Robust modelling of DTARCH models

YER VAN HUI* AND JIANCHENG JIANG†,

*Department of Management Sciences, City University of Hong Kong, Hong Kong
†LMAM & School of Mathematical Sciences, Peking University, China
E-mail: jiang@math.pku.edu.cn

Received: February 2004

Summary  Autoregressive conditional heteroscedastic (ARCH) models and its extensions are widely used in modelling volatility in financial time series. One of the variants, the double-threshold autoregressive conditional heteroscedastic (DTARCH) model, has been proposed to model the conditional mean and the conditional variance that are piecewise linear. The DTARCH model is also useful for modelling conditional heteroscedasticity with nonlinear structures such as asymmetric cycles, jump resonance and amplitude-frequency dependence. Since asset returns often display heavy tails and outliers, it is worth studying robust DTARCH modelling without specific distribution assumption. This paper studies DTARCH structures for conditional scale instead of conditional variance. We examine $L_1$-estimation of the DTARCH model and derive limiting distributions for the proposed estimators. A robust portmanteau statistic based on the $L_1$-norm fit is constructed to test the model adequacy. This approach captures various nonlinear phenomena and stylized facts with desirable robustness. Simulations show that the $L_1$-estimators are robust against innovation distributions and accurate for a moderate sample size, and the proposed test is not only robust against innovation distributions but also powerful in discriminating the delay parameters and ARCH models. It is noted that the quasi-likelihood modelling approach used in ARCH models is inappropriate to DTARCH models in the presence of outliers and heavy tail innovations.

Keywords: Conditional heteroskedasticity, Double-threshold, Median regression, Model diagnostic checking, Robust portmanteau statistic.

1. INTRODUCTION

Modelling volatility is important in financial data analysis. One of the most widely used tools in modelling the changing volatility is the autoregressive conditional heteroscedasticity (ARCH) model pioneered by Engle (1982). ARCH models and its extensions have been widely applied in finance and econometrics (Bollerslev et al. 1992, 1994; Bera and Higgins 1993; Fan and Yao 2003). Li and Li (1996) proposed a double-threshold autoregressive conditional heteroscedastic (DTARCH) model to study the piecewise linear patterns of the conditional mean and the conditional variance. They studied model identification, estimation and diagnostic check based on the maximum likelihood principle. This approach is useful for detecting nonlinear structures such
as asymmetric behaviour in the mean and the volatility of an asset return, and heteroscedasticity with clustering in the volatility. In practice, it is observed that financial returns tend to have thicker tails than normal distributions. Note that misspecification of the conditional distribution in the likelihood approach may create serious problems in parameter estimation. Moreover, likelihood-based testing methods may fail in detecting false structures in the conditional variance of asset return. It is worth investigating robust modelling techniques without specific distribution assumptions. This motivates us to consider DTARCH models for conditional scale based on $L_1$-norm fitting. The advantage of such an approach was discussed in Jiang et al. (2001) for ARCH models.

$L_1$-norm-based estimation has received a great deal of attention in time series analysis (Davis and Dunsmuir 1997; Peng and Yao 2003). However, the least absolute deviation fitting approach has difficulties in deriving the distributional properties of the estimators and checking the model adequacy while closed-form expressions of the estimators are not available. To solve this problem, we give a Bahadur representation for the proposed estimators in the study. Common portmanteau statistics (Box and Pierce 1970) can be useful in many applications, but they could be inappropriate to $L_1$-norm-fitted models where outliers give large residuals and affect the $p$-value of the test statistic. We derive a robust portmanteau test for DTARCH models in model diagnostic. This test statistic has a known asymptotic distribution for a large variety of error distributions when the model is correct. Simulations show that the proposed portmanteau test is not only robust against outliers and innovation distributions, but also powerful in discriminating the delay parameters and ARCH models. It is also observed that the quasi-likelihood approach used in ARCH models is inappropriate to DTARCH models in the presence of outliers and heavy-tail innovations.

This paper is organized as follows. Section 2 introduces the $L_1$-estimation procedure and gives the Bahadur representations of the estimators. Section 3 discusses the asymptotic properties of the standardized absolute $L_1$-residual autocorrelation and the robust portmanteau statistic. Empirical data analysis and simulations are presented to illustrate the proposed method in Section 4. Section 5 summarizes the research findings and offers conclusions. Technical proofs are presented in the Appendix.

2. $L_1$-ESTIMATION OF DTARCH MODELS

Given a time series $y_t$, $t = 1, \ldots, n$, let $\mathcal{F}_t$ be the $\sigma$-field generated from the realized value $\{y_t, y_{t-1}, \ldots\}$ at time $t$. For each $t$, $\epsilon_t$ is a random variable with median zero and conditional absolute scale $h_t$ given $\mathcal{F}_{t-1}$. Assume that $y_t$ is generated by
\begin{equation}
y_t = X'_{t,j} \alpha^{(j)} + \epsilon_t \quad \text{if} \quad r_{j-1} < y_{t-d} \leq r_j,
\end{equation}
where $j = 1, \ldots, m$; the delay parameter $d$ is a positive integer; the threshold parameters $r_j$ satisfy $-\infty = r_0 < r_1 < r_2 < \cdots < r_m = \infty$; $X_{t,j} = (1, y_{t-1}, \ldots, y_{t-p_j})'$ is a $(p_j + 1) \times 1$ vector of lagged variables; $\alpha^{(j)} = (\alpha_0^{(j)}, \alpha_1^{(j)}, \ldots, \alpha_{p_j}^{(j)})'$ is a $(p_j + 1) \times 1$ parameter vector. The stochastic error satisfies $\epsilon_t = h_t \epsilon_t$ with
\begin{equation}
h_t = \sum_{j=1}^m I_{t,j} [\beta_0^{(j)} + \beta_1^{(j)} |\epsilon_{t-1}| + \cdots + \beta_{q_j}^{(j)} |\epsilon_{t-q_j}|] = \sum_{j=1}^m I_{t,j} Z'_{t,j} \beta^{(j)},
\end{equation}
where $I_{t,j} = I(r_{j-1} < y_{t-d} \leq r_j)$; the parameters in the conditional scales satisfy $\beta_0^{(j)} > 0$, $\beta_i^{(j)} \geq 0$ ($i = 1, \ldots, q_j$) and the innovations $\{\epsilon_t\}$ are independently identically distributed random
variables with an unknown distribution $F(x)$ and a density function $f(u)$. Without loss of generality, we assume the median of $u_t$ is zero and $|u_t| = 1$. For convenience, as in Tsay (1989) and Li and Li (1996), we refer to the model in (1) and (2) as a DTARCH$(p_1, \ldots, p_m; q_1, \ldots, q_m)$ model, where the first $m$ integers $p$’s represent the AR orders in the $m$ regimes and the last $m$ integers $q$’s denote the ARCH orders. The interval $r_{j-1} < y_{t-d} \leq r_j$ is the $j$th regime of $y_t$. The proposed model is similar to that given in Li and Li (1996) where the conditional scale instead of the conditional variance is specified as the ARCH structure.

The distinguished features of the model in (1) and (2) are (i) the conditional scale $h_t$ is a piecewise linear function of the absolute values of the lagged errors and each piece has an ARCH structure, which depicts the clustering of deviations at different regions of the lagged variable $y_{t-d}$; (ii) the double-threshold structure extends Tong’s threshold model in a natural way and is capable to capture nonlinear phenomena such as asymmetric cycles, jump resonance and amplitude-frequency dependence (see Tong and Lim 1980) and (iii) the conditional median zero assumption on the error enables robust inference for the model.

We now consider the $L_1$-estimation of the model. Let $\alpha = \text{vec}(\alpha^{(1)}, \ldots, \alpha^{(m)})$ and $\beta = \text{vec}(\beta^{(1)}, \ldots, \beta^{(m)})$. Denote $X_t = \text{vec}(I_{t,1}X_{t,1}, \ldots, I_{t,m}X_{t,m})$ and $Z_t = \text{vec}(I_{t,1}Z_{t,1}, \ldots, I_{t,m}Z_{t,m})$. Following Jiang et al. (2001), we estimate the parameter vectors $\alpha$ and $\beta$ as follows:

1. Note that the conditional median of $y_t$ given $F_{t-1}$ is $\text{median}(y_t | F_{t-1}) = \sum_{j=1}^{m} I_{t,j}X_j'\alpha^{(j)} = X'_t\tilde{\alpha}$. The median regression provides a direct approach in the estimation of $\alpha$. We have

$$\tilde{\alpha} = \arg \min_{\alpha} \sum_{t=s+1}^{n} |y_t - X_t'\alpha|, \quad (3)$$

where $s = \max(p_1, \ldots, p_m)$.

2. Observe that $\text{median}(|\varepsilon_t| | F_{t-1}) = h_t = \sum_{j=1}^{m} I_{t,j}Z_j'\beta^{(j)} = Z'_t\tilde{\beta}$, this suggests the median regression estimator

$$\tilde{\beta} = \arg \min_{\beta} \sum_{t=s'+1}^{n} ||\hat{\varepsilon}_t| - Z'_t\beta||, \quad (4)$$

where $s' = \max(p_1, \ldots, p_m, q_1, \ldots, q_m)$, $\hat{\varepsilon}_t = y_t - X'_t\tilde{\alpha}$ and $\hat{Z}_t = \text{vec}(I_{t,1}\hat{Z}_{t,1}, \ldots, I_{t,m}\hat{Z}_{t,m})$ with $\hat{Z}_{t,j} = (1, |\hat{\varepsilon}_{t-1}|, \ldots, |\hat{\varepsilon}_{t-q_j}|)'$.

3. For the DTARCH model with conditional scales varying over time, we anticipate that the weighted median regression with estimated weights can improve the unweighted estimator. Specifically, we consider the weighted median regression estimator. Let $\hat{h}_t = \hat{Z}_t'\hat{\beta}$, then the refined estimator of $\beta$ is obtained from

$$\hat{\beta} = \arg \min_{\beta} \sum_{t=s'+1}^{n} ||\hat{\varepsilon}_t| - Z'_t\beta|\hat{h}_t^{-1}. \quad (5)$$

The conditional scale estimates are given by $\hat{h}_t = \hat{Z}_t'\hat{\beta}$.

Note that $\hat{h}_t$ and $\hat{h}_t$ are positive so that the conditional scale cannot be misspecified. The accuracy of the $\beta$ estimators is improved by repeating the minimization in (5) with $\hat{h}_t$ replaced by $\hat{h}_t$.

The computation of the proposed method is simple and easy to implement. The simplex algorithm for linear models developed by Koenker and Bassett (1978) can be used in the
minimization problems in (3)–(5). One can easily compute the estimates using the $l1$ fit-function in Splus.

The following notations and assumptions are made in the study.

(A1) $median(u_i) = 0$ and $median(|u_i|) = 1$.

(A2) $E|y_t|^{2+\delta} < +\infty$, for $\delta > 0$ and $y_t$ is strictly stationary and ergodic.

(A3) $f(u)$ is symmetric and continuous at $u = 0$ and $u = 1$, with $f(0) > 0$ and $f(1) > 0$.

(A4) $E(X_tX_ih_t^{-r}) = D_r$ is positive definite for $r = 0, 1$; $E(Z_iZ_ih_t^{-r}) = G_r$ is positive definite for $r = 0, 1, 2$.

(A5) $u_t$ is independent of $\mathcal{F}_{t-1}$.

The stated assumptions are mild and can be fulfilled in many applications. Condition (A1) is used to identify parameters $\alpha$ and $\beta$. The continuity of $f(u)$ at $u = 0$ and $u = 1$ is required. It appears to be the minimal smoothness assumption on $f(u)$. The symmetry of $f(u)$ in (A3) is not essential. It is stated only for simplicity. Conditions (A2) and (A4) are usual assumptions on $L_1$-norm fitting in ARCH models. A necessary condition for (A4) is that $\lim n_j/n \rightarrow E(I_{i,j}) > 0$, for all $j$’s (see Tsay 1989).

**Theorem 1** Suppose that the threshold and the delay parameters are known. Let $\psi(u) = 1/2 - I(u < 0)$. Under assumptions (A1)–(A5),

(i) $\sqrt{n}(\hat{\alpha} - \alpha) = (D_1^{-1}/f(0))n^{-1/2} \sum_{t=s+1}^{n} X_t\psi(u_t) + o_p(1)$ and

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{D} \mathcal{N} \left( 0, \frac{D_1^{-1}D_0D_1^{-1}}{4f^2(0)} \right);$$

(ii) $\sqrt{n}(\hat{\beta} - \beta) = G_1^{-1}(f(-1) + f(1))^{-1}n^{-1/2} \sum_{t=s+1}^{n} Z_t\psi(|u_t| - 1) + o_p(1)$ and

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} \mathcal{N} \left( 0, \frac{G_1^{-1}G_0G_1^{-1}}{4(f(-1) + f(1))^2} \right);$$

(iii) $\sqrt{n}(\hat{\beta} - \beta) = G_2^{-1}(f(-1) + f(1))^{-1}n^{-1/2} \sum_{t=s+1}^{n} Z_t h_t^{-1}\psi(|u_t| - 1) + o_p(1)$ and

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} \mathcal{N} \left( 0, \frac{G_2^{-1}}{4(f(-1) + f(1))^2} \right).$$

Note that $G_1^{-1}G_0G_1^{-1} - G_2^{-1}$ is non-negative definite, which assures that the weighted estimator $\hat{\beta}$ is asymptotically more efficient than the unweighted estimator $\hat{\beta}$. $\hat{\beta}$ and $\hat{\beta}$ are asymptotically equivalent when there is no ARCH effect. Also note that the asymptotic distribution of $\hat{\beta}$ does not depend on $\alpha$ and the regression covariates $X_t$. This indicates that the conditional scale can be well estimated without the knowledge of the autoregressive parameters.

**Remark 1** The proposed $L_1$-estimators are $\sqrt{n}$-consistent and asymptotically normal. Note that we do not assume the normality of the innovation and the inequality of the parameter vectors in different regimes as given in Li and Li (1996). Here, robust inference for $\beta$ is achieved under general conditions. In particular, the tail-weight of the innovation is irrelevant because there is no condition on the moments of the innovation beyond $E|y_t|^{2+\delta} < \infty$. In contrast, the asymptotic normality for the maximum likelihood estimation in ARCH models requires a finite
fourth moment. This phenomenon is also observed for $L_1$-estimation in ARCH and GARCH models in Hall and Yao (2003). The likelihood estimators (MLE) of model (1) and (2) maximize the following conditional log likelihood (apart from a constant):

$$
\ell(\alpha, \beta) = -\frac{1}{n} \sum_{t=s+1}^{n} \left[ \log h_t + \frac{(y_t - X_t'\alpha)^2}{2h_t^2\sigma^2} \right],
$$

(6)

where $\sigma^2 = E[u_t^2]$ is used to normalize the scale of $u_t$ such that the parameter $\beta$ can be identified. For non-normal innovations, the above likelihood serves as the quasilikelihood advocated by Weiss (1986). We can estimate the parameters by maximizing the quasilikelihood. Denote the resulting estimator as QMLE. As noted in Jiang et al. (2001), the LAD estimator is less efficient than the MLE under conditional normal distributions. However, the $L_1$-estimator generally becomes more efficient than the QMLE when the innovation distributions deviate from normal distributions.

**Remark 2** An easy identification of the DTARCH model is desired in practice. Let $E\ lvert u_t \rvert = v$ and $v_t = \lvert \epsilon_t \rvert - vh_t$. Observing the autoregressive form of (2), the absolute error is

$$
|\epsilon_t| = vh_t + v_t = v \sum_{j=1}^{m} [\beta_0 + \beta_1|\epsilon_{t-1}| + \cdots + \beta_q|\epsilon_{t-q}|] I_{t,j} + v_t.
$$

(7)

By taking iterative expectation, $v_t$ has median zero and is serially uncorrelated. Note that $E(v_t|\mathcal{F}_{t-1}) = 0$ and $\{v_t\}$ is a sequence of martingale differences. The conditional scale at each threshold region can be written in the form of linear regression and Tsay’s (1989) arranged autoregression can be extended to the DTARCH models based on the $L_1$-norm fitting. We can preliminarily identify the DTARCH model using the identification procedure in Li and Li (1996, p. 257) with the MLE estimation replaced by the proposed $L_1$-norm fitting. This approach will be further explored in the empirical study.

### 3. DIAGNOSTIC CHECKING

Residual autocorrelations are useful for checking the assumption of white noise disturbances in the specified model (Box and Jenkins 1970). An important goodness-of-fit test is the portmanteau test (Box and Pierce 1970). Li and Li (1996) suggested a portmanteau statistic to check the adequacy of DTARCH models based on maximum likelihood estimation. The result applies to conditional heteroscedastic time series models. However, a drawback of this approach is the lack of robustness because it is sensitive to outliers and error distributions. We introduce the following robust portmanteau test based on $L_1$-norm fitting. A similar test for ARCH models has been considered in Jiang et al. (2001).

Denote the empirical version of the robust lag $k$ absolute innovation autocorrelation by

$$
r_k = \frac{n}{\sum_{t=k+1}^{n} \psi(|\epsilon_t|h_t^{-1} - \bar{\epsilon}) \psi(|\epsilon_{t-k}|h_{t-k}^{-1} - \bar{\epsilon})}{\sum_{t=1}^{n} \psi^2(|\epsilon_t|h_t^{-1} - \bar{\epsilon})},
$$

where $\bar{\epsilon} = \text{median}_{1 \leq t \leq n}(|\epsilon_t|h_t^{-1})$. Since $\bar{\epsilon} = 1$ by assumption,

$$
r_k = 4n^{-1} \sum_{t=k+1}^{n} \psi(|\epsilon_t|h_t^{-1} - 1) \psi(|\epsilon_{t-k}|h_{t-k}^{-1} - 1).
$$
Then $r_k$ can be estimated by $\hat{r}_k = 4n^{-1} \sum_{t=k+1}^n \psi(|\hat{\epsilon}_t|\hat{h}_t^{-1} - 1)\psi(|\hat{\epsilon}_{t-k}|\hat{h}_{t-k}^{-1} - 1)$. Intuitively, if the $L_1$-norm-fitted model is adequate, $\hat{r}_k$ would be a good estimator of $r_k$. It can be used to construct a robust portmanteau statistic to assess the fitted model. The estimator $\hat{r}_k$ is essentially the autocorrelation function of the signs of centralized and standardized absolute residuals, and is clearly robust. Other robust autocorrelation estimators such as the truncated estimator may be considered. Nevertheless, the proposed estimators seem intuitive, appealing and complimentary to those in Busto and Yohai (1986).

Let $\mathbf{r} = (r_1, \ldots, r_M)'$ and $\hat{\mathbf{r}} = (\hat{r}_1, \ldots, \hat{r}_M)'$ for any positive integer $M$. Following Jiang et al. (2001), we obtain the following asymptotic representation for $\hat{r}_k$:

**Theorem 2** Assume that (A1)–(A5) hold. Then,
\[
\sqrt{n}\hat{r}_k = \sqrt{n}r_k - 4(f(-1) + f(1))U_k'\sqrt{n}(\hat{\beta} - \beta) + o_p(1)
\]
and
\[
\sqrt{n}\hat{r} = \sqrt{n}\mathbf{r} - 4(f(-1) + f(1))\mathbf{U}\sqrt{n}(\hat{\beta} - \beta) + o_p(1),
\]
where $U_k = E[\mathbf{Z}_t h_t^{-1}\psi(|u_{t-k}|-1)]$ and $\mathbf{U} = (\mathbf{U}_1, \ldots, \mathbf{U}_M)'$.

Note that $\sqrt{n}(\hat{\beta} - \beta)$ is asymptotically normally distributed. We obtain the following theorem from Theorems 1(iii) and 2 using the Mann-Wald device and the martingale central limit theorem (Brown 1971):

**Theorem 3** Under assumptions (A1)–(A5), $\sqrt{n}\hat{r}$ is asymptotically normally distributed with mean zero and covariance $\mathbf{V} = \mathbf{I} - 4\mathbf{U}\mathbf{G}_2^{-1}\mathbf{U}'$. Therefore, $\hat{Q}_M = n\hat{r}'\hat{V}^{-1}\hat{r}$ is asymptotically chi-squared distributed with $M$ degrees of freedom.

The robust residual correlations give smaller standard errors compared to $1/\sqrt{n}$, the standard error of the squared residual autocorrelations (see Box and Pierce 1970). Note that the asymptotic covariance $\mathbf{V}$ of $\sqrt{n}\hat{r}$ can be replaced by any consistent estimator without changing the asymptotic distribution of $\hat{Q}_M$. We define the robust portmanteau statistic as $\hat{Q}_M = n\hat{r}'\hat{V}^{-1}\hat{r}$, where $\hat{V}$ is a consistent estimator of $\mathbf{V}$. Then $\hat{Q}_M$ is asymptotically chi-squared distributed with $M$ degrees of freedom.

Note that $U_k = 0$ when $k > q = \max(q_1, \ldots, q_n)$. For $M > q$, $\mathbf{U}$ has zero entries from the $(q + 1)$th row onwards. This implies that the statistic $\hat{Q}_M(q) = n \sum_{q+1}^M \hat{r}_i^2$ is chi-squared distributed with $(M-q)$ degrees of freedom. When the conditional scales $h_t$ are homogeneous, $U_k$ equals zero. Then the asymptotic standard error of $\hat{r}_k$ is $1/\sqrt{n}$. Therefore, large $\hat{Q}_M$ suggests the inadequacy of the model, especially the misspecification of $h_t$.

**Remark 3** In practice, a convenient choice for $\hat{V}$ is the sample averages. Let $\hat{G}_2 = n^{-1} \sum_{t=1}^n \hat{Z}_t \hat{Z}_t' h_t^{-2}$ and $\hat{\mathbf{U}} = (\hat{U}_1, \ldots, \hat{U}_M)'$, where $\hat{U}_k = n^{-1} \sum_{t=k+1}^n \hat{Z}_t h_t^{-1} \psi(|\hat{\epsilon}_{t-k}|\hat{h}_{t-k}^{-1} - 1)$. Then $\hat{V} = \mathbf{I} - 4\hat{\mathbf{U}}\hat{G}_2^{-1}\hat{\mathbf{U}}'$ is a consistent estimator of $\mathbf{V}$. We can also use the trimmed means instead of the averages to achieve robustness. In our simulations, we will use the 1%-trimmed mean.

The robust portmanteau statistic provides a useful tool in checking the homogeneity of the conditional scales of DTARCH models. For the model given in (1) and (2), we can also use the maximum likelihood in (6) to estimate the parameters and then use the squared residual autocorrelation to construct the portmanteau statistic as in Li and Li (1996). When the innovation
is non-normal, the quasi-likelihood estimation applies to ARCH\((p,q)\) models and the resulting portmanteau statistic is useful for checking model adequacy (see Jiang et al. 2001). However, the quasi-likelihood method does not work well for DTARCH models. This can be seen in the simulation study.

4. NUMERICAL STUDIES

4.1. Empirical results

Distributions of financial returns tend to exhibit asymmetry, heavy tails and other stylized facts. This phenomenon could be described by the conditional median and conditional scale specification in the DTARCH models.

We study the daily S&P 500 Index series in 2001. The return series \(\{y_t\}\) is defined as the log difference of the price, which is shown on the left panel of Figure 1. As the market experienced the 911 incidence and the index was volatile during the selected period, we are interested in the asymmetry of the conditional mean and conditional variance. The proposed robust inference is applied to the data set with 248 observations. We set \(r = 0\) and \(d = 1\), which is consistent with observations in the stock market.

We consider the following DTARCH model for the return series \(\{y_t\}\) after preliminary identification (Remark 2):

\[
y_t = \begin{cases} 
\alpha_1 y_{t-1} + \varepsilon_t & \text{if } y_{t-1} \leq 0 \\
\alpha_2 y_{t-1} + \varepsilon_t & \text{if } y_{t-1} > 0 
\end{cases}
\]

where \(\varepsilon_t = h_t u_t\), with

\[
h_t = \begin{cases} 
\beta_0^{(1)} + \beta_1^{(1)} |\varepsilon_{t-1}| & \text{if } y_{t-1} \leq 0 \\
\beta_0^{(2)} + \beta_1^{(2)} |\varepsilon_{t-1}| & \text{if } y_{t-1} > 0 
\end{cases}
\]
Using the proposed $L_1$-norm fitting, we obtain the estimated parameters $(\hat{\alpha}_1, \hat{\alpha}_2) = (0.06344, -0.06994)$ and $(\hat{\beta}_0^{(1)}, \hat{\beta}_1^{(1)}, \hat{\beta}_0^{(2)}, \hat{\beta}_1^{(2)}) = (0.00734, 0.05511, 0.00679, 0.05589)$. The $p$-value is 0.173 for the DTARCH model, which suggests that the fitted DTARCH model captures the conditional median and conditional scale structures reasonably well. The sample skewness and kurtosis of the original return series $\{y_t\}$ are 1.5783 and 3.4574, respectively. The skewness and the kurtosis of the residuals from the fitted model ($-0.0906$ and $1.2443$) are much smaller in magnitude than those of $\{y_t\}$, reflecting that the DTARCH model successfully reveals the asymmetry in the time-varying conditional median and the conditional scale of the return series.

The Q–Q-plots for the residuals based on $L_1$-norm fitting is reported on the right panel of Figure 1. The residuals appear to be white noise as indicated from the nearly straight line pattern. It is possible to refine the fitted model using the model selection rules in Tsay (1989), but this is beyond the scope of this study. The results suggest that the proposed robust DTARCH modelling approach is useful in describing the changing patterns of the conditional median and the conditional scale for financial data.

4.2. Robustness of the estimation and approximation to null distribution

We compare the proposed robust estimation approach with that based on the MLE when the innovation is normal. For non-normal innovations, we compare the robust approach with the quasi-likelihood based method.

**Example 1.** Consider the following model:

\[
y_t = \begin{cases} 
\alpha_1^{(1)} y_{t-1} + \varepsilon_t, & \text{if } y_{t-1} \leq 0 \\
\alpha_2^{(2)} y_{t-1} + \varepsilon_t, & \text{if } y_{t-1} > 0,
\end{cases}
\]

\[
h_t = \begin{cases} 
\beta_0^{(1)} + \beta_1^{(1)}|\varepsilon_{t-1}|, & \text{if } y_{t-1} \leq 0 \\
\beta_0^{(2)} + \beta_1^{(2)}|\varepsilon_{t-1}|, & \text{if } y_{t-1} > 0,
\end{cases}
\]

where $(\alpha_1^{(1)}, \alpha_2^{(2)}) = (0.20, 0.35)$, $(\beta_0^{(1)}, \beta_1^{(1)}) = (0.002, 0.30)$ and $(\beta_0^{(2)}, \beta_1^{(2)}) = (0.004, 0.25)$. This model belongs to the same type of models for the S&P 500 index series. Three sets of innovation variables—i.e. mixed normal distribution $(0.9N(0, 1^2) + 0.1N(0, 5^2))$, standard normal distribution, and Student’s $t$-distribution with 5 degrees of freedom—are considered. The innovation variables are all scaled such that median$(|u_t|) = 1$. There are 1,000 independent replications for each error distribution. Parameters $\alpha$ and $\beta$ are estimated from the $L_1$-estimation procedure discussed in Section 2. The means and the standard errors of $\hat{\alpha}$ and $\hat{\beta}$ in the 1,000 replications are reported in Table 1.

It can be seen that $L_1$-estimators are robust and accurate against the error distributions. When the innovation is normal, both MLEs and $L_1$-estimators of the parameters are rather close to the true values even though $L_1$-estimators have larger standard errors. This verifies that MLEs are more efficient than $L_1$-estimators. However, QMLEs are quite unsatisfactory in the two non-normal innovation cases since the averages of $\hat{\beta}$ estimates are far away from the true values even though the parameters $\alpha$ are well estimated. This result queries the use of QMLE for DTARCH models, which is different for ARCH models (see Jiang et al. 2001).

The robust standardized absolute $L_1$-residual autocorrelations and the robust portmanteau statistics $\hat{Q}_M(M = 6)$ with trim = 0.05 are computed. Figures 2–4 display the densities of the $\hat{Q}_M(M = 6)$ statistic and the quantile–quantile plot for each innovation distribution. These plots demonstrate the approximations to the $\chi^2_{(0)}$ distribution are quite satisfactory for different innovations, which supports the result of Theorem 2. The portmanteau test results also indicate the adequacy of the $L_1$-norm fitted model. Since the numerical results of QMLE are quite
Table 1. The QMLE and \( L_1 \)-estimates of DTARCH model parameters in 1,000 replications.

<table>
<thead>
<tr>
<th>Innovation</th>
<th>Estimation</th>
<th>( \alpha_1^{(1)} )</th>
<th>( \alpha_1^{(2)} )</th>
<th>( \beta_0^{(1)} )</th>
<th>( \beta_1^{(1)} )</th>
<th>( \beta_0^{(2)} )</th>
<th>( \beta_1^{(2)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N(0, 1) )</td>
<td>Mean of QMLE</td>
<td>0.195</td>
<td>0.337</td>
<td>0.002</td>
<td>0.292</td>
<td>0.004</td>
<td>0.236</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.0792</td>
<td>0.093</td>
<td>0.0002</td>
<td>0.056</td>
<td>0.0004</td>
<td>0.0743</td>
</tr>
<tr>
<td></td>
<td>Mean of L1</td>
<td>0.193</td>
<td>0.334</td>
<td>0.002</td>
<td>0.283</td>
<td>0.004</td>
<td>0.221</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.1046</td>
<td>0.1128</td>
<td>0.0004</td>
<td>0.0835</td>
<td>0.0006</td>
<td>0.1188</td>
</tr>
<tr>
<td>Mixed normal</td>
<td>Mean of QMLE</td>
<td>0.196</td>
<td>0.335</td>
<td>0.003</td>
<td>0.442</td>
<td>0.006</td>
<td>0.390</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.1035</td>
<td>0.1119</td>
<td>0.0010</td>
<td>0.1973</td>
<td>0.0015</td>
<td>0.2693</td>
</tr>
<tr>
<td>( t(5) )</td>
<td>Mean of L1</td>
<td>0.190</td>
<td>0.335</td>
<td>0.002</td>
<td>0.277</td>
<td>0.004</td>
<td>0.222</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.1164</td>
<td>0.1108</td>
<td>0.0004</td>
<td>0.0831</td>
<td>0.0006</td>
<td>0.1051</td>
</tr>
</tbody>
</table>

unsatisfactory for non-normal innovations, it is not necessary to report the squared residual autocorrelations and the portmanteau test results.

4.3. Power of the portmanteau statistic

For simplicity, \( \alpha \) is set to zero and only simulations for the volatility part are considered. We also examine the discriminating power of the test on the delay parameter. The asymptotic null distribution stated in Theorem 3 is applied to determine the \( p \)-value for large samples. We use simulations to compute the null distribution of the test statistic for a finite sample. The computational algorithm is given as follows:

1. Obtain the volatility estimate \( \hat{h}_t \) under the null hypothesis and the residual vectors \( \hat{u}_t \) under the alternative hypothesis, for \( t = s' + 1, \ldots, n \).
2. Compute the test statistic \( \hat{Q}_M \).
3. Draw a bootstrap residual vector \( \hat{u}^*_t \) of length \( n - s' \) from the centred empirical distribution of \( \hat{u}_t \) and compute \( y^*_t = \hat{h}_t \hat{u}^*_t \) for \( t = s' + 1, \ldots, n \).
4. Use the generated random sample \( y^*_t \) and the initial values \( y_t (t = 1, \ldots, s') \) to obtain the test statistic \( \hat{Q}_{M}^* \).
5. Repeat steps 3 and 4 \( B \) times to obtain the bootstrap statistics \( \hat{Q}_{M1}^*, \ldots, \hat{Q}_{MB}^* \).
6. The \( p \)-value of the test statistic \( \hat{Q}_M \) is the percentage of the bootstrap statistics \( \{ \hat{Q}_{M1}^*, \ldots, \hat{Q}_{MB}^* \} \) that exceed \( \hat{Q}_M \).

A similar simulation approach to determine the \( p \)-value of a test statistic is given in Fan and Jiang (2005) for additive models.

Example 2. The following ARCH(0,1) model is chosen as the null hypothesis:

\[
y_t = h_t u_t \quad \text{and} \quad h_t = \beta_0 + \beta_1 |y_{t-1}|. \tag{9}
\]

As the ARCH model is a special case of the DTARCH model, the proposed \( \hat{Q}_M \) statistic also applies to ARCH models. Here the power of the proposed test is evaluated for a sequence of alternatives ranging from the DTARCH model to reasonably distant models. Let \( \hat{\beta}_i^{(j)} \) be the corresponding estimators of the parameters \( \beta_i^{(j)} (i = 0, 1; j = 1, 2) \) under the alternative hypothesis with
Figure 2. Left panel: Densities of the statistic $\hat{Q}_M$ with $M = 6$ for normal innovation (solid: real density, dash-dotted: estimated density). Right panel: Q–Q plot for the statistic $\hat{Q}_M$ with $M = 6$ for normal innovation.

$\theta = 1$ in (10). We simulate 600 samples from the alternative with $\theta = 1$ and fit the data under the null model (9). The parameter values of the null model is given by the averages of the parameter estimates in the 600 samples. We evaluate the power of the portmanteau test for a sequence of alternative models indexed by $\theta$:

$$y_t = h_t^{(\theta)} u_t, \quad h_t^{(\theta)} = h_t + \theta (h_t^{(1)} - h_t),$$

(10)

where

$$h_t^{(1)} = \begin{cases} 
\beta_0^{(1)} + \beta_1^{(1)} |y_{t-1}|, & \text{if } y_{t-1} \leq 0 \\
\beta_0^{(2)} + \beta_1^{(2)} |y_{t-1}|, & \text{if } y_{t-1} > 0 
\end{cases}$$

with $(\beta_0^{(1)}, \beta_1^{(1)}) = (0.02, 0.30)$ and $(\beta_0^{(2)}, \beta_1^{(2)}) = (0.06, 0.90)$. © Royal Economic Society 2005
Robust DTARCH modelling

Figure 3. Left panel: Densities of the statistic $\hat{Q}_M$ with $M = 6$ for mixed normal innovation (solid: real density, dash-dotted: estimated density). Right panel: Q–Q plot for the statistic $\hat{Q}_M$ with $M = 6$ for mixed normal innovation.

Figure 4. Left panel: Densities of the statistic $\hat{Q}_M$ with $M = 6$ for $t(5)$ innovation (solid: real density, dash-dotted: estimated density). Right panel: Q–Q plot for the statistic $\hat{Q}_M$ with $M = 6$ for $t(5)$ innovation.

For a given value of $\theta$, we simulate data series with lengths 200 and 400 from the alternative model (10) for three different innovation distributions: $N(0, 1)$, $0.9N(0, 1) + 0.1N(0, 5^2)$ and $t(3)$. The rejection rate of the test statistic $\hat{Q}_M$ with $M = 6$ is computed. Note that when $\theta = 0$, the alternative model becomes the ARCH model so that the power should be roughly 5% at the significance level 0.05. This is indeed the case given in Table 2. As the index $\theta$ increases, the
Table 2. Simulated powers of the proposed test at significance level 5%. $H_0$: ARCH vs $H_1$: DTARCH ($d = 1$).

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Innovation</th>
<th>$\theta = 0.0$</th>
<th>$\theta = 0.2$</th>
<th>$\theta = 0.4$</th>
<th>$\theta = 0.6$</th>
<th>$\theta = 0.8$</th>
<th>$\theta = 1.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 200$</td>
<td>$N(0, 1)$</td>
<td>0.045</td>
<td>0.072</td>
<td>0.135</td>
<td>0.253</td>
<td>0.362</td>
<td>0.620</td>
</tr>
<tr>
<td></td>
<td>$0.9N(0, 1) + 0.1N(0, 5^2)$</td>
<td>0.050</td>
<td>0.0717</td>
<td>0.160</td>
<td>0.275</td>
<td>0.475</td>
<td>0.725</td>
</tr>
<tr>
<td></td>
<td>$t(3)$</td>
<td>0.052</td>
<td>0.088</td>
<td>0.210</td>
<td>0.348</td>
<td>0.520</td>
<td>0.677</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>$N(0, 1)$</td>
<td>0.050</td>
<td>0.122</td>
<td>0.310</td>
<td>0.548</td>
<td>0.753</td>
<td>0.918</td>
</tr>
<tr>
<td></td>
<td>$0.9N(0, 1) + 0.1N(0, 5^2)$</td>
<td>0.050</td>
<td>0.222</td>
<td>0.527</td>
<td>0.747</td>
<td>0.902</td>
<td>0.965</td>
</tr>
<tr>
<td></td>
<td>$t(3)$</td>
<td>0.043</td>
<td>0.208</td>
<td>0.420</td>
<td>0.662</td>
<td>0.850</td>
<td>0.960</td>
</tr>
</tbody>
</table>

Table 3. Simulated powers of the proposed test at significance level 5%. $H_0$: DTARCH ($d = 1$) vs $H_1$: DTARCH ($d = 2$).

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Innovation</th>
<th>$\theta = 0.0$</th>
<th>$\theta = 0.2$</th>
<th>$\theta = 0.4$</th>
<th>$\theta = 0.6$</th>
<th>$\theta = 0.8$</th>
<th>$\theta = 1.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 200$</td>
<td>$N(0, 1)$</td>
<td>0.047</td>
<td>0.088</td>
<td>0.140</td>
<td>0.243</td>
<td>0.348</td>
<td>0.650</td>
</tr>
<tr>
<td></td>
<td>$0.9N(0, 1) + 0.1N(0, 5^2)$</td>
<td>0.043</td>
<td>0.092</td>
<td>0.202</td>
<td>0.363</td>
<td>0.522</td>
<td>0.775</td>
</tr>
<tr>
<td></td>
<td>$t(3)$</td>
<td>0.050</td>
<td>0.127</td>
<td>0.262</td>
<td>0.382</td>
<td>0.555</td>
<td>0.742</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>$N(0, 1)$</td>
<td>0.045</td>
<td>0.123</td>
<td>0.310</td>
<td>0.567</td>
<td>0.755</td>
<td>0.932</td>
</tr>
<tr>
<td></td>
<td>$0.9N(0, 1) + 0.1N(0, 5^2)$</td>
<td>0.038</td>
<td>0.263</td>
<td>0.600</td>
<td>0.827</td>
<td>0.923</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>$t(3)$</td>
<td>0.052</td>
<td>0.208</td>
<td>0.468</td>
<td>0.680</td>
<td>0.880</td>
<td>0.977</td>
</tr>
</tbody>
</table>

alternative hypothesis deviates further away from the null and one expects the rejection rates to increase. The simulation result confirms that the test is very powerful. When $\theta = 1$ and $n = 400$, we reject approximately over 90% of the time (correct decision) for the three innovations cases. This suggests that the proposed test has a high discriminating power in differentiating the ARCH model (9) from the DTARCH model (10).

Example 3. Consider the test of the null model $y_t = h_t^{(1)}u_t$ with $h_t^{(1)}$ specified in Example 2, against the sequence of the alternative models $y_t = h_t^{(0)}u_t$, where $h_t^{(0)} = h_t^{(1)} + \theta(h_t^{(2)} - h_t^{(1)})$ with

$$h_t^{(2)} = \begin{cases} 
\beta_0^{(1)} + \beta_1^{(1)}|y_{t-1}|, & \text{if } y_{t-2} \leq 0 \\
\beta_0^{(2)} + \beta_1^{(2)}|y_{t-1}|, & \text{if } y_{t-2} > 0.
\end{cases}$$

The parameters $\beta_i^{(j)} (i = 0, 1; j = 1, 2)$ are the same as those in Example 2. The null is a DTARCH model with the delay parameter $d = 1$ and the alternative is a DTARCH model with the delay parameter $d = 2$. The only difference between $h_t^{(1)}$ and $h_t^{(2)}$ is the delay parameter. This setup is designed for examining the discriminating power of the robust portmanteau test on the delay parameter.

Table 3 reports the computed powers at the 5% significance level. It gives the same results as observed in Example 2. This reflects the fact that the proposed test is not only robust against outliers and innovation distributions but also of high discriminating power for the delay parameter.

5. DISCUSSION

We have proposed a robust $L_1$-estimation method in estimating the DTARCH models where the innovation distribution is unknown. Bahadur representations of the $L_1$-estimators are given under
mild conditions. The robust $L_1$-residual autocorrelations and the robust portmanteau statistic are easy to compute. They are demonstrated to be useful tools in model diagnostic. Empirical data analysis illustrates the use of the $L_1$-norm-based estimation approach. The results can be easily adapted to the TAR models considered in Tsay (1989). The quasi-likelihood approach has been shown to be an efficient estimation method for ARCH models, but our simulations query the use of the QMLE-based method for DTARCH models.

ACKNOWLEDGEMENTS

This work was partially supported by a grant from City University of Hong Kong (Project No. 7001490) and Chinese NSF Grants 10471006 and 79800017.

REFERENCES


© Royal Economic Society 2005


**APPENDIX**

Theorem 1(iii) follows the same argument as given in Koenker and Zhao (1996). Since Theorems 1(ii) and (iii) can be derived in the same way and the proof of Theorem 1(ii) is much easier than that of Theorem 1(iii), we omit the proof of Theorem 1(ii) and only discuss the proofs of Theorem 1(iii) and Theorem 2.

We adopt the following notations given in Jiang et al. (2001). Let \( q = \sum_{j=1}^{m}(q_j + 1) \) and \( L \) be a positive number. \( E_{t-1} \) denotes the expectation conditional on \( F_{t-1} \). Denote \( \Delta = (\Delta_1, \Delta_2)' \), \( \alpha(\Delta_1) = \alpha + n^{1/2} \Delta_1 \), and \( \beta(\Delta_2) = \beta + n^{-1/2} \Delta_2 \). Let

\[
    \varepsilon_t(\Delta_1) = y_t - X_t'\alpha(\Delta_1) = \varepsilon_t - n^{-1/2} X_t' \Delta_1,
    
    \mathbf{Z}_t(\Delta_1) = (1, |\varepsilon_{t-1}(\Delta_1)|, \ldots, |\varepsilon_{t-q}(\Delta_1)|)',
    
    h_t(\Delta) = \mathbf{Z}_t(\Delta_1)'\beta(\Delta_2),
    
    \xi_t(\Delta) = I(|\varepsilon_t| < h_t) - I(|\varepsilon_t(\Delta_1)| < h_t(\Delta)),
    
    \mathbf{B}_t = (0, X_{t-1} \text{sgn}(\varepsilon_{t-1}), \ldots, X_{t-q} \text{sgn}(\varepsilon_{t-q}))',
    
    \mathbf{B}_{t1} = (0, \|X_{t-1}\|, \ldots, \|X_{t-q}\|)',
    
    \mathbf{H}'_t = h_t^{-1}(X_t \text{sgn}(\varepsilon_t) - \beta' \mathbf{B}_t, \mathbf{Z}_t),
    
    \hat{\Delta}_n = (\hat{\Delta}_{1n}', \hat{\Delta}_{2n}') = (\sqrt{n}(\hat{\alpha} - \alpha)', \sqrt{n}(\hat{\beta} - \beta)')',
    
    \hat{\Delta}_{2n} = \sqrt{n}(\hat{\beta} - \beta)'.
\]

Then, \( \varepsilon_t(\hat{\Delta}_{1n}) = \tilde{\varepsilon}_t, \varepsilon_t(0) = \varepsilon_t, h_t(\hat{\Delta}_n) = \hat{h}_t, h_t(\hat{\Delta}_{1n}, \hat{\Delta}_{2n}) = \hat{h}_t, h_t(0) = h_t, \mathbf{Z}_t(\hat{\Delta}_{1n}) = \hat{\mathbf{Z}}_t \), and \( \mathbf{Z}_t(0) = \mathbf{Z}_t \).

**Proof of Theorem 1(iii)** Let \( V_n(\Delta_2) = \frac{-1}{2} \sum_{t=1}^{n} \hat{Z}_t \hat{h}_t^{-1} \psi(|\hat{\varepsilon}_t| - \hat{\mathbf{Z}}_t' \beta) \). Using Lemma A.2 in Ruppert and Carroll (1980), we obtain

\[
    \|V_n(\Delta_{2n})\| = \left\| n^{-1/2} \sum_{t=1}^{n} \hat{Z}_t \hat{h}_t^{-1} \psi(|\hat{\varepsilon}_t| - \hat{\mathbf{Z}}_t' \beta) \right\| \leq 2 \dim(\mathbf{Z}_t)n^{-1/2} \max_{t \leq n} \left( \|\hat{Z}_t \hat{h}_t^{-1}\| \right) = o_p(1). \tag{A.1}
\]

Using the identity \(|a - b| = |a| - \text{sgn}(a)b - 2(|a| - \text{sgn}(a)b)I(|a| < |b|, ab \geq 0)\), we have

\[
    \sup_{t \leq n, |\Delta_1| \leq L} \left| \varepsilon_t(\Delta_1) - |\varepsilon_t| + n^{-1/2} X_t' \Delta_1 \text{sgn}(\varepsilon_t) \right| = o_p \left( n^{-1/2} \right) \tag{A.2}
\]

and

\[
    \sup_{t \leq n, |\Delta_1| \leq L} \left| \mathbf{Z}_t(\Delta_1) - \mathbf{Z}_t + n^{-1/2} \mathbf{B}_t \Delta_1 \right| = o_p \left( n^{-1/2} \right). \tag{A.3}
\]
Robust DTARCH modelling

157

Note that Theorem 1(i) gives $\hat{\Delta}_{1n} = O_p(1)$. It follows from (A.2) and (A.3) that

$$
\sup_{t \leq n, ||\Delta|| \leq L} \left| \hat{\xi}_t - \hat{Z}_t \beta (\Delta_2) - h_t \left[ |u_t| - 1 - n^{-\frac{1}{2}} H \left( \frac{\hat{\Delta}_{1n}}{\Delta_2} \right) \right] \right| = o_p \left( n^{-\frac{1}{2}} \right).
$$

(A.4)

Then

$$
\sup_{||\Delta_2|| \leq L} \| V_n (\Delta_2) - \hat{V} (\hat{\Delta}_{1n}, \Delta_2) \| = o_p(1). \tag{A.5}
$$

Using the same argument as in Lemma 2 of Jiang et al. (2001), we obtain

$$
\sup_{||\Delta_2|| \leq L} \| \hat{V} (\hat{\Delta}_{1n}, \Delta_2) - V(0) + (f(-1) + f(1)) G_2 \Delta_2 \| = o_p(1). \tag{A.6}
$$

This together with (A.5) gives

$$
\sup_{||\Delta_2|| \leq L} \| V_n (\Delta_2) - V(0) + (f(-1) + f(1)) G_2 \Delta_2 \| = o_p(1). \tag{A.7}
$$

Note that $\sum_t \hat{h}_t^{-1} || \hat{\xi}_t - \hat{Z}_t (\beta + \lambda n^{-\frac{1}{2}} \Delta_2) ||$ is a convex function in $\lambda$ and the gradient of the function, $-\Delta_2 V_n (\lambda \Delta_2)$, is non-decreasing in $\lambda$. Therefore,

$$
-\Delta_2^T V_n (\lambda \Delta_2) \geq -\Delta_2^T V_n (\Delta_2), \text{ for } \lambda \geq 1. \tag{A.8}
$$

Simple algebra gives $V(0) = O_p(1)$. Applying Lemma A.4 in Koenker and Zhao (1996) with (A.1), (A.7) and (A.8), we complete the proof of the theorem. \hfill \Box

Proof of Theorem 2

Let

$$
A_n (\Delta) = 4n^{-\frac{1}{2}} \sum_{t=k+1}^n \psi \left( |\xi_t (\Delta_1)| \hat{h}_t^{-1} (\Delta) - 1 \right) \left( |\xi_{t-k} (\Delta_1)| \hat{h}_{t-k}^{-1} (\Delta) - 1 \right).
$$

Then $A_n (\Delta_n) = \sqrt{n} \hat{h}_k$ and $A_n (0) = \sqrt{n} h_k$. Note that

$$
A_n (\Delta) - A_n (0) = 4n^{-\frac{1}{2}} \sum_{t=k+1}^n \psi \left( |\xi_{t-k} (\Delta_1)| \hat{h}_{t-k}^{-1} (\Delta) - 1 \right) \xi_t (\Delta) \\
+ 4n^{-\frac{1}{2}} \sum_{t=k+1}^n \psi \left( |\xi_t | \hat{h}_t^{-1} - 1 \right) \xi_{t-k} (\Delta)
$$

$$
\equiv I_n (\Delta) + J_n (\Delta). \tag{A.9}
$$

Using the same argument as in Lemma 4 of Jiang et al. (2001), we have

$$
\sup_{||\Delta|| \leq L} | A_n (\Delta) - A_n (0) - I_n (\Delta) | = o_p(1). \tag{A.10}
$$

Note that

$$
I_n (\Delta) = 4n^{-\frac{1}{2}} \sum_{t=k+1}^n \psi (|\xi_{t-k} (\Delta_1)| - h_{t-k} (\Delta)) [\xi_t (\Delta) - E_{t-1} \xi_t (\Delta)] \\
+ 4n^{-\frac{1}{2}} \sum_{t=k+1}^n \psi (|\xi_t (\Delta_1)| - h_t (\Delta)) E_{t-1} \xi_t (\Delta)
$$

$$
\equiv T_n (\Delta) + 4n^{-\frac{1}{2}} \sum_{t=k+1}^n \psi (|\xi_{t-k} (\Delta_1)| - h_{t-k} (\Delta)) E_{t-1} \xi_t (\Delta). \tag{A.11}
$$

© Royal Economic Society 2005
Taking iterative expectation, we have

\[ ET_n^2(\Delta) = 4n^{-1} \sum_{t=k+1}^{n} E[\xi_t(\Delta) - E_{t-1}\xi_t(\Delta)]^2 \]

\[ \leq 4n^{-1} \sum_{t=k+1}^{n} E[\sup_{||\Delta|| \leq L} |\xi_t(\Delta)|]. \]  \hspace{1cm} (A.12)

Then

\[
\sup_{||\Delta|| \leq L} |\xi_t(\Delta)| \leq \sup_{||\Delta|| \leq L} I\left(|u_t - 1| < n^{-\frac{1}{2}}|X_i\Delta| h_i^{-1} + |h_i(\Delta) - h_{i-1}| h_{i-1}^{-1}\right) \\
\leq I\left(|u_t - 1| < n^{-\frac{1}{2}}|X_i| L h_i^{-1} + n^{-\frac{1}{2}}|Z_i| L h_i^{-1} + n^{-\frac{1}{2}}|^B_1||\beta|| + n^{-\frac{1}{2}} L h_i^{-1}\right) \\
\equiv I(|u_t - 1| < \eta_t),
\]

where \( \sup_{t \leq n} |\eta_t| = o_p(1) \) by the same argument as that in Lemma 1(i) of Jiang et al. (2001). The mean value theorem gives

\[ E_{t-1}[\sup_{||\Delta|| \leq L} |\xi_t(\Delta)|] \leq 2\eta_t f(1 + \theta\eta_t), \]  \hspace{1cm} (A.13)

where \( \theta \in [-1, 1] \). Therefore, \( T_n(\Delta) = o_p(1) \). Using the chaining argument in Bickel (1975), we have \( \sup_{||\Delta|| \leq L} |T_n(\Delta)| = o_p(1) \). This, together with (A.10) and (A.11), gives

\[
\sup_{||\Delta|| \leq L} |A_n(\Delta) - A_n(0) - 4n^{-\frac{1}{2}} \sum_{t=k+1}^{n} \psi(|u_{t-k}(\Delta)| - h_{t-k}(\Delta)) E_{t-1}\xi_t(\Delta)| = o_p(1). \]  \hspace{1cm} (A.14)

Using the same argument as in Lemma 3 of Jiang et al. (2001), we obtain

\[
\sup_{||\Delta|| \leq L} |A_n(\Delta) - A_n(0) + 4(f(-1) + f(1))n^{-1} \sum_{t=k+1}^{n} \psi(|u_{t-k} - 1|)Z_i\Delta h_i^{-1}| = o_p(1). \]

Note that Theorems 1(i) and (iii) imply \( \hat{\Delta}_n = O_p(1) \). It follows that

\[ A_n(\hat{\Delta}_n) = A_n(0) - 4(f(-1) + f(1))n^{-1} \sum_{t=k+1}^{n} \psi(|u_{t-k} - 1|)Z_i\hat{\Delta}_2 h_i^{-1} + o_p(1). \]

Therefore,

\[ \sqrt{n}\hat{\epsilon}_k = \sqrt{n}\epsilon_k - 4(f(-1) + f(1))\hat{U}_k\sqrt{n}(\hat{\beta} - \beta) + o_p(1). \]  \hspace{1cm} (A.15)

By Stout (1974, theorem 3.48), \( \{\psi(|u_{t-k} - 1|) Z_i h_i^{-1}\} \) is strictly stationary and ergodic. Applying the ergodic theorem and assumption (A2), we have

\[ \hat{U}_k = U_k + o_p(1). \]  \hspace{1cm} (A.16)

Combining (A.15) and (A.16) completes the proof of the theorem. \( \square \)