under certain conditions. Bickel, Herzberg and Schilling (1981) showed numerical results corresponding to the asymptotic theory.

Wiens and Zhou (1996) studied minimax regression designs for approximately linear models with autocorrelated errors. They assume that the spectral density $g(\omega)$ of the error process is of the form

$$g(\omega) = (1-\alpha)g_0(\omega) + \alpha g_1(\omega),$$

where $g_0(\omega)$ is uniform, $\alpha \in [0, 1)$ is fixed, and $g_1(\omega)$ is arbitrary. The main results in the paper are that "*a design which is asymptotically optimal for uncorrelated errors retains its optimality under autocorrelation if the design points are a random sample, or a random permutation, of points from this distribution*". Wiens and Zhou (1999) also introduced minimax designs for

approximately linear models with AR(1) errors, where they found an asymptotically optimal procedure for AR(1) models consists of selecting points from the design measure which is optimal for uncorrelated errors, and then implementing them in an appropriate order.

Zhou (2001) proposed a criterion to minimize the absolute value of the change of the variance function, over all possible run orders, due to possible correlation between the observations. Specifically, the criterion starting with a design that is exact optimal, efficient, or robust against departures in the regression response, and then find a permutation of the design points as a solution of

$$min_{\xi}|CVF_a(\xi, \mathbf{P})|,$$

with respect to a certain autocorrelation matrix $\mathbf{P}$, where the minimum is over all permutations of the order of the design points in ξ. And

$$CVF_a(\xi, \mathbf{P}) = \frac{\partial \mathbf{a^T V}((\mathbf{1}-\mathbf{t})\mathbf{I}+\mathbf{tP})\mathbf{a})/\partial \mathbf{t}|_{\mathbf{t=0}}}{\mathbf{a^T V(I)a}},$$

where ξ denotes the design points $(x_1, ..., x_n)$, a is a constant vector, and their ordering and $\partial/\partial t$ denotes the partial derivative with respect to t.

Particular attention in this dissertation is paid to models as at equations (1.5) - (1.10), with correlation structures expressed through autocorrelation matrices

$$\mathbf{P}_\gamma = (1-\gamma)\mathbf{I} + \gamma\mathbf{P}, \tag{1.14}$$

for some unknown autocorrelation matrix $\mathbf{P}$.

1.4 Illustation: Toy Sales Example

Suppose one toy company would like to predict sales amount with respect to price. For example, there is toy A with price 36 dollars each. The minimum sale price may be 18 dollars (so that the company will not be broken). The company anticipates predicted sales based on the linear relationship between sales and price when price drops from $36 to $18.

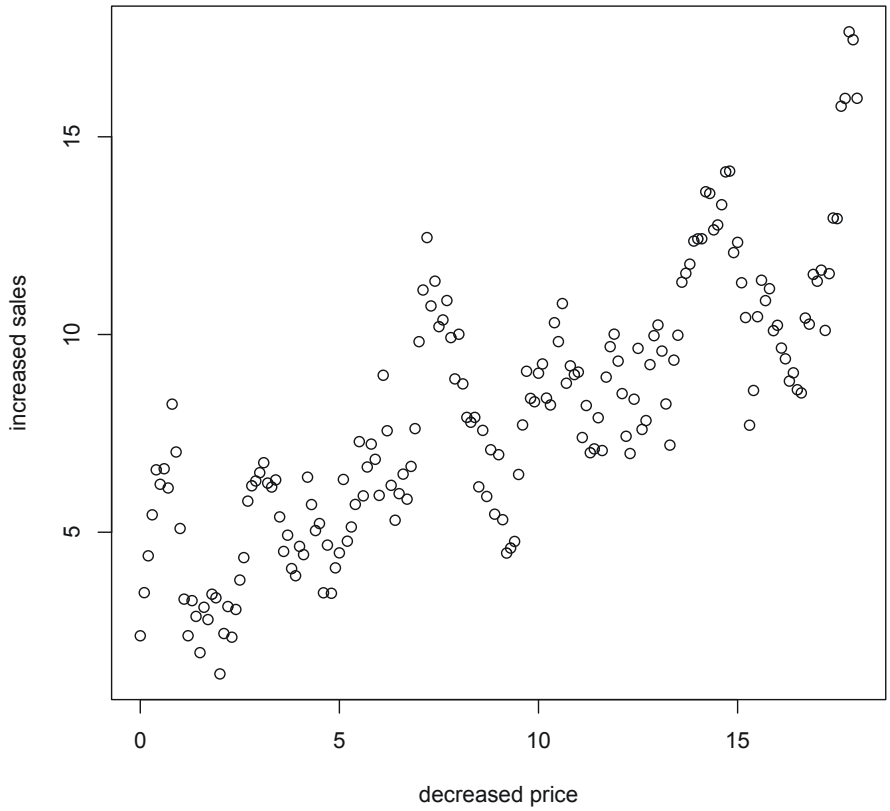

Figure 1.1: Relation between sales and prices

In order to find the model, the company may set a price in one period, record the sales amount at the end of the period, then set another price and

record at the end of that period, etc. It is helpful to look for some experimental design which would give a good prediction. It is obvious that sales amount would be also affected by other factors, such as market demand, toy quality, brand effect, etc. The effect of these factors on previous sales will affect current sales in a very similar way, providing there is no dramatic change in the market or the company. Therefore, it seems reasonable to assume autocorrelated error for this sale-price model.

Let us assume the true model between increased sales and decreased prices is

$$y_t = 0.6x_t + 3 + \varepsilon_t,$$

where y is increased sales amount with 10,000 per unit, x is price decreased from \$36; and $\varepsilon_t = e_{1,t} + e_{2,t}$, where $\{e_{1,t}\}$ follows AR(1) with $\rho = 0.9, \sigma = 1$, and $\{e_{2,t}\}$ follows i.i.d uniform distribution $U(-0.5, 0.5)$.

The sample data may look like Figure 1.1. It may be a good example to apply V-robust designs to, which we will revisit in later section.

Chapter 2

V-Robust Designs for MA(1) Processes

2.1 V-Robust Design and Most V-Robust Designs for MA(1)

This section illustrates and extends the work done by Wiens and Zhou (1997). Suppose a random variable Y is observed at locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in a q-dimensional space $\mathcal{S}$. Specifically, assume the response $E[Y|\mathbf{x}]$ is approximately linear in the parameters, with regressors $p \times 1$ vectors $\mathbf{z}(\mathbf{x})$ and is observed subject to possibly serially correlated errors. With $f(\mathbf{x}) = E[Y|\mathbf{x}] - \mathbf{z^T}(\mathbf{x})\theta$, the observations satisfy

$$Y_i = \mathbf{z}^T(\mathbf{x}_i)\theta + f(\mathbf{x}_i) + \varepsilon_i, \qquad i = 1, \ldots, n \tag{2.1}$$

where θ is a $p \times 1$ vector with unknown parameters, and ε_i's are zero-mean random errors, which are autocorrelated with covariance matrix $\sigma^2\mathbf{P}$ for some

autocorrelation matrix $\mathbf{P}$.

Let ξ be the design measure, and define $\mathbf{b}_{f,\xi}$ same as (1.12)

$$\mathbf{b}_{f,\xi} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i),$$

then define

$$\mathbf{B}_{\xi}(m) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n-m} \mathbf{z}(\mathbf{x_i})\mathbf{z^T}(\mathbf{x_{i+m}}) & 0 \le m \le n-1, \\ \mathbf{B}_{\xi}^{T}(-m) & -(n-1) \le m < 0. \end{cases} \tag{2.2}$$

Denote the $n \times p$ model matrix with rows $\mathbf{z^T}(\mathbf{x}_i)$ by $\mathbf{Z}$. Then the determinant of the MSE matrix of $n^{1/2}\hat{\theta}$ is

$$\mathcal{D}(f, \xi, \mathbf{P}) = \sigma^{2p} \left|\mathbf{B}_{\xi}(0)\right|^{-2} \left|\frac{\mathbf{Z}^T\mathbf{P}\mathbf{Z}}{n}\right| \times \left(1 + \frac{n}{\sigma^2}\mathbf{b}_{f,\xi}^{T}\left(\frac{\mathbf{Z}^T\mathbf{P}\mathbf{Z}}{n}\right)^{-1}\mathbf{b}_{f,\xi}\right).$$

Let $\mathcal{P}$ be a convex class of autocorrelation matrices $\mathbf{P}$ containing $\mathbf{P}_0 = \mathbf{I}$.

Wiens and Zhou (1997) considered the case that $\mathbf{P}$ belongs to MA(1) processes. They define the *change-of-variance function* (CVF) for ξ at $\mathbf{P}_0$, in the direction $\mathbf{P} \in \mathcal{P}$, by

$$CVF(\xi, \mathbf{P}) = \frac{\frac{d}{dt}\mathcal{D}(f_0, \xi, (1-t)\mathbf{P_0} + t\mathbf{P})_{|t=0}}{\mathcal{D}(f_0, \xi, \mathbf{P_0})}. \tag{2.3}$$

CVF corresponds to the derivative of $\ln \mathcal{D}$, that is, basing the loss on $\ln \mathcal{D}$. The suprema of CVF over $\mathcal{P}$ is defined as *change-of-variance sensitivity* (CVS). The main motivation was to investigate the *local robustness* of the asymptotic variance. The framework of CVF with all details can be found in Hampel et al. (1986).

Denote $\mathbf{P}_{i,j}$ by $\rho(|i-j|)$ for some autocorrelation function ρ, and straightforward calculations give

$$\begin{aligned} CVS(\xi, \mathcal{P}) &= \sup_{\mathbf{P}\in\mathcal{P}} \left\{ trace\left(\frac{\mathbf{Z}^T(\mathbf{P}-\mathbf{I})\mathbf{Z}}{n} \mathbf{B}_\xi^{-1}(0) \right) \right\} \\ &= \sup_{\mathbf{P}\in\mathcal{P}} \left\{ \sum_{0\le|s|\le n-1} \rho(s) trace(\mathbf{B}_\xi(s)\mathbf{B}_\xi^{-1}(0)) \right\}. \end{aligned} \tag{2.4}$$

For a given α, we say that a design ξ is V robust if it minimizes $\mathcal{D}(f_0, \xi, \mathbf{P}_0)$; that is, maximizes $|\mathbf{B}_\xi(0)|$ subject to the constraint

$$CVS(\xi, \mathcal{P}) \le \alpha, \tag{2.5}$$

and is *most V robust* if α is the infimum of the CVS over a given class of designs.

In other words, the *most* V-robust design is the design which minimizes, over a given class of designs, the maximum value of CVF over some class of autocorrelation structures.

Wiens and Zhou (1997) obtained V-robust and most V-robust designs for the classes

$$\mathcal{P}_1 = \{\mathbf{P} | \rho(s) = 0 \text{ for } |s| \ge 2; c_0 \le \rho(1) < 1 \text{ with } c_0 > 0\};$$

$$\mathcal{P}_2 = \{\mathbf{P} | \rho(s) = 0 \text{ for } |s| \ge 2; -1 < \rho(1) \le -c_1 \text{ with } c_1 > 0\}.$$

These classes correspond to MA(1) processes with positive and negative lag-1 correlations bounded away from 0. They consider the multiple linear regression model which is restricted to the class $\mathscr{X}_{n,q}$ of n-point designs for

which $\mathbf{B}_\xi(0)$ is a diagonal matrix.

Using (2.4), they obtain that

$$CVS(\xi, \mathcal{P}) = \sup_{\rho(1)} 2\rho(1) \left(\frac{n-1}{n} + \sum_{j=1}^{q} \mathcal{Q}(\mathbf{x}_{(j)}) \right),$$

where $\mathbf{x}_{(j)}$ is the $j+1$th column of $\mathbf{Z}$,

$$\mathcal{Q}(\mathbf{x}) := \frac{\sum_{i=1}^{n-1} x_i x_{i+1}}{\sum_{i=1}^{n} x_i^2} = \frac{\mathbf{x^T Q x}}{\mathbf{x^T x}}, \tag{2.6}$$

and $\mathbf{Q}$ is the tridiagonal matrix with (i,j)th element $q_{ij} = \frac{1}{2} I(|i-j| = 1)$. The theorem below shows that if α is sufficiently large then (2.5) imposes no restriction.

Theorem 2.1: *(Wiens and Zhou 1997) Let $q = 1$ and $\mathcal{S} = [-1, 1]$. Define*

$$\alpha_{1,n} = \begin{cases} 0, & n \text{ even} \\ \frac{2}{n(n-1)}, & n \text{ odd} \end{cases} \tag{2.7}$$

and

$$\alpha_{2,n} = \begin{cases} -4c_1 \frac{n-2}{n}, & n \text{ even} \\ -2c_1 \frac{2n^2-5n+1}{n(n-1)}, & n \text{ odd} \end{cases} \tag{2.8}$$

- *For $\alpha \geq \alpha_{1,n}$, V-robust designs for $\mathcal{P}_1$ are*

$$\mathbf{x}_{(1)} = \begin{cases} < 1, -1, 1, -1, \ldots, 1, -1 > & n \text{ even} \\ < 1, -1, 1, -1, \ldots, 1, -1, 0 > & n \text{ odd} \end{cases}$$

with $CVS(\xi, \mathcal{P}_1) = \alpha_{1,n}$.

- *For $\alpha \geq \alpha_{2,n}$, V-robust designs for $\mathcal{P}_2$ are*

$$\mathbf{x}_{(1)} = \begin{cases} <\underbrace{1,\ldots,1}_{n/2},\underbrace{-1,\ldots,-1}_{n/2}> & n \text{ even} \\ <\underbrace{1,\ldots,1}_{n/2},0,\underbrace{-1,\ldots,-1}_{n/2}> & n \text{ odd} \end{cases}$$

with $CVS(\xi, \mathcal{P}_1) = \alpha_{2,n}$.

It needs some definitions to get the *most* V-robust designs for multiple regression.

Define constants and $n \times 1$ vectors $(\mu_j, \mathbf{r}_j), 1 \leq j \leq n$, and $(\nu_j, \mathbf{s}_j), 1 \leq j \leq [(n-1)/2]$, with the μ_j and ν_j ordered from largest to smallest, by

$$\mu_j = \cos\frac{j\pi}{n+1}, \qquad (\mathbf{r}_j)_k = \sqrt{2/(n+1)}\sin\frac{kj\pi}{n+1}, \tag{2.9}$$

and

$$\nu_j = \cos\phi_j, \qquad (\mathbf{s}_j)_k = \sqrt{\frac{2(n+1)}{n(n+2)}}\cot\frac{n+1}{2}\phi_j\left(1 - \frac{\cos(k-\frac{n+1}{2})\phi_j}{\cos\frac{n+1}{2}\phi_j}\right), \tag{2.10}$$

where ϕ_j is the solution, in $(2j\pi/(n+1), (2j+1)\pi/(n+1))$, to the equation

$$\tan\frac{n+1}{2}\phi - (n+1)\tan\frac{\phi}{2} = 0. \tag{2.11}$$

Place the $\mathbf{r}_{2j}$ and the $\mathbf{s}_j$ into a matrix $\mathbf{X}$, and define a corresponding sequence $\{\lambda_j\}$ by

$$\mathbf{X}_{n\times(n-1)} = (\mathbf{r}_2, \mathbf{s}_1, \mathbf{r}_4, \mathbf{s}_2, \ldots, \mathbf{r}_{2[(n-1)/2]}, \mathbf{s}_{[(n-1)/2]}, \mathbf{r}_n) \quad \text{(if n is even)}$$

and

$$\{\lambda_j\}_{j=1}^{n-1} =< \mu_2, \nu_1, \mu_4, \nu_2, \ldots, \mu_{2[(n-1)/2]}, \nu_{[(n-1)/2]}, \mu_n > \quad \text{(if n is even)}. \tag{2.12}$$

Theorem 2.2: *(Wiens and Zhou 1997) The most V-robust designs in $\mathscr{Z}_{n,q}$ for $\mathcal{P}_k, k = 1, 2$, have model matrices $\mathbf{Z} = (\mathbf{1}\vdots\mathbf{X}_{(q;k)})$, where $\mathbf{X}_{(q;k)}$ consist of the last $(k = 1)$ or first $(k = 2)$ q columns of $\mathbf{X}$ and $\mathbf{D}_k$ is a diagonal matrix chosen to have maximum determinant, subject to the constraint that the rows of $\mathbf{X_{(q;k)}D_k}$ belong to $\mathcal{S}$. The corresponding covariance matrices of $\hat{\theta}$ at $\mathbf{P}_0 = \mathbf{I}$ are $\sigma^2(\mathbf{Z^TZ})^{-1} = \sigma^2(n^{-1} \oplus \mathbf{D}_k^{-2})$. The CVS are*

$$CVS(\xi, \mathcal{P}_k) = \begin{cases} 2\left(\frac{n-1}{n} + \sum_{j=1}^{q} \lambda_{n-j}\right), & k = 1 \\ -2c_1\left(\frac{n-1}{n} + \sum_{j=1}^{q} \lambda_j\right), & k = 2 \end{cases}$$

Proof. This is a sketch of the proof, which will be referred to later on. The complete proof can be found in Wiens and Zhou (1997).

There are two claims before the formal proof.

Claim 1. The matrix $\mathbf{Q}$ has eigenvalues μ_j and corresponding orthonormal eigenvectors $\mathbf{r}_j$, given by (2.9). The eigenvectors $\mathbf{r}_j$ are orthogonal to $\mathbf{1} = (1, \ldots, 1)^T$ iff j is even.

Claim 2. $\mathbf{X}^T\mathbf{X} = \mathbf{I}_{n-1}$ and

$$\min_{Z_{n,q}} \sum_{j=1}^{q} \mathcal{Q}(\mathbf{x}_{(j)}) = \sum_{j=1}^{q} \lambda_{n-j}, \qquad \max_{Z_{n,q}} \sum_{j=1}^{q} \mathcal{Q}(\mathbf{x}_{(j)}) = \sum_{j=1}^{q} \lambda_j.$$

These extrema are attained at arbitrary nonzero multiples of the first q and last q columns of $\mathbf{X}$.

Now, if $\xi \in \mathscr{Z}_{n,q}$ has model matrix $\mathbf{Z} = (\mathbf{1}\vdots\mathbf{X}_0)$, then for $\mathcal{P}_1$ we are to

maximize $|\mathbf{Z}^T\mathbf{Z}| = n|\mathbf{X}_0^T\mathbf{X}_0|$, subject to $\sum_{j=1}^q \mathcal{Q}(\mathbf{x}_{(j)})$ being a minimum. But by Claim 2, any $\mathbf{X}_0$ whose columns minimize $\sum_{j=1}^q \mathcal{Q}(\mathbf{x}_{(j)})$ must be of the form $\mathbf{X}_{(q;1)}\mathbf{D}_1$ for a diagonal matrix $\mathbf{D}_1$, and then $|\mathbf{X}_0^T\mathbf{X}_0| = |\mathbf{D}_1|^2$ is maximized by maximizing $|\mathbf{D}_1|$. The proof for $\mathcal{P}_2$ is entirely analogous. □

Proof of Claim 2. (Wiens and Zhou 1997)

Proof. Straightforward calculation gives

$$\max_{Z_{n,q}} \sum_{j=1}^{q} \mathcal{Q}(\mathbf{x}_{(j)}) = \max\left\{\text{trace } \mathbf{Q}\mathbf{X}\mathbf{X}^T \mid \mathbf{X}^T\mathbf{X} = \mathbf{I}_q, \mathbf{X}^T\mathbf{1} = \mathbf{0}\right\},$$

where $\mathbf{X}$ has columns $\{\mathbf{x}_{(j)}\}_{i=1}^q$. If $\mathbf{J}_{n\times n-1}$ satisfies $\mathbf{J}^T\mathbf{J} = \mathbf{I}_{n-1}$ and $\mathbf{J}\mathbf{J}^T = \mathbf{I}_n - (\frac{1}{n})\mathbf{1}\mathbf{1}^T$, then the conditions on $\mathbf{X}$ are equivalent to " $\mathbf{X} = \mathbf{J}\mathbf{H}$ for some $\mathbf{H}_{n-1\times q}$ with $\mathbf{H}^T\mathbf{H} = \mathbf{I}_q$". With $\mathbf{R} = \mathbf{J}^T\mathbf{Q}\mathbf{J}$, the desired maximum is thus $\max\{\text{trace } \mathbf{R}\mathbf{H}\mathbf{H}^T \mid \mathbf{H}^T\mathbf{H} = \mathbf{I}_q\}$. Theorem 1.10.2 of Srivastava and Khatri (1979) states that this maximum is $\sum_{j=1}^q \lambda_j$, and similarly the minimum is $\sum_{j=1}^q \lambda_{n-j}$, where $\lambda_1 \geq \ldots \geq \lambda_{n-1}$ are the eigenvalues of $\mathbf{R}$. These extrema are attained if $\mathbf{H}$ consists of the q corresponding orthonormalized eigenvectors of $\mathbf{R}$, and then $\mathbf{x}_{(j)}$ is (any nonzero multiple of) $\mathbf{J}\mathbf{h}_{(j)}$.

The eigenvectors of $\mathbf{R}$ with roots $\lambda \neq 0$ are of the form $\mathbf{h} = \mathbf{J}^T\mathbf{z}$, where $\mathbf{z}$ is an eigenvector of $\widetilde{\mathbf{R}} := \mathbf{J}\mathbf{J}^T\mathbf{Q}$ with root λ. There is an extraneous eigenvector of $\widetilde{\mathbf{R}}$ with root 0, which is useless to us. If n is odd, then $\mathbf{Q}$ has an eigenvector $\mathbf{r}_{(n+1)/2}$ with root $\lambda = 0$; this provides an additional eigenvector of $\widetilde{\mathbf{R}}$.

The equations $\widetilde{\mathbf{R}}\mathbf{z} = \lambda\mathbf{z}$ may be written as

$$(\mathbf{Q} - \lambda\mathbf{I})\mathbf{z} = c\mathbf{1}, \qquad c = \mathbf{1}^T\mathbf{Q}\mathbf{z}/n. \tag{2.13}$$

Premultiplying by $\mathbf{1}^T$ gives $\lambda\mathbf{1}^T\mathbf{z} = 0$, so that if $\lambda \neq 0$, we have $\mathbf{z} = \mathbf{J}\mathbf{J}^T\mathbf{z} = \mathbf{J}\mathbf{h}$.

Thus the set $\mathcal{X}$ of unordered vectors $\mathbf{x}_{(j)}$ consists of the eigenvectors $\mathbf{z}$ of $\widetilde{\mathbf{R}}$ corresponding to nonzero eigenvalues, plus possibly a vector arising from $\mathbf{r}_{(n+1)/2}$.

Case 1. If $c = 0$ in (2.13), then $\mathbf{z}$ is a eigenvector of $\mathbf{Q}$. By Claim 1, there are $[n/2]$ such vectors that are orthogonal to $\mathbf{1}$. These include $\mathbf{r}_{(n+1)/2}$ iff $n+1$ is odd and $(n+1)/2$ is even. Thus this case contributes the vectors $\mathbf{r}_{2j},\ j = 1, \ldots, [n/2]$ to $\mathcal{X}$, with corresponding roots μ_{2j}.

Case 2. Let $c \neq 0,\ \lambda \neq 0$ and assume that λ is not a eigenvalue of $\mathbf{Q}$. Then the first equation in (2.13) gives $\mathbf{z} = c(\mathbf{Q} - \lambda\mathbf{I})^{-1}\mathbf{1}$, and the second yields

$$\mathbf{1}^T(\mathbf{Q} - \lambda\mathbf{I})^{-1}\mathbf{1} = 0. \tag{2.14}$$

Writing (2.14) as

$$\sum_{j=1}^{[(n+1)/2]} (\mathbf{1}^T\mathbf{r}_{2j-1})^2/(\mu_{2j-1} - \lambda) = 0$$

shows that there are $[(n-1)/2]$ solutions $\lambda = \nu_j$ which when ordered satisfy $\mu_{2j-1} > \nu_j > \mu_{2j+1}$. The remaining elements of $\mathcal{X}$ are then (multiples of) the vectors $\mathbf{s}_j = (\mathbf{Q} - \nu_j\mathbf{I})^{-1}\mathbf{1}$.

Of the n equations given by $(\mathbf{Q} - \lambda\mathbf{I})\mathbf{s} = \mathbf{1}$, $n-2$ are of the form

$$(v_{k+1} - \frac{1}{1-\lambda}) = 2\lambda(v_k - \frac{1}{1-\lambda}) - (v_{k-1} - \frac{1}{1-\lambda}), \quad k = 2, \ldots, n-1.$$

This recursion, when solved and combined with the remaining two equations and with (2.14), yields

$$v_k = \frac{\left(1 - \frac{cos(k - \frac{n+1}{2})\phi}{cos\frac{n+1}{2}\phi}\right)}{2\sin^2\frac{\phi}{2}},$$

where $\phi = \cos^{-1}\lambda$ satisfies (2.11). From this, we calculate

$$\sum_{k=1}^{n} v_k^2 = n(n+1)(n+2)/(2sin^2\phi),$$

whence normalizing $\mathbf{s}$ to have unit length gives (2.10). It then remains only to establish that the terms in (2.12) are in decreasing order. Because both μ_{2j} and ν_j are in (μ_{2j+1}, μ_{2j-1}), we require $\mu_{2j} > \nu_j$. This follows from the observation that the function on the left of (2.11) is strictly increasing where it is nonnegative and is negative at $\cos^{-1}\mu_{2j}$. □

2.2 Other V-Robust Designs for MA(1) Processes

The V-robust design problem is to maximize $|B_\xi(0)|$ subject to the constraint

$$CVS\,(\xi,\ \mathcal{P}) \leq \alpha,$$

and is most V robust if α is the infimum of the CVS over a given class of designs.

For convenience, the value of CVS is denoted as α_{mv} if the design is the most V-robust design.

Wiens and Zhou (1997) have already obtained V-robust designs if α is sufficiently large, and the most V-robust designs for MA(1) processes. However, there may be other situations where α does not have to be sufficiently large nor attain the infimum. For example, the experimenter may anticipate robustness against bias as well as autocorrelation, which may cause α to be greater than the infimum. So this section takes a quick glance in V-robust designs for the α between $\alpha_{1,n}$ (or $\alpha_{2,n}$) and infimum. The results are stated for $q = 1$ (i.e. simple regression) and finite design space $\mathcal{S} = [-1, -i_1, -i_2, \ldots, -i_N, 0, i_N, \ldots, i_2, i_1, 1]$, for some N and $i_j \in (0, 1)$ for all j.

Now, the problem becomes

$$\max_{x_i \in S} \sum_{i=1}^{n} x_i^2 \qquad s.t. \quad CVS\,(\xi,\ \mathcal{P}) \leq \alpha,$$

where x_i are the observations to be made and

$$CVS\,(\xi,\ \mathcal{P}) = \sup_{\rho(1)} 2\rho(1)\left(\frac{n-1}{n} + \frac{\sum_{i=1}^{n-1} x_i x_{i+1}}{\sum_{i=1}^{n} x_i^2}\right).$$

Since α is less than $\alpha_{1,n}$ or $\alpha_{2,n}$ given in the Theorem 2.1, the only choice is to observe at other points rather than boundary points 1 or -1 to make CVS smaller.

In fact, V-robust designs have observations forced in the subset of the finite design space, $\mathcal{S}_0 = (1, -1, 0)$, while *most* V-robust designs have observations that can be in the entire design space $\mathcal{S}$. Hence, obtaining V-robust designs for some given α between $\alpha_{1,n}$ (or $\alpha_{2,n}$) and α_{mv} can be achieved by finding designs with CVS as small as possible, $\sum_{i=1}^{n} x_i^2$ as large as possible, where points are forced to be observed in some design space $\mathcal{S}'$ where $\mathcal{S}_0 \subset \mathcal{S}' \subset \mathcal{S}$. The cases discussed below will focus on finding such designs, and then the corresponding V-robust designs will be obtained.

CASE 1. $\rho \in \mathcal{P}_1$, $\mathcal{S}_1 = (1, -1, i_1, -i_1, 0)$.

The easiest case is that $i_1, -i_1$ are to be observed only once, and n is even.

The designs can be

$$< i_1, -1, 1, -1, \ldots, -1, 1, -i_1 >,$$

$$< 1, -1, i_1, -1, \ldots, 1, -1, 1, -i_1, 1, -1 >,$$

$$< i_1, -i_1, 1, -1, \ldots, -1, 1, -1 >,$$

...

It is tedious to list all designs and find the one minimizing CVS. In fact, to minimize

$$CVS = \frac{\sum_{i=1}^{n-1} x_i x_{i+1}}{\sum_{i=1}^{n} x_i^2} + \frac{n-1}{n},$$

while $\sum_{i=1}^{n} x_i^2$ is unchanged, is equivalent to minimizing $\sum_{i=1}^{n-1} x_i x_{i-1}$. And in the $n-1$ terms of $\sum_{i=1}^{n-1} x_i x_{i-1}$, the less $-i_1$ and $-i_1^2$ appear, the smaller $\sum_{i=1}^{n-1} x_i x_{i-1}$ is.

Denote $T = (A_1, A_2)$, where A_1 represents the number of term $-i_1$, and A_2 represents the number of term $-i_1^2$ in $\sum_{i=1}^{n-1} x_i x_{i+1}$. So, $T = (2,0), (4,0), (1,1)$ for the 3 designs above, respectively. Since A_1, A_2 the smaller the better, $T = (2,0)$ is better than $T = (4,0)$. And by direct calculation, $T = (2,0)$ is better than $T = (1,1)$. All other design's T would either bigger than $T = (1,1)$ or $T = (2,0)$. That means, nothing can beat $T = (2,0)$. So,

$$< i_1, -1, 1, -1, 1, -1, \ldots, 1, -1, 1, -i_1 >$$

is the 'best' design among them.

Now consider making two observations at i_1 and two at $-i_1$.

The only design that with $A_1 = 1$ is,

$$< i_1, -i_1, i_1, -i_1, 1, -1, \ldots, 1, -1 >$$

So, $T = (1,3)$.

For those designs with $A_1 = 2$, they have the same $T = (2,2)$. And $T = (2,2)$ is better than $T = (1,3)$. For those designs with $A_2 = 1$, the best design has $T = (4,1)$. And $T = (2,2)$ is better than $T = (4,1)$.

Therefore, the design with $T = (2,2)$ is the 'best'. And one of these designs

is

$$< i_1, -i_1, 1, -1, 1, \ldots, -1, i_1, -i_1 > .$$

The rule can be applied for more than 2 observations at i_1, the 'best' design is

$$< i_1, -i_1, \ldots, i_1, -i_1, 1, -1, 1, -1, \ldots, i_1, -i_1, i_1, -i_1 > .$$

It is obvious that the 'best' design with i_1 observed once has larger determinant than the 'best' design with m observations at i_1, where $1 < m \leq n/2$, and it turns out the CVS is also smaller.

The 'best' design with i_1 observed once has CVS_1:

$$CVS_1 = \frac{\sum_{i=1}^{n-1} x_i x_{i+1}}{\sum_{i=1}^{n} x_i^2} + \frac{n-1}{n} = -\frac{2i_1 + n - 3}{2i_1^2 + n - 2}.$$

The 'best' design with m observations at i_1 has CVS_m:

$$CVS_m = \frac{\sum_{i=1}^{n-1} x_i x_{i+1}}{\sum_{i=1}^{n} x_i^2} + \frac{n-1}{n} = -\frac{2(m-1)i_1^2 + 2i_1 + (n - 2m - 1)}{2mi_1^2 + (n - 2m)}.$$

Then straightforward calculation gives

$$CVS_1 \leq CVS_m.$$

Thus, the design

$$< i_1, -1, 1, -1, 1, -1, \ldots, 1, -1, 1, -i_1 >$$

has the largest determinant and smallest CVS, with $CVS = 2(\frac{n-1}{n} - \frac{2i_1+n-3}{2i_1^2+n-2})$.

If n is odd, it is necessary to add one observation at 0. The discussion is

same as above and the design

$$< i_1, -1, 1, -1, 1, -1, \ldots, 1, -1, 1, -i_1, 0 >$$

has the largest determinant and smallest CVS, with $CVS = 2(\frac{n-1}{n} - \frac{2i_1+n-4}{2i_1^2+n-3})$.

CASE 2. $\rho \in \mathcal{P}_1$, $\mathcal{S}_k = (-1, -i_1, -i_2, \ldots, -i_k, 0, i_k, \ldots, i_2, i_1, 1)$.

It is easy to extend the argument in $CASE$ 1, and obtain the 'best' design:

$< i_k, -i_{k-1}, \ldots, -i_2, i_1, -1, 1, -i_1, i_2, \ldots, i_{k-1}, -i_k >$ if n is even;

$< i_k, -i_{k-1}, \ldots, -i_2, i_1, -1, 1, -i_1, i_2, \ldots, i_{k-1}, -i_k, 0 >$ if n is odd.

Example 1. Let design space $\mathcal{S} = [-1, -0.9, -0.8, \ldots, -0.1, 0, 0.1, \ldots, 0.8, 0.9, 1]$, n = 16, $\rho \in \mathcal{P}_1$.

$CVS = -0.02$ for the design $< 0.9, -1, 1, \ldots, -1, 1, -0.9 >$. Thus, it is the V-robust design for $-0.02 \leq \alpha < 0$.

$CVS = -0.036$ for the design $< 0.8, -1, 1, \ldots, -1, 1, -0.8 >$. And $CVS = -0.0364$ for the design $< 0.8, -0.9, 1, \ldots, -1, 0.9, -0.8 >$. Thus, the design $< 0.8, -0.9, 1, \ldots, -1, 0.9, -0.8 >$ is the V-robust design for $-0.0364 \leq \alpha < -0.036$.

CASE 3. $\rho \in \mathcal{P}_2$, $\mathcal{S}_k = (-1, -i_1, -i_2, \ldots, -i_k, 0, i_k, \ldots, i_2, i_1, 1)$

Similar discuss as first two cases, the 'best' design:

If n is even,

$$< i_2, i_4, i_6, \ldots, i_k, 1, 1, 1, \ldots, i_{k-1}, i_{k-3}, \ldots, i_3, i_1,$$

$$-i_1, -i_3, \ldots, -i_{k-3}, -i_{k-1}, -1, -1, -1, \ldots, -i_k, \ldots, -i_4, -i_2 >;$$

If n is odd,

$$< i_2, i_4, i_6, \ldots, i_k, 1, 1, 1, \ldots, i_{k-1}, i_{k-3}, \ldots, i_3, i_1, 0$$

$$-i_1, -i_3, \ldots, -i_{k-3}, -i_{k-1}, -1, -1, -1, \ldots, -i_k, \ldots, -i_4, -i_2 > .$$

Example 2. Let design space $\mathcal{S} = [-1, -0.9, -0.8, \ldots, -0.1, 0, 0.1, \ldots, 0.8, 0.9, 1]$, n = 16, $\rho \in \mathcal{P}_2$. Then $\alpha_{2,n} = -4c_1\frac{n-2}{n} = -3.5c_1$.

$CVS = -3.54c_1$ for the design $< 1, 1, \ldots, 1, 0.9, -0.9, -1, -1, \ldots, -1 >$. Thus, it is the V-robust design for $-3.54c_1 \leq \alpha < -3.5c_1$.

Therefore, for finite design space $\mathcal{S} = [-1, -i_1, -i_2, \ldots, -i_N, 0, i_N, \ldots, i_2, i_1, 1]$, it is theoretically accomplishable to obtain the *most* V-robust design. If $\Delta i = i_{j-1} - i_j$ in $\mathcal{S}$ is small, then the most V-robust design can be the approximation for that in infinite design space $\mathcal{S} = [-1, 1]$. However, the calculation will be tremendous even if n is not large.

Next section the attention is paid to the *most* V-robust designs for AR(1) processes.

Chapter 3

V-Robust Designs for AR(1) Processes

3.1 Most V-Robust Design for AR(1)

Consider AR(1) processes, where the correlation $\rho(k) = \rho^{|k|}$, for $\rho \in (-1, 1)$. So the CVS can be obtained from (2.4):

$$CVS(\xi, \mathcal{P}) = sup_{P \in \mathcal{P}} \left\{ \sum_{0 \le |s| \le n-1} \rho^{|s|} trace(B_\xi(s) B_\xi^{-1}(0)) \right\}.$$

Because of the definition of $\mathbf{B}_\xi(m)$ in equation (2.2), and $B_\xi(0)$ is assumed to be diagonal, we have

$$\begin{aligned} trace(B_\xi(s) B_\xi^{-1}(0)) &= trace(B_\xi^{-1}(0) B_\xi^T(s)) \\ &= trace(B_\xi^{-1}(0) B_\xi(-s)) \\ &= trace(B_\xi(-s) B_\xi^{-1}(0)). \end{aligned}$$

Then CVS becomes

$$CVS(\xi,\mathcal{P}) = sup_{P\in\mathcal{P}}\left\{\rho^0 trace(B_\xi(0)B_\xi^{-1}(0)) + 2\sum_{k=1}^{n-1}\rho^{|k|}trace(B_\xi(k)B_\xi^{-1}(0))\right\}.$$

And

$$\rho^0 trace(B_\xi(0)B_\xi^{-1}(0)) = \rho^0 trace(\mathbf{I}_{1+q}) = 1+q,$$

$$\rho^i trace(B_\xi(k)B_\xi^{-1}(0)) = \rho^i\left(\frac{n-i}{n} + \sum_{j=1}^{q}\mathcal{Q}_i(x_{(j)})\right).$$

Thus

$$CVS(\xi,\mathcal{P}) = sup_{\rho\in(-1,1)}\left\{1+q+2\sum_{i=1}^{n-1}\rho^i\frac{n-i}{n} + 2\sum_{k=1}^{n-1}\rho^k\sum_{j=1}^{q}\mathcal{Q}_k(x_{(j)}))\right\}.$$

where, q is the number of parameters, $\mathcal{Q}_k(x) = \frac{\sum_{i=1}^{n-k}x_i x_{i+k}}{\sum_{i=1}^{n}x_i^2}$, and $x_{(j)}$ is the $j+1^{th}$ column of Z.

So

$$CVS(\xi,\mathcal{P}) = sup_{\rho\in(-1,1)}\left\{1+q+2\sum_{i=1}^{n-1}\rho^i\frac{n-i}{n} + \sum_{j=1}^{q}\sum_{k=1}^{n-1}2\rho^k\mathcal{Q}_k(x_{(j)}))\right\},$$

that is,

$$CVS(\xi,\mathcal{P}) = sup_{\rho\in(-1,1)}\left\{1+q+2\sum_{i=1}^{n-1}\rho^i\frac{n-i}{n} + \sum_{j=1}^{q}\mathbf{Q}(x_{(j)})\right\}$$

or

$$CVS(\xi,\mathcal{P}) = sup_{\rho\in(-1,1)}\left\{1+q+\frac{2}{n}\frac{\rho^{n+1}-\rho^2}{(\rho-1)^2} - \frac{n-1}{n}\frac{\rho}{\rho-1} + \sum_{j=1}^{q}\mathbf{Q}(x_{(j)})\right\}. \tag{3.1}$$

where

$$\mathbf{Q}(x) = \frac{x^T Q x}{x^T x},$$

and

$$Q = \begin{pmatrix} 0 & \rho & \rho^2 & \rho^3 & \dots & \rho^{n-1} \\ \rho & 0 & \rho & \rho^2 & \dots & \rho^{n-2} \\ \rho^2 & \rho & 0 & \rho & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \rho^{n-5} & \dots & \rho \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \dots & 0 \end{pmatrix}.$$

Evaluating the behaviour of the eigenvalues and eigenvectors of Q is one of the key steps to get the most V-robust designs for this multiple regression model.

Let $R = Q + I$, then R becomes the correlation matrix of AR(1) processes. And

$$\lambda_R = \lambda_Q + 1,$$

the corresponding eigenvector

$$\mathbf{v}_R = \mathbf{v}_Q,$$

that is because of

$$(Q + I)\mathbf{v} = \lambda_{Q+I}\mathbf{v} = (\lambda_Q + 1)\mathbf{v} \Leftrightarrow Q\mathbf{v} = \lambda_Q \mathbf{v}.$$

So instead of finding the eigenvalues and eigenvectors of Q, we find them of R firstly.

Theorem 3.1: *The kth eigenvalue of Q is*

$$\lambda_k = \frac{1-\rho^2}{1-2\rho\cos(\omega_k)+\rho^2} - 1, \quad 1 \le k \le n,$$

(note: decreasing order if $\rho \in (0,1)$, increasing order if $\rho \in (-1,0)$.)

where $\{\omega_k\}$ are the roots in $(0,\pi)$ of the equation

$$\tan(n\omega_k) = -\frac{(1-\rho^2)\sin(\omega_k)}{\cos(\omega_k) - 2\rho + \rho^2\cos(\omega_k)}.$$

And the corresponding k^{th} eigenvector (unstandardized),

$$\mathbf{v}_{k,i} = \sin[\omega_k(i - \frac{n+1}{2}) + k\pi/2].$$

Proof. According to Grenander and Szegö (1958), the entries of matrix R

$$R = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \rho^{n-5} & \dots & \rho \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \dots & 1 \end{pmatrix}$$

is the coefficient of the Fourier series $f(x)$,

$$f(x) = \sum_{n=-\infty}^{\infty} \rho^{|n|} e^{inx} = \frac{1-\rho^2}{1-2\rho\cos(x)+\rho^2}, \qquad -1 < \rho < 1.$$

They proved that the eigenvalues are of the form,

$$\lambda_k = f(\omega_k),$$

where, according to Ray and Driver (1970), the ω_k in $(0, \pi)$ are the roots of

$$\tan(n\omega_k) = -\frac{(1-\rho^2)\sin(\omega_k)}{\cos(\omega_k) - 2\rho + \rho^2\cos(\omega_k)}.$$

and

$$\omega_1 < \omega_2 < \cdots < \omega_k.$$

Akansu and Torum(2012) obtained the explicit form of eigenvectors for R, which is

$$\mathbf{v}_{k,j} = \sin[\omega_k(j - \frac{n+1}{2}) + k\pi/2].$$

□

Proposition 3.1: *The eigenvector* $\mathbf{v}_k$ *is orthogonal to* $\mathbf{1} = (1, \ldots, 1)^T$ *when* k *is even.*

Proof. It is sufficient to prove that

$$\sum_{i=1}^{n} \sin[\omega_k(i - \frac{n+1}{2})] = 0, \qquad if\ k\ is\ even$$

And because of the identity

$$\sum_{i=1}^{n} \cos[\alpha + k\beta] = \frac{\sin(\alpha + \frac{2n+1}{2}\beta) - \sin(\alpha + \frac{\beta}{2})}{2\sin\frac{\beta}{2}},$$

we obtain,

$$\begin{aligned}\sum_{i=1}^{n}\sin[\omega_k(i-\frac{n+1}{2})] &= \sum_{i=1}^{n}\sin[\omega_k(i-\frac{n+1}{2}-\frac{\pi}{2}+\frac{\pi}{2})]\\ &= \sum_{i=1}^{n}\cos[\omega_k(i-\frac{n+1}{2}-\frac{\pi}{2})]\\ &= \frac{\sin(\frac{n}{2}\omega_k-\frac{\pi}{2})+\sin(\frac{n}{2}\omega_k+\frac{\pi}{2})}{2\sin(\omega_k/2)}\\ &= \frac{-\cos(\frac{n}{2}\omega_k)+\cos(\frac{n}{2}\omega_k)}{2\sin(\omega_k/2)}\\ &= 0.\end{aligned}$$

□

Proposition 3.2: *Q is invertible except for $n-1$ values of ρ.*

Proof. It is sufficient to prove that except for $n-1$ values of ρ, 1 is not the eigenvalue of $R(=Q+I)$, which leads to 0 is not an eigenvalue of Q.

If 1 is one of the eigenvalues of R, then $f(\omega)=\frac{1-\rho^2}{1-2\rho\cos(\omega)+\rho^2}=1$, for some $\omega\in(0,\pi)$.

That is,

$$1-\rho^2 = 1-2\rho\cos(\omega)+\rho^2,$$

So $\rho=\cos(\omega)$.

Also, ω should be the root of the polynomial,

$$\tan(n\omega) = -\frac{(1-\rho^2)\sin(\omega)}{\cos(\omega)-2\rho+\rho^2\cos(\omega)}.$$

Simplify it, we have

$$\tan(n\omega)=\tan(\omega),$$

the roots are

$$\omega = \frac{k\pi}{n-1}, \qquad k = 1, 2, \ldots, n-1$$

for $\omega \in (0, \pi)$.

Thus, except for the above corresponding $n-1$ values of ρ, Q is invertible.

□

Now, follow the proof of *Theorem 2.2* in Section 2.1, we can find the *most* V-robust design similarly. Proof would be the same until the equation (2.14). Since now Q is different, ν and $\mathbf{s}$ will be different and have to be determined.

Let $\mathbf{s} = (v_1, v_2, \ldots, v_n)^T$, then because $(\mathbf{Q} - \nu\mathbf{I})\mathbf{s} = \mathbf{1}$, we have

$$\begin{pmatrix} -\nu & \rho & \rho^2 & \rho^3 & \ldots & \rho^{n-1} \\ \rho & -\nu & \rho & \rho^2 & \ldots & \rho^{n-2} \\ \rho^2 & \rho & -\nu & \rho & \ldots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\ \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \rho^{n-5} & \ldots & \rho \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \ldots & -\nu \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Assume ν is not the eigenvalue of $\mathbf{Q}$, then $|\mathbf{Q} - \nu\mathbf{I}| \neq 0$. So $(\mathbf{Q} - \nu\mathbf{I})\mathbf{s} = \mathbf{1}$ has the unique root, and $\mathbf{Q} - \nu\mathbf{I}$ is invertible. By binomial inverse theorem, we can find $(\mathbf{Q} - \nu\mathbf{I})^{-1}$ via $\mathbf{Q}^{-1}$,

$$(\mathbf{Q} - \nu\mathbf{I})^{-1} = \mathbf{Q}^{-1} + \nu\mathbf{Q}^{-1}(\mathbf{I} - \nu\mathbf{Q}^{-1})^{-1}\mathbf{Q}^{-1}. \tag{3.2}$$

Since we have already known the eigenvalues and eigenvectors of $\mathbf{Q}$, and

$$\mathbf{Q} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{\mathbf{T}},$$

$$\mathbf{Q}^{-1} = \mathbf{P}\mathbf{\Lambda}^{-\mathbf{1}}\mathbf{P}^{\mathbf{T}},$$

we can easily calculate $\mathbf{Q}^{-1}$. Thus, we can rewrite equation(3.2) as

$$(\mathbf{Q} - \nu\mathbf{I})^{-1} = \mathbf{P}(\mathbf{\Lambda}^{-\mathbf{1}} + \nu\mathbf{\Lambda}^{-\mathbf{1}}(\mathbf{I} - \nu\mathbf{\Lambda}^{-\mathbf{1}})^{-1}\mathbf{\Lambda}^{-\mathbf{1}})\mathbf{P}^{T}.$$

One last thing is to find the expression for ν numerically, which is the root of

$$\sum_{j=1}^{[(n+1)/2]} (\mathbf{1}^T\mathbf{v}_{2j-1})^2/(\lambda_{2j-1} - \nu) = 0.$$

3.1.1 Toy Sales Example: Revisit

Now, let us revisit the sales example introduced in Chapter 1. We carry out most V-robust design for AR(1) processes, with comparison to uniform design (equally spaced, ascendingly ordered). These two design strategies will give different design points. For example, if there are 5 periods that can be used for experiment, the uniform design will give $x =< 0, 4.5, 9, 13.5, 18 >$, and most V-robust design will give $x =< 5.5, 17, 0, 17, 5.5 >$. Recall that the underlying model is

$$y_i = 0.6x_i + 3 + \varepsilon.$$

SSPE (Sum of Squared Prediction Error) is used to measure the goodness of designs,

$$SSPE = \sum(\hat{y}_i - 0.6x_i - 3)^2,$$

where $x_i \in (0, 1, 2, \ldots, 17, 18)$.

Generalized least squares is used in predicting $\hat{y}$ from *most* V-robust design, where it turns out that the estimator of ρ for autocorrelation matrix only makes

a subtle difference in SSPE. For the uniform design, ordinary least squares is used because it ignores the correlation of errors.

SSPEs of two designs are displayed in Figure 3.1.1, which shows that most V-robust design of AR(1) processes performs much better in prediction of sales.

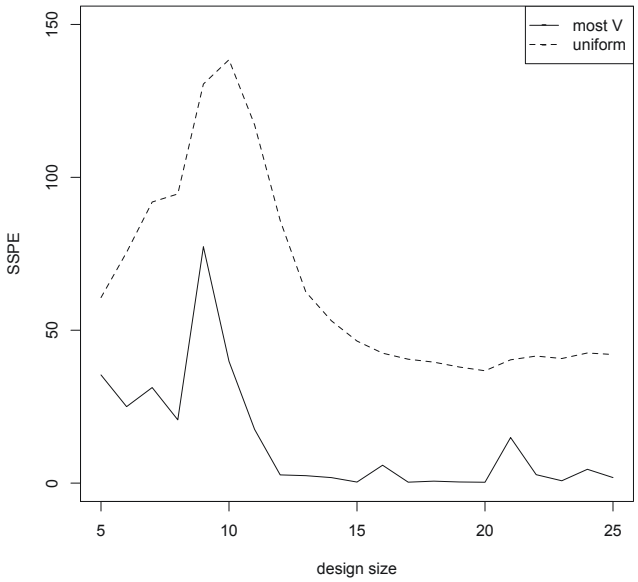

Figure 3.1: SSPE of Toy Sales Example

3.2 V-Robust design for AR(1)

Consider V-robust designs for the classes

$$\mathcal{P}_3 = \{\mathbf{P}|\rho(s) = \rho^s \text{ for } c_0 \leq \rho < 1 \text{ with some } c_0 > 0\};$$

$$\mathcal{P}_4 = \{\mathbf{P}|\rho(s) = \rho^s \text{ for } -1 < \rho \leq -c_1 \text{ with some } c_1 > 0\}.$$

Here the case $q = 1$ is considered since the extension to $q > 1$ is rather evident.

The CVS in equation (3.1) can be simplified

$$CVS(\xi, \mathcal{P}) = sup_{\rho\in(-1,1)} \left\{ 2 + 2\sum_{i=1}^{n-1} \rho^i \frac{n-i}{n} + 2\sum_{i<j} x_i x_j \rho^{j-i} \right\}. \tag{3.3}$$

If α is sufficiently large, then observations can only be made at 1 or -1, plus one made at 0 if n is odd, it is easy to determine the order of observations to minimize $\mathbf{Q}(x)$(i.e. to minimize $\sum_{i<j} x_i x_j \rho^{j-i}$). It turns out the design is exactly the same as these of MA(1) processes.

- V-robust designs for $\mathcal{P}_3$ are

$$\mathbf{x}_{(1)} = \begin{cases} < 1, -1, 1, -1, \ldots, 1, -1 > & \text{n even,} \\ < 1, -1, 1, -1, \ldots, 1, -1, 0 > & \text{n odd.} \end{cases}$$

- V-robust designs for $\mathcal{P}_4$ are

$$\mathbf{x}_{(1)} = \begin{cases} \underbrace{< 1, \ldots, 1,}_{n/2} \underbrace{-1, \ldots, -1 >}_{n/2} & \text{n even,} \\ \underbrace{< 1, \ldots, 1,}_{n/2} 0, \underbrace{-1, \ldots, -1 >}_{n/2} & \text{n odd.} \end{cases}$$

And the corresponding CVS can be calculated by equation(3.3).

3.3 Comparision

The example is illustrated because "(a) this f is least favourable, in a minimax sense, for straight line regression and (b) a quadratic disturbance represents the most common and worrisome departure from linearity in most applications", according to Section 6 (Page 1507) in Wiens and Zhou (1997). The true model is given by

$$Y_i = \mathbf{z}^T(x_i)\theta + f(x_i) + \varepsilon_i, \qquad i = 1, \ldots, n$$

with $f(x) = \eta(45/8)^{1/2}(x^2 - 1/3)$, and $\varepsilon = \sigma_0 \mathbf{P}^{1/2}\mathbf{w}$, where $\mathbf{w}$ is a vector of white noise with variance σ_w^2, $\mathbf{P}$ is an autocorrelation matrix to be specified, and $\sigma_0^2 = var(\varepsilon_t/\sigma_w)$.

Denoting σ_w^{-2} times the MSE matrix of $n^{1/2}\hat{\theta}$ by $\mathbf{C}$, Wiens and Zhou (1997) find that

$$\mathbf{C} = \sigma_0^2 \begin{pmatrix} 1 & 0 \\ 0 & \tau_2^{-1} \end{pmatrix} \frac{\mathbf{Z}^T\mathbf{P}\mathbf{Z}}{n} \begin{pmatrix} 1 & 0 \\ 0 & \tau_2^{-1} \end{pmatrix} + \frac{45}{8}\nu \begin{pmatrix} \tau_2 - \frac{1}{3} \\ \frac{\tau_3}{\tau_2} \end{pmatrix} \left(\tau_2 - \tfrac{1}{3}, \tfrac{\tau_3}{\tau_2} \right), \qquad (3.4)$$

where $\mathbf{Z} = (\mathbf{1}\vdots\mathbf{x})$ is the model matrix, $\tau_k = \sum_{i=1}^n x_i^k/n$ is the kth moment of the design, and $\nu = n\eta^2/\sigma_w^2$. The value of ν may be viewed as reflecting the relative importance of bias versus variance in the mind of the experimenter. $\nu = 0$ indicates that the fitted model is exactly correct, and $\nu = 1$ indicates that $(\int f^2)^{1/2} (= \eta)$ is of the same magnitude as a standard error. $\mathbf{P} = \mathbf{P}_j(\rho)$ is taken to be the autocorrelation matrix of one of the following error processes ε_t with lag-one autocorrelation ρ: (a) $j = 1$: MA(1) with $\rho \geq 0$; (b)$j = 2$: MA(1) with $\rho \leq 0$; (c) $j = 3$: AR(1) with $\rho \geq 0$ or (d) $j = 4$: AR(1) with $\rho \leq 0$. For

$j = 1, 2$ we have $\sigma_0^2 = 1 + \theta^2$, where $\theta \in [-1, 1]$ satisfies $\rho = -\theta/(1 + \theta^2)$. For $j = 3, 4$, we have $\sigma_0^2 = (1 - \rho^2)^{-1}$.

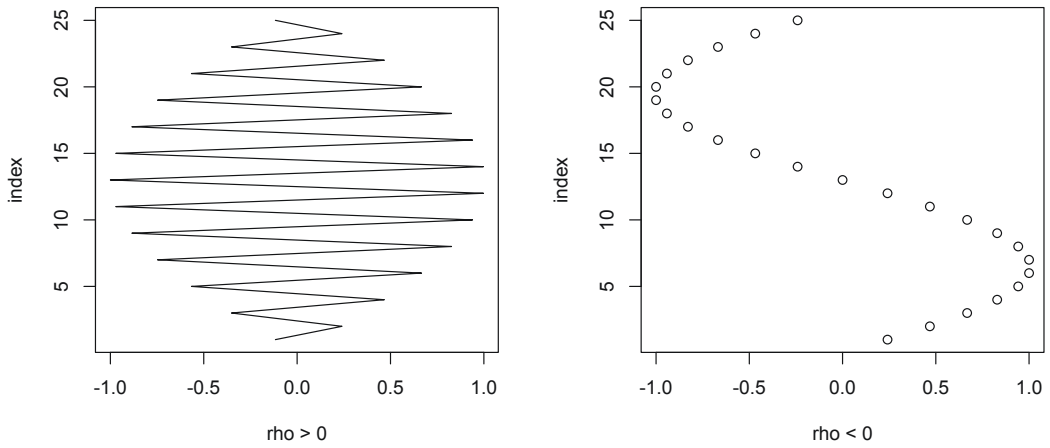

Figure 3.2: Most V-robust designs of MA(1), n = 25.

The *most* V-robust designs for MA(1) processes only depend on the sign of ρ as showed in the Figure 3.3. However, the *most* V-robust designs for AR(1) processes, as showed in Figure 3.3, depend on the the magnitude of ρ.

It seems when $n = 25$, the designs for AR(1) and MA(1) are almost the same if $\rho > 0$. And when $\rho < 0$, the difference between them seems to be the orders of the same design points. Hence, it is likely that the performances of these designs are similar if n is not small.

Same as Wiens and Zhou (1997), several loss functions are considered here (a) $det = |\mathbf{C}|$; (b) $trace = trace(\mathbf{C})$; and (c) IMSE = $trace(\mathbf{CA})$, where $\mathbf{A} = \int_{-1}^{1}(1, x)^T(1, x)dx = diag(2, 2/3)$. It turns out the performances of the *most* V-robust designs for MA(1) and AR(1) are almost indistinguishable when $n = 25$, if the true model is MA(1) or AR(1).

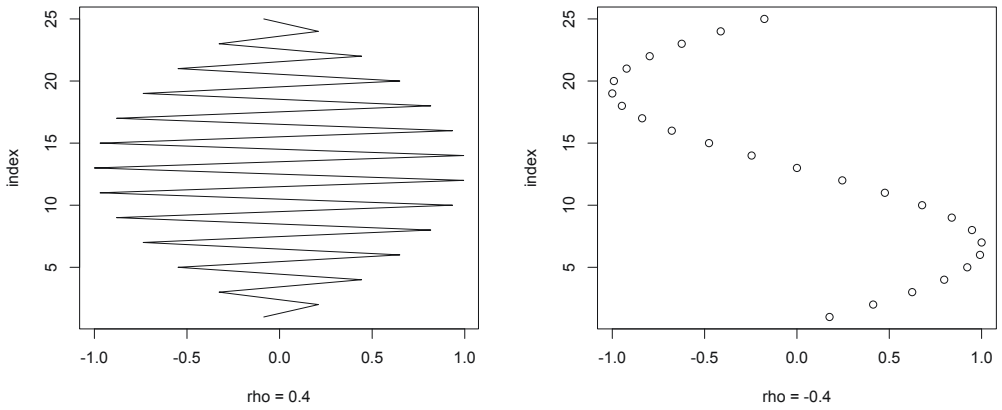

Figure 3.3: Most V-robust designs of AR(1), n = 25.

The performances are the same for $\nu = 0, n = 25$.

If the sample size is changed to a small value, the performances will be much different. The results where $n = 5$, $n = 10$ and $n = 15$ are displayed in Figure 3.6 to Figure 3.11. It turns out that the V-robust designs for AR(1) perform better if $\nu = 1$, but the V-robust designs for MA(1) perform better if $\nu = 0$.

The designs are displayed below($n = 15, 25$ are omitted because designs are almost the same):

- For $n = 5$, most V-robust designs of MA(1) processes are

$$\begin{cases} < 1, 1, 0, -1, -1 > & \rho \in \mathcal{P}_2; \\ < -0.45, 0.95, -1, 0.95, -0.45 > & \rho \in \mathcal{P}_1. \end{cases}$$

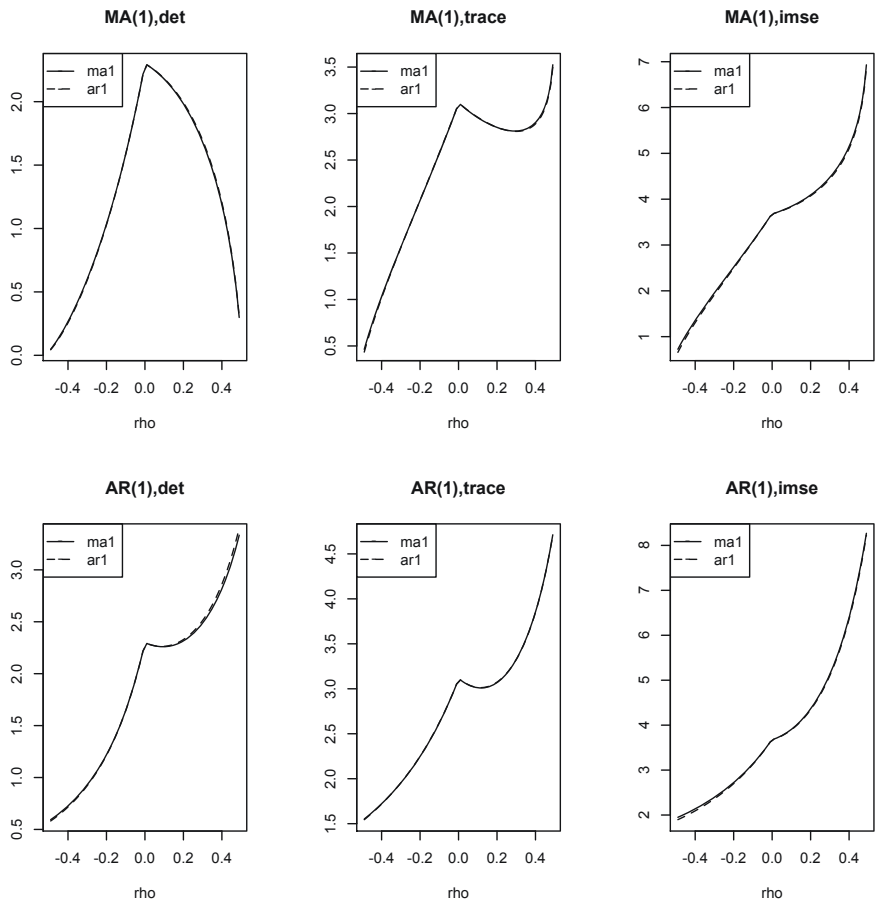

Figure 3.4: Comparison of different loss functions, when $n = 25, \nu = 1$. For example, MA(1), det: the true model with autocorrelation matrix of MA(1) processes, determinant as loss function.

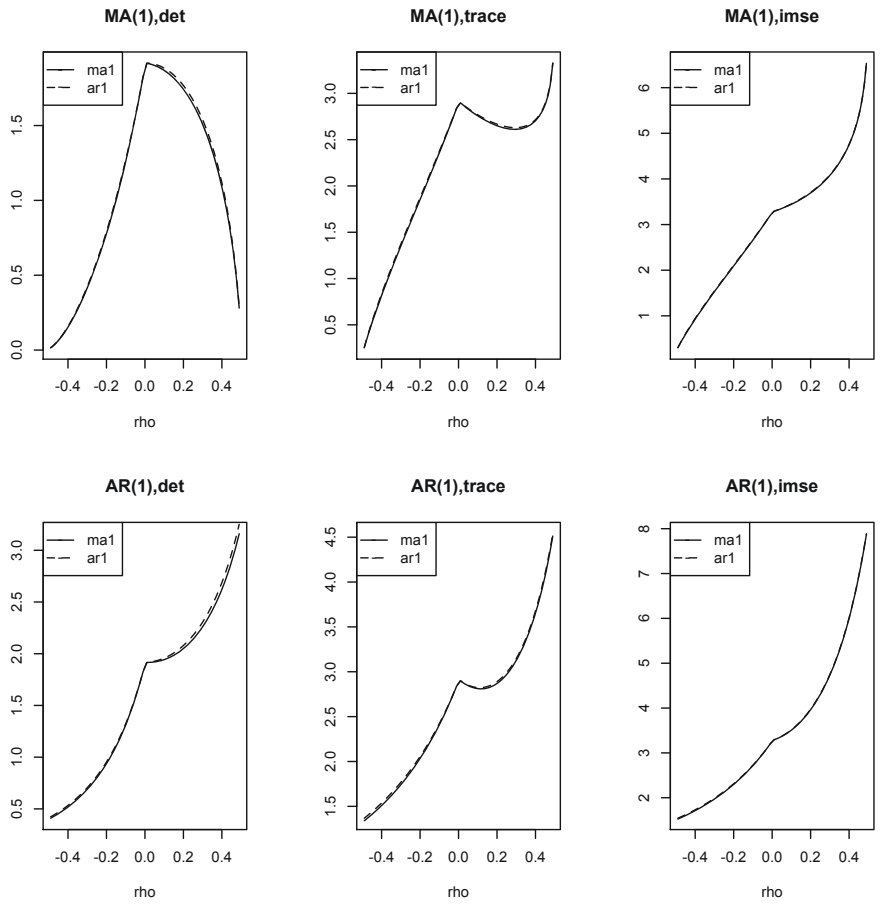

Figure 3.5: Comparison of different loss functions, when $n = 25, \nu = 0$.

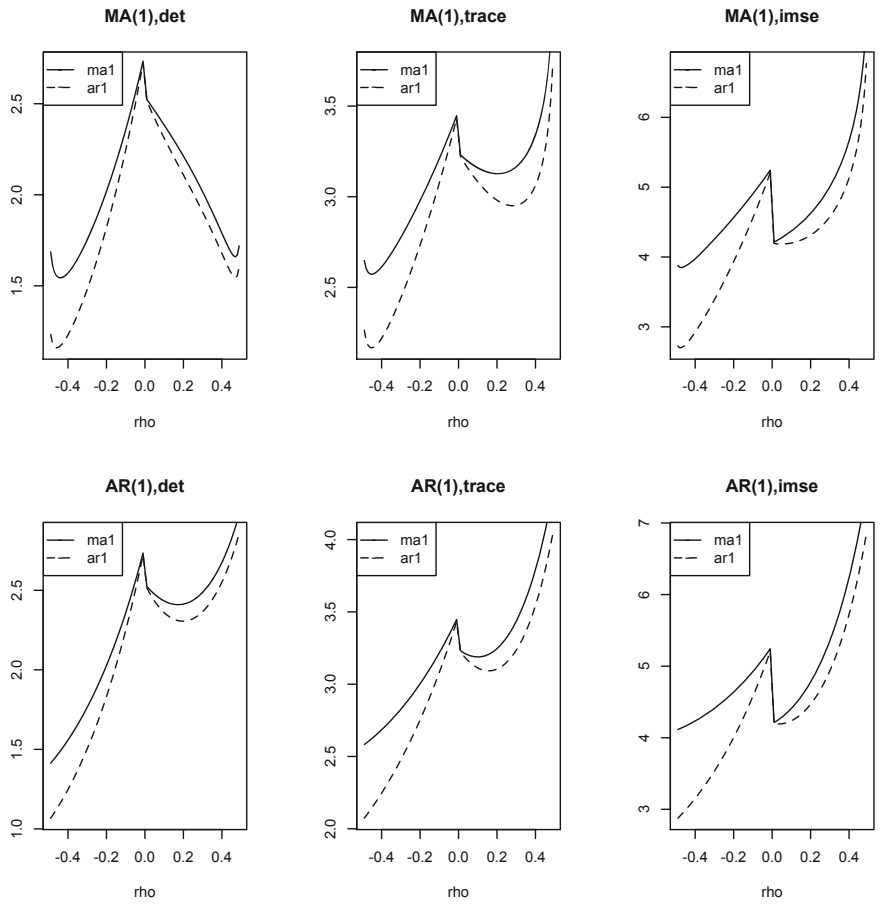

Figure 3.6: Comparison of different loss functions, when $n = 5, \nu = 1$.

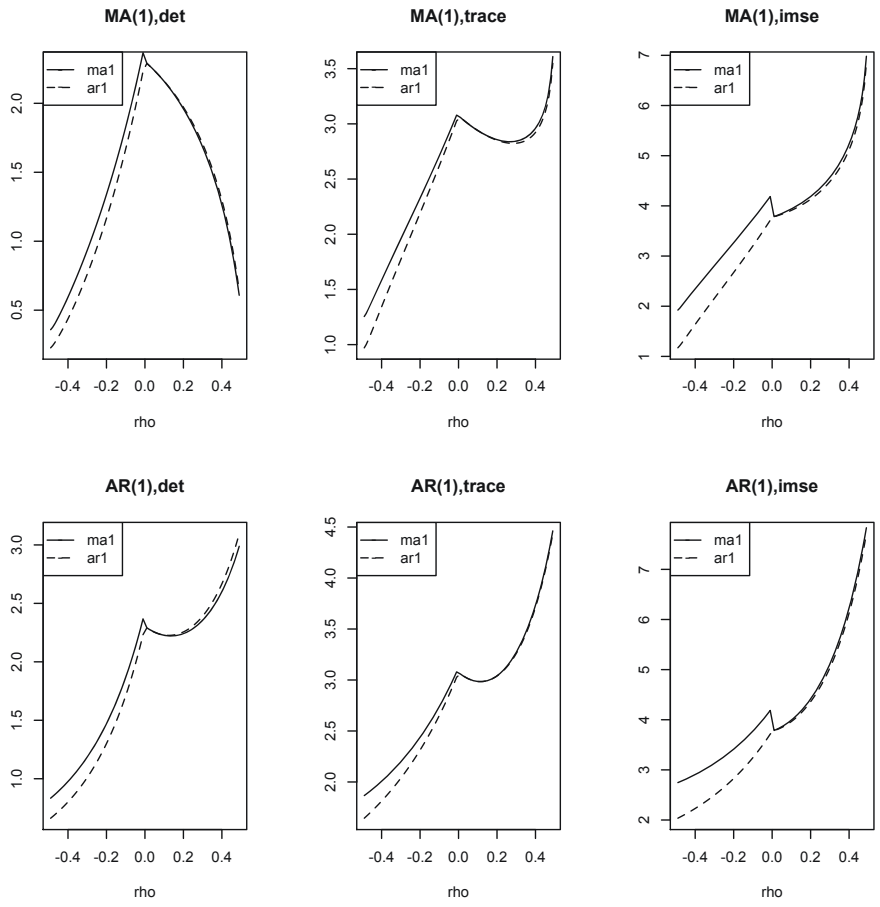

Figure 3.7: Comparison of different loss functions, when $n = 10, \nu = 1$.

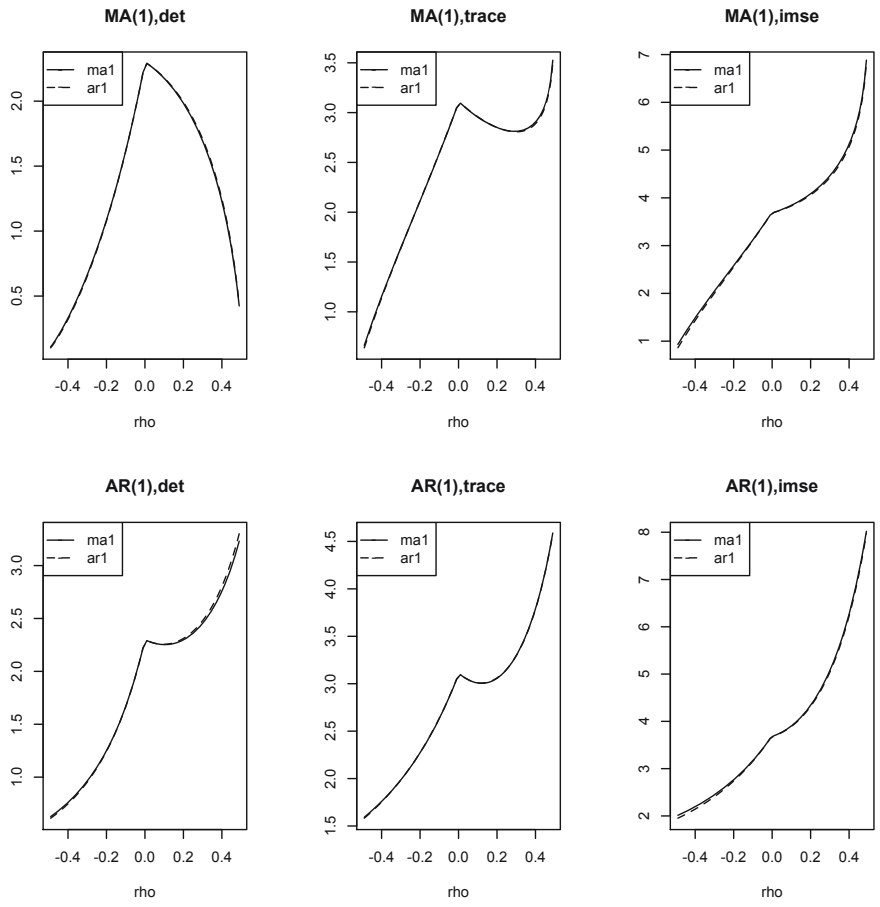

Figure 3.8: Comparison of different loss functions, when $n = 15, \nu = 1$.

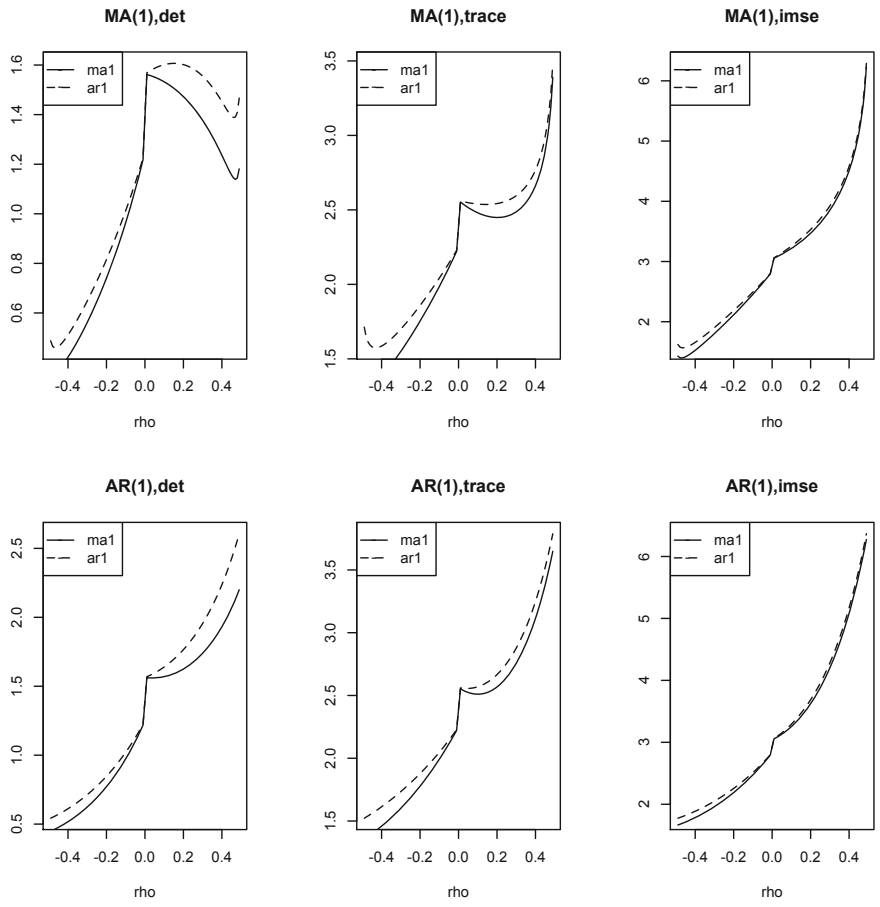

Figure 3.9: Comparison of different loss functions, when $n = 5, \nu = 0$.

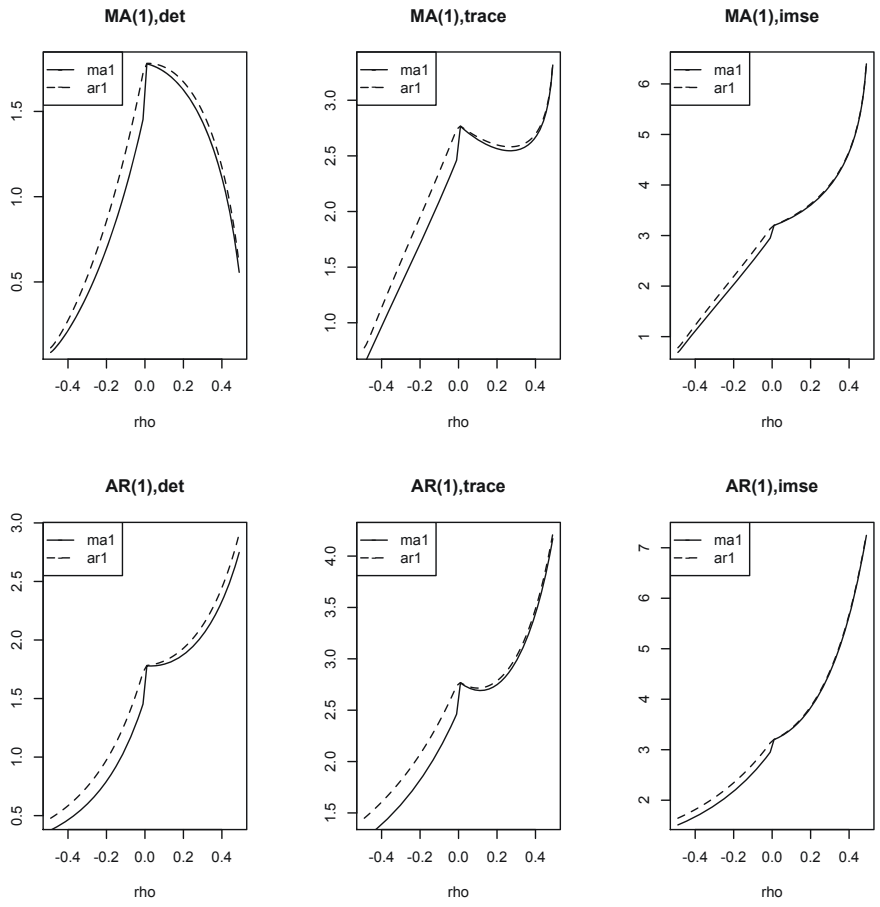

Figure 3.10: Comparison of different loss functions, when $n = 10, \nu = 0$.

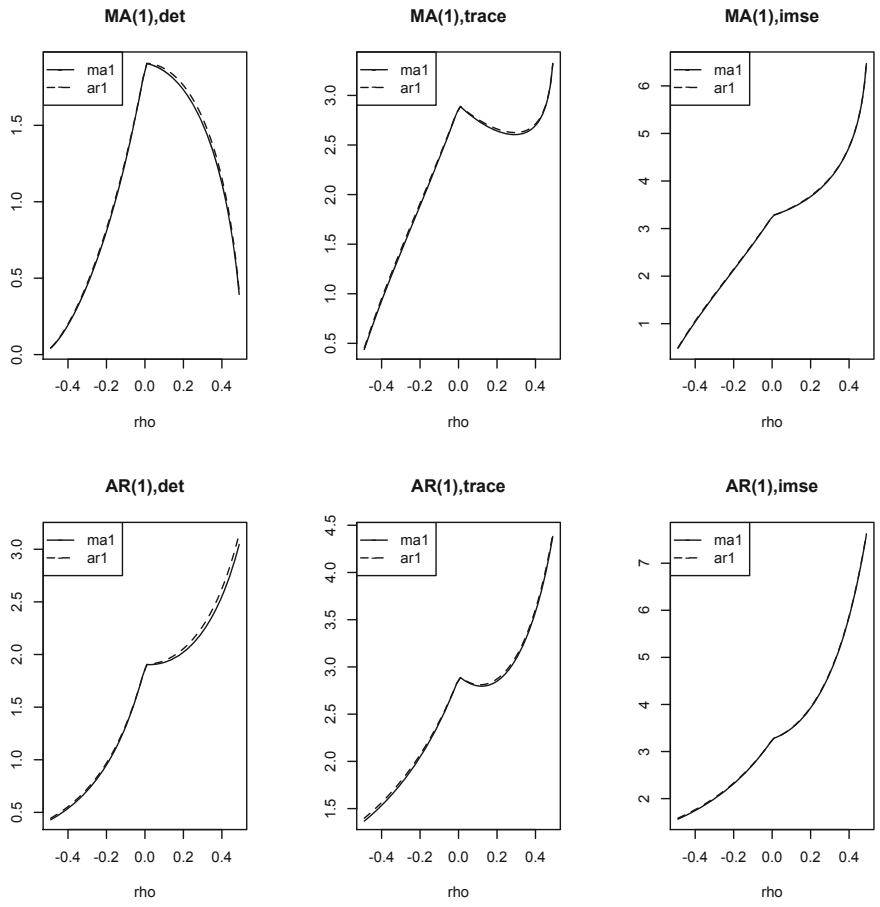

Figure 3.11: Comparison of different loss functions, when $n = 15, \nu = 0$.

- For $n = 5$, most V-robust designs of AR(1) processes are

$$\begin{cases} < 0.82, 1, 0, -1, -0.82 > & \rho = -0.4, \\ < -0.90, -1, 0, 1, 0.90 > & \rho = -0.2, \\ < -0.40, 0.90, -1, 0.90, -0.40 > & \rho = 0.2, \\ < -0.36, 0.86, -1, 0.86, -0.36 > & \rho = 0.4. \end{cases}$$

- For $n = 10$, most V-robust designs of MA(1) processes are

$$\begin{cases} < 0.55, 0.92, 1, 0.76, 0.28, -0.28, -0.76, -1, -0.92, -0.55 > & \rho \in \mathcal{P}_2, \\ < 0.28, -0.55, 0.76, -0.92, 1, -1, 0.92, -0.76, 0.55, -0.28 > & \rho \in \mathcal{P}_1. \end{cases}$$

- For $n = 10$, most V-robust designs of AR(1) processes are

$$\begin{cases} < 0.42, 0.86, 1, 0.79, 0.30, -0.30, -0.79, -1, -0.86, -0.42 > & \rho = -0.4, \\ < 0.48, 0.89, 1, 0.78, 0.29, -0.29, -0.78, -1, -0.89, -0.48 > & \rho = -0.2, \\ < 0.25, -0.52, 0.75, -0.91, 1, -1, 0.91, -0.75, 0.52, -0.25 > & \rho = 0.2, \\ < 0.22, -0.50, 0.74, -0.91, 1, -1, 0.91, -0.74, 0.50, -0.22 > & \rho = 0.4. \end{cases}$$

3.3.1 Comparison for true model as MA(2) or AR(2)

This subsection discusses the situation where the true model has autocorrelation matrix of MA(2) or AR(2) processes. Either MA(2) or AR(2) has two parameters, θ_1, θ_2.

By calculation,

$$\sigma_0^2 = 1 + \theta_1^2 + \theta_2^2$$

for MA(2);

$$\sigma_0^2 = \frac{1}{1 - \theta_2(\frac{\theta_1}{1-\theta_2}) - \theta_2(\frac{\theta_1^2+\theta_2-\theta_2^2}{1-\theta_2})}$$

for AR(2). Let $\rho_1 = \rho$, where ρ (the parameter in MA(1) or AR(1) processes) is determined in advance to obtain the designs, and then solve θ_1, θ_2 to determine the autocorrelation matrix $\mathbf{P}$ of MA(2) or AR(2). Here ρ_1 is chosen from $(-0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.4)$ and for each particular value of ρ_1, we assign values to ρ_2 by assigning values to θ_2. So there will be jump from one value to another value of ρ_1. This is not a problem since the attention is paid to the difference of performances of designs.

The most V-robust design of AR(1) or MA(1) can be applied to the models with autocorrelation matrix of AR(2) or MA(2) processes. Again, the situation where $\nu = 1, 0$ and $n = 5, 25$ are discussed in Figure 3.3.1 and Figure 3.3.1. We can see that the most V-robust design for AR(1) performs better than the most V-robust design for MA(1) when $\nu = 1$.

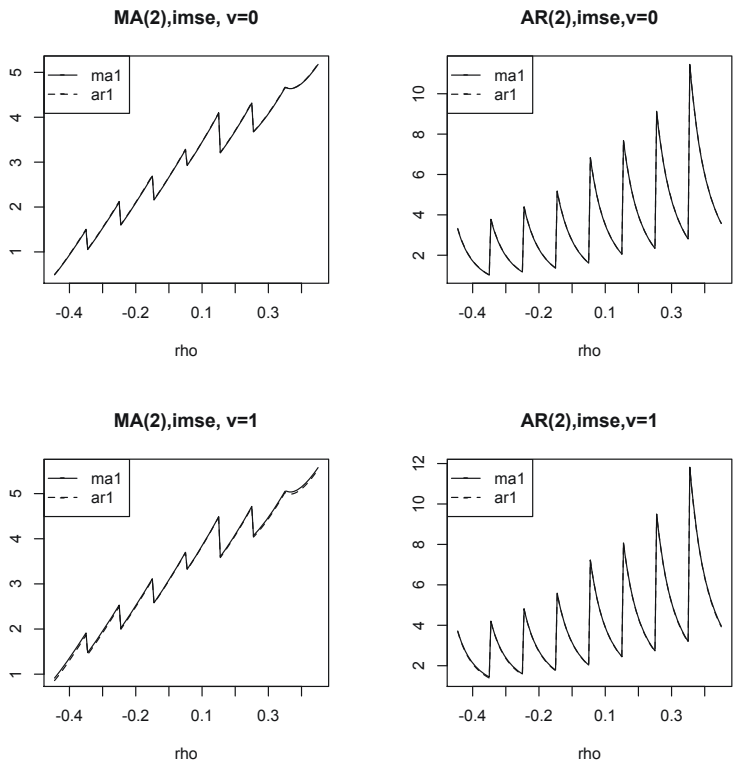

Figure 3.12: Comparison when model is MA(2), AR(2), $n = 25$.

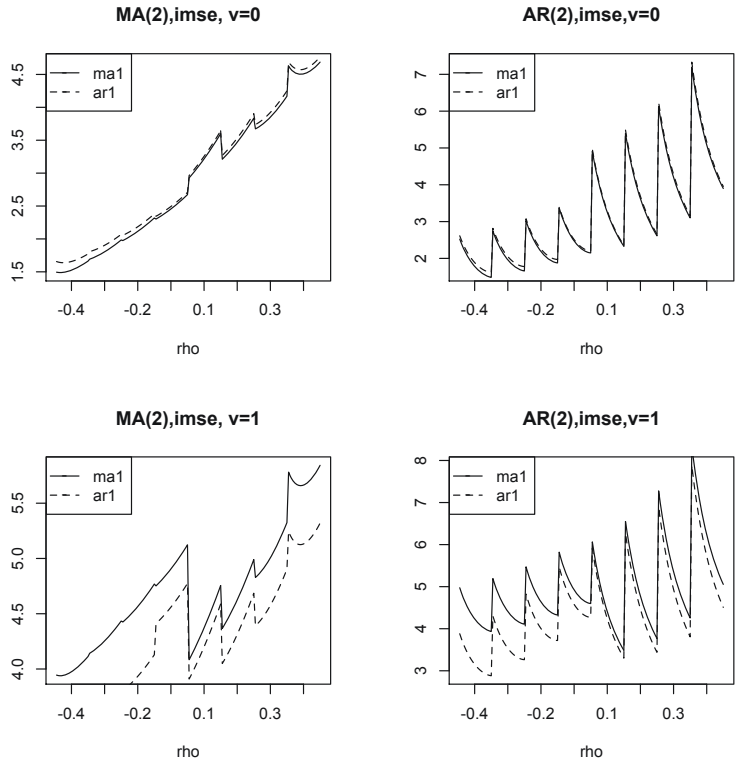

Figure 3.13: Comparison when model is MA(2), AR(2), $n = 5$.

3.3.2 Conclusion

In conclusion, the most V-robust design for AR(1) processes performs better than the most V-robust design for MA(1) processes if n is small and $\nu = 1$ while MA(1) performance better if n is small and $\nu = 0$. However, when n is moderately large, they are indistinguishable.

3.4 Conclusions and Future Research

V-robust designs should be applied to cases where autocorrelated errors are anticipated or detected. V-robust designs for MA(1) processes and for AR(1)

processes have indistinguishable performance in some situations. However, it is showed most V-robust design for AR(1) processes can be more powerful if the number of observations is small and there is some bias in the fitted model. It seems reasonable to expect most V-robust design for AR(1) processes to perform better than most V-robust design for MA(1) processes in reality because the fitted model would not be very accurate at most of the time, and limited cost would lead to limited observations. It would be good to see its power in industrial applications.

Bibliography

Akansu, Ali N. and Torum, Mustafa U. (2012), "Toeplitz Approximation to Empirical Correlation Matrix of Asset Returns: A Signal Processing Perspective", *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no.4, pp.319-326.

Bickel, P. J., and Herzberg, A. M. (1979), "Robustness of Design Against Autocorrelation in Time I: Asymptotic Theory, Optimality for Location and Linear Regression", *The Annals of Statistics*, 7, 77-95.

Bickel, P. J., Herzberg, A. M., and Schilling, M. F. (1981), "Robustness of Design Against Autocorrelation in Time II: Optimality, Theoretical and Numerical Results for the First-Order Autoregressive Process", *Journal of the American Statistical Association*, 76, 870-877.

Box, G. E. P., and Draper, N. R. (1959), "A Basis for the Selection of a Response Surface Design", *Journal of the American Statistical Association*, 54, 622-654.

Chernoff, H. (1953), "Locally Optimal Designs for Estimating Parameters", *Annals of Mathematical Statistics*, 29, 586-602.

Ehrenfeld, E. (1955), "On the Efficiency of Experimental Design", *Annals of Mathematical Statistics*, 26, 247-255.

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

Huber, P.J. (1975), "Robustness and Designs" in: *A Survey of Statistical Design and Linear Models*, ed. J. N. Srivastava, North Holland, pp. 287-303.

Grenander, U. and Szegö, G (1958), *Toeplitz Forms and Their Application*, Berkeley: University of California Press.

Kiefer, J. (1959), *Optimal Experimental Designs*, Journal of the Royal Statistical Society, B 21, 273-319.

Ray W. and Driver R. (1970), *"Further decomposition of the Karhunen-Loeve series representation of a stationary random process"*, IEEE Trans. Inf. Theory, vol. IT-16, no.6, pp.663-668,Nov. 1970.

Ray W. and Driver R. (1970), *"Further decomposition of the Karhunen-Loeve series representation of a stationary random process"*, IEEE Trans. Inf. Theory, vol. IT-16, no.6, pp.663-668,Nov. 1970.

Rousseeuw, P. J. (1981), *"A new infinitesimal approach to robust estimation"*, Z. Wahrsch. verw. Geb. 56, 127-132.

Tsai, Y.-L., and Zhou, J. (2005), "Discrete M-Robust Designs for Regression Models", *Journal of Statistical Planning and Inference*, 131, 393-406.

Wiens, D. P. (1992), "Minimax Designs for Approximately Linear Regression", *Journal of Statistical Planning and Inference*, 31, 353-371.

Douglas P. Wiens (2013) *"Robustness of Design",* to appear in *"Handbook of Design and Analysis of Experiments"*, Chapman Hall.

Wiens, D.P. and Zhou, J. (1996), "Minimax Regression Designs for Approximately Linear Models With Autocorrelated Errors," *Journal of Statistical Planning and Inference*, 55, 95-106.

Wiens, D.P. and Zhou, J. (1997), "Robust Designs Based on the Infinitesimal Approach", *Journal of the American Statistical Association*, Vol. 92, No.440, Theory and Methods.

Wiens, D.P. and Zhou, J. (1999), "Minimax Designs for Approximately Linear Models with AR(1) Errors", *The Canadian Journal of Statistics*, 27, 781-794.

Zhou, J. (2001) "A Robust Criterion for Experimental Designs for Serially Correlated Observations", *Technometrics*, 43, 462-467.