Journal of the Korean Statistical Society (2007), **36***: 3, pp 387–395*

A DOUBLY ROBUSTIFIED ESTIMATING FUNCTION FOR ARCH TIME SERIES MODELS†

SAHM KIM[1] AND S. Y. HWANG[2]

ABSTRACT

We propose a doubly robustified estimating function for the estimation of parameters in the context of ARCH models. We investigate asymptotic properties of estimators obtained as solutions of robust estimating equations. A simulation study shows that robust estimator from specified doubly robustified estimating equation provides better performance than conventional robust estimators especially under heavy-tailed distributions of innovation errors.

AMS 2000 subject classifications. Primary 62M10; Secondary 62F10.
Keywords. ARCH model, doubly robustified estimating function, Huber's function.

1. INTRODUCTION

It is well known that usual estimation methods such as maximum likelihood, least squares may be very sensitive to outliers in the observations. In this case, one of the methods which resolves the problem is so called the robust estimation method. There are many sources in robust estimation literature such as, Denby and Martin (1979), Huber (1981) and Basawa *et al.* (1985). Also, Kulkarni and Heyde (1987) proposed an optimal robust estimation for time series models based on the framework of Godambe (1985)'s estimating function.

Chan and Cheung (1994) studied robust estimation for threshold autoregressive models. They investigated effects of additive outliers (AO) based on generalized-M (GM) estimate for models. They argued that GM approach may be preferable to least squares estimation method under the existence of additive outliers (AO) in the observations.

Received January 2007; accepted April 2007.
†This work was supported by a grant from KRF (2005-070-C00022).
[1]Department of Statistics, Chung-Ang University, Seoul 155-756, Korea
[2]Corresponding author. Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea (e-mail: shwang@sookmyung.ac.kr)

Muler and Yohai (2002) proposed two robust estimates such as τ and filterd τ-estimates for ARCH(p) models. They presented that proposed estimates work better than maximum likelihood estimate even under the small percentage of outlier contamination. Hui and Jiang (2005) investigated robust estimation based L_1-norm for double-threshold autoregressive conditional heteroscedastic (DTARCH) model which was originally proposed by Li and Li (1996).

In this paper, following similar ideas of Basawa *et al.* (1985), we will propose a doubly robustified estimating function which is relatively new in the ARCH context. We will investigate asymptotic properties of estimators obtained as solutions of the doubly robustified estimating functions. Small simulation study shows that the proposed estimators outperform others.

2. DOUBLY ROBUSTIFIED ESTIMATING FUNCTIONS

In this section, we are dealing with the following AR-ARCH model which was proposed by Engle (1982). This model will be called by (M).

$$(M) \quad y_t = \phi y_{t-1} + \varepsilon_t,$$

$$\varepsilon_t = \sqrt{h_t} e_t, \quad h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \quad e_t \sim i.i.d.\ N(0,1), \quad |\phi| < 1.$$

Consider the following doubly robustified estimating function (DREF, for short) given by

$$S_n(\theta) = \sum_{t=1}^{n} \Psi_1 \left(\frac{\varepsilon_t(\theta)}{\sqrt{h_t}} \right) \Psi_2 \left(\frac{\partial}{\partial \theta} \frac{\varepsilon_t(\theta)}{\sqrt{h_t}} \right), \tag{2.1}$$

where $\theta = (\phi, \alpha_0, \alpha_1)$ and Ψ_1 and Ψ_2 are bounded and continuous functions on θ. Typically Ψ_1 and Ψ_2 are chosen as standard Huber's function. The term "doubly" refers to two robust functions Ψ_1 and Ψ_2 instead of single robust function commonly used in the literature. When $\Psi_1(x) = \Psi_2(x) = x$, $S_n(\theta)$ is the same as the least squares (LS) estimating function. If $\Psi_2(x) = x$, and $\Psi_1(x)$ is a bounded function, then (2.1) reduces to those of Denby and Martin (1979). As Kulkarni and Heyde (1987) pointed out, DREF defined in (2.1) is of special interest to compare the performance of the proposed robust EF with that of Denby and Martin (1979).

The following conditions are imposed for deriving asymptotic results.

1. $\{y_t\}$ is stationary and ergodic with $Ey_t^2 < \infty$,

2. Ψ_1 and Ψ_2 are bounded and continuously differentiable with respect to $\theta \in \Theta$,

3. $E\Psi_i(\cdot) = 0$, $i = 1, 2$,

4. $|A(\tilde{\theta_n}) - A(\theta)| \xrightarrow{p} 0$ and

5. $\left|-\frac{1}{n}S_n(\tilde{\theta_n}) - A(\theta)\right| \xrightarrow{p} 0$,

where $\tilde{\theta_n}$ is a $\sqrt{n}$-consistent estimator of θ and $A(\theta) = E(S_n'(\theta))$ where the prime ($'$) indicates differentiation. Conditions 4 and 5 refer to smoothness conditions on $S_n(\theta)$.

We are now in a position to present asymptotic normality of DREF $S_n(\theta)$ in (2.1).

LEMMA 2.1. *Under conditions 1–3, we have as sample size n goes to infinity,*

$$\frac{1}{\sqrt{n}}S_n(\theta) \xrightarrow{d} N(0, B(\theta)),$$

where $B(\theta) = E(S_n(\theta)S_n(\theta)^T)$ *with* T *denoting transpose of the vector.*

PROOF. Notice that $S_n(\theta)$ is a sum of martingale differences due to condition 3. It follows from conditions 1 and 2 that a sequence of martingale differences

$$\left\{\Psi_1\left(\frac{\varepsilon_t(\theta)}{\sqrt{h_t}}\right)\Psi_2\left(\frac{\partial}{\partial\theta}\frac{\varepsilon_t(\theta)}{\sqrt{h_t}}\right)\right\}$$

forms a stationary and ergodic time series. Hence martingale central limit theorem tells us

$$\frac{1}{\sqrt{n}}S_n(\theta) \xrightarrow{d} N(0, B(\theta)),$$

where $B(\theta) = E(S_n(\theta)S_n(\theta)^T)$ of which the existence is ensured by condition 1, completing the proof. □

Let $\widehat{\theta_n}$ denote the solution of $S_n(\theta) = 0$. The following theorem establishes consistency and asymptotic normality of $\widehat{\theta_n}$ which is the solution of DREF.

THEOREM 2.1. *Under conditions 1–5, as n tends to infinity,* *(1)* $\widehat{\theta}_n \xrightarrow{p} \theta$,

(2) $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, \{A(\theta)B^{-1}(\theta)A(\theta)\}^{-1})$.

PROOF. It will be assumed that $S_n(\theta) = 0$ provides a unique solution $\widehat{\theta}_n$. This will be relaxed later. Using the Taylor's expansion of $S_n(\theta)$ about $\theta = \widehat{\theta}_n$, one can write

$$-S_n(\theta) = S_n'(\theta_n^*)\left(\widehat{\theta}_n - \theta\right), \tag{2.2}$$

where $|\theta_n^* - \theta| \leq |\widehat{\theta}_n - \theta|$. Consequently

$$\sqrt{n}(\widehat{\theta}_n - \theta) = \frac{\left[-n^{-1}S_n'(\theta_n^*)\right] S_n(\theta)}{\sqrt{n}}.$$

It then follows from smoothness conditions 4 and 5 that

$$\sqrt{n}(\widehat{\theta}_n - \theta) = \frac{[-n^{-1}S_n'(\theta)]S_n(\theta)}{\sqrt{n}} + o_p(1). \tag{2.3}$$

Combining Lemma 2.1 and condition 4, assertion (2) readily follows. The result (1) is an immediate consequence of (2). Even for the case when $\widehat{\theta}_n$ is a near zero of $S_n(\theta) = 0$, *i.e.*, $S_n(\widehat{\theta}_n) = o_p(1)$, (2.2) continues to hold asymptotically, *i.e.*,

$$-S_n(\theta) = S_n'(\theta_n^*)(\widehat{\theta}_n - \theta) + o_p(1)$$

and in turn (2.3) remains valid. Thus, results (1) and (2) again hold. □

3. SIMULATION STUDY

To evaluate the performance of robust estimators from DREF, we consider three different estimating functions given below. We compare proposed estimator $\widehat{\theta}_n$ as a solution of $S_{33}(\theta) = 0$ defined in (3.3) with other estimators based the first two estimating functions appeared in (3.1) and (3.2).

$$S_{11}(\theta) = \sum \frac{(y_t - \phi y_{t-1})}{h_t}\left(\frac{\partial}{\partial\theta}\varepsilon_t(\theta)\right), \tag{3.1}$$

$$S_{22}(\theta) = \sum \Psi_1\left(\frac{y_t - \phi y_{t-1}}{\sqrt{h_t}}\right)\left(\frac{\partial}{\partial\theta}\frac{\varepsilon_t(\theta)}{\sqrt{h_t}}\right), \tag{3.2}$$

$$S_{33}(\theta) = \sum \Psi_1\left(\frac{y_t - \phi y_{t-1}}{\sqrt{h_t}}\right)\Psi_2\left(\frac{\partial}{\partial\theta}\frac{\varepsilon_t(\theta)}{\sqrt{h_t}}\right). \tag{3.3}$$

Here Ψ_1 and Ψ_2 are chosen as standard Huber function given by

$$\Psi_1(x) = \Psi_2(x) = \begin{cases} x, & \text{if } |x| \le k, \\ k, & \text{if } x > k,\ k > 0, \\ -k, & \text{if } x < -k. \end{cases}$$

Estimation of $\theta = (\phi, \alpha_0, \alpha_1)$ is carried out based on the following one step scoring equation.

$$\widehat{\theta_n} = \widehat{\theta_{LS}} - \left(\frac{S_n'(\widehat{\theta_{LS}})}{n} \right)^{-1} \frac{S_n(\widehat{\theta_{LS}})}{n}, \tag{3.4}$$

where $\widehat{\theta_{LS}}$ is a preliminary $\sqrt{n}$-consistent LS estimator of θ which can be used as initial estimator for parameter of interest. In this simulation study, we focus on ϕ and other parameters are treated as nuisance parameters. For each estimating functions (3.1) to (3.3), sample size is given by 200 and number of iterations to get sample means and standard deviations is set to be 30. After some calculations, we use

$$S_{33}(\theta) = \begin{pmatrix} S_1(\theta) \\ S_2(\theta) \\ S_3(\theta) \end{pmatrix} = \begin{pmatrix} \Psi_1\left(\frac{\varepsilon_t(\theta)}{\sqrt{h_t}}\right) \Psi_2\left(\frac{\partial}{\partial \phi} \frac{\varepsilon_t(\theta)}{\sqrt{h_t}}\right) \\ \Psi_1\left(\frac{\varepsilon_t(\theta)}{\sqrt{h_t}}\right) \Psi_2\left(\frac{\partial}{\partial \alpha_0} \frac{\varepsilon_t(\theta)}{\sqrt{h_t}}\right) \\ \Psi_1\left(\frac{\varepsilon_t(\theta)}{\sqrt{h_t}}\right) \Psi_2\left(\frac{\partial}{\partial \alpha_1} \frac{\varepsilon_t(\theta)}{\sqrt{h_t}}\right) \end{pmatrix}$$

and

$$S_{33}'(\theta) = \begin{pmatrix} \frac{\partial S_1}{\partial \phi} & \frac{\partial S_1}{\partial \alpha_0} & \frac{\partial S_1}{\partial \alpha_1} \\ \frac{\partial S_2}{\partial \phi} & \frac{\partial S_2}{\partial \alpha_0} & \frac{\partial S_2}{\partial \alpha_1} \\ \frac{\partial S_3}{\partial \phi} & \frac{\partial S_3}{\partial \alpha_0} & \frac{\partial S_3}{\partial \alpha_1} \end{pmatrix}.$$

3.1. Case 1: Normal errors

Consider the case when $\{e_t\}$ is *i.i.d.* N(0,1). Based on the simulated time series of length $n = 200$, we obtained $\widehat{\phi_n}$ for each estimating functions S_{11}, S_{22} and S_{33}. In this simulation, fix $\alpha_0 = 0.2$ and $\alpha_1 = 0.2$ and they are treated as the nuisance parameters. Next, mean and standard deviation (S.D) are calculated based on 30 replications. One can notice from Tables 3.1 and 3.3 that DREF

TABLE 3.1 $n = 200, k = 1.5$ *(normal distribution)*

EF	S_{11}		S_{22}		S_{33}	
ϕ	*mean*	*S.D*	*mean*	*S.D*	*mean*	*S.D*
0.3	0.2910	0.0738	0.2978	0.0752	0.2917	0.0668
0.5	0.4887	0.0502	0.4907	0.0640	0.4868	0.0527
0.7	0.6894	0.0324	0.6988	0.0440	0.6878	0.0416
0.9	0.8931	0.0369	0.9008	0.0407	0.8953	0.0370

TABLE 3.2 $n = 200, k = 2$ *(normal distribution)*

EF	S_{11}		S_{22}		S_{33}	
ϕ	*mean*	*S.D*	*mean*	*S.D*	*mean*	*S.D*
0.3	0.2910	0.0738	0.2888	0.0711	0.2924	0.0645
0.5	0.4887	0.0502	0.4878	0.0634	0.4890	0.0511
0.7	0.6894	0.0324	0.6969	0.0426	0.6886	0.0385
0.9	0.8931	0.0349	0.9000	0.0366	0.8959	0.0351

TABLE 3.3 $n = 200, k = 2.5$ *(normal distribution)*

EF	S_{11}		S_{22}		S_{33}	
ϕ	*mean*	*S.D*	*mean*	*S.D*	*mean*	*S.D*
0.3	0.2910	0.0738	0.2946	0.0743	0.2917	0.0641
0.5	0.4887	0.0502	0.4892	0.0618	0.4899	0.0487
0.7	0.6894	0.0324	0.6967	0.0376	0.6892	0.0369
0.9	0.8931	0.0369	0.8988	0.0375	0.8952	0.0342

S_{33} provides more accurate $\widehat{\phi_n}$ in the sense of the smaller S.D. Also, across all parameter values $\phi = 0.3, 0.5, 0.7$ and 0.9, as k increases, biases of $\widehat{\phi_n}$ from DREF S_{33} tend to be smaller than those for S_{11} and S_{22}. It is however noticed that S_{33} gains slight edge over S_{11} and S_{22} for small k, say, $k = 1.5$.

3.2. Case 2: t-distributions

Consider the case where $\{e_t\}$ is *i.i.d.* t-distribution with degrees of freedom $3[t(3)$, see Tables 3.4 to 3.6] and $5[t(5)$, refer to Tables 3.7 to 3.9], respectively. Variances of $t(3)$ and $t(5)$ are standardized as unity. It is noted for heavy-tailed distribution $t(3)$ that a remarkable improvement of the accuracy (smaller S.D) is achieved for S_{33}, compared to S_{11} and S_{22}. Also, for most values of $\phi = 0.3$ to $\phi = 0.9$, $\widehat{\phi_n}$ for S_{33} reveals the smaller bias than S_{11} and S_{22}. To summarize,

TABLE 3.4 $n = 200, k = 1.5$ *(t(3) distribution)*

EF	S_{11}		S_{22}		S_{33}	
ϕ	*mean*	*S.D*	*mean*	*S.D*	*mean*	*S.D*
0.3	0.2789	0.0884	0.2768	0.0878	0.2832	0.0518
0.5	0.4866	0.0647	0.4880	0.0660	0.4868	0.0419
0.7	0.6936	0.0450	0.6839	0.0554	0.6933	0.0353
0.9	0.8960	0.0206	0.8993	0.0220	0.8932	0.0206

TABLE 3.5 $n = 200, k = 2$ *(t(3) distribution)*

EF	S_{11}		S_{22}		S_{33}	
ϕ	*mean*	*S.D*	*mean*	*S.D*	*mean*	*S.D*
0.3	0.2789	0.0884	0.2897	0.0870	0.2824	0.0529
0.5	0.4866	0.0647	0.4879	0.0720	0.4881	0.0419
0.7	0.6936	0.0450	0.6916	0.0422	0.6922	0.0361
0.9	0.8960	0.0266	0.8994	0.0290	0.8938	0.0261

TABLE 3.6 $n = 200, k = 2.5$ *(t(3) distribution)*

EF	S_{11}		S_{22}		S_{33}	
ϕ	*mean*	*S.D*	*mean*	*S.D*	*mean*	*S.D*
0.3	0.2789	0.0884	0.2900	0.0839	0.2846	0.0564
0.5	0.4866	0.0647	0.4816	0.0707	0.4881	0.0450
0.7	0.6936	0.0450	0.6949	0.0408	0.6923	0.0373
0.9	0.8960	0.0266	0.8999	0.0256	0.8941	0.0250

TABLE 3.7 $n = 200, k = 1.5$ *(t(5) distribution)*

EF	S_{11}		S_{22}		S_{33}	
ϕ	*mean*	*S.D*	*mean*	*S.D*	*mean*	*S.D*
0.3	0.3198	0.1145	0.3143	0.0665	0.3163	0.0604
0.5	0.5088	0.1004	0.5059	0.0484	0.5087	0.0469
0.7	0.6999	0.0773	0.7014	0.0375	0.7021	0.0371
0.9	0.8939	0.0449	0.8959	0.0272	0.8943	0.0285

proposed DREF S_{33} outperforms conventional functions S_{11} and S_{22} especially when the error distribution is heavy-tailed such as $t(3)$ and $t(5)$.

TABLE 3.8 $n = 200, k = 2$ *(t(5) distribution)*

EF	S_{11}		S_{22}		S_{33}	
ϕ	*mean*	*S.D*	*mean*	*S.D*	*mean*	*S.D*
0.3	0.3198	0.1145	0.3163	0.0572	0.3157	0.0603
0.5	0.5088	0.1004	0.5083	0.0430	0.5092	0.0430
0.7	0.6999	0.0773	0.7021	0.0366	0.7022	0.0364
0.9	0.8939	0.0449	0.8966	0.0269	0.8950	0.0259

TABLE 3.9 $n = 200, k = 2.5$ *(t(5) distribution)*

EF	S_{11}		S_{22}		S_{33}	
ϕ	*mean*	*S.D*	*mean*	*S.D*	*mean*	*S.D*
0.3	0.3198	0.1145	0.3171	0.0577	0.3163	0.0605
0.5	0.5088	0.1004	0.5100	0.0434	0.5103	0.0407
0.7	0.6999	0.0773	0.7037	0.0361	0.7037	0.0329
0.9	0.8939	0.0449	0.8971	0.0277	0.8957	0.0278

4. CONCLUSIONS

We propose a doubly robustified estimating function for ARCH models and investigate asymptotic properties of estimators which are the solutions of robust estimating equation. Under the normal distribution of innovation errors, LS estimator works fine, but we can see better performance of proposed estimators than LS estimator as well as partially robust estimator under heavy-tailed distributions of innovation errors. This study can be extended to cover more complicated models.

ACKNOWLEDGEMENTS

We thank the referee for careful reading of the paper. The contents of the paper were presented at the Autumn Conference, 2006 of the Korean Statistical Society. This work was supported by a grant from KRF (2005 -070-C00022).

REFERENCES

BASAWA, I. V., HUGGINS, R. M. AND STAUDTE, R. G. (1985). "Robust tests for time series with an application to first-order autoregressive processes", *Biometrika*, **72**, 559–571.

CHAN, W.-S. AND CHEUNG, S.-H. (1994). "On robust estimation of threshold autoregressions", *Journal of Forecasting*, **13**, 37–49.

DENBY, L. AND MARTIN, R. D. (1979). "Robust estimation of the first-order autoregressive parameter", *Journal of the American Statistical Association*, **74**,140–146.

ENGLE, R. F. (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, **50**, 987–1007.

GODAMBE, V. P. (1985). "The foundations of finite sample estimation in stochastic processes", *Biometrika*, **72**, 419–428.

HUBER, P. J. (1981). *Robust Statistics*, John Wiley & Sons, New York.

HUI, Y. V. AND JIANG, J. (2005). "Robust modelling of DTARCH models", *The Econometrics Journal*, **8**,143–158.

KULKARNI, P. M. AND HEYDE, C. C. (1987). "Optimal robust estimation for discrete time stochastic processes", *Stochastic Processes and their Applications*, **26**, 267–276.

LI, C. W. AND LI, W. K. (1996). "On a double-threshold autoregressive heteroscedastic time series model", *Journal of Applied Econometrics*, **11**, 253–274.

MULER, N. AND YOHAI, V. J. (2002). "Robust estimates for ARCH processes", *Journal of Time Series Analysis*, **23**, 341–375.