KoulMde: An R Package for Koul’s Minimum Distance Estimation

Jiwoong Kim
Michigan State University
June 15, 2016

Abstract
This article provides a full description of the R package KoulMde which is designed for Koul’s minimum distance estimation method. When we encounter estimation problems in the linear regression and autoregressive models, this package provides more efficient estimators than other R packages.

Keywords: minimum distance estimation; linear regression; autoregressive error

1 Introduction
Minimum distance (MD) estimation method refers to the technique that obtains estimators by minimizing a difference between a function obtained from the sample of observations and the one from the assumed model. The most common and popular distance used in the literature of the MD estimation methodology is Cramér-von Mises (CM) type distance. Wolfowitz (1957) used CM type distance which measures difference between empirical distribution function and assumed model distribution function. Parr and Schucany (1980) empirically showed the robustness of these MD estimators of location parameters in the one and two sample location models. Departing from the one sample model, Koul and De Wet (1986) extended domain of applications of these MD estimators to the linear regression model, where the regression parameters of interest are estimated by minimizing a CM type distance between weighted empirical residual process and its expectation. All of these works assume that error distribution in these models is known, which is not a practical assumption. Koul (1985) weakened this assumption by assuming that the error distribution in the multiple linear regression model is symmetric around the origin. He defined a class of $L_2$ distances between weighted empiricals of residuals and negative residuals and a class of estimators that minimize these distances. Koul (1986) broadened the domain of applications of this MD estimation methodology to the linear autoregression models. R package KoulMde is based on his work (Koul (1985) and Koul (1986)). The package is available from Comprehensive R Archive Network (CRAN) at https://cran.r-project.org/web/packages/KoulMde/index.htm.

This article is organized as follows. In Section 2 we provide the detailed description of the package. We also provide a comparison between KoulMde and other existing estimation methods or R packages; ordinary least squares (OLS) vs. KoulMde in Section 2.1 arima (R package) vs. KoulMde in Section 2.2 and orcutt (R package) vs. KoulMde in Section 2.3. We conclude this article with a brief summary of the package in Section 3.

2 KoulMde package
The KoulMde package contains three functions: KoulLrMde, KoulArMde, and Koul2StageMde. The function KoulLrMde estimates the regression parameter vector in the multiple linear regression model; KoulArMde deals with the estimation of the parameter vector in linear autoregressive model of order $q$, a known positive integer. Wrapping up these two functions, Koul2StageMde provides consistent estimators of both regression and autoregressive parameters when we consider the linear regression model with autoregressive errors.
2.1 KoulLrMde

Consider the linear regression model for

\[ Y_i = x_i'\beta + \varepsilon_i, \quad i = 1, \ldots, n, \]  

(1)

where \( x_i = (x_{i1}, \ldots, x_{ip})' \in \mathbb{R}^p \) and \( \beta = (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^p \) is the parameter vector of interest. Let \( \varepsilon_i \)'s be independently and identically distributed random variables. In addition, \( \varepsilon_1 \) is assumed to be symmetric around zero. As in Koul (1985), we introduce the distance function for \( b \in \mathbb{R}^p \),

\[
T(b) := \sum_{k=1}^{p} \int \left[ \sum_{i=1}^{n} d_{ik} \left\{ I(Y_i - x_i'b \leq y) - I(-Y_i + x_i'b < y) \right\} \right]^2 dH(y),
\]

where \( H \) is a \( \sigma \)-finite measure on \( \mathbb{R} \) and symmetric around 0, i.e., \( dH(-x) = -dH(x) \), for \( x \in \mathbb{R} \), \( d_{ik}, 1 \leq i \leq n, 1 \leq k \leq p, \) are some real numbers. Define the class of MD estimators \( \hat{\beta} \), one for each \( H \), as

\[
T(\hat{\beta}) := \inf_b T(b).
\]

If \( H \) is continuous, then from Koul (2002, p. 149),

\[
T(b) = \sum_{k=1}^{p} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ik}d_{jk} \left[ |H(Y_i - x_i'b) - H(-Y_j + x_j'b)| \right. \\
\left. - |H(Y_i - x_i'b) - H(Y_j - x_j'b)| \right].
\]

Note that for degenerate \( H \),

\[
T(b) = \sum_{k=1}^{p} \left[ \sum_{i=1}^{n} d_{ik} \text{sgn}(Y_i - x_i'b) \right]^2,
\]

where

\[
\text{sgn}(x) := \begin{cases} 1, & \text{if } x > 0; \\
0, & \text{if } x = 0; \\
-1, & \text{if } x < 0. 
\end{cases}
\]

Choosing optimal measure combined with optimal \( d_{ik} \)'s will give a rise to well-celebrated estimators. For example, minimizing \( T \) with degenerate measure combined with \( d_{ik} = x_{ik} \) will yield least absolute deviation (LAD) estimators; see, e.g., Chapter 5.3 in Koul (2002) for the detail. The function KoulLrMde estimates \( \beta \) in the regression model (1) by minimizing distance function \( T \) after selecting \( d_{ik} \) and \( H \). Table 1 summarizes its arguments and return values. \( D \) is a \( n \times p \) matrix whose \( (i, k) \)th entry is \( d_{ik} \). Koul (2002) showed that

### Table 1: Summary of KoulLrMde

| Usage         | KoulLrMde(Y, X, D, IntMeasure). |
|---------------|---------------------------------|
| Arguments     |                                 |
| Y             | Vector of response variable in linear regression model. |
| X             | Design matrix of explanatory variable in linear regression model. |
| D             | Weight matrix. Dimension of D should match that of X. |
| IntMeasure    | “default” uses \( XA \) where \( A = (X'X)^{-1}/2. \). |
|               | Symmetric and \( \sigma \)-finite measure. |
| Return Values |                                 |
| betahat       | Minimum distance estimator of \( \beta \). |
| residual      | Residuals after minimum distance estimation. |

\( D := XA \) gives the most efficient estimator of \( \beta \); see Chapter 5.6 therein.

The following example shows the usage of KoulLrMde. First, generate model (1) with normal error where \( n = 50, \ p = 3, \) and \( \beta = (-2, 0.3, 1.5)' \).
n <- 50
p <- 3
X <- matrix(runif(n*p, 0,50), nrow=n, ncol=p)
beta <- c(-2, 0.3, 1.5)
eps <- rnorm(n, 0, 5)
Y <- X %*% beta + eps

Next, determine D and IntMeasure. We use default value and Lebesgue measure, respectively.

D <- "default"
Lx <- function(x){return(x)}

Finally, use KoulLrMde to obtain m.d. estimator of β and residuals.

MDEResult <- KoulLrMde(Y,X,D,Lx)
betahat <- MDEResult$betahat
resid <- MDEResult$residual

We finish this section by comparing the performance of OLS and KoulLrMde. Let \( f_N, f_{La}, f_{Lo} \) denote density functions of normal, Laplace, and logistic random variables, respectively. Then,

\[
\begin{align*}
  f_N(x) & := (2\sigma_1)^{-1/2} \exp\left(-\frac{(x - \mu_1)^2}{2\sigma_1^2}\right), \\
  f_{La}(x) & := (2\sigma_2)^{-1} \exp\left(-\frac{|x - \mu_2|}{\sigma_2}\right), \\
  f_{Lo}(x) & := \sigma_3^{-1} \exp\left(-\frac{(x - \mu_3)}{\sigma_3}\right)/(1 + \exp(-\frac{(x - \mu_2)}{\sigma_3}))^2.
\end{align*}
\]

When we generate errors or innovations in the subsequent sections, we set \( \mu_i = 0, i = 1, 2, 3 \); we use 5 for \( \sigma_i, i = 1, 2, 3 \). We repeat above example 1000 times and obtain OLS and MD estimators each time. Table 2 reports bias, standard error (SE), and mean squared error (MSE) of OLS and MD estimators corresponding to normal, Laplace, and logistic errors. To obtain MD estimator from KoulLrMde, we chose \( H(x) \equiv x \). As shown in Table 2 the corresponding MD method is superior to OLS in terms of MSE when the regression error distribution is Laplace or logistic, while the opposite is true for normal errors. If we judge superiority in terms of bias and SE separately, a similar conclusion is made; MD estimators for all \( \beta_i \)'s display smaller SE's when error is Laplace or Logistic; the opposite is again true when error is normal. In terms of the bias, it is hard to judge the superiority when error is Laplace or logistic. However, not surprisingly, it is easily seen that OLS displays smaller bias for all \( \beta_i \)'s for the case of normal error.

2.2 KoulArMde

Consider the following autoregressive model of known order \( q \)

\[ X_i = Z'_i \rho + \xi_i, \]

where \( \rho = (\rho_1, ..., \rho_q)' \in \mathbb{R}^q \), and the innovations \( \xi_i \)'s are identically and independently distributed random variables. Assume that \( \xi_i \) is independent of \( Z_i = (X_{i-1}, ..., X_{i-q})' \in \mathbb{R}^q \). Furthermore, assume that \( \xi_0 \) is symmetric around 0. The OLS estimation method gives an estimator \( \left( \sum_{i=1}^{n} Z_i Z_i' \right)^{-1} \sum_{i=1}^{n} Z_i X_i \), which is consistent for \( \rho \). However, this estimator displays poor efficiency for contaminated Gaussian innovations: see, e.g., Fox (1972) and Denby and Martin (1979). Seeking alternative estimation methods, we propose KoulArMde, an analogue of KoulLrMde, which provides MD estimators of autoregressive parameters in the model \( Z \). We compare KoulArMde with R package arima-which gives another alternative method of estimation- at the end of this section; we empirically verify that KoulArMde is superior to arima.
| Innovation Parameter | OLS | KoullLrMde |
|----------------------|-----|-----------|
|                      | bias | SE | MSE | bias | SE | MSE |
| Normal               | 0.0389 | 1.9404 | 3.7666 | 0.0629 | 2.0054 | 4.0257 |
|                      | 0.0015 | 0.0513 | 0.0026 | 0.0011 | 0.0532 | 0.0028 |
|                      | -0.0007 | 0.0505 | 0.0025 | -0.0013 | 0.0519 | 0.0027 |
| Laplace              | 0.0034 | 2.7448 | 7.5338 | -0.0064 | 2.3735 | 5.6334 |
|                      | -0.0017 | 0.0684 | 0.0047 | -0.0011 | 0.0595 | 0.0035 |
|                      | 0.0010 | 0.0710 | 0.0050 | 0.0010 | 0.0617 | 0.0038 |
| Logistic             | 0.0147 | 3.4372 | 11.8147 | 0.0377 | 3.3857 | 11.4643 |
|                      | -0.0041 | 0.0917 | 0.0084 | -0.0043 | 0.0901 | 0.0081 |
|                      | 0.0009 | 0.0918 | 0.0084 | 0.0005 | 0.0897 | 0.0080 |

Table 2: OLS vs KoullLrMde: Bias, SE, and MSE

Let $g \in \mathcal{R}$ be a measurable function. Define the class of distances and the corresponding MD estimators, respectively, to be

$$
M_g(r) := \sum_{k=1}^{q} \int \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_{i-k}) \left\{ I(X_i - Z_i r \leq y) - I(-X_i + Z_i r < y) \right\} \right]^2 dH(y),
$$

$$
M_g(\hat{\rho}_g) := \inf_r M_g(r), \quad r \in \mathcal{R}^q
$$

Among a class of estimators $\{\hat{\rho}_g : g \in \mathcal{R}\}$, Koul (1986) showed MD estimator obtained by taking $g(x) \propto x$ has the smallest asymptotic variance, for every given $H$ satisfying the assumed conditions. Hence, we preset $g(x) := x$ in KoullArMde for estimating $\rho$. Similar to $T$, choosing optimal $\sigma$-finite and symmetric measure $H$ results in well-celebrated estimators. The estimators $\hat{\rho}_g$’s corresponding to the $H$ equivalent to Lebesgue measure and the degenerate measure at 0 are analogues of the Hodges-Lehmann and LAD estimators, respectively. See, e.g., Chapter 7 of Koul (2002) for the detail. Table 3 shows the summary of KoullArMde: its arguments and return values.

| Usage          | KoullArMde(X, AR_Order, IntMeasure). |
|----------------|--------------------------------------|
| Arguments      |                                       |
| $X$            | Vector of $n$ observed values.        |
| AR_Order       | Order of the autoregression model.    |
| IntMeasure     | Symmetric and $\sigma$-finite measure.|
| Return Values  |                                       |
| rhohat         | Minimum distance estimators of $\rho$.|
| residual       | Residuals after minimum distance estimation. |

Table 3: Summary of KoullArMde

The following example describes the usage of KoullArMde with degenerate measure at 0. Define degenerate measure $\text{degenx}$ and pass it to the function.

```r
> degenx = function(x){
> if(x==0){return(1)}
> else{return(0)}
> }
```

Generate model (2) with $n = 100$, $q = 4$, and logistic innovation.

```r
> n <- 100
> q <- 4
```
> rho <- c(-0.2, 0.8, 0.4, -0.7)
> eps <- rlogis(n, 0, 5)
>
> X <- rep(0, times=n)
> for (i in 1:n){
>   tempCol <- rep(0, times=q)
>   for (j in 1:q){
>     if(i-j<=0){
>       tempCol[j] <- 0
>     }else{
>       tempCol[j] <- X[i-j]
>     }
>   }
>   X[i] <- t(tempCol) %*% rho + eps[i]
> }
>
> MDEResult <- KoulArMde(X, q, degenx)
> rhohat <- MDEResult$rhohat
> rhohat
>
> Corresponding to normal, Laplace, and logistic innovations, we repeat the above example 1000 times and obtain estimators of $\rho = (\rho_1, \rho_2, \rho_3, \rho_4)'$ from arima and KoulArMde with $H(x) \equiv x$ and with $H$=degenerate measure at 0. Table 4 compares arima and KoulArMde in terms of bias, SE, and MSE. As shown in the table, KoulArMde outperforms arima at all innovations. Regardless of innovations, KoulArMde corresponding to the degenerate $H$ is superior to all of the other estimators considered here in terms of SE and MSE. Putting it aside, KoulArMde corresponding to $H(x) \equiv x$ is still superior to arima for all innovations. In terms of bias, KoulArMde corresponding to $H(x) \equiv x$ shows smaller bias for all $\rho_i$'s than one corresponding to $H =$degenerate measure when innovations is Laplace; the opposite is true for the case of normal and logistic innovations. arima displays lack of competition due to both poor bias and SE. Therefore, we finish this section by concluding there is no doubt about the superiority of KoulArMde to arima.

2.3 Koul2StageMde

Consider the linear regression model (1) where $\varepsilon_i$'s obey the autoregressive model (2) with known order $q$. The OLS method gives unbiased and consistent estimators of regression parameters for a large class of error distributions, but at non-Gaussian errors, they are inefficient and hence can lead to inaccurate inference procedures. When the presence of autoregressive errors is suspected, alternative methods such as generalized
least squares (GLS) are recommended to use. If \( \rho \) is known, i.e., if there is no need to estimate variance-covariance matrix of error, direct application of GLS yields an efficient estimator. In practice, \( \rho \) is rarely known, and hence, it needs to be estimated along with \( \beta \). In econometrics literature, Cochrane-Orcutt (CO) iterative estimation procedure has been the most popular method; it provides estimators of both \( \rho \) and \( \beta \). In this section, we propose \texttt{Koul2StageMde} as a competing method and show that it still remains competitive under various autoregressive errors. \texttt{Koul2StageMde} is an analogue of the CO procedure but gives more efficient estimators of \( \rho \) and \( \beta \) at some error distributions.

To describe these MD estimators, rewrite the given regression-autoregressive model as

\[
Y_i - \rho_1 Y_{i-1} - \cdots - \rho_q Y_{i-q} = (x_i - \rho_1 x_{i-1} - \cdots - \rho_q x_{i-q})' \beta + \xi_i. \tag{3}
\]

\texttt{Koul2StageMde} replaces \( \rho \) in (3) with a consistent estimator and applies MD estimation method to (3) while CO procedure is an application of OLS method to (3). Since MD estimation method provides more efficient estimator than OLS in linear regression model with logistic or Laplace innovations as shown in Section 2.1 \texttt{Koul2StageMde} is expected to yield more efficient estimators than CO procedure at these error distributions. The following describes two stage algorithm of the function \texttt{Koul2StageMde}.

Stage 1:

(i) Obtain \( \hat{\beta}^{(1)} \in \mathbb{R}^p \) by \texttt{KoulLrMde}.

(ii) Obtain residuals \( \tilde{\varepsilon}^{(1)} = (\tilde{\varepsilon}_1^{(1)}, \ldots, \tilde{\varepsilon}_q^{(1)})' \in \mathbb{R}^q \).

(iii) Pass \( \tilde{\varepsilon}^{(1)} \) to \texttt{KoulArMde} and obtain \( \hat{\rho}^{(1)} = (\hat{\rho}_1^{(1)}, \ldots, \hat{\rho}_q^{(1)})' \in \mathbb{R}^q \).

Stage 2:

(i) Define \( \tilde{Y} := (\tilde{Y}_1, \ldots, \tilde{Y}_n-q)' \in \mathbb{R}^{n-q} \) and \( \tilde{X} := [\tilde{x}_1 \ \tilde{x}_2 \ \cdots \ \tilde{x}_{n-q}]' \in \mathbb{R}^{(n-q) \times p} \) where

\[
\tilde{Y}_i := Y_{i+q} - (\hat{\rho}_1^{(1)} Y_{i+q-1} + \hat{\rho}_2^{(1)} Y_{i+q-2} + \cdots + \hat{\rho}_q^{(1)} Y_i), \\
\tilde{x}_i := x_{i+q} - (\hat{\rho}_1^{(1)} x_{i+q-1} + \hat{\rho}_2^{(1)} x_{i+q-2} + \cdots + \hat{\rho}_q^{(1)} x_i),
\]

for \( 1 \leq i \leq (n-q) \).

(ii) Pass \( \tilde{Y} \) and \( \tilde{X} \) to \texttt{KoulLrMde} and obtain \( \hat{\beta}^{(2)} \).

(iv) Obtain residuals \( \tilde{\varepsilon}^{(2)} := (\tilde{\varepsilon}_1^{(2)}, \ldots, \tilde{\varepsilon}_n^{(2)})' \in \mathbb{R}^n \).

(v) Pass \( \tilde{\varepsilon}^{(2)} \) to \texttt{KoulArMde} and obtain \( \hat{\rho}^{(2)} \).

\texttt{Koul2StageMde} combines \texttt{KoulLrMde} and \texttt{KoulArMde} together, and hence, estimates both regression and autoregression parameters through two stages. Table 5 summarizes \texttt{Koul2StageMde}.

The following example shows the usage of \texttt{Koul2StageMde}.

```r
> n <- 50
> p <- 4
> X <- matrix(runif(n*p, 0,50), nrow=n, ncol=p) #### Generate n-by-p design matrix X
> beta <- c(-2, 0.3, 1.5, -4.3) #### true beta = (-2, 0.3, 1.5, -4.3)'
> q <- 1
> rho <- 0.4 #### true rho = 0.4
> Xi <- rnorm(n, 0,5)
> eps <- rep(0, times=n)
> for (i in 1:n){
>   tempCol <- rep(0, times=q)
>   for (j in 1:q){
```
Usage

Koul2StageMde(Y, X, D, RegIntMeasure, AR_Order, ArIntMeasure).

Arguments

Y
Vector of response variables in linear regression model.
X
Design matrix of explanatory variables in linear regression model.
D
Weight matrix. Dimension of D should match that of X.
"default" uses XD where A=(X’X)^{-1/2}.
RegIntMeasure
Symmetric and \(\sigma\)-finite measure used for estimating \(\beta\).
AR_Order
Order of the autoregressive error.
ArIntMeasure
Symmetric and \(\sigma\)-finite measure used for estimating autoregressive coefficients.

Return Values

MDE1stage
The list of betahat1stage, residual1stage, and rhohat1stage.
betahat1stage
The first stage minimum distance estimators of regression coefficients.
residual1stage
Residuals after the first stage minimum distance estimation.
rhohat1stage
The first stage minimum distance estimators of autoregressive coefficients.

MDE2stage
The list of betahat2stage, residual2stage, and rhohat2stage.
betahat2stage
The second stage minimum distance estimators of regression coefficients.
residual2stage
Residuals after the second stage minimum distance estimation.
rhohat2stage
The second stage minimum distance estimators of autoregressive coefficients.

Table 5: Summary of Koul2StageMde

We finish this section with reporting our findings in the simulation. We use \texttt{R} package \texttt{orcutt} for CO procedure. Similar to previous sections, we repeat the above example 1000 times and obtain estimators from Koul2StageMde and \texttt{orcutt}. Table 6 shows a comparison between Koul2StageMde and \texttt{orcutt} in terms of bias, SE, and MSE. The simulation result is consistent with earlier literature \cite{Econometrics} in the \texttt{MDE} procedure in the case of normal innovation.
| Innovation Parameter | orcutt        | MSE | Kou12StageMde  | MSE |
|----------------------|--------------|-----|----------------|-----|
|                      | bias         | SE  | bias           | SE  |
| Normal               | -0.1008      | 2.4235 | 5.8837    | -0.0404 | 2.5159 | 6.3312 |
|                      | 0.0111       | 0.0498 | 0.0025   | 0.0007 | 0.0521 | 0.0027 |
|                      | 0.0010       | 0.0494 | 0.0024   | 0.0005 | 0.0521 | 0.0027 |
|                      | 0.0015       | 0.0475 | 0.0023   | 0.0009 | 0.0497 | 0.0025 |
|                      | -0.0334      | 0.1521 | 0.0242  | -0.0390 | 0.1568 | 0.0261 |
| Laplace              | -0.0619      | 3.3641 | 11.3208  | -0.0512 | 2.9165 | 8.5087 |
|                      | 0.0051       | 0.0691 | 0.0048   | 0.0036 | 0.0599 | 0.0036 |
|                      | -0.0032      | 0.0669 | 0.0045   | -0.0030 | 0.0601 | 0.0036 |
|                      | 0.0006       | 0.0679 | 0.0046   | 0.0011 | 0.0597 | 0.0036 |
|                      | -0.0284      | 0.1332 | 0.0185  | -0.0314 | 0.1227 | 0.0160 |
| Logistic             | 0.1422       | 4.2534 | 18.1119  | 0.1338 | 4.1827 | 17.5131 |
|                      | -0.0028      | 0.0896 | 0.0080   | -0.0032 | 0.0895 | 0.0080 |
|                      | -0.0024      | 0.0847 | 0.0072   | -0.0020 | 0.0838 | 0.0070 |
|                      | 0.0005       | 0.0866 | 0.0075   | 0.0011 | 0.0865 | 0.0075 |
|                      | -0.0299      | 0.1416 | 0.0209   | -0.0349 | 0.1392 | 0.0206 |

Table 6: orcutt vs Kou12StageMde: Bias, SE, and MSE

3 Conclusion

This article discussed R package Kou1Mde which performs the MD estimation for linear regression and autoregressive models. This package contains three functions: Kou1LrMde, Kou1ArMde, and Kou12StageMde. The former two provide MD estimators of parameters in linear regression and autoregressive models, respectively. Kou12StageMde deals with linear regression model with autoregressive errors and estimates both regression and autoregressive parameters. Compared to other estimation methods and R packages, this package remains competitive in that it outperforms them; our findings in simulation studies show all Kou1LrMde, Kou1ArMde, and Kou12StageMde display better MSE than competing methods or R packages.

References

[1] Denby, L. and Martin, R. (1979). Robust estimation of the rst-order autoregressive parameter. J. Amer. Statist. Assoc., 74, 140-146.

[2] Fox, A. J. (1972). Outliers in time series. J. Roy. Statist. Soc. Ser. B., 34 350-363.

[3] Koul, H. L. (1985). Minimum distance estimation in linear regression with unknown error distributions. Statist. Probab. Lett., 3 1-8.

[4] Koul, H. L. (1986). Minimum distance estimation and goodness-of-fit tests in first-order autoregression. Ann. Statist., 14 1194-1213.

[5] Koul, H. L. (2002). Weighted empirical process in nonlinear dynamic models. Springer, Berlin, Vol. 166.

[6] Koul, H. L. and De Wet (1983). Minimum distance estimation in a linear regression model. Ann. Statist., 11 921-932.

[7] Parr, W. C. and Schucany, W. R. (1979). Minimum distance and robust estimation. J. Amer. Statist. Assoc., 75 616-624.

[8] Woffowitz, J. (1957). Minimum distance method. Ann. Math. Statist., 28 75-88.