1998

Semiparametric Maximum Likelihood Estimation of GARCH Models

Jian Yang

Follow this and additional works at: https://ir.lib.uwo.ca/economicsresrpt

Part of the Economics Commons

Citation of this paper:
RESEARCH REPORT 9816

Semiparametric Maximum Likelihood Estimation of GARCH Models

by

Jian Yang ECONOMICS REFERENCE CENTRE

FEB 15 2000

UNIVERSITY OF WESTERN ONTARIO

September 1998

Department of Economics
Social Science Centre
University of Western Ontario
London, Ontario, Canada
N6A 5C2
econref@julian.uwo.ca
Semiparametric Maximum Likelihood Estimation of GARCH Models

Jian Yang
Department of Economics
University of Western Ontario
London, Ontario
Canada, N6A 5C2
yang@julian.uwo.ca

September 1998

ABSTRACT

We consider a semiparametric GARCH model where the functional form for the conditional density of the errors is unknown. Adaptive conditions of the parameters are examined. Semiparametric Maximum Likelihood (SML) estimators are constructed by maximizing the nonparametric pseudo log-likelihood function computed using the residuals from initial $\sqrt{n}$-consistent estimates. SML estimators are shown to be adaptive for the adaptively estimable parameters and consistent for all identifiable parameters. Monte Carlo results suggest that SML estimators outperform quasi maximum likelihood estimators and the adaptive maximum likelihood estimators in finite samples.

Key words: GARCH models; Adaptive estimation; Semiparametric estimation

JEL classification: C13; C14
1. Introduction

Models of autoregressive conditional heteroskedasticity (ARCH), first proposed by Engle (1982) and generalized by Bollerslev (1986), have been widely used for modeling the behavior of financial time series. The conditional error terms in the Generalized ARCH (GARCH) models are usually assumed normally distributed in empirical research. However, considerable evidence has been provided in the literature over the last several years indicating that the normality assumption fits poorly in many applications to financial data.

In the literature, there are already Gaussian Quasi-Maximum Likelihood (QML) results in the work of Weiss (1986) and Bollerslev and Wooldridge (1988), who show that consistent estimators of the parameters of the model can be obtained by maximizing a Gaussian likelihood function, even though the true likelihood function could be some other. However, Gaussian QMLE is no longer asymptotically efficient. It has a loss of efficiency due to falsely assuming normality. The relative efficiency loss of QMLE for GARCH models has been calculated by Engle and González-Rivera (1991) when the true conditional distribution follows a t-distribution or a gamma-distribution.

To recover efficiency, researchers studying financial time series have proposed alternative parametric specifications to the Gaussian-GARCH specification. For example, t distributions and mixture of normal distributions have been found to fit some financial time series better than a Gaussian distribution.

In principle, a researcher can always search through a large number of non-normal parametric distributions until a good fit is found for a particular data set, but the computation would be intensive. Moreover, when the likelihood is assumed to be a specific non-Gaussian distribution, the so-constructed non-Gaussian QMLE is generally not consistent if the true error distribution is not the assumed non-Gaussian distribution (see Newey and Steigerwald: 1994). Lastly, an estimator constructed from a distribution found in this way suffers from pre-test bias.

In this paper, we consider a nonparametric approach without assuming any specific
function form of the conditional distribution. We examine the adaptive conditions for these types of GARCH models, construct asymptotically efficient (adaptive) estimators of structural parameters when the adaptive conditions are satisfied, and study the finite-sample performance of the proposed estimators.

In the framework of models with conditional heteroskedasticity, Engle and González-Rivera (1991) considered semiparametric GARCH models. The authors estimated the parameters by maximizing nonparametric estimator of the log-likelihood function and reported Monte-Carlo simulation results when the true error density was assumed to be either a $t$-distribution or a gamma-distribution. They found that, although these estimators outperformed the Gaussian QML estimators, they appeared to be considerably less efficient than the ML estimators. They suggested that the semiparametric estimators were not adaptive even when the error density was symmetric.

Linton (1993) reparametrized Engle's semiparametric ARCH specification and examined the adaptive condition for the reparametrized ARCH model under the symmetry condition on the unknown error density. He found that all the identifiable parameters of the reparametrization were adaptively estimable, although it was not clear whether the parameters in Engle's specification were adaptive or not. He then constructed adaptive estimators of the identifiable parameters. However, Linton (1993) did not study the generalized ARCH models, which were more widely used in the financial literature. Secondly, the unknown error density in Linton (1993) was assumed to be symmetric, which simply limited the application of his theoretical results. Moreover, the approach used in his paper to construct adaptive estimator was the so-called Adaptive Maximum Likelihood (AML) estimation method. As argued by Yang (1997), AML estimator behaved poorly in small samples since it was a one-step procedure and utilized the outer-product of the gradient (OPG) to estimate the information matrix. Drost and Klaassen (1997) considered a generalization of the reparametrized GARCH(1,1) model without the symmetry condition on the unknown error density. They found that all the identifiable parameters, which
characterized the conditional variance, were adaptively estimable. Although their results
generalized Linton’s finding regarding the first two aforementioned aspects, they studied
only the adaptation of the parameters in the second moment but left the parameters char-
acterizing the first moment out of consideration. More importantly, Drost and Klaassen
(1997) applied also the AML approach to construct adaptive estimator, which appeared
less efficient in small samples as found by Yang (1997) in a linear regression context.

We consider in this paper a semiparametric linear regression model with conditional
variance governed by a GARCH(1, 1) process and also a special case with a free inter-
cept in the regression function. Generalizing Linton (1993), we reparametrize Bollerslev’s
GARCH(1, 1) model. The new findings are as follows.

(1) In the reparametrization, we find that all the identifiable parameters are adaptively
estimable with symmetric conditional error densities and all the identifiable parameters
except the free intercept in the special case are adaptively estimable without the symmetry
condition.

(2) We propose adaptive estimators for the adaptively estimable parameters using the
SML approach proposed in Yang (1997). They are constructed in a way similar to that
of Engle and González-Rivera (1991), but the nonparametric estimator of the likelihood
function is obtained using trimming techniques. With excessive contributions trimmed out,
SML estimators are proven to be adaptive for all the adaptively estimable parameters. In
addition, they are found to be consistent for all the identifiable parameters that we can
apply the SML method quite confidently even the symmetry condition cannot be well
justified.

(3) In finite samples, as Yang (1997) found, SML estimator could outperform AML
estimator. This is simply because AML method is only a one-step procedure and utilizes
the OPG matrix to estimate the information matrix, and the estimator of the information
matrix constructed using the OPG matrix behaves poorly in finite samples. The Monte
Carlo results in this paper provide further evidence in favor of SML method.
The rest of this article is structured as follows. The GARCH model and its reparametrization are specified and the adaptive condition for the reparametrized GARCH model are examined in Section 2. In section 3, we define semiparametric maximum likelihood estimators of the identifiable parameters and examine the asymptotic properties of SML estimators. Section 4 conducts Monte Carlo studies on the finite sample performances of SML estimators.

2. The GARCH Model and the Stein-Bickel Necessary Conditions for Adaptation

2.1 The GARCH Model

We consider a linear regression model with conditional heteroskedasticity of the following form,

\[ y_t = x_t \beta + u_t; \quad u_t = \varepsilon_t v_t(\phi), \quad (2.1) \]

where \( x_t \in \mathbb{R}^k \) is assumed weakly exogenous; \( \beta \) and \( \phi \) are parameters and \( v_t(\phi) \) is defined below; the returns innovations, \( \{\varepsilon_t\}_{t=1}^n \), are independent and identically distributed with unknown density \( g_0, g_0(\varepsilon) > 0, \forall \varepsilon \in \mathbb{R} \). A special case of (2.1) is a linear regression model with a free intercept,

\[ y_t = \beta_1 + x_{2t} \beta_2 + u_t; \quad u_t = \varepsilon_t v_t(\phi), \quad (2.2) \]

where \( (\beta_1, \beta_2)^T = \beta \) and \( x_{2t} \in \mathbb{R}^{k-1} \).

Bollerslev (1987) proposed a GARCH\((p, q)\) process of conditional heteroskedasticity,

\[ v_t^2 = a + \sum_{i=1}^{p} b_i u_{t-i}^2 + \sum_{j=1}^{q} c_j v_{t-j}^2. \quad (2.3) \]

We shall assume that the process \( v_t^2 \) is stationary, which requires that

\[ \sum_{i=1}^{p} b_i + \sum_{j=1}^{q} c_j < 1. \]

If all \( c_j \)'s are known to be zero, (2.3) becomes Engle's ARCH\((p)\) specification, i.e.

\[ v_t^2 = a + \sum_{i=1}^{p} b_i u_{t-i}^2. \quad (2.3') \]
Linton (1993) argued that the parameters in (2.3') were not jointly identifiable with the unknown density $g_0$. Actually, this identifiability problem also arises in all the adaptive estimation literature, where the variance parameter of the disturbance term can not be estimated semiparametrically since we assumed that $g_0$ is unknown. To deal with the identifiability problem, Engle and González-Rivera (1991) employed an approach by normalizing the variance of $g_0$ to be one. As argued by Linton (1993), this introduced a nonlinear constraint on the class of allowable error densities which was difficult to incorporate in information bound calculations. It was not clear whether this approach would give rise to adaptively estimable parameters in (2.3).

Linton (1993) then suggested the transformed specification of ARCH($p$), which was as follows,

$$v_t^2 = e^{\alpha}(1 + \sum_{i=1}^{p} \gamma_i u_{t-i}^2),$$  \hspace{1cm} (2.4)

where the overall scale parameter $e^{\alpha}$ in the conditional heteroskedasticity was separated from the slope parameters $\gamma_i$. The identifiability problem becomes clearer in this setting. When $g_0$ is unknown, we have no information about $\alpha$, which is the scale parameter of the conditional variance. It then can not be identified by a semiparametric approach. However, $\gamma \equiv (\gamma_1, \cdots, \gamma_p)$ are found to satisfy the adaptive condition, that is, $\gamma$ is in principle adaptively estimable. Linton (1993) also found that the $\beta$ were adaptively estimable when $g_0$ was assumed to be symmetric around zero. He then constructed adaptive estimators of these parameters, which were adaptively estimable, using the AML approach. Since the scale parameter $\alpha$ was separated from the other parameters, the constructions of these adaptive estimators did not require an estimation of $\alpha$. This was done by assuming that $e^{\alpha}$ was absorbed by the density function $g_0$.

Generalizing Linton (1993), we consider a transformed GARCH($p,q$) model and we take $p = q = 1$ for notational simplicity. As Drost and Klaassen (1997) pointed out, to obtain the corresponding results for the general case (with any fixed integers $p$ and $q$) a careful replacement of coefficients by vectors suffices. Then the transformed GARCH(1,1)
process can be described as follows.

\[ v_t^2(\phi) = e^{\alpha} \sigma_t^2(\psi), \quad \sigma_t^2(\psi) = 1 + \gamma u_{t-1}^2 + \theta \sigma_{t-1}^2(\psi). \]  

(2.5)

We denote \( \delta = (\gamma, \theta)^T, \psi = (\beta^T, \delta^T)^T, \phi = (\alpha, \psi^T)^T \); and assume that \( \gamma \geq 0, \theta \geq 0 \), and \((e^{\alpha}\gamma) + \theta < 1\).

In specification (2.5), the conditional variance \( v_t^2 \) is the product of two terms, \( e^{\alpha} \) and \( \sigma_t^2 \). And \( \sigma_t^2 \) is separated from \( \alpha \) in the sense that \( \sigma_t^2 \) can be determined without the knowledge of \( \alpha \). Therefore, all the parameters specifying \( \sigma_t^2 \) can be identified by semiparametric approaches, although \( \alpha \) can not. The true log-likelihood for a sample of size \( n \) is

\[ \ell_n(y|x; \phi, g_0) = \ell_0(Y_0; \phi) - \frac{1}{2} \sum_{t=1}^{n} \log v_t^2(\phi) + \sum_{t=1}^{n} \log g_0(u_t/v_t(\phi)), \]  

(2.6)

where \( \ell_0 \) is the unconditional log-density function of the initial observation with the initial condition \( Y_0 = (u_0, v_0^2) \) observed. \( \ell_0(Y_0; \phi) \) is assumed asymptotically negligible and not included in the following calculations.

2.2. The necessary conditions for adaptation

The GARCH specification (2.1) \& (2.5) is not generally adaptively estimable. We consider a special case where \( g_0 \) is assumed to be symmetric.

To examine the adaptation of this reparametrized GARCH(1,1) setting, we begin by recalling the Stein-Bickel necessary conditions for adaptation. In particular, we consider the adaptation of a given sub-vector of parameters instead of the whole vector of parameters of interest. Let \( \phi \) be the vector of parameters of interest, \( \eta \) the vector of nuisance parameters, \( \phi = (\phi_1^T, \phi_2^T)^T, \phi_1 \in \mathbb{R}^{k_1}, \phi_2 \in \mathbb{R}^{k_2} \), and \( k_1 + k_2 = k \), then the information matrix associated with joint estimation of \( I(\phi, \eta) \) is

\[ I(\phi, \eta) \equiv \begin{pmatrix} I(\phi_1, \phi_2) & I_{\phi_1 \eta} \\ I_{\eta \phi} & I_{\eta \eta} \end{pmatrix} \equiv \begin{pmatrix} I_{\phi_1 \phi_1} & I_{\phi_1 \phi_2} & I_{\phi_1 \eta} \\ I_{\phi_2 \phi_1} & I_{\phi_2 \phi_2} & I_{\phi_2 \eta} \\ I_{\eta \phi_1} & I_{\eta \phi_2} & I_{\eta \eta} \end{pmatrix}. \]
The smallest possible asymptotic variance for an estimator of \( \phi_1 \) with known \( \eta \) is the \( k_1 \times k_1 \) upper left sub-matrix of \( I^{-1}(\phi_1, \phi_2) \); and the smallest possible asymptotic variance for estimator of \( \phi_1 \) with unknown \( \eta \) is the \( k_1 \times k_1 \) upper left sub-matrix of \( I^{-1}(\phi, \eta) \). They are equal to each other if and only if

\[
I_{\phi_1 \eta} - I_{\phi_1 \phi_2} I_{\phi_2 \phi_2}^{-1} I_{\phi_2 \eta} = 0.
\]  

(2.7)

To verify condition (2.7) for our reparametrized GARCH(1,1) setting (2.1) and (2.5), we denote the sample mean of the score function for any parameter vector \( \xi \) as

\[
S^n_{\xi} = \frac{1}{n} \sum_{t=1}^{n} S_t \xi,
\]  

(2.8)

where \( \xi \) is any of \( \phi_1, \phi_2, \) or, \( \eta; S_t \xi \) is the score of \( \xi \) for observation \( t \). Then condition (2.7) can be rewritten as

\[
E\{[S^n_{\phi_1} - I_{\phi_1 \phi_2} \phi_1^{-1} S^n_{\phi_2} S^n_{\phi_2}]S^n_{\eta}\} = 0,
\]  

(2.9)

where \( S^n_{\phi_1} - I_{\phi_1 \phi_2} \phi_1^{-1} S^n_{\phi_2} \) is the projection of \( S^n_{\phi_1} \) off the span of \( S^n_{\phi_2} \). When we only consider the adaptation of \( \phi_1 \) with unknown \( \phi_2 \) rather than the adaptation of both \( \phi_1 \) and \( \phi_2 \), \( S^n_{\phi_1} - I_{\phi_1 \phi_2} \phi_1^{-1} S^n_{\phi_2} \) is called the **efficient score** function of \( \phi_1 \) with unknown \( \phi_2 \). Then condition (2.9) can be interpreted as saying that the efficient score of \( \phi_1 \) is orthogonal to the score function of \( \eta \).

We are now ready to examine the adaptation of GARCH specification (2.1) & (2.5). In this model, we do not consider the adaptive condition for \( \alpha \) since it cannot be identified by a semiparametric approach. We only examine the adaptive condition for \( \beta, \gamma, \) and \( \theta \) with unknown \( \alpha \). Therefore, we need to verify that the efficient scores of \( \beta, \gamma, \) and \( \theta \) with unknown \( \alpha \) are orthogonal to the sample mean of score function for any nuisance parameter \( \eta \). To verify these orthogonality conditions, we first calculate the score functions for all the parameters including \( \alpha \). We denote the score functions, for observation \( t \), of \( \phi, \delta, \alpha, \beta, \gamma \) and \( \theta \) under \( g_0 \) as respectively \( S_t(\phi, g_0), S_t(\delta, g_0), S_t(\alpha, g_0), S_t(\beta, g_0), S_t(\gamma, g_0) \)
and $S_t(\theta, g_0)$, and the sample mean of score functions of $\phi$, $\delta$, $\alpha$, $\beta$, $\gamma$ and $\theta$ under $g_0$ as respectively $S^n(\phi, g_0)$, $S^n(\delta, g_0)$, $S^n(\alpha, g_0)$, $S^n(\beta, g_0)$, $S^n(\gamma, g_0)$, and $S^n(\theta, g_0)$. Then, we have

$$S^n(\alpha, g_0) = \frac{1}{n} \sum_{t=1}^{n} S_t(\alpha, g_0) = -\frac{1}{2n} \sum_{t=1}^{n} A_t(\phi),$$

$$S^n(\beta, g_0) = \frac{1}{n} \sum_{t=1}^{n} S_t(\beta, g_0) = -\frac{1}{2n} \sum_{t=1}^{n} A_t(\phi)D_t(\psi) - \frac{1}{2n} \sum_{t=1}^{n} \frac{1}{\epsilon_t(\phi)} x_t g_0(\epsilon_t(\phi)), 
$$

$$S^n(\gamma, g_0) = \frac{1}{n} \sum_{t=1}^{n} S_t(\gamma, g_0) = -\frac{1}{2n} \sum_{t=1}^{n} A_t(\phi)B_t(\psi),$$

$$S^n(\theta, g_0) = \frac{1}{n} \sum_{t=1}^{n} S_t(\theta, g_0) = -\frac{1}{2n} \sum_{t=1}^{n} A_t(\phi)C_t(\psi),$$

and

$$S^n(\phi, g_0) = (S^n(\alpha, g_0)^T, S^n(\beta, g_0)^T, S^n(\delta, g_0)^T)^T,$$

$$S^n(\delta, g_0) = (S^n(\gamma, g_0)^T, S^n(\theta, g_0)^T)^T,$$

where

$$A_t(\phi) = 1 + \frac{\epsilon_t(\phi) g_0(\epsilon_t(\phi))}{g_0(\epsilon_t(\phi))},$$

$$B_t(\psi) = \frac{1}{\sigma_t^2(\psi)} \left[ u_t^2(\psi) + \theta \frac{\partial \sigma_t^2(\psi)}{\partial \gamma} \right],$$

$$C_t(\psi) = \frac{1}{\sigma_t^2(\psi)} \left[ \sigma_t^2(\psi) + \theta \frac{\partial \sigma_t^2(\psi)}{\partial \theta} \right],$$

$$D_t(\psi) = \frac{1}{\sigma_t^2(\psi)} \left[ -2\gamma x_t u_{t-1}(\psi) + \theta \frac{\partial \sigma_t^2(\psi)}{\partial \beta} \right],$$

and

$$\epsilon_t(\phi) = \frac{u_t(\psi)}{v_t(\phi)} = \frac{y_t - x_t \beta}{v_t(\phi)}. $$
Notice that \( A_t \) is independent of \( B_t \) and \( C_t \) since \( B_t \) and \( C_t \) depend solely on pre-determined variables. In addition, \( B_t, C_t \) are stationary ergodic processes and bounded from above.

For any scalar nuisance parameter \( \eta \) which parameterizes \( g_0 \), the score vector for \( \eta \) is

\[
S^n(\eta, g_0) = \frac{1}{n} \sum_{t=1}^{n} S_t(\eta, g_0) = \frac{1}{n} \sum_{t=1}^{n} \frac{g'(t)}{g_0(t)}(\varepsilon_t).
\]

(2.21)

When \( \varepsilon_t \) is i.i.d with zero mean and its density \( g_0 \) symmetric about zero, we have the following properties.

- \( S_t(\eta, g_0) \) is symmetric w.r.t. \( \varepsilon_t \);
- \( g_0(\varepsilon_t) \) is anti-symmetric w.r.t. \( \varepsilon_t \);
- \( A_t(\phi) \) is symmetric w.r.t. \( \varepsilon_t \);
- \( B_t(\psi) \) and \( C_t(\psi) \) are symmetric w.r.t. past \( \varepsilon \)'s;
- \( D_t(\psi) \) is anti-symmetric w.r.t. past \( \varepsilon \)'s.

Given these properties, we consider the covariance of \( S_t(\beta, g_0) \) and \( S_t(\eta, g_0) \), which is computed by the following two steps. We first take the expectation of the product of \( S_t(\beta, g_0) \) and \( S_t(\eta, g_0) \) conditional on the past information of time \( t \) and then take the expectations with respect to the past \( \varepsilon \)'s. In the first step, the contribution from the second term of (2.11) is zero since it is antisymmetric (w.r.t. \( \varepsilon_t \)), however the contribution from the first term is nonzero. It is only when we take expectations with respect to the past \( \varepsilon \)'s that the antisymmetry of \( D_t(\psi) \) w.r.t. the past \( \varepsilon \)'s gives us zero. Then the score of \( \beta \) is orthogonal to the score of \( \eta \). However, to show that \( \beta \) is adaptively estimable, we have to prove that the efficient score of \( \beta \) with unknown \( \alpha \) is orthogonal to the score of \( \eta \).

Observing that \( A_t(\phi) \) is symmetric w.r.t. \( \varepsilon_t \), \( S_t(\alpha, g_0) \) is also symmetric w.r.t. \( \varepsilon_t \). Then the covariance of \( S_t(\beta, g_0) \) and \( S_t(\alpha, g_0) \) is zero by similar calculations as that of the covariance of \( S_t(\beta, g_0) \) and \( S_t(\eta, g_0) \), which is also true for the covariance of \( S_t(\beta, g_0) \) and \( S_t(\delta, g_0) \) since \( B_t(\psi) \) and \( C_t(\psi) \) are symmetric w.r.t. the past \( \varepsilon \)'s. In other words, \( I_{\beta\alpha} = 0 \), then the efficient score of \( \beta \) is exactly the same as the ordinary score of \( \beta \). Therefore, the efficient score of \( \beta \) is orthogonal to \( S^n(\eta, g_0) \) and \( \beta \) is adaptively estimable.
In the following, we are going to show that the orthogonality condition also holds for \( \delta \), i.e. for both \( \gamma \) and \( \theta \).

The efficient score functions for \( \gamma \) and \( \theta \) in the presence of unknown \( \alpha \) are obtained by projecting \( S^n(\gamma, g_0) \) and \( S^n(\theta, g_0) \) off \( S^n(\alpha, g_0) \). By the definition, the efficient score of \( \gamma \) is

\[
S^{n*}(\gamma, g_0) \equiv \frac{1}{n} \sum_{t=1}^{n} S^*_t(\gamma, g_0) \\
\equiv S^n(\gamma, g_0) - I_{\gamma\alpha} I_{\alpha\alpha}^{-1} S^n(\alpha, g_0).
\]

Observing that \( A_t \) is a scalar and \( B_t \) is independent of \( A_t \), we then have

\[
I_{\gamma\alpha} = E[S_t(\gamma, g_0)S_t(\alpha, g_0)] = \frac{1}{4} \{ E[ A_t(\phi)]^2 \} E[ B_t(\psi)] = I_{\alpha\alpha} E[ B_t(\psi)],
\]

and the efficient score of \( \gamma \)

\[
S^{n*}(\gamma, g_0) = -\frac{1}{2n} \sum_{t=1}^{n} A_t(\phi) (B_t(\psi) - \bar{B}(\psi)). \tag{2.22}
\]

Similarly, we have

\[
S^{n*}(\theta, g_0) \equiv \frac{1}{n} \sum_{t=1}^{n} S^*_t(\theta, g_0) \\
\equiv S^n(\theta, g_0) - I_{\theta\alpha} I_{\alpha\alpha}^{-1} S^n(\alpha, g_0) \\
= -\frac{1}{2n} \sum_{t=1}^{n} A_t(\phi) (C_t(\psi) - \bar{C}(\psi)), \tag{2.23}
\]

where \( \bar{B}(\psi) = E[B_t(\psi)] \) and \( \bar{C}(\psi) = E[C_t(\psi)] \). It is not hard to see that \( S^{n*}(\gamma, g_0) \) and \( S^{n*}(\theta, g_0) \) are orthogonal to the score function \( S^n(\eta, g_0) \) by recalling that \( B_t \) and \( C_t \) are independent of \( A_t \) and \( S_t(\eta, g_0) \). Therefore, \( \gamma \) and \( \theta \) are adaptively estimable without the information of any nuisance parameter \( \eta \). It is worth noting that the orthogonality conditions of \( \gamma \) and \( \theta \) do not require the symmetry condition.

In the above consideration, we assume that the density function of the error term is symmetric, which is fairly restrictive. We then examine model (2.2) and (2.5), where there is a free intercept in equation (2.2). Although this model is a special case of (2.1), it is
definitely not restrictive at all since many linear regression models include an intercept.

As found by Manski (1984), \( \beta_2 \) in model (2.2) can be adaptively estimable even without
the symmetry condition. If we denote \( \psi_2 = (\beta_2^T, \delta^T)^T \), then all the identifiable parameters
in model (2.2) and (2.5) except \( \beta_1 \), i.e. \( \psi_2 \), remain adaptively estimable without the
symmetry condition on the error density.

3. Semiparametric Maximum Likelihood Estimation

3.1 The Definition of the Estimator

In Section 2.2, we verified the adaptive condition for \( \psi \). In principle, we should be able to
construct adaptive estimator for \( \psi \) with symmetry condition and for \( \psi_2 \) without symme-
try condition. For ARCH types of models, Linton (1993) and Drost and Klaassen (1997)
adopted the AML approach to construct the adaptive estimators for the adaptively es-
timable parameters in their models. We could, of course, use AML approach to construct
adaptive estimators for the adaptively estimable parameters in our GARCH(1,1) setting.
And all the asymptotic properties obtained by those authors for their models could also
be obtained for our GARCH model without any difficulty. However, as argued by Yang
(1997), AML estimators for a linear regression model had disadvantages in finite samples.
Therefore in this section, we attempt to use the SML approach to construct adaptive
estimators for the adaptively estimable parameters.

Most researchers in the adaptive estimation literature started from the discussion of
Local Asymptotic Normality (LAN). Instead of starting from primitive conditions, we fol-
low Manski (1984) and simply assume that the estimation problem is sufficiently regular
that, for \( g_0 \), the maximum likelihood estimator of \( \phi \) exists. We then attempt to construct
adaptive estimators of the adaptively estimable parameters using the semiparametric max-
imum likelihood approach.

Since \( \alpha \) and \( g_0 \) cannot be jointly identified, we assume that \( \alpha \) is absorbed by \( g_0 \) and
define \( w_t(\psi) = (y_t - x_t\beta)/\sigma_t(\psi) = e^{\alpha_0/2}\varepsilon_t \), where \( \alpha_0 \) denotes the true value of \( \alpha \). The
density function for $w_t$ is $g(w_t)$, where

$$g(w_t) = e^{-\alpha / 2} g_0(e^{-\alpha / 2} w_t)$$

and the log-likelihood under $g$ and $\psi$ becomes

$$\ell_n(y|x; \psi, g) = -\frac{1}{2} \sum_{t=1}^{n} \log \sigma_t^2(\psi) + \sum_{t=1}^{n} \log g(w_t(\psi)). \quad (3.1)$$

Without the knowledge of $g_0$, MLE is not available but the least-squares estimators and Gaussian QML estimators of the parameters $\phi$, which includes an estimator of $\alpha$, are proved to be $\sqrt{n}$-consistent with correct specifications of the first and second order moments and with the fourth moment assumptions (see Bollerslev and Wooldridge: 1988) described in the following.

**The fourth moment assumptions:**

$$\int \varepsilon^4 g_0(\varepsilon) d\varepsilon < +\infty, \quad E(\sigma_t^4) < +\infty,$$

$$\int s_1^4(\varepsilon) g_0(\varepsilon) d\varepsilon < +\infty, \quad \int s_2^4(\varepsilon) g_0(\varepsilon) d\varepsilon < +\infty,$$

where

$$s_1 = \frac{g'(\varepsilon)}{g_0(\varepsilon)}; \quad s_2 = \varepsilon \frac{g'(\varepsilon)}{g_0(\varepsilon)} + 1.$$  

Given the $\sqrt{n}$-consistent estimates of $\psi$, which are denoted as $\psi_n \equiv (\beta_n^T, \delta_n^T)^T$, we then have residuals $\tilde{u}_{tn} = y_t - x_t \beta_n$ and $\tilde{w}_{tn} = \tilde{u}_{tn}/\sigma_t(\psi_n)$, $t = 1, \ldots, n$. These residuals $\{\tilde{w}_{1n}, \ldots, \tilde{w}_{nn}\}$ will be used to construct nonparametric density function estimates of $g$. The Kernel method with normal kernel has been chosen for this purpose. Define the convolution of $g_0$ and $\phi_{\lambda_n}$ as

$$g_{\lambda_n}(w) = g_0 * \phi_{\lambda_n}(w) = \int_{-\infty}^{+\infty} \phi_{\lambda_n}(w - z) g_0(z) dz,$$

where $\phi_{\lambda_n}$ is the $N(0, \lambda_n^2)$ density. We consider the idealized situation in which we are able to access the true disturbance terms. The density estimate of $g_0(w_t)$ with the leave-one-out is then of the following form,

$$g_{e\lambda_n}(w_t) \equiv (n - 1)^{-1} \sum_{s \neq t}^{n} \phi_{\lambda}(w_t - w_s). \quad (3.2)$$
If \( g_0 \) is symmetric, we then define \( g_{e\lambda_n}(w_t) \) as

\[
g_{e\lambda_n}(w_t) = (2(n - 1))^{-1} \sum_{s \neq t}^{n} \{ \phi_{\lambda_n}(w_t - w_s) + \phi_{\lambda_n}(w_t + w_s) \},
\]

where \( \lambda_n \) is the bandwidth parameter. The naive log-density estimate is then

\[
\ell_{e\lambda_n}(w_t) = \log(g_{e\lambda_n}(w_t)).
\]

We suppress dependence on \( w_1, \ldots, w_n \). For given trimming parameters \( n, b_n, c_n, d_n \) and \( e_n \) for the same purpose described in Section 3.2 of Yang (1997), we define the non-parametric log-density with trimming

\[
\ell_{n\lambda_n}(w; \psi) =
\begin{cases} 
\ell_{e\lambda_n}(w) & \text{if } g_{e\lambda_n}(w) \geq d_n, |w| \leq e_n, |g'_{e\lambda_n}(w)| \leq c_n g_{e\lambda_n}(w) \text{ and } |g^{(2)}_{e\lambda_n}(w)| \leq b_n g_{e\lambda_n}(w) \\
1 & \text{otherwise.}
\end{cases}
\]

If \( w \) does not satisfy the four conditions in (3.5), we simply trim out its contribution to the log-likelihood, in other words, we set the weight on the contribution by \( w \) to the log-likelihood as 0. For notational purposes, we define \( g_{n\lambda_n}(w; \psi) =
\begin{cases} 
g_{e\lambda_n}(w) & \text{if } g_{e\lambda_n}(w) \geq d_n, |w| \leq e_n, |g'_{e\lambda_n}(w)| \leq c_n g_{e\lambda_n}(w) \text{ and } |g^{(2)}_{e\lambda_n}(w)| \leq b_n g_{e\lambda_n}(w) \\
1 & \text{otherwise.}
\end{cases}
\]

Noting that \( g_{n\lambda_n}(w; \psi) \) is no longer a density function, we find, however,

\[
\ell_{n\lambda_n}(w; \psi) = \log(g_{n\lambda_n}(w; \psi)).
\]

Actually, \( w_1, \ldots, w_n \) in our model are unobservable, we then construct \( \tilde{g}_{n\lambda_n}(w; \psi_n) \) and \( \tilde{\ell}_{n\lambda_n}(w; \psi_n) \) in the same way as (3.5) and (3.6) with \( w_1, \ldots, w_n \) replaced by \( \tilde{w}_1, \ldots, \tilde{w}_n \).

They are called pseudo nonparametric density function and pseudo nonparametric log-density respectively. We then define SML estimator of model (2.1) & (2.5).

**Definition 3.1.**
The SMLE of $\psi$ in model (2.1) & (2.5) is defined as

$$
\hat{\psi}_n \equiv \arg_{\psi} \max \left\{ -\frac{1}{2} \sum_{t=1}^{n} \log \sigma_t^2(\psi) + \sum_{t=1}^{n} \tilde{\ell}_{n\lambda_n}(\frac{y_t - x_t^\top \beta}{\sigma_t(\psi)}; \psi_n) \right\},
$$

where $\tilde{\ell}_{n\lambda_n}(\frac{y_t - x_t^\top \beta}{\sigma_t(\psi)}; \psi_n)$ is the pseudo nonparametric log-likelihood function contributed by observation $t$.

This estimator is constructed using almost the same procedure as described in Yang (1997) with some modifications regarding the heteroskedasticity presented in the model. Under regularity assumptions, we can prove that this estimator is adaptive or asymptotically equivalent to ML estimator with the knowledge of $g_0$ for all the adaptively estimable parameters.

### 3.2 Asymptotic Properties of SMLE Under Symmetry Condition

Following Davidson and MacKinnon (1993, Chapter 8), we define

$$
\bar{\ell}(\phi; \phi_0) \equiv \text{plim}_0 \ell^n(\phi; \phi_0) \equiv \text{plim}_0 \left( \frac{1}{n} \ell_n(y|x; \phi, g_0) \right),
$$

where $\ell_n(y|x; \phi, g_0)$ is the nonparametric log-likelihood function contributed by observation $t$, i.e.

$$
\ell_{n\lambda_n}(y_t|x_t; \psi) = -\frac{1}{2} \log \sigma_t^2(\psi) + \ell_{n\lambda_n}(w_t(\psi)).
$$

The functions $\bar{\ell}(\phi; \phi_0)$ and $\bar{\ell}(\psi; \psi_0)$ are respectively the limiting values of $n^{-1}$ times the log-likelihood function under $g_0$ and under $g$, where in the latter case $\alpha$ takes the
true value $\alpha_0$. $\tilde{L}_{n, \lambda_n}(\psi)$ is the limiting value of \(n^{-1}\) times the nonparametric log-likelihood functions associated with \(g\) rather than with \(g_0\). We also make some other assumptions.

**Assumption 3.1**

The density \(g_0\) is bounded and absolutely continuous with respect to Lebesgue measure. It has zero mean and finite Fisher information for both scale and location parameters, i.e.

\[
0 < I_1(g_0) = \int s_1^2(\varepsilon)g_0(\varepsilon)d\varepsilon < +\infty, \quad 0 < I_2(g_0) = \int s_2^2(\varepsilon)g_0(\varepsilon)d\varepsilon < +\infty
\]

where

\[
s_1(\varepsilon) = \frac{g_0'(\varepsilon)}{g_0(\varepsilon)}, \quad s_2(\varepsilon) = 1 + \varepsilon \frac{g_0'(\varepsilon)}{g_0(\varepsilon)}.
\]

Here the assumption that \(g_0\) is bounded is required to prove the consistency of SMLE. It will not restrict the class of densities since most commonly used density functions are bounded from above. The other conditions in Assumption 1 are only regularity conditions on GARCH models for the existence of Maximum Likelihood Estimators.

**Assumption 3.2**

\(\phi_0 \in \Phi, \Phi\) is compact and has a nonempty interior.

**Assumption 3.3**

\[E\sup_{\phi \in \Phi} |\ell(y|x; \phi)| < +\infty, \ell(y|x; \phi)\) is continuously twice differentiable function of \(\phi, and

\[
\max_{\phi \in \Phi} \ell(\phi, g_0)
\]

has an unique solution.

The first two parts of this assumption ensure that \(\{\ell(y_t|x_t; \phi), t = 1, \cdots, n\}\) satisfies the uniform weak law of large numbers. The second part of this assumption is the asymptotic identification condition. Given this assumption and assumptions 3.1, 3.2, and 3.4, the MLE exists with the knowledge of \(g_0\) and the unique solution of (3.12) should be \(\phi_0\). Since \(g\) combines \(g_0\) and \(e^{\alpha}\), then we have unique solution for the following maximization
problem,

$$\max_{\psi \in \Psi} \tilde{\ell}(\psi; g).$$  \hfill (3.13)

**Assumption 3.4**

$$\partial^2 \ell(y_t | x_t; \phi) / \partial \phi \partial \phi^T$$ and $$(\partial \ell(y_t | x_t; \phi) / \partial \phi)(\partial \ell(y_t | x_t; \phi) / \partial \phi)^T$$ satisfy the uniform weak law of large numbers.

**Assumption 3.5**

The trimming parameters satisfy:

a. $b_n, d_n, \lambda_n \to 0$, $c_n, e_n \to +\infty$.

b. $nd_n^2 \lambda_n^2 \to +\infty$, $c_n \lambda_n \to 0$, $b_n \lambda_n \to 0$, $e_n \lambda_n^{-3} = o(n)$, $n \lambda_n \to +\infty$ and $ne_n^{-2} \lambda_n^3 c_n^{-2} \to +\infty$ as $n \to +\infty$.

Given the contiguity of the sequence $\{w_{tn}\}, t = 1, \ldots, n$ to the sequence $\{w_t\}, t = 1, \ldots, n$, we can consider the idealized situation where the true disturbances $w_1, \ldots, w_n$ are observed.

To show the uniform consistency of $\hat{\psi}_n$, it suffices to prove, with the asymptotic identification condition, that

$$\tilde{\ell}_{n\lambda_n}(\psi; \psi_0) = \ell(\psi; \psi_0),$$  \hfill (3.14)

in other words, the difference between $\ell_{n\lambda_n}^n(\psi; \psi_0)$ and $\ell^n(\psi; \psi_0)$ goes to zero uniformly in probability when $n$ goes to infinity. If we look at $\ell_{n\lambda_n}^n(\psi; \psi_0)$ and $\ell^n(\psi; \psi_0)$ more closely, we find that each of them has two components and the same first component, the sample mean of $(-1/2) \log(\sigma_t^2(\psi_0))$. Then we only need to show that the difference between the sample mean of $\log(g_{n\lambda_n}(w_t; \psi_0))$ and the sample mean of $\log(g(w_t))$ goes to zero in probability. As found from Theorem 1.1 in Yang (1997), we are able to control the convergence of the sample mean of the nonparametric log-likelihood to the sample mean of the true log-likelihood by using some suitable trimming rules.

**Theorem 3.1**:
Under Assumptions 3.1, 3.2, 3.3, and 3.5, 
\( \ell_{n\lambda_n}^{m}(\psi; \psi_0) - \ell^{m}(\psi; \psi_0) = o_p(1) \), uniformly in \( \psi \), then \( \hat{\psi}_n \) is a uniformly consistent estimator of \( \psi_0 \).

Proof: See appendix A.

Theorem 3.1 proves that SMLE is consistent for models specified by (2.1) & (2.5). It is worth noting that the consistency of \( \hat{\psi}_n \) does not require the symmetry assumption and only requires the mean absolute convergence of the sample mean of the nonparametric log-likelihood to the sample mean of the true log-likelihood since we have the asymptotic identification assumption. This can be satisfied for models with more general conditional density functions without the symmetry condition.

In section 2.2, we showed that the necessary conditions of adaptation are satisfied for \( \beta \) and \( \delta \) when the conditional distribution is assumed to be symmetric, we should be able to estimate them adaptively under regularity conditions. We will show that SMLE is an adaptive estimator. In our model, an estimator of \( \psi \) is adaptive if this estimator is asymptotically as efficient as the ML estimator with the knowledge of \( g_0 \) rather than with the knowledge of \( g \), where the latter combines the knowledge of both \( g_0 \) and \( e^{\alpha_0} \). Therefore, the information bound of \( \psi \) given the knowledge of \( g_0 \) is quite different from that given the knowledge of \( g \), simply because in the former case the parameter \( \alpha \) is unknown. Although the SML estimation of \( \psi \) does not require an estimation or knowledge of \( \alpha \), we still need to keep in mind that \( \alpha \) is an unknown parameter of interest.

To find out the best attainable precision or the information bound of \( \psi \) in the presence of unknown \( \alpha \), we begin by considering the information matrix for \( \phi \) under \( g_0 \) with the symmetry assumption. Following the discussion in section 2.2, the information matrix of \( \phi \), which combines \( \alpha \) and \( \psi \), is as follows,

\[
I_{\phi\phi} = \begin{pmatrix}
I_{\alpha\alpha} & 0 & I_{\alpha\delta} \\
0 & I_{\beta\beta} & 0 \\
I_{\delta\alpha} & 0 & I_{\delta\delta}
\end{pmatrix},
\]

(3.15)

\( I_{\phi\phi} \) is assumed strictly positive definite and \( I_{\phi\phi}^{-1} \) is the best attainable precision of the
estimation of $\phi$, which can be reached by the ML estimator of $\phi$ with the knowledge of $g_0$. However, if $\alpha$ is known, in other words, if we have the knowledge of $g$ rather than only the knowledge of $g_0$, then the information matrix of $\psi$ would be

$$I_{\psi\psi} = \begin{pmatrix} I_{\beta\beta} & 0 \\ 0 & I_{\delta\delta} \end{pmatrix}.$$ 

$I_{\psi\psi}^{-1}$ is the best attainable precision of the estimation of $\psi$ with known $\alpha$, which can be reached by the ML estimator of $\psi$ with the knowledge of $g$.

Since the parameter $\alpha$ is unknown in our setting, the information bound for $\psi$ in the presence of unknown $\alpha$ is $I_{\psi\psi}^{-1}$, where

$$I_{\psi\psi}^* = \begin{pmatrix} I_{\beta\beta} & 0 \\ 0 & I_{\delta\delta}^* \end{pmatrix},$$ (3.16)

and

$$I_{\delta\delta}^* = I_{\delta\delta} - I_{\delta\alpha}I_{\alpha\alpha}^{-1}I_{\alpha\delta}.$$ (3.17)

$I_{\delta\delta}^{-1}$ is actually the lower right sub-matrix of $I_{\psi\psi}^{-1}$. To show that $\hat{\psi}_n$ is adaptive, we need to prove that the $\sqrt{n} \cdot (\hat{\psi}_n - \psi_0)$ converges in distribution to $N(0, I_{\psi\psi}^{-1})$.

In the idealized situation, if we had the knowledge of $g_0$, we could obtain the ML estimators of $\psi$ as well as $\alpha$. If we denote the ML estimator of $\phi$ with the knowledge of $g_0$ as $\phi_n = (\alpha_n, \psi_n^T)^T$, then $\sqrt{n} \cdot (\phi_n - \phi_0)$ converges in distribution to $N(0, I_{\psi\psi}^{-1})$ and $\sqrt{n} \cdot (\psi_n - \psi_0)$ converges in distribution to $N(0, I_{\psi\psi}^{-1})$ given the regularity conditions on $g_0$. The problem then becomes that $\sqrt{n} \cdot \hat{\psi}_n$ and $\sqrt{n} \cdot \psi_n$ have the same asymptotic distributions. We will verify the adaptive property of $\hat{\psi}_n$ by making a comparison of $\sqrt{n} \cdot (\hat{\psi}_n - \psi_0)$ and $\sqrt{n} \cdot (\psi_n - \psi_0)$.

Let the score function in (3.1) for observation $t = 1, ..., n$ under $g$ be denoted as $S_t(\psi, g)$, which is defined in the same way as in (2.11) to (2.14) with $g$, $\sigma_t$ and $w_t$ instead of $g_0$, $v_t$ and $\varepsilon_t$. The sample mean of score vector is $S^n(\psi, g) = \frac{1}{n} \sum_{i=1}^{n} S_t(\psi, g)$. Then the first-order conditions for $\hat{\psi}_n$ can be written as,

$$S^n(\hat{\psi}_n, \bar{g}_n \lambda_n) = 0.$$ (3.18)
Taking a Taylor expansion, we have that

\[ 0 = \sqrt{n} S^n(\hat{\psi}_n, \tilde{g}_{n\lambda_n}) \]

\[ = \sqrt{n} S^n(\psi_0, \tilde{g}_{n\lambda_n}) + \frac{\partial S^n(\psi^*, \tilde{g}_{n\lambda_n})}{\partial \psi} \sqrt{n}(\hat{\psi}_n - \psi_0), \quad (3.19) \]

where \( \psi^* \) lies on the line segment joining \( \hat{\psi}_n \) and \( \psi_0 \).

To make the comparison of \( \sqrt{n} \cdot \hat{\psi}_n \) and \( \sqrt{n} \cdot \psi_n \), we also present in the following the first order conditions of ML estimator \( \phi_n \) given the knowledge of \( g_0 \), which are

\[
\begin{pmatrix} 
S^n(\alpha_n, g_0) \\
S^n(\phi_n, g_0)
\end{pmatrix} = 0. \quad (3.20)
\]

Taking a Taylor expansion of \( S^n(\phi_n, g_0) \), then we have

\[ 0 = \sqrt{n} S^n(\psi_n, g_0) \]

\[ = \sqrt{n} S^n(\psi_0, g_0) + \frac{\partial S^n(\psi^{**}, g_0)}{\partial \psi} \sqrt{n}(\psi_n - \psi_0), \quad (3.21) \]

where \( \psi^{**} \) lies on the line segment joining \( \psi_n \) and \( \psi_0 \).

Under suitable assumptions, we can show that

\[
\lim_{p} \left\{ \frac{\partial S^n(\psi^*, \tilde{g}_{n\lambda_n})}{\partial \psi} - \frac{\partial S^n(\psi^{**}, g_0)}{\partial \psi} \right\} \rightarrow o_p(1). \quad (3.22)
\]

The symmetry condition of \( g_0 \) is not required here since we are proving the mean absolute convergence condition rather than the mean square convergence condition. However, we verify the mean square convergence condition to show that

\[
\sqrt{n} S^n(\psi_0, \tilde{g}_{n\lambda_n}) - \sqrt{n} S^n(\psi_0, g_0) \rightarrow o_p(1). \quad (3.23)
\]

It is worth pointing out that the symmetry condition is not required for \( \sqrt{n} S^n(\delta_0, \tilde{g}_{n\lambda_n}) - \sqrt{n} S^n(\delta_0, g_0) \rightarrow o_p(1) \) but for \( \sqrt{n} S^n(\beta_0, \tilde{g}_{n\lambda_n}) - \sqrt{n} S^n(\beta_0, g_0) \rightarrow o_p(1) \). However, to show this convergence, the symmetry condition is exploited. Given that \( \sqrt{n}(\psi_n - \psi_0) \xrightarrow{d} N(0, I^{-1}(\psi, \psi)) \), we then have the following theorem.
**Theorem 3.2:** Under Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5, we then have \( \sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{d} N(0, I^{-1}(\psi_0, \psi_0)) \), the SML estimator \( \hat{\psi}_n \) is adaptive.

Proof: See Appendix B.

Theorem 3.2 proves that SMLE of \( \psi_0 \) is adaptive under the symmetry condition on the conditional distribution function of the error terms. Therefore, we are able to estimate \( \psi_0 \) efficiently even without the knowledge of the conditional distribution, as long as the conditional distribution is symmetric around zero.

### 3.3 Asymptotic Properties of SMLE Without Symmetry Condition

In Section 3.2, we developed adaptive estimators for linear GARCH models under the symmetry condition, which is required for the adaptation of structural parameters in the regression function. However, this fairly strict assumption is not necessary for all the sub-families of linear GARCH models. One of them is a linear GARCH model with free intercept described by (2.2) and (2.5). For this model, we will assume all the same assumptions as in section 3.2 except the symmetry of \( g_0 \), which is no longer assumed. Then \( \psi_2 \) satisfies the adaptive condition and is then adaptively estimable although \( \beta_1 \) is not in this case. In the following, we will show that the estimator of \( \psi_2 \) obtained by the procedure proposed in Section 3.1 is also adaptive.

Let \( \hat{\beta}_{1n}, \hat{\beta}_{2n}, \) and \( \hat{\psi}_{2n} \) denote respectively the estimators, with sample size \( n \), of \( \beta_{10}, \beta_{20}, \) and \( \psi_{20} \) obtained from the procedure in Section 3.1. We have

\[
S_{t\beta}(\psi, g) \equiv \begin{pmatrix} S_{t\beta_1}(\psi, g) \\ S_{t\beta_2}(\psi, g) \end{pmatrix}.
\]  
(3.24)

Since model (2.2) is a special form of model (2.1) and the consistency of \( \hat{\beta}_n \) in model (2.1) does not require the symmetry assumption, the consistency of \( \hat{\beta}_{1n} \) and \( \hat{\psi}_{2n} \) simply follows.

**Corollary 3.1** Under Assumption 3.1, 3.2, 3.3 and 3.5, \( \hat{\beta}_{1n} \) and \( \hat{\psi}_{2n} \) are uniformly consistent estimators of \( \beta_{10} \) and \( \psi_{20} \).

If we further denote the ML estimator of \( \phi \) with the knowledge of \( g_0 \) as \( \phi_n = (\alpha_n, \beta_{1n}, \psi_{2n})^\top \). It then suffices to show that \( \sqrt{n} \cdot \hat{\psi}_{2n} \) and \( \sqrt{n} \cdot \psi_{2n} \) have the same asympt-
totic distributions. We will verify the adaptive property of $\hat{\psi}_{2n}$ by making a comparison of $\sqrt{n} \cdot (\hat{\psi}_{2n} - \psi_20)$ and $\sqrt{n} \cdot (\psi_{2n} - \psi_20)$. The same calculations as in (3.19), (3.21) and (3.22) can be done without any change since they do not require the symmetry condition. However, (3.23) is no longer valid since the symmetry condition is exploited when proving it and $\beta_1$ is not adaptively estimable under asymmetry. However, $\hat{\psi}_{2n}$ remains adaptive estimator of $\psi_20$, which can be shown by verifying the mean square convergence condition, i.e.

$$\sqrt{n}S^n(\psi_{20}, \hat{g}_{n\lambda_n}) - \sqrt{n}S^n(\psi_{20}, g_0) \to o_p(1).$$  

(3.24)

Recall that it does not require the symmetry condition to show that $\sqrt{n}S^n(\delta_0, \tilde{g}_{n\lambda_n}) - \sqrt{n}S^n(\delta_0, g_0) \to o_p(1)$, the problem becomes

$$\sqrt{n}S^n(\beta_{20}, \hat{g}_{n\lambda_n}) - \sqrt{n}S^n(\beta_{20}, g_0) \to o_p(1).$$  

(3.25)

**Theorem 3.3:** Under Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5, we then have $\sqrt{n}(\hat{\psi}_{2n} - \psi_{20})$ has the same asymptotic distribution as $\sqrt{n}(\psi_{2n} - \psi_{20})$, the SML estimator $\hat{\psi}_n$ is adaptive.

**Proof:** See Appendix C.
4. Monte Carlo studies

4.1. Introduction

In section 3, we showed that SML estimators are adaptive for the identifiable parameters in the transformed GARCH specification with symmetry assumption on density and for the identifiable parameters except the free intercept in a GARCH model with a free intercept without the symmetry condition on density. However, the theoretical results leave questions and doubts regarding the implementation and the finite sample performance of SML estimators of the GARCH models. The concern is typically the extent of the finite sample efficiency gains by which the SMLE outperforms the other estimators, especially AMLE, since part of the motivation for suggesting SMLE rather than AMLE is because we believe that SMLE has better finite sample behavior than AMLE.

Yang (1997) compared the finite sample performance of SMLE with AMLE and OLSE in the context of linear regression models. It is found that SMLE performs the best and OLSE the worst for the adaptively estimable parameters and that the more the true density function differs from the normal density, the greater the relative efficiency gains SMLE and AMLE achieve over OLSE. In this section, we perform comparisons of SMLE, AMLE, and QMLE for a GARCH model in terms of relative efficiency and our primary criterion for evaluating relative efficiency is root mean squared error (MSE), which is usually used by researchers.

4.2 The Simulated Models

A simple GARCH(1, 1) model is studied in this experiment, which is as follows.

\[
y_t = \beta_{10} + \beta_{20} x_t + u_t, \tag{4.1}
\]

where \( x_t \) is i.i.d. and generated from an uniform distribution on \([0, 1]\), \( u_t \) follows GARCH(1,1) specification to capture the characteristics of volatility clustering. The transformed GARCH(1,1) model is described as

\[
u_t = \varepsilon_t \nu_t, \tag{4.2}\]
\[ v_t^2 = e^{\alpha_0 \sigma_t^2} = e^{\alpha_0} (1 + \gamma_0 u_{t-1}^2 + \theta_0 \sigma_{t-1}^2), \] (4.3)

where the initial values of \( u_0 \) and \( v_0 \) are set equal to zero for simplicity.

The parameter values for the simulations are:

\[ (\alpha_0, \beta_{10}, \beta_{20}, \gamma_0, \theta_0) = (0, 0.1, 0.2, 0.2, 0.7). \] (4.4)

It is easy to verify that they satisfy all the constraints described in section 3. The true conditional density of \( \epsilon_t, g_0 \), is assumed to be either leptokurtic or asymmetric. Leptokurtic densities differ from a normal density in that they have higher tail probabilities. These densities capture both the large number of outliers and the shape of many daily stock returns and daily exchange rate series. Asymmetric densities differ from a normal density in that they are skewed and have higher tail probabilities. These densities capture the outliers and skewness that characterize many stock returns. The specific densities we consider are

(a) \( t \) with five degrees of freedom \( (t_5) \);

(b) bimodal symmetric mixture of two normals, \( 0.5(N(-2, 1) + N(2, 1)) \ (BN) \);

(c) log normal, \( \log(\epsilon) \) is \( N(0, 1) \) \( (LN) \).

\( t_5 \) and \( BN \) are leptokurtic, and \( LN \) asymmetric. In order to identify the intercept parameter \( \beta_{10} \), we need to center, where necessary, the various distributions at mean zero. For the purposes of comparability, we also rescale the above distributions so that they all have unit variance.

When \( g_0 \) takes on any of the forms \( t_5 \) and \( BN \), either SMLE or AMLE for \( \psi_0 \) is adaptive. They should theoretically outperform QMLE of \( \psi_0 \). When \( g_0 \) takes on the form \( LN \), SMLE for all the parameters of \( \psi_0 \) except \( \beta_{10} \) are adaptive.

4.3 Implementation of the Study

One concern, which appears in this Monte Carlo study, is how to select the smoothing and trimming parameters called for in the nonparametric density estimator. Hsieh
and Manski (1987) conducted Monte Carlo experiments to compare the small sample properties of AMLE and other estimators for a linear regression model. For sample size 50, they selected \( \lambda_n \) from a number of values ranging from 0.01 to 2.00 and \( [e_n = \rho, d_n = \exp(-\rho^2/2), c_n = \rho] \), where \( \rho \) was allowed to take on the values of 3, 4 and 8. They observed the best performances of AMLE appearing when \( \rho \) was 8 and \( \lambda \) was between 0.05 and 0.5 varying across distributions.

Since how to optimally select these parameters is not our main concern, we simply follow Hsieh and Manski (1987) by choosing \( [e_n = \rho, d_n = \exp(-\rho^2/2), c_n = \rho] \). However, with reasonable large sample sizes, 200 and 1000 in our experiments, we allow \( \rho \) to take the value 8, and \( \lambda_n \) to take the value 0.125. This is because, when \( n \to \infty \), \( \lambda_n \) and \( d_n \) should go to zero, and \( e_n \) and \( c_n \) should go to infinity according to theoretical considerations. Therefore we set a relative big value for \( \rho \) and relative small value for \( \lambda_n \). We set \( b_n = \infty \) for purposes of comparability simply because there is no estimate involved for the second order derivative of density function for AMLE. We use normal density as a kernel to estimate the nonparametric density function. In each case, 1000 replications will be performed. The number of replications is quite small because the computation of GARCH model is burdensome, especially when the sample size is as large as 1000.

4.4 The Results

Table 1-6 report results on root MSE for the SML, AML and QML estimators and the ratios of root MSE of ML estimator to the other estimators.

The findings can be summarized below.

Relative Performance of SMLE, AMLE and QMLE in the Symmetric Cases. In the symmetric cases, both SMLE and AMLE are adaptive; they should in principle outperform the Gaussian QMLE. From our results, we find that SMLE outperforms both AMLE and QMLE, typically for the parameters in the regression function. Although AMLE outperforms QMLE for the parameters in the regression function, it behaves poorly for the parameters characterizing the conditional variance. We have included MLE into our
comparison, which allows us to examine how much asymptotic efficiency is achieved by the adaptive estimators in finite samples. The results show that SML estimator is sometimes very close to ML estimator in terms of efficiency.

Relative Performance of SMLE, AMLE and QMLE in the log-normal Case.

In the log-normal case, $\beta_{10}$ is not adaptively estimable by either SML approach or AML approach. Our Monte Carlo results show that SMLE still outperforms AMLE and QMLE for the adaptively estimable parameters except for $\theta_0$, where SMLE is slightly worse than QMLE. Although $\beta_{10}$ is not adaptively estimable, SMLE of $\beta_{10}$ may or may not be worse than QMLE. Nevertheless, in these asymmetric cases, AMLE behaves the worst, especially for the parameters in the conditional variance, which suggests that AML approach is probably not very reliable to estimate these parameters. This is not surprising. As we argued before, AML estimator could behave poorly because it is a one-step procedure and utilizes the OPG matrix to estimate the information matrix. There is a numerical problem which could possibly also cause the poor finite sample performance of AML estimator.

When we implement ML, SML, or QML estimations, we are able to impose the constraint on the parameters in the conditional variance by using a constrained optimization procedure, while this constraint is impossible to be maintained when obtaining AML estimates. This finding is somehow inconsistent with that in the simulation studies of Drost and Klaassen (1997), where AML estimators behaved quite well. However, we observe the simplicity of their model, where no specification of the first moment is included.

The Effect of the Sample Size

We have conducted Monte Carlo experiments for the GARCH model with sample size 200 and 1000. For the purpose of comparability, we set all the smoothing parameter and the trimming parameters the same values. We find the ranking orders of alternative estimators are consistent with the two sample sizes.
### Table 1. Root MSE and Relative Efficiency, t5 errors (n=200)

<table>
<thead>
<tr>
<th>parameter</th>
<th>$\beta_{10}$</th>
<th>$\beta_{20}$</th>
<th>$\gamma_0$</th>
<th>$\theta_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.30401</td>
<td>0.54684</td>
<td>0.11246</td>
<td>0.16690</td>
</tr>
<tr>
<td>SMLE</td>
<td>0.33977</td>
<td>0.60319</td>
<td>0.11290</td>
<td>0.20107</td>
</tr>
<tr>
<td>AMLE</td>
<td>0.36341</td>
<td>0.65642</td>
<td>0.22083</td>
<td>0.28458</td>
</tr>
<tr>
<td>QMLE</td>
<td>0.35346</td>
<td>0.63799</td>
<td>0.11506</td>
<td>0.20121</td>
</tr>
<tr>
<td>MLE/SMLE</td>
<td>0.89476</td>
<td>0.90658</td>
<td>0.99606</td>
<td>0.83008</td>
</tr>
<tr>
<td>MLE/AMLE</td>
<td>0.83656</td>
<td>0.83306</td>
<td>0.50924</td>
<td>0.58647</td>
</tr>
<tr>
<td>MLE/QMLE</td>
<td>0.86011</td>
<td>0.85713</td>
<td>0.97733</td>
<td>0.82949</td>
</tr>
</tbody>
</table>

### Table 2. Root MSE and Relative Efficiency, BN errors (n=200)

<table>
<thead>
<tr>
<th>parameter</th>
<th>$\beta_{10}$</th>
<th>$\beta_{20}$</th>
<th>$\gamma_0$</th>
<th>$\theta_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.21167</td>
<td>0.35881</td>
<td>0.07897</td>
<td>0.09270</td>
</tr>
<tr>
<td>SMLE</td>
<td>0.27343</td>
<td>0.40152</td>
<td>0.07995</td>
<td>0.10042</td>
</tr>
<tr>
<td>AMLE</td>
<td>0.32127</td>
<td>0.52357</td>
<td>0.12826</td>
<td>0.18406</td>
</tr>
<tr>
<td>QMLE</td>
<td>0.39071</td>
<td>0.67977</td>
<td>0.08180</td>
<td>0.10152</td>
</tr>
<tr>
<td>MLE/SMLE</td>
<td>0.77413</td>
<td>0.89364</td>
<td>0.98773</td>
<td>0.92314</td>
</tr>
<tr>
<td>MLE/AMLE</td>
<td>0.65885</td>
<td>0.68532</td>
<td>0.61570</td>
<td>0.50363</td>
</tr>
<tr>
<td>MLE/QMLE</td>
<td>0.54175</td>
<td>0.52784</td>
<td>0.96536</td>
<td>0.91313</td>
</tr>
</tbody>
</table>

### Table 3. Root MSE and Relative Efficiency, LN errors (n=200)

<table>
<thead>
<tr>
<th>parameter</th>
<th>$\beta_{10}$</th>
<th>$\beta_{20}$</th>
<th>$\gamma_0$</th>
<th>$\theta_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.09516</td>
<td>0.13711</td>
<td>0.13005</td>
<td>0.22116</td>
</tr>
<tr>
<td>SMLE</td>
<td>0.12696</td>
<td>0.16691</td>
<td>0.15500</td>
<td>0.31374</td>
</tr>
<tr>
<td>AMLE</td>
<td>0.30786</td>
<td>0.28486</td>
<td>1.19580</td>
<td>0.94878</td>
</tr>
<tr>
<td>QMLE</td>
<td>0.13435</td>
<td>0.22388</td>
<td>0.51839</td>
<td>0.31061</td>
</tr>
<tr>
<td>MLE/AMLE</td>
<td>0.74952</td>
<td>0.82148</td>
<td>0.83901</td>
<td>0.70491</td>
</tr>
<tr>
<td>MLE/QMLE</td>
<td>0.30909</td>
<td>0.48133</td>
<td>0.10876</td>
<td>0.23310</td>
</tr>
<tr>
<td>MLE/QMLE</td>
<td>0.70827</td>
<td>0.61244</td>
<td>0.25087</td>
<td>0.71201</td>
</tr>
</tbody>
</table>
Table 4. Root MSE and Relative Efficiency, t5 errors (n=1000)

<table>
<thead>
<tr>
<th>parameter</th>
<th>$\beta_{10}$</th>
<th>$\beta_{20}$</th>
<th>$\gamma_0$</th>
<th>$\theta_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.14386</td>
<td>0.25170</td>
<td>0.05734</td>
<td>0.06286</td>
</tr>
<tr>
<td>SMLE</td>
<td>0.15421</td>
<td>0.26475</td>
<td>0.06514</td>
<td>0.07782</td>
</tr>
<tr>
<td>AMLE</td>
<td>0.15608</td>
<td>0.26997</td>
<td>0.08822</td>
<td>0.12486</td>
</tr>
<tr>
<td>QMLE</td>
<td>0.17309</td>
<td>0.30010</td>
<td>0.06594</td>
<td>0.07922</td>
</tr>
<tr>
<td>MLE/SMLE</td>
<td>0.93284</td>
<td>0.95071</td>
<td>0.88032</td>
<td>0.80773</td>
</tr>
<tr>
<td>MLE/AMLE</td>
<td>0.92167</td>
<td>0.93232</td>
<td>0.64999</td>
<td>0.50345</td>
</tr>
<tr>
<td>MLE/QMLE</td>
<td>0.83110</td>
<td>0.83781</td>
<td>0.86967</td>
<td>0.79352</td>
</tr>
</tbody>
</table>

Table 5. Root MSE and Relative Efficiency, BN errors (n=1000)

<table>
<thead>
<tr>
<th>parameter</th>
<th>$\beta_{10}$</th>
<th>$\beta_{20}$</th>
<th>$\gamma_0$</th>
<th>$\theta_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.09218</td>
<td>0.16110</td>
<td>0.04107</td>
<td>0.04226</td>
</tr>
<tr>
<td>SMLE</td>
<td>0.12235</td>
<td>0.16686</td>
<td>0.04191</td>
<td>0.04244</td>
</tr>
<tr>
<td>AMLE</td>
<td>0.14605</td>
<td>0.21506</td>
<td>0.05617</td>
<td>0.09929</td>
</tr>
<tr>
<td>QMLE</td>
<td>0.17460</td>
<td>0.29906</td>
<td>0.04198</td>
<td>0.04360</td>
</tr>
<tr>
<td>MLE/SMLE</td>
<td>0.75344</td>
<td>0.96551</td>
<td>0.98011</td>
<td>0.99560</td>
</tr>
<tr>
<td>MLE/AMLE</td>
<td>0.63118</td>
<td>0.74909</td>
<td>0.73120</td>
<td>0.42560</td>
</tr>
<tr>
<td>MLE/QMLE</td>
<td>0.52796</td>
<td>0.53869</td>
<td>0.97836</td>
<td>0.96922</td>
</tr>
</tbody>
</table>

Table 6. Root MSE and Relative Efficiency, LN errors (n=1000)

<table>
<thead>
<tr>
<th>parameter</th>
<th>$\beta_{10}$</th>
<th>$\beta_{20}$</th>
<th>$\gamma_0$</th>
<th>$\theta_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.03973</td>
<td>0.05677</td>
<td>0.06916</td>
<td>0.14286</td>
</tr>
<tr>
<td>SMLE</td>
<td>0.06360</td>
<td>0.07079</td>
<td>0.11468</td>
<td>0.31346</td>
</tr>
<tr>
<td>AMLE</td>
<td>0.30293</td>
<td>0.10834</td>
<td>0.46830</td>
<td>0.85769</td>
</tr>
<tr>
<td>QMLE</td>
<td>0.05787</td>
<td>0.10174</td>
<td>0.18630</td>
<td>0.27348</td>
</tr>
<tr>
<td>MLE/SMLE</td>
<td>0.62612</td>
<td>0.80525</td>
<td>0.59320</td>
<td>0.45899</td>
</tr>
<tr>
<td>MLE/AMLE</td>
<td>0.13195</td>
<td>0.52705</td>
<td>0.14913</td>
<td>0.16756</td>
</tr>
<tr>
<td>MLE/QMLE</td>
<td>0.68645</td>
<td>0.56243</td>
<td>0.37376</td>
<td>0.52276</td>
</tr>
</tbody>
</table>
5. Conclusion

This paper examines the adaptive condition for a GARCH model. We reparametrize Bollerslev’s GARCH specification, and find that the parameters, which can be identified by semiparametric approaches, are adaptively estimable when the conditional density is assumed to be symmetric. In a model with a free intercept in the regression function and asymmetric conditional density, all the identifiable parameters except the intercept are adaptively estimable. We propose Semiparametric Maximum Likelihood Estimator with trimming techniques and show that it is adaptive for all the adaptively estimable parameters and consistent for all the identifiable parameters.

One of the motivations for proposing SMLE rather than AMLE is that we expect SMLE to have better finite sample performance than AMLE. This is because the latter is an one-step procedure and utilizes the nonparametric OPG matrix to estimate the information matrix.

Monte Carlo experiments show that, when the true density function is symmetric, SML estimators of the identifiable parameters perform the best based on the root MSE standard. For asymmetric cases, SML estimators could be slightly worse than QML estimators for some parameters.
Appendix A

We define the following metrics used in the appendix to prove our theorems. \( L^1(z) = E(|z|) \), \( L^2(z) = [E(z^2)]^{1/2} \). The relationship between them is

\[
L^1(z) \leq L^2(z), \tag{A.1}
\]

where \( z \) is a random variable.

Proof of Theorem 3.3.1:

We begin by recalling the following Lemma.

**Lemma A.1.**

*Let \( g \) is bounded and uniformly continuous, there exist positive constants \( c_1 \) and \( c_2 \) such that*

\[
P\{\sup_y |g(w) - g_{e\lambda_n}(w)| > \varepsilon_n\} \leq c_1 \exp\left(-c_2 n\varepsilon_n^2 \lambda_n^{-2}\right), \tag{A.2}
\]

*where \( \lambda_n = o(\varepsilon_n) \).*

Proof: See Prakasa Rao (1983), Section 4.1.

To prove Theorem 3.1, we use \( L^1 \) metric and let

\[
I = \int \cdots \int \cdots \int _n^{n-1} \sum_{t=1}^n |\ell_{n\lambda_n}(w_t) - \log[g(w_t)]| g(w_1) \cdots g(w_n) f(x_1) \cdots f(x_n) dw_1 \cdots dw_n dx_1 \cdots dx_n, \tag{A.3}
\]

where \( f \) is the density function of \( x \), then

\[
I = \int \int |\ell_{n\lambda_n}(w) - \log[g(w)]| g(w) f(x) dwdx. \tag{A.4}
\]

Denote the conditions in (3.4) by \( A, B, C, D \), and the right hand side of (A.4) by \( I_1 + I_2 \), where
\[ I_1 = \int \int_{ABCD} \left| \log(g_{e\lambda_n}(w)) - \log(g(w)) \right| g(w)f(x) dwdx, \quad (A.5) \]

\[ I_2 = \int \int_{[ABCD]^c} \left| \log(g(w)) \right| g(w)f(x) dwdx. \quad (A.6) \]

Because \( |\log(z)| \leq \left| \frac{1}{z} - 1 \right| + |z - 1| \) for all \( z > 0 \), then \( I_1 \) is bounded by

\[ \int \int_{ABCD} \left\{ \frac{g_{e\lambda_n}(w) - g(w)}{g(w)} + \frac{|g_{e\lambda_n}(w) - g(w)|}{g_{e\lambda_n}(w)} \right\} g(w)f(x) dwdx. \quad (A.7) \]

And \( (A.7) \) is bounded by

\[ \sup_w |g_{e\lambda_n}(w) - g(w)| \left( \frac{1}{d_n} + 2e_n \right). \quad (A.8) \]

Given the conditions \( nd_n^2 \lambda_n^2 \to \infty \) and and \( ne_n^{-2} \lambda_n^3 c_n^{-2} \to +\infty \) as \( n \to +\infty \), then \( I_1 \to o_p(1) \) uniformly in \( \psi \) following Lemma A.1.

\( I_2 \) is bounded by

\[ I_2 \leq \int \int |\log(g(w))|[P\{|g_{\lambda_n}^{(2)}(w)| > b_n g_{\lambda_n}(w)\}]
+ P\{|g_{\lambda_n}(w)| > c_n g_{\lambda_n}(w)\} \]
\[ + P\{g_{\lambda_n}(w) < d_n\} + P\{|w| > e_n\}] g(w)f(x) dwdx. \quad (A.9) \]

Using the elementary estimates noted in Stone (1975). For \( \kappa_i \) universal constants and all \( w \), then

\[ \text{Var} g_{n\lambda_n}^{(i)}(w) \leq \kappa_i \lambda_n^{-(2i+1)} \nu^{-1} g_{\lambda_n}(w), \quad i = 0, 1, \ldots. \quad (A.10) \]

Under Assumption (2), we claim that \(^1\):

\[ g_{\lambda_n}(w) \to g(w) \text{ in probability for all } w \text{ if } n\lambda_n \to +\infty, \]

\(^1\) see Bickel (1982): Section 6, Lemma 6.1 and also see Yang (1997), Appendix A.
\( g'_{\lambda_n}(w) \to g'(w) \) in probability a.e.w if \( n\lambda_n^3 \to +\infty \),

\( g^{(2)}_{\lambda_n}(w) \to g^{(2)}(w) \) in probability a.e.w if \( n\lambda_n^5 \to +\infty \).

It follows Assumption (1) that \( I_2 \to o_p(1) \). The Theorem is proved.

**Appendix B**

Proof of Theorem 3.3.2:

**Lemma B.1**

Let

\[
I_3 = E_0 \left| \frac{\partial S_n(\psi_{n,}^*, \tilde{g}_{n,\lambda})}{\partial \psi^*} - \frac{\partial S_n(\psi_{n,}^*, g_0)}{\partial \psi^*} \right|,
\]

where \( E_0 \) means the expectation is calculated under the true DGP with the density \( g_0 \), then

\( I_3 \to o_p(1) \).

Proof:

\[
I_3 = E_0 \left| \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{\partial^2 \log[\tilde{g}_{n,\lambda}(w_t(\psi^*))]}{\partial \psi \partial \psi^*} - \frac{\partial^2 \log[g(w_t(\psi_0))]}{\partial \psi \partial \psi^*} \right\} + o_p(1) \right|
\]

\[
= E_0 \left| \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{\partial^2 \log[\tilde{g}_{n,\lambda}(w_t(\psi_0))]}{\partial \psi \partial \psi^*} - \frac{\partial^2 \log[g(w_t(\psi_0))]}{\partial \psi \partial \psi^*} \right\} + o_p(1) \right|
\]

\[
= E_0 \left| \frac{1}{n} \sum_{t=1}^{n} \left\{ \left[ \frac{\tilde{g}'_{n,\lambda}}{\tilde{g}_{n,\lambda}} - \frac{g'}{g} \right] (w_t(\psi_0)) \frac{\partial^2 w(\psi_0)}{\partial \psi \partial \psi^*} + \frac{\tilde{g}^{(2)}_{n,\lambda} \tilde{g}_{n,\lambda} - \left[ \tilde{g}'_{n,\lambda} \right]^2}{\tilde{g}_{n,\lambda}^2} - \frac{g^{(2)}g - [g']^2}{g^2} \right\} (w_t(\psi_0)) \frac{\partial w(\psi_0)}{\partial \psi} \frac{\partial w(\psi_0)}{\partial \psi^*} \right| + o_p(1).
\]

Following Appendix B in Yang (1997), we have that

\[
E_0 \left\{ \frac{g'_{n,\lambda}}{g_{n,\lambda}} - \frac{g'}{g} \right\}^2 (w_t(\psi_0)) \to o_p(1),
\]

and

\[
E_0 \left\{ \frac{\tilde{g}^{(2)}_{n,\lambda} g_{n,\lambda} - \left[ \tilde{g}'_{n,\lambda} \right]^2}{\tilde{g}_{n,\lambda}^2} - \frac{g^{(2)}g - [g']^2}{g^2} \right\} (w_t(\psi_0)) \right| \to o_p(1).
\]
These equations will hold with $\tilde{g}_{n\lambda_n}$ instead of $g_{n\lambda_n}$ by contiguity.

Since $\frac{1}{n} \sum_{t=1}^{n} E_0 \left| \frac{\partial^2 w_t(\psi_0)}{\partial \psi \partial \psi^t} \right|$ and $\frac{1}{n} \sum_{t=1}^{n} E_0 \left| \frac{\partial w_t(\psi_0)}{\partial \psi} \frac{\partial w_t(\psi_0)}{\partial \psi^t} \right|$ are bounded from above under the moments assumptions, the result follows.

**Lemma B.2**

$\sqrt{n} S^n(\psi_0, g_{n\lambda_n}) - \sqrt{n} S^n(\psi_0, g_0) \rightarrow o_p(1)$.

**Proof:** Let

$$I_4 = E_0 \left| \sqrt{n} [S^n(\psi_0, g_{n\lambda_n}) - S^n(\psi_0, g_0)] \right|,$$

then $I_4 = (I_5, I_6)^T$, where

$$I_5 = E_0 \left| \sqrt{n} [S^n(\beta_0, g_{n\lambda_n}) - S^n(\beta_0, g_0)] \right|,$$

$$I_6 = E_0 \left| \sqrt{n} [S^n(\delta_0, g_{n\lambda_n}) - S^n(\delta_0, g_0)] \right|.$$

To bound $I_5$, we consider $I_7 = E_0 \left\| \sqrt{n} [S^n(\beta_0, g_{n\lambda_n}) - S^n(\beta_0, g_0)] \right\|^2$.

$$I_7 \leq 2E_0 \left\| n^{-1/2} \sum_{t=1}^{n} \left[ \frac{g'_{n\lambda_n}(w_t(\psi_0))}{g_{n\lambda_n}(w_t(\psi_0))} - \frac{g'_0(\varepsilon_t(\phi_0))}{g_0(\varepsilon_t(\phi_0))} \right] \frac{x_t}{\nu_t(\phi_0)} \right\|^2 + 2E_0 \left\| n^{-1/2} \sum_{t=1}^{n} \left[ \frac{\varepsilon_t(\phi_0) g'_{n\lambda_n}(w_t(\psi_0))}{g_{n\lambda_n}(w_t(\psi_0))} - \frac{\varepsilon_t(\phi_0) g'_0(\varepsilon_t(\phi_0))}{g_0(\varepsilon_t(\phi_0))} \right] D_t(\psi_0) \right\|^2.$$

Simply observing that $\frac{g'_{n\lambda_n}(w_t(\psi_0))}{g_{n\lambda_n}(w_t(\psi_0))}$, $\frac{g'_0(\varepsilon_t(\phi_0))}{g_0(\varepsilon_t(\phi_0))}$ and $D_t(\psi_0)$ are all antisymmetric, however, $\varepsilon_t(\phi_0) \frac{g'_{n\lambda_n}(w_t(\psi_0))}{g_{n\lambda_n}(w_t(\psi_0))}$ and $\varepsilon_t(\phi_0) \frac{g'_0(\varepsilon_t(\phi_0))}{g_0(\varepsilon_t(\phi_0))}$ are symmetric when $g_0$ is assumed to be symmetric, the cross products drop out in the last calculation.

By a minor modification of Linton's (1993) arguments, we have that

$$E_0 \left\| \sqrt{n} [S^n(\beta_0, g_{n\lambda_n}) - S^n(\beta_0, g_0)] \right\| \rightarrow o_p(1).$$

Notice that the conclusion for the estimated scores for $\beta$ does not require any sample splitting. However, as Linton (1993) argues, when we examine the estimated scores for $\delta$, we are unable to exploit symmetry properties and the argument becomes considerably
more involved. Following Linton (1993), we adopt sample splitting to provide a simple proof. We split the sample into two subsamples.

\[ J_1 = \{ t : t = 1, 2, \ldots, n_1 \}; \quad J_2 = \{ t : t = n_1 + 1, \ldots, n \}, \]

where \( n_1(n) \to \infty; \quad n_1/n \to 0 \) as \( n \to \infty \).

The first subsample is used to estimate the pseudo density function, while the remaining observations are used to construct the estimator.

\[ I_6 = E_0 \left| n^{-1/2} \sum_{t \in J_2} \frac{\varepsilon_t(\phi_0) g'_{n \lambda_n}(w_t(\psi_0))}{g_{n \lambda_n}(w_t(\psi_0))} - \frac{\varepsilon_t(\phi_0) g'_0(\varepsilon_t(\phi_0))}{g_0(\varepsilon_t(\phi_0))} \right| \left( B_t(\psi_0)^\top, C_t(\psi_0)^\top \right)^\top. \]

Let \( F_t(\psi_0) = (B_t(\psi_0)^\top, C_t(\psi_0)^\top)^\top \) and \( \bar{F}(\psi_0) = (B(\psi_0)^\top, C(\psi_0)^\top)^\top \), then we have

\[ I_6 \leq E_0 \left| n^{-1/2} \sum_{t \in J_2} \left[ \frac{\varepsilon_t(\phi_0) g'_{n \lambda_n}(w_t(\psi_0))}{g_{n \lambda_n}(w_t(\psi_0))} - \frac{\varepsilon_t(\phi_0) g'_0(\varepsilon_t(\phi_0))}{g_0(\varepsilon_t(\phi_0))} \right] \left[ F_t(\psi_0) - \bar{F}(\psi_0) \right] \right| \]

\[ + E_0 \left| n^{-1/2} \sum_{t \in J_2} \left[ \frac{\varepsilon_t(\phi_0) g'_{n \lambda_n}(w_t(\psi_0))}{g_{n \lambda_n}(w_t(\psi_0))} - \frac{\varepsilon_t(\phi_0) g'_0(\varepsilon_t(\phi_0))}{g_0(\varepsilon_t(\phi_0))} \right] \bar{F}(\psi_0) \right| = I_8 + I_9. \]

To bound \( I_8 \), we consider

\[ I_{10} = E_0 \left\| n^{-1/2} \sum_{t \in J_2} \left[ \frac{\varepsilon_t(\phi_0) g'_{n \lambda_n}(w_t(\psi_0))}{g_{n \lambda_n}(w_t(\psi_0))} - \frac{\varepsilon_t(\phi_0) g'_0(\varepsilon_t(\phi_0))}{g_0(\varepsilon_t(\phi_0))} \right] \left[ F_t(\psi_0) - \bar{F}(\psi_0) \right] \right\|^2. \]

Since \( F_t(\psi_0) - \bar{F}(\psi_0) \) is zero mean and independent of \( \frac{\varepsilon_t(\phi_0) g'_{n \lambda_n}(w_t(\psi_0))}{g_{n \lambda_n}(w_t(\psi_0))} - \frac{\varepsilon_t(\phi_0) g'_0(\varepsilon_t(\phi_0))}{g_0(\varepsilon_t(\phi_0))} \), then the cross products drop out. But \( F_t(\psi_0) \) is bounded, and hence \( I_{10} \to o_p(1) \).

\( I_9 \) is bounded by the following argument. We apply directly the results of Kreiss (1987). We have

\[ \max_{1 \leq t \leq n} E_0 \int \left\{ \frac{g'_{n \lambda_n}(w_t(\psi_0))}{g_{n \lambda_n}(w_t(\psi_0))} - \frac{g'_0(\varepsilon_t(\phi_0))}{g_0(\varepsilon_t(\phi_0))} \right\}^2 g_0(\varepsilon) d\varepsilon \to 0. \]
Since \( \varepsilon_t, t = 1, \ldots, n \), are i.i.d, then
\[
n^{-1/2} \sum_{t \in J_2} \varepsilon_t(\phi_0) \rightarrow N(0, 1).
\]

Given any small number \( \nu_1 \rightarrow 0^+ \) and \( \nu_2 \rightarrow 0^+ \), then there exist \( N_1 \) and \( M \), such that
\[
P\{n^{-1/2} \sum_{t \in J_2} \varepsilon_t(\phi_0) \geq M\} \leq \nu_2/2, \text{ when } n \geq N_1. \] There exists \( N_2 \), such that
\[
P\left\{ \max_{1 \leq t \leq n} \left| \frac{g_n^\prime(\lambda_n)(w_t(\psi_0))}{g_n(\lambda_n)(w_t(\psi_0))} - \frac{g_0(\varepsilon_t(\phi_0))}{g_0(\varepsilon_t(\phi_0))} \right| \geq \frac{\nu_1}{M} \right\} \leq \nu_2/2
\]
when \( n \geq N_2 \). For \( n \geq \max\{N_1, N_2\} \),
\[
P\left\{ \left| n^{-1/2} \sum_{t \in J_2} \left[ \frac{\varepsilon_t(\phi_0)g_n^\prime(\lambda_n)(w_t(\psi_0))}{g_n(\lambda_n)(w_t(\psi_0))} - \frac{\varepsilon_t(\phi_0)g_0(\varepsilon_t(\phi_0))}{g_0(\varepsilon_t(\phi_0))} \right] \right| \geq \nu_1 \right\} \leq \nu_2,
\]
then \( I_0 \rightarrow o_p(1) \).

Lemma B.3:
\[\sqrt{n}(\phi_n - \phi_0) \xrightarrow{d} N(0, I_{\phi\phi}^{-1}) \text{ where } I_{\phi\phi} \text{ is the information matrix of } \phi \text{ under } g_0.\]

Proof: See Davidson and MacKinnon (1993), Chapter 8.

Given these results, we conclude that
\[\sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{d} N(0, I_{\psi\psi}^{-1}(\psi, \psi)).\]

**Appendix C**

Proof of Theorem 3.3:

According to equation (2.19) and (2.11), we have
\[
D_t(\psi) = \frac{1}{\sigma_t^2(\psi)}(-2\gamma \sum_{i=1}^{\infty} \theta^{i-1} x_{t-i} u_{t-i}(\psi)), \tag{C.1}
\]
\[
S^n(\beta_1, g) = -\frac{1}{2n} \sum_{t=1}^{n} \frac{A_t(\phi)}{\sigma_t^2(\psi)}(-2\gamma \sum_{i=1}^{\infty} \theta^{i-1} u_{t-i}(\psi)) - \frac{1}{n} \sum_{t=1}^{n} \frac{1}{\sigma_t^2(\psi)} \frac{g'(\varepsilon_t(\phi))}{g(\varepsilon_t(\phi))}, \tag{C.2}
\]

34
\[ S^n(\beta_2, g) = -\frac{1}{2n} \sum_{t=1}^{n} \frac{A_t(\phi)}{\sigma_t^2(\psi)} (-2\gamma \sum_{i=1}^{\infty} \theta^{i-1} x_{2,t-i} u_{t-i}(\psi)) - \frac{1}{n} \sum_{t=1}^{n} \frac{1}{v_t^2(\phi)} \frac{x_{2t}}{g(\varepsilon_t(\phi))}. \]  

(C.3)

Then

\[ I_{11} = \sqrt{n} S^n(\beta_{20}, \tilde{g}_{\lambda n}) - \sqrt{n} S^n(\beta_{20}, g_0) = \]

\[-\frac{1}{2n} \sum_{t=1}^{n} \frac{\tilde{A}_t(\phi)}{\sigma_t^2(\psi)} (-2\gamma \sum_{i=1}^{\infty} \theta^{i-1} [x_{2,t-i} - E(x_{2,t-i})] u_{t-i}(\psi)) - \frac{1}{n} \sum_{t=1}^{n} \frac{1}{v_t^2(\phi)} \frac{x_{2t} - E(x_{2t})}{\tilde{g}_{\lambda n}(\varepsilon_t(\phi))} \]

\[+ \frac{1}{2n} \sum_{t=1}^{n} \frac{A_t(\phi)}{\sigma_t^2(\psi)} (-2\gamma \sum_{i=1}^{\infty} \theta^{i-1} [x_{2,t-i} - E(x_{2,t-i})] u_{t-i}(\psi)) + \frac{1}{n} \sum_{t=1}^{n} \frac{1}{v_t^2(\phi)} \frac{x_{2t} - E(x_{2t})}{g_0(\varepsilon_t(\phi))}, \]  

(C.4)

where \( \tilde{A}_t(\phi) \) is calculated in the same way as \( A_t(\phi) \) in equation (3.16) with \( g_0(\cdot) \) replaced by \( \tilde{g}_{\lambda n} \).

Notice that, in this case, we do not assume the symmetry condition of \( g_0 \), however, we find that \( x_{2,t-i} - E(x_{2,t-i}), i = 1, 2, \cdots \), have zero mean and are independent of \( \left\{ \frac{\tilde{g}_{\lambda n}'}{\tilde{g}_{\lambda n}} - \frac{g_0'}{g_0} \right\}(\varepsilon_t) \). Let \( I_{12} = E_0(I_{11}^2) \), then the cross products drop out.

Using the same calculation in Bickel (1982), we have

\[ I_{12} \to o_p(1), \]

and Theorem 3.3 simply follows.

\[ \text{Acknowledgements.} \] This paper is partly drawn from my thesis at Queen’s University. I would like to thank my supervisors, Russell Davidson and James MacKinnon, for their patience and guidance. Thanks also go to Gregor Smith and the seminar participants at University of Waterloo for helpful comments. All remaining errors are mine. I gratefully acknowledge financial support from Queen’s University and the Social Sciences and Humanities Research Council of Canada.
Literature


