A General Approach to Conditional Moment Specification Testing with Projections

Xuexin WANG*

November 23, 2010

Abstract

This paper develops a general approach for model specification analysis within the conditional moment specification testing framework. We focus on the class of tests proposed by Newey (1985a) and Tauchen (1985), who consider a finite number of unconditional moment restrictions, given a conditional moment restriction. The main feature of our methodology is to exploit the nature of the conditional moment specification testing to remove the non-negligible estimation effect via a martingale transformation-like projection, using additional unconditional moment restrictions. This general methodology includes Wooldridge (1990)’s modified statistic as a special case. This approach is robust to the departures from distributional assumptions that are not being tested, moreover only a $\sqrt{T}$-consistent estimator is needed, and the transformation is distribution free. At the same time, the transformed statistics could reach asymptotically efficiency in the sense of GMM under the mild condition, when the additional unconditional moment restrictions are properly chosen. The robustness and efficiency are obtained without paying too much price. In the light of our general framework,
Wooldridge (1990)’s statistic appears too restrictive in the sense that the additional unconditional moments used to remove the parameters estimation effect are predetermined (they are just the score functions of the conditional moment restriction). Furthermore, these score functions are not necessarily the optimal instrument in the sense of GMM, which would lead to a potential loss of efficiency. Our general framework provides alternative ways to overcome the shortcomings of Wooldridge (1990)’s modified statistic. As examples, we apply our methodology to test the adequacy of the estimated GARCH model. The simulation results show that our new statistic has very good size properties and nontrivial power, comparing with Lagrange multiplier (LM) type tests and Wooldridge (1990)’s modified statistic.

**JEL Classification:** C12 C22

**Keywords:** Conditional Moment (CM) Test; Estimation Effect Removal; GARCH Model Specification Testing
1 Introduction

Statistically speaking, an econometric model is a set of restrictions on the joint distribution of the dependent and independent variables. Without the full knowledge of the joint distribution, econometric models are usually defined via conditional moment restrictions. In the context of maximum likelihood models, conditional moment restrictions appear in the score functions when exogenous variables are present. So it becomes very crucial to check the validity of the conditional moment restrictions. The conditional moment restriction implied by a parametric model, theoretically, is equivalent to some infinite set of unconditional moment restrictions. One line of research exploits this point to gain the “omnibus” testing property, for example Bierens (1982), de Jong and Bierens (1994). On the other hand, the conditional moment tests proposed by Newey (1985a) and Tauchen (1985) aim at testing the “directional” validity of conditional moment restrictions via a finite number of unconditional moment conditions implied by the conditional moment restriction. This idea includes many specification testing procedures as special cases, examples are Hausman’s statistic, White’s statistic, portmanteau test, and so on.

This paper attempts to develop a general approach to the conditional moment tests with unknown parameters. To illustrate the idea, it is convenient to introduce mathematical notations and some examples. Let \( \{(Y_t, X_t) : t = 1, 2, \cdots \} \) be a sequence of observations where \( Y_t \) is a scalar and \( X_t \) is a \( 1 \times K \) vector. Let \( I_{t-1} = \{X_t, X_{t-1}, \cdots ; Y_{t-1}, Y_{t-2}, \cdots \} \) represents the information set at time \( t \). The interest lies in explaining some features of the conditional distribution of \( Y_t \) in terms of the information set \( I_{t-1} \). In dynamic model analysis, one often considers the conditional expectation:

\[
E(Y_t | I_{t-1}) = f_t(I_{t-1}), \ t = 1, 2, \cdots .
\]

This framework is quite general, including both nonparametric and parametric models.
One typically considers parametric models such that

\[ E(Y_t|I_{t-1}) = f_t(I_{t-1}, \theta_0), \quad t = 1, 2, \ldots \]

where \( f_t(I_{t-1}, \theta_0) \) is the parametric specification for \( f_t(I_{t-1}) \), \( \theta_0 \subset \Theta \subset \mathbb{R}^P \), where \( \Theta \) is the parameter space.

Suppose that the interest lies in testing parametric model hypotheses about the conditional expectation of \( Y_t \) given the information set \( I_{t-1} \). By the definition of the model, the null hypothesis is

\[ H_0 : E(Y_t|I_{t-1}) = f_t(I_{t-1}, \theta_0), \quad \text{for some } \theta_0 \in \Theta, \quad t = 1, 2, \ldots \]

which is equivalent to test the conditional moment restriction:

\[ H_0 : E(e_t(Y_t, I_{t-1}, \theta_0)|I_{t-1}) = 0, \quad \text{for some } \theta_0 \in \Theta, \quad t = 1, 2, \ldots \]

where \( e_t(Y_t, I_{t-1}, \theta) = Y_t - f_t(I_{t-1}, \theta) \).

In the framework of conditional maximum likelihood estimation, suppose that the parameter conditional likelihood function is \( L(Y_t|I_{t-1}, \theta) \). Under some regularity conditions, by differentiating the identity \( \int L(Y_t|I_{t-1}, \theta) dy = 1 \), at \( \theta_0 \), we can get

\[ E[\partial l(Y_t|I_{t-1}, \theta_0)\partial \theta|I_{t-1}] = 0, \quad t = 1, 2, \ldots, \quad (1) \]

where \( l(Y_t|I_{t-1}, \theta_0) = \ln[L(Y_t|I_{t-1}, \theta_0)] \). Then the null hypothesis of correct specification is the same as (1).

In this paper, we consider general null hypothesis with the form

\[ H_0 : E(\phi_t(Y_t, I_{t-1}, \theta_0)|I_{t-1}) = 0, \quad \text{for some } \theta_0 \in \Theta, \quad t = 1, 2, \ldots. \quad (2) \]
for some (smooth) function $\phi_t$, which is allowed to change with $t$. The alternative is

$$H_1 : P(E(\phi_t(Y_t, I_{t-1}, \theta_0)|I_{t-1}) = 0) < 1, \text{ for } \theta_0 \in \Theta, \ t = 1, 2, \cdots. \quad (3)$$

For simplicity, we assume that $\phi_t(Y_t, I_{t-1}, \theta)$ is a scalar random function.$^1$

It is very difficult to test (2) directly, if not impossible. Most specification tests are based on the fact that any function of $I_{t-1}$ should be uncorrelated with $\phi_t(Y_t, I_{t-1}, \theta_0)$. More specifically, the null becomes

$$H'_{0} : E(\Lambda_t(I_{t-1})\phi_t(Y_t, I_{t-1}, \theta_0)) = 0, \text{ for some } \theta_0 \in \Theta, \ t = 1, 2, \cdots,$$

where $\Lambda_t(\cdot)$ is an $S \times 1$ vector $\sigma(I_{t-1})$ measurable function, in which $\sigma(I_{t-1})$ denotes the sigma-field generated by information set $I_{t-1}$. When $\Lambda_t(I_{t-1})$ is properly chosen, and $S$ goes to infinity, under some circumstances, the countable number of unconditional moment restrictions is equivalent to the conditional moment restriction, see Donald et al. (2003). The tests for $H'_{0}$, exploiting this point, can achieve consistency against $H_1$. In most cases, $S$ is a finite number, and $\Lambda_t(I_{t-1})$ has the form $\Lambda_t(I_{t-1}, \theta_0)$, which means that it is a function of $\theta_0$ in addition to the information set, the null becomes

$$H'_{0} : E(\Lambda_t(I_{t-1}, \theta_0)\phi_t(Y_t, I_{t-1}, \theta_0)) = 0, \text{ for some } \theta_0 \in \Theta, \ t = 1, 2, \cdots. \quad (4)$$

Newey (1985a) refers to this kind of moment tests based on conditional moment restrictions as conditional moment (CM) tests, which includes many testing procedures as special cases. One classical example is the diagnostic test for an ARMA($p_0, q_0$) model, which is proposed by Box and Pierce (1970) and Ljung and Box (1978) (BPL):

$$BPL(S) = T(T + 2)\sum_{j=1}^{S} (T - j)^{-1} \hat{\rho}^2(j),$$

$^1$In Wooldridge (1990), so called $L \times 1$ "generalized residual vector" $\phi_t$ is considered. Although the setting is more general in some degree, it complicates the presentation of the idea.
where \( \hat{\rho}(j) \) is the sample autocorrelation function of \( \{\hat{e}_t\}_{t=1}^T \), \( \hat{e}_t = e(Y_t, I_t, \hat{\theta}_T) = Y_t - f(I_{t-1}, \hat{\theta}_T) \), and \( f(I_{t-1}, \hat{\theta}_T) \) is an estimated ARMA\((p_0, q_0)\) model. The BPL statistic could be regarded as the quadratic form of the conditional moment test, choosing

\[
\Lambda_t(I_{t-1}, \theta) = (e(Y_{t-1}, I_{t-2}, \theta), \ldots, e(Y_{t-S}, I_{t-S-1}, \theta))' / E[e(Y_t, I_{t-1}, \theta)^2]
\]

and \( \phi_t(Y_t, I_{t-1}, \theta) = e(Y_t, I_{t-1}, \theta) \).

For the CM test (4), when \( \theta_0 \) is known, the test statistic is based on the \( S \times 1 \) vector

\[
\xi_T = T^{-1/2} \sum_{t=1}^T \Lambda_t \phi_t,
\]

where \( \xi_T = \xi_T(I_{t-1}, \theta_0), \Lambda_t = \Lambda_t(I_{t-1}, \theta_0), \phi_t = \phi_t(Y_t, I_{t-1}, \theta_0) \) for \( t = 1, \ldots, T \). In this case, some central limit theory usually could be applied directly, which makes it quite straightforward to obtain the quadratic form of \( \xi_T \) with an asymptotic chi-square distribution.

But the reality is that \( \theta_0 \) is unknown, and has to be estimated firstly. Assume that there is an estimator \( \hat{\theta}_T \) such that \( \sqrt{T}(\hat{\theta}_T - \theta_0) = O_p(1) \), then a computable test statistic of the CM test is the \( S \times 1 \) vector

\[
\hat{\xi}_T = T^{-1/2} \sum_{t=1}^T \hat{\Lambda}_t \hat{\phi}_t,
\]

where \( \hat{\xi}_T = \hat{\xi}_T(I_{t-1}, \hat{\theta}_T), \hat{\Lambda}_t = \Lambda_t(I_{t-1}, \hat{\theta}_T), \hat{\phi}_t = \phi_t(Y_t, I_{t-1}, \hat{\theta}_T) \) for \( t = 1, \ldots, T \). Now it becomes more difficult to obtain the quadratic form of \( \hat{\xi}_T \) with an asymptotic chi-square distribution, since there exists an “estimation effect”, when model parameters
have to be estimated. To see this point, denote

\[ \Xi_T = T^{-1} \sum_{t=1}^{T} E [\Lambda_t(I_{t-1}, \theta_0) \nabla_{\theta} \phi_t(Y_t, I_{t-1}, \theta_0)] \]

\[ = T^{-1} \sum_{t=1}^{T} E [\Lambda_t(I_{t-1}, \theta_0) E(\nabla_{\theta} \phi_t(Y_t, I_{t-1}, \theta_0)|I_{t-1})] \]

and \( \Xi = \lim_{T \to \infty} \Xi_T. \) Under some regularity conditions, it could be shown that (the result will be formally presented in Theorem 1 in next section)

\[ \hat{\xi}_T = \xi_T + \Xi \sqrt{T} (\hat{\theta}_T - \theta_0) + o_p(1). \]

(7)

There are rare cases when \( \Xi = 0 \) holds. When \( \Xi \neq 0 \), the existence of the term \( \Xi \sqrt{T}(\hat{\theta}_T - \theta_0) \), which is called “estimation effect”, makes the asymptotic inference complicated. In order to derive the covariance matrix correctly, the asymptotic joint distribution of \( \sqrt{T}(\hat{\theta}_T - \theta_0) \) and \( \xi_T \) has to be considered, which depends on the model and DGP characteristics, the method of estimating \( \hat{\theta}_T \) and even the unknown parameter \( \theta_0 \).

There are several ways to obtain the asymptotic variance of \( \hat{\xi}_T \), depending on the different estimation methods applied. One classical way of achieving this is based on the standard LM (Lagrange Multiplier) approach. Even though it is quite easy to compute by artificial regression in many simple setups, this approach lacks robustness, as is pointed out by Wooldridge (1990) in nonlinear least square (NLS) framework: it relies on certain auxiliary assumptions holding in addition to the relevant null hypothesis, or has poor finite-sample properties in the maximum likelihood framework: it tends to reject the null hypothesis too often when it is true. In the framework of (conditional) moment testing framework, Newey (1985a) and Tauchen (1985) consider the statistic based on maximum-likelihood estimator. Newey (1985b) considers the statistic based

---

\(^2\)Following convention, if function \( a(\theta) \) is a \( S \times 1 \) function with \( \theta \in \mathbb{R}^P \), \( \nabla_{\theta} a(\theta) \) denotes the \( S \times P \) Jacobian matrix.
on GMM estimator. White (1994) incorporates both MLE and GMM into a general framework.

Instead of computing the consistent covariance matrix, Kuan and Lee (2006) propose robust moment tests using some random normalizing matrix. To circumvent the estimation effect, it has to be “created” in the normalizing matrix ironically. The statistic is asymptotically pivotal, but it does not follow a chi-square distribution, and Monte Carlo study shows that this approach typically has no much power in time series scenarios.

Contrary to Kuan and Lee (2006)’s method of creating estimation effect, another strategy is to remove the estimation effect by transformation on the test statistic. One methodology to achieve this is proposed by Wooldridge (1990). Given a CM test statistic as (6), assume \( \theta_0 \in \text{Int}(\Theta) \), where \( \text{Int}(\Theta) \) denotes interior of \( \Theta \), and \( \phi_t(Y_t, I_{t-1}, \theta) \) is differentiable on \( \theta \in \text{Int}(\Theta) \). Define

\[
\Phi_t = \Phi_t(I_{t-1}, \theta_0) = E[\nabla_\theta \phi_t(Y_t, I_{t-1}, \theta_0)|I_{t-1}],
\]

the modified statistic of Wooldridge (1990) is

\[
\bar{\xi}_T = T^{-1/2} \sum_{t=1}^{T} (\hat{\Lambda}_t - (\hat{\Phi}_t \hat{B}_T)' \hat{\Phi}_t),
\]

where \( \hat{\Phi}_t = \Phi_t(I_{t-1}, \hat{\theta}_T) \) is a consistent estimator of \( \Phi_t(I_{t-1}, \theta_0) \), and

\[
\hat{B}_T = \left( \sum_{t=1}^{T} \hat{\Phi}_t' \hat{\Phi}_t \right)^{-1} \sum_{t=1}^{T} \hat{\Phi}_t' \hat{\Lambda}_t,
\]

which is the \( P \times S \) matrix of regression coefficients from the matrix regression of \( \hat{\Lambda}_t \) on \( \hat{\Phi}_t \).
It has been shown in Wooldridge (1990) that, under some regularity conditions,

$$\bar{\xi}_T = T^{-1/2} \sum_{t=1}^{T} (\Lambda_t - (\Phi_t B_T)' \phi_t + o_p(1),$$

where $B_T = \left( \sum_{t=1}^{T} E[\Phi_t' \Phi_t] \right)^{-1} \sum_{t=1}^{T} E[\Phi_t' \Lambda_t]$. The intuition is that $\Lambda_t - (\Phi_t B_T)'$ is orthogonal to $\Phi_t$. The quadratic form of the modified statistic $\bar{\xi}_T$ follows a $\chi^2$ distribution asymptotically. One advantage of Wooldridge’s approach is that the test statistic could also be conveniently computed via linear regressions, but the new statistic may lose consistency because of the transformation.

As far as the method of removing estimation effect is concerned, among the parametric empirical process literature, the martingale transformation is introduced by Khmaladze (1981) to tackle it. Stute, Thies, and Zhu (1998) and Bai (2003) focus on the martingale transformation of parametric empirical process to gain the pivotal distribution of the testing statistic. On the other hand, Delgado and Velasco (2009) introduce the idea of “recursive residuals”, which is proposed by Brown, Durbin and Evans (1975), into the test of residual autocorrelation in time series models to remove the estimation effect, the transformation being quite similar to the martingale transformation in discrete case.

This paper follows the line of removing the estimation effect by transformation. The transformation is in the spirit of the martingale transformation in parametric empirical process or “recursive residuals”. The idea is that testing the null (4) is only a compromise of testing the null (2)–as it just tests the “directional” validity of conditional moment restrictions (2), while leaving other possible misspecification directions untouched. Under the conditional moment restriction (2), not only the “directional” unconditional moment restrictions (4) hold, but any other unconditional moment restrictions. We use some of these additional unconditional moment restrictions to conduct the projection of our testing statistic. It turns out that our general framework includes
Woodridge (1990)’s modified statistic as a special case. This approach is robust to the departures from distributional assumptions that are not being tested, moreover only a $\sqrt{T}$-consistent estimator is needed, and the transformation is distribution free. At the same time, the transformed statistics could reach asymptotically efficiency in the sense of GMM under the mild conditions, when the additional unconditional moment restrictions are properly chosen. The robustness and efficiency are obtained without paying too much price. In the light of our new framework, Wooldridge (1990)’s statistic is too restrictive in the sense that the additional unconditional moments used to remove the parameters estimation effect are predetermined (they are just the score functions of the conditional moment restriction). Furthermore, these score functions are not necessarily the optimal instrument in the sense of GMM, which would bring about a potential loss of efficiency. Our general framework provides alternative ways to overcome the shortcomings of Wooldridge (1990)’s modified statistic. Monte Carlo simulation study shows that when the transformation vectors are properly chosen, our new statistic has very good empirical size, and nontrivial power, comparing to the LM type statistic and Wooldridge’s modified statistic.

The outline of the paper is following. In Section 2, we define the new general testing framework and study its properties. Section 3 discusses its relation with other tests. Section 4 discusses the optimal test. Section 5 applies the new methodology to specification testing of GARCH models. Section 6 concludes.

2 Test Statistics

Similar to Wooldridge (1990), we focus on the conditional moment restriction testing framework (2) and the CM test null hypothesis (4). Then the statistic is (5) when $\theta_0$ is known. We denote

$$\xi_T[0] = \xi_T, \Lambda_t[0] = \Lambda_t, \Xi[0] = \Xi,$$
to simplify the notation. In the following, $\xi_T$, $\Lambda_t$, $\Xi$, and $\xi_T[0]$, $\Lambda_t[0]$, $\Xi[0]$ are used interchangeably.

When $\theta_0$ is unknown, assume that we have an estimator $\hat{\theta}_T$ such that $\sqrt{T}(\hat{\theta}_T - \theta_0) = O_p(1)$, the statistic is the same as (6). Our aim is to transform $\hat{\xi}_T$ so as to remove the estimation effect. Note that even though we are testing (4), the conditional moment restriction (2) holds firstly. Under (2) there exist other unconditional moment restrictions in addition to the “directional” unconditional moment restrictions (4). It is possible to find a vector of unconditional moment restrictions with $L \times 1$ dimension such that

$$E(\Lambda_t[1](I_{t-1}, \theta_0)\phi_t(Y_t, I_{t-1}, \theta_0)) = 0, \text{ for some } \theta_0 \in \Theta, \ t = 1, 2, \cdots,$$

where $\Lambda_t[1](I_{t-1}, \theta_0)$ is $L \times 1$ vector function of $I_{t-1}$ and $\theta_0$, for $t = 1, 2, \cdots$. Accordingly, we could form the statistics:

$$\xi_T[1] = T^{-1/2} \sum_{t=1}^T \Lambda_t[1] \phi_t,$$

where $\xi_T[1] = \xi_T[1](I_{t-1}, \theta_0)$, $\Lambda_t[1] = \Lambda_t[1](I_{t-1}, \theta_0)$, for $t = 1, \cdots, T$, when $\theta_0$ is known. The computable statistic is

$$\hat{\xi}_T[1] = T^{-1/2} \sum_{t=1}^T \hat{\Lambda}_t[1] \hat{\phi}_t,$$

where $\hat{\xi}_T[1] = \xi_T[1](I_{t-1}, \hat{\theta})$, $\hat{\Lambda}_t[1] = \Lambda_t[1](I_{t-1}, \hat{\theta}_T)$, for $t = 1, \cdots, T$, when we just have a $\sqrt{T}$-consistent estimator $\hat{\theta}_T$. One simple example could be the autocorrelation testing of Box and Pierce (1970), in which only the null $(\rho(1), \cdots, \rho(S))' = 0$ is considered, but under the conditional moment restriction, unconditional moment restrictions $(\rho(S + 1), \cdots, \rho(S + L))' = 0$ also hold, where $\rho(j)$ is the autocorrelation function of the error term of some ARMA model.
Note that similar to $\hat{\xi}_T$, $\hat{\xi}_T[1]$ is also affected by the estimation effect, i.e

$$\hat{\xi}_T[1] = \xi_T[1] + \Xi[1]\sqrt{T}(\hat{\theta}_T - \theta_0) + o_p(1).$$

(8)

where $\Xi[1] = \lim_{T \to \infty} \Xi_T[1]$,

$$\Xi_T[1] = T^{-1} \sum_{t=1}^{T} E[\Lambda_t[1](I_{t-1}, \theta_0)\nabla \phi_t(Y_t, I_{t-1}, \theta_0)] = T^{-1} \sum_{t=1}^{T} E[\Lambda_t[1](I_{t-1}, \theta_0)\Phi_t(Y_t, I_{t-1}, \theta_0)]$$

The extra unconditional moment restrictions or the extra statistics $\xi_T[1]$ form our bricks of the removal of the estimation effect of the statistic $\hat{\xi}_T$.

The transformed statistic is then

$$\tilde{\xi}_T = \hat{\xi}_T - \hat{\Xi}_T[0]\left(\hat{\Xi}_T[1]'\hat{\Xi}_T[1]\right)^{-1}\hat{\Xi}_T[1]'\hat{\xi}_T[1],$$

(9)

where $\hat{\Xi}_T[m]$ is a consistent estimator of $\Xi[m]$, for $m = 0, 1$. It will be shown that

$$\tilde{\xi}_T = \xi_T - \Xi[0]\left(\Xi[1]'\Xi[1]\right)^{-1}\Xi[1]'\xi_T[1] + o_p(1),$$

(10)

i.e. the estimation effect is canceled out asymptotically.

On the face of it, the transformation is quite similar to the discrete case of the martingale transformation of Khmaladze (1981), or the idea of “recursive residuals” used by Delgado and Velasco (2009), but there exist some essential differences. One difference is that we execute the transformation on a vector function directly. The martingale transformation of Khmaladze (1981) is just conducted on the parametric empirical process, which is a scalar function. Delgado and Velasco (2009) do consider a vector of transformed autocorrelations, but they first transform the autocorrelations individually and use a different transformation to each. Another difference is that we do not exploit the i.i.d property of some stochastic processes. In previous works, i.i.d
condition is the critical condition in studying parametric empirical process. Khmaladze (1981) just assumes the i.i.d condition; Bai (2003) exploits the fact that, when the true parameter value is known, the dependent data could be transformed into an i.i.d sequence of uniformly distributed random variables by integral transformation. Delgado and Velasco (2009), in their simplest case, exploit the asymptotic i.i.d property of the residual autocorrelation under the true parameter. Our framework instead assumes some generic central limit theorem holds, as Assumption 7 will show below. The price of the generality is that the previous works could utilize infinite dimensional elements in the transformation, our transformation just explicitly utilizes finite dimensional elements. But infinite dimensional transformation just has limited theoretical value – as in practice, it is intractable, and has to be truncated in a finite sample.

In the following we present the assumptions for our analysis.

**Assumption 1.** $\Theta \subset \mathbb{R}^P$ is compact parameter space, and $\theta_0 \in \text{Int}(\Theta)$

Assumption 1 contains the requirements regarding the parameter space and the true parameter.

**Assumption 2.** \[ \{ \Lambda_t[m](I_{t-1}, \theta)\phi_t(Y_t, I_{t-1}, \theta), \theta \in \Theta \} \]

is a sequence of vector random functions such that $\Lambda_t[m](\cdot, \theta)\phi_t(\cdot, \cdot, \theta)$ are Borel measurable for each $\theta \in \Theta$, and $\Lambda_t[m](I_{t-1}, \cdot)\phi_t(Y_t, I_{t-1}, \cdot)$ are continuously differentiable on the interior of $\Theta$ for all $Y_t, \sigma(I_{t-1}), t = 1, 2, \cdots, m = 0, 1$.

Assumption 2 makes it relatively easy to prove our main result on estimation effect via Taylor expansion, furthermore, this assumption makes sure that the integral and differential operators are interchangeable.

**Assumption 3.** \[ \{ \frac{1}{T} \sum_{t=1}^{T} E[\Lambda_t[m](I_{t-1}, \theta)\phi_t(Y_t, I_{t-1}, \theta)] : \theta \in \Theta, T = 1, 2, \cdots \} \] is $O(1)$ and continuous on $\Theta$ uniformly in $T$,

$$
\sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^{T} [\Lambda[m]_t(I_{t-1}, \theta)\phi_t(Y_t, I_{t-1}, \theta) - E(\Lambda_t[m](I_{t-1}, \theta)\phi_t(Y_t, I_{t-1}, \theta))]) \right| \overset{p}{\to} 0,
$$
where $|| \cdot ||$ denotes Euclidean norm, for $m = 0, 1$.

**Assumption 4.** \( \{ T^{-1} \sum_{t=1}^{T} E \nabla_{\theta} [\Lambda_t(I_{t-1}, \theta) \phi_t(Y_t, I_{t-1}, \theta)] : \theta \in \Theta, T = 1, 2, \cdots \} \) is $O(1)$ and continuous on $\Theta$ uniformly in $T$,

\[
\sup_{\theta \in \Theta} || \frac{1}{T} \sum_{t=1}^{T} \{ \nabla_{\theta} [\Lambda[m]_t(I_{t-1}, \theta) \phi_t(Y_t, I_{t-1}, \theta)] - E \nabla_{\theta} [\Lambda_t(I_{t-1}, \theta) \phi_t(Y_t, I_{t-1}, \theta)] \} ||_p \to 0,
\]

for $m = 0, 1$.

Assumptions 3 and 4 are about uniformly weak convergence laws of large numbers (UWLLN). Andrews (1987) and Pötscher and Prucha (1989) provide requirements that can be used to establish UWLLN in a wide variety of situations, for example, stationary and ergodic process, and $\alpha$— or $\phi$—mixing process with mixing coefficients declining at a proper rate.

**Assumption 5.** $\sqrt{T}(\hat{\theta}_T - \theta_0) = O_p(1)$.

Assumption 5 show that the testing statistic needs only a $\sqrt{T}$-consistent estimator, and this could be regarded as another aspect of the robustness.

**Assumption 6.** Assume that $\hat{\Xi}_T[1]'\hat{\Xi}_T[1]$ is full rank with probability 1, and $L \geq P$.

Assumption 6 levies some conditions on the choice of $\Lambda_t[1]$ for $t = 1, 2, \cdots$, and the dimension $L$. To understand this point, notice that this transformation is in effect a detrending operation. We first regress $\hat{\xi}_T[1]$ on $\hat{\Xi}_T[1]$, which requires that $\Xi[1]'\Xi[1]$ is full rank with probability 1, then the least squares estimator is given by $\left(\hat{\Xi}_T[1]'\hat{\Xi}_T[1]\right)^{-1}\hat{\Xi}_T[1]'\hat{\xi}_T[1]$, notice that $\hat{\xi}_T[1]$ is a $L \times 1$ vector, $\hat{\Xi}_T[1]$ is a $L \times P$ matrix, so it at least requires that $L \geq P$. Finally, multiplying this estimator by $\hat{\Xi}_T$, we obtain the predicted value of $\hat{\xi}_T$, and the residual is given by (9). Given $S$, $L \geq P$ defines the lower bound of $L$, there is no other restriction on it, $L$ could be chosen as an increasing function of the sample size, although $L$ is regarded as a fixed number given a sample size.
**Assumption 7.** Define the $S + L$ vector

$$
\Pi_t(I_{t-1}, \theta) = (\Lambda_t^\prime(0)(I_{t-1}, \theta), \Lambda_t^\prime[1](I_{t-1}, \theta))',
$$
and assume that under $H_0$

$$
T^{-1/2} \sum_{t=1}^{T} \Pi_t \phi_t \to_d N(0, \Gamma),
$$
where $\Pi_t = \Pi_t(I_{t-1}, \theta_0), \Gamma = \lim_{T \to \infty} \text{Var}(T^{-1/2} \sum_{t=1}^{T} \Pi_t \phi_t) > 0$.

Note that $\sum_{t=1}^{T} \Pi_t \phi_t$ is the sum of a vector martingale difference sequence under $H_0$, its limiting distribution is generally derivable from a central limit theorem. For example there are the central limit theorems in the case of ergodic stationary martingale difference process and $\alpha-$ or $\phi-$mixing process.

**Theorem 1.** When Assumption 1-5 holds, under $H_0$, (7) and (8) hold.

**Proof.** See Appendix

With Theorem 1, we present our main result in Theorem 2.

**Theorem 2.** When Assumption 1-7 holds, under $H_0$, (10) holds, and

$$
\tilde{\xi}_T \to_d N(0, \Sigma),
$$
where

$$
\Sigma = \Upsilon \Gamma \Upsilon^\prime
$$

and therefore

$$
\tilde{\xi}_T^\prime \hat{\Sigma}^{-1} \tilde{\xi}_T \to_d \chi^2(S),
$$
where $\hat{\Sigma} = \hat{\Upsilon} \hat{\Gamma} \hat{\Upsilon}^\prime$ is a consistent estimate of $\Sigma$, in which $\hat{\Gamma} = T^{-1} \sum_{t=1}^{T} \hat{\Pi}_t \hat{\Pi}_t \hat{\phi}_t^2$.
2.1 Local Alternatives

We consider the following class of local alternative hypothesis

\[ H_{1T} : T^{-1} \sum_{t=1}^{T} E(\Lambda_t[m]\phi_t) = \frac{\delta[m]}{\sqrt{T}}, \text{ for some } \theta_0 \in \Theta, \ m = 0, 1, \]

where \( \delta[1] \) is an \( L \times 1 \) nonrandom constant vector, with \( \delta[0] \) is an \( S \times 1 \) nonrandom constant vector.

Assumption 8. Assume that under \( H_{1T} \)

\[ T^{-1/2} \sum_{t=1}^{T} \Pi_t \phi_t \rightarrow_d N(\delta, \Gamma), \]

where \( \delta = (\delta[0]', \delta[1]')', \Gamma = \lim_{T \rightarrow \infty} \text{Var}(T^{-1/2} \sum_{t=1}^{T} (\Pi_t \phi_t)) > 0. \)

The following theorem provides the limiting distribution of \( \hat{\xi}_T \) under the local alternative.

Theorem 3. When Assumption 1-6 and 8 hold, under \( H_{1T} \),

\[ \hat{\xi}_T \Sigma^{-1} \hat{\xi}_T \rightarrow_d \chi^2(S, \kappa' \Sigma^{-1} \kappa), \]

where \( \kappa = \delta[0] - \Xi[0] (\Xi[1]'\Xi[1])^{-1} \Xi[1]'\delta[1], \Sigma = \Upsilon \Gamma \Upsilon', \Upsilon \) is defined in (15), and \( \hat{\Sigma} \rightarrow_p \Sigma. \)

Proof. See Appendix
3 Relation to Other Tests

As mentioned before, the framework of this paper is so general that it includes Wooldridge (1990)’s modified statistic as a special case. We can rewrite Wooldridge’s transformation as

\[
\tilde{\xi}_T = \xi_T - T^{-1/2} \sum_{t=1}^T \hat{\Lambda}_t \hat{\Phi}_t \left( \sum_{t=1}^T \hat{\Phi}_t' \hat{\Phi}_t \right)^{-1} \sum_{t=1}^T \hat{\Phi}_t' \hat{\phi}_t
\]

\[
= \hat{\xi}_T - T^{-1/2} \sum_{t=1}^T \hat{\Lambda}_t \hat{\Phi}_t \left( \left( \sum_{t=1}^T \hat{\Phi}_t' \hat{\Phi}_t \right)' \left( \sum_{t=1}^T \hat{\Phi}_t' \hat{\phi}_t \right) \right)^{-1} \left( \sum_{t=1}^T \hat{\Phi}_t' \hat{\phi}_t \right)' \sum_{t=1}^T \hat{\Phi}_t' \hat{\phi}_t.
\]

Note that it is because of the symmetry of \( \sum_{t=1}^T \hat{\Phi}_t' \hat{\Phi}_t \) that the second equation holds. So Wooldridge (1990)’s modified statistic \( \bar{\xi}_T \) turns out to be equivalent to \( \tilde{\xi}_T \), using \( \xi_T[1] = T^{-1/2} \sum_{t=1}^T \hat{\Phi}_t' \hat{\phi}_t \) with \( \Lambda_t[1] = \Phi_t' \). As claimed by Wooldridge (1990), the test statistic maintains asymptotic efficiency under ideal conditions. Later we will show its limitations in the light of our general framework.

Our tests can be regarded as Neyman (1959) \( C(\alpha) \) tests in the conditional moment testing framework. It is asymptotically equivalent to the White (1994) type testing statistic which is evaluated at the estimator that is obtained from using the instrument \( \Lambda_t[1](I_{t-1}, \theta) \) as we now argue. Following Wooldridge (1990), let \( \theta^*_T \) be nonstochastic sequences in the local alternative such that \( H_{2T} : \sqrt{T} (\theta^*_T - \theta_0) = o_p(1) \), assume that

\[
E_{\theta^*_T} \left[ T^{-1} \sum_{t=1}^T G_t(Y_t, I_{t-1}, \theta^*_T) \right] - E_{\theta_0} \left[ T^{-1} \sum_{t=1}^T G_t(Y_t, I_{t-1}, \theta_0) \right] \rightarrow 0
\]

as \( T \rightarrow \infty \) for various functions \( G_t \). The arguments of Theorem 3 can be used to show that under the local alternatives \( H_{2T} \),

\[
\tilde{\xi}_T = T^{-1/2} \sum_{t=1}^T L_t^* \phi^*_t - T^{-1/2} \sum_{t=1}^T \hat{L}_t^* \phi^*_t \left( \left( \sum_{t=1}^T \hat{L}_t^*[1]|\phi^*_t \right)' \left( \sum_{t=1}^T \hat{L}_t^*[1]|\phi^*_t \right) \right)^{-1} \left( \sum_{t=1}^T \hat{L}_t^*[1]|\phi^*_t \right)' \sum_{t=1}^T \hat{L}_t^*[1]|\phi^*_t + o_p(1)
\]

where the values with “*” superscript are evaluated at \( \theta^*_T \). Therefore if \( \hat{\theta}^{(1)}_T, \hat{\theta}_T \) are
both $\sqrt{T}$ consistent estimators of $\theta^*_T$, under $H_{2T}$, then

$$\hat{\xi}^{(1)}_T - \tilde{\xi}_T = o_p(1)$$

where $\hat{\xi}^{(1)}_T$ is evaluated at $\hat{\theta}^{(1)}_T$, and $\tilde{\xi}_T$ is evaluated at $\tilde{\theta}_T$.

If $\tilde{\theta}_T$ is chosen to satisfy that

$$T^{1/2} \left( \frac{1}{T} \sum_{t=1}^{T} \Lambda_t[1](\tilde{\theta}_T)\Phi_t(\tilde{\theta}_T) \right)' \frac{1}{T} \sum_{t=1}^{T} \Lambda_t[1](\tilde{\theta}_T)\phi_t(\tilde{\theta}_T) = o_p(1),$$

then

$$\hat{\xi}_T - \tilde{\xi}_T = o_p(1)$$

where $\hat{\xi}_T$ is evaluated at $\hat{\theta}_T$. Note that when $\hat{\theta}_T$ satisfies

$$\hat{\theta}_T = \arg\min_{\theta} \left( \frac{1}{T} \sum_{t=1}^{T} \Lambda_t[1](\theta)\phi_t(\theta) \right)' \frac{1}{T} \sum_{t=1}^{T} \Lambda_t[1](\theta)\phi_t(\theta)$$

(17) holds, which is the GMM estimation objective function with identity weighting matrix $I_L$. Note that $\left( \frac{1}{T} \sum_{t=1}^{T} \Lambda_t[1](\tilde{\theta}_T)\Phi_t(\tilde{\theta}_T) \right)' \frac{1}{T} \sum_{t=1}^{T} \Lambda_t[1](\tilde{\theta}_T)\phi_t(\tilde{\theta}_T)$ is the main part of the first derivative of (18). This means our new statistic is asymptotically equivalent to the testing statistic which is evaluated at the estimator that is obtained from using the instrument $\Lambda_t[1](I_{t-1}, \theta)$.

### 4 Optimal Statistic

The analysis of the previous section shows that our new methodology is asymptotically equivalent to using GMM estimation to handle the estimation effect. Note that so far we have not discussed the optimal weighting matrix for GMM parameter estimation. As we know, in GMM framework, in order to obtain the efficient estimate, the optimal weighting function has to be chosen. So analogically, we could introduce the optimal weighting matrix into our framework. By considering the optimal weighting matrix...
\[ \hat{W}_T = \left( \frac{1}{T} \sum_{t=1}^{T} \hat{\Lambda}_t[1] \hat{\phi}_t \right)^{-1} \], our statistic becomes

\[ \tilde{\xi}_T = T^{-1/2} \sum_{t=1}^{T} \hat{\Lambda}_t \hat{\phi}_t - T^{-1/2} \sum_{t=1}^{T} \hat{\Lambda}_t \hat{\Phi}_t \left( \left( \sum_{t=1}^{T} \hat{\Lambda}_t[1] \hat{\phi}_t \right)' \hat{W}_T \sum_{t=1}^{T} \hat{\Lambda}_t[1] \hat{\phi}_t \right)^{-1} \left( \sum_{t=1}^{T} \hat{\Lambda}_t[1] \hat{\Phi}_t \right)' \hat{W}_T \sum_{t=1}^{T} \hat{\Lambda}_t[1] \hat{\phi}_t \]

It could be shown that \( \tilde{\xi}_T \) is asymptotically equivalent to \( \hat{\xi}_T \), where \( \hat{\xi}_T \) is evaluated at \( \hat{\theta}_T \), which satisfies

\[ \hat{\theta}_T = \arg\min_{\theta} \left( \frac{1}{T} \sum_{t=1}^{T} \Lambda_t[1](\theta) \phi_t(\theta) \right)' \hat{W}_T \left( \frac{1}{T} \sum_{t=1}^{T} \Lambda_t[1](\theta) \phi_t(\theta) \right). \]

This argument sheds some light on the limitations of the Wooldridge (1990)'s modified statistic. Wooldridge (1990)'s modified statistic uses just identified unconditional moment restrictions, and the instrument used is not necessarily the optimal instrument in most cases. Furthermore, comparing to our general framework, Wooldridge (1990)'s statistic is too restrictive in the sense that the additional unconditional moments used to remove the parameters estimation effect are predetermined (they are just the score functions \( \Phi_t \) of the conditional moment restriction). The possible choices of \( \Lambda[0] \) will be restrained by the form of \( \Phi'_t \).

Our general approach deviates from the estimation-tests paradigm of Newey (1985a, b), Tauchen (1985) and White (1994). Newey (1985b) considers specification testing in the framework of GMM, in his case (in the notation of this paper), a linear combination of the moment conditions \( E(\Lambda_t[0]' \Lambda_t[1]' \phi_t = 0 \) is considered, and the estimator is based on GMM estimation of \( E(\Lambda_t[0]' \Lambda_t[1]' \phi_t = 0 \) under some weighting function. But without any specification of the form of the linear combination and further assumptions, generalized inverse to standardize the sample moment conditions evaluated at the parameter estimators has to be handled. Moreover, Newey (1985b) derives the optimal GMM test in the case that the distribution information is assumed to be known, and the optimality is about testing for the linear combination of the moment conditions \( E(\Lambda_t[0]' \Lambda_t[1]' \phi_t = 0 \). On the other hand our framework assumes condi-
tional moment restrictions hold, and assume that the variance of \((\Lambda_t[0]'\Lambda_t[1]')'\phi_t\) is full rank. Moreover, Our framework explicitly focuses on testing the validity of moment conditions \(E(\Lambda_t[0]\phi_t) = 0\). It is impossible to obtain the same optimality as Newey (1985b) generally. In some special cases, we can establish the asymptotically equivalence between our new framework and Newey (1985b)'s optimal test. More specifically, denote \(Z_t = (Y_t, X_t)'\), and suppose \(Z_t, t = 1, \cdots, T\), consists of random vectors which are the first \(T\) elements of s strictly stationary stochastic process \(\{Z_t; t = 1, 2, \cdots\}\), and has a measurable joint density function and suppose \(f(Z_1, \cdots, Z_T, C_T)\) with respect to a measure of \(\sum_{t=1}^{T} v\), where \(v\) is a \(\sigma\)-finite measure. \(C_T\) is a \(Q \times 1\) vector of parameters, which satisfies
\[
C_T = C_0 + \delta/\sqrt{T}.
\]

Suppose that \(\Lambda_t[0]\phi_t\) is a \(Q \times 1\) vector and
\[
E(\Lambda_t[0]'\Lambda_t[1]')'\phi_t = (C', 0)'
\]

Separate \(\Gamma\), which is defined in (12), into
\[
\begin{pmatrix}
  \Gamma_{00} & \Gamma_{01} \\
  \Gamma_{10} & \Gamma_{11}
\end{pmatrix}
\]

Define \(\hat{\Lambda}[0]\phi_t = \Lambda[0]\phi_t - \Gamma_{10}\Gamma_{00}^{-1}\Lambda[1]\phi_t\), it could be shown that \(\check{\xi}_T\) with \(\hat{\Lambda}[0]\phi_t\) and extra moment conditions \(E(\Lambda_t[1]\phi_t) = 0\) with optimal weighting function is asymptotically equivalent to Newey (1985b)'s optimal GMM test (Proposition 3) in this setting. Note that the alternative is quite special, there is only a drift in moment \(E(\Lambda_t[0]\phi_t)\). The optimal test need partial out the uncontaminated moment conditions \(E(\Lambda_t[1]\phi_t)\) by \(\Lambda_t[0]\phi_t - \Gamma_{10}\Gamma_{00}^{-1}\Lambda[1]\phi_t\), while GMM estimation uses only moment conditions \(E(\Lambda_t[1]\phi_t)\).

The efficiency of test depends on the choices of the unconditional moment restrictions which are used to purge the estimation effect. For example, in linear model with
heteroskedasticity, we may follow Cragg (1983) to choose extra moment conditions under the heteroskedasticity in addition to the moment conditions which are used to obtain the initial estimator \( \hat{\theta}_T \). It is possible to choose unconditional moment restrictions to reach semiparametric efficiency bound.

**Example 1.** Consider the conditional moment restriction

\[
H_0 : E(\phi(z, \theta_0) | x) = 0, \text{ for some } \theta_0 \in \Theta, \ t = 1, 2, \cdots .
\]  

(19)

where \( z \) denotes a single observation, \( \theta \) a \( p \times 1 \) vector of parameter, \( \phi(z, \theta) \) a scalar function, \( x \) is a subvector of \( z \), acting as conditional variables.

For each positive integer \( K \) let \( q^K(x) = (q_{1K}(x), \cdots , q_{KK}(x))' \) be a \( K \times 1 \) vector of approximating functions, satisfying the following assumption:

For all \( K \), \( E[q^K(x)' q^K(x)] \) is finite, and for any \( a(x) \) with \( E[a(x)^2] < \infty \), there are \( K \times 1 \) vector \( \gamma_K \), such that as \( K \to \infty \),

\[
E[a(x) - q^K(x)' \gamma_K]^2 \to 0
\]

In this case, we could properly choose \( \Lambda[1] = q^K(x) \) to get optimal testing statistic under some conditions. Since the estimator reaches semiparametric lower bound, for more details see Donald et al (2003).

**Example 2.** Consider a univariate AR(\( p \)) model,

\[
y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + \epsilon_t = X_t' \theta + \epsilon_t
\]

Where \( X_t = (y_{t-1}, \cdots , y_{t-p})' \); \( \theta = (\theta_1, \cdots , \theta_p)' \), which satisfy the condition that the roots of the associated lag polynomial lie outside of the unit circle; and \( \epsilon_t \) satisfies the
martingale difference assumption

\[ E(\epsilon_t|I_{t-1}) = 0 \]

where \( I_{t-1} = (y_{t-1}, y_{t-2}, \cdots) \) is the information set at \( t-1 \). So by the notation of our testing framework, we have \( \phi_t = \epsilon_t, \Phi_t = X'_t \).

If we assume that conditional homoskedasticity: \( E(\epsilon_t^2|I_{t-1}) = E(\epsilon_t^2) \), The LM test is optimal, since OLS or QMLE estimator reaches efficiency. But the assumption of conditional homoskedasticity is too strong, and when it does not hold, Then OLS or QMLE estimator is not optimal. There exists more efficient or most efficient GMM estimator. In this case, if we choose instrument variable \( \Lambda[1] = (y_{t-1}, \cdots, y_{t-L})' \), where \( L > p \), we will obtain more efficient estimator. If we assume that \( E(\epsilon_t^2\epsilon_{t-j}\epsilon_{t-k}) = 0 \), when \( j \neq k \), it is possible to choose

\[
\left( \sum_{t=1}^{T} \Lambda_t[1]\Phi_t \right)' W_T \Lambda_t[1] = \sum_{j=1}^{\infty} (E\epsilon_{t-j}X_t/E\epsilon_t^2\epsilon_{t-j})\epsilon_{t-j}
\]

to get the optimal test. Since \( \sum_{j=1}^{\infty} (E\epsilon_{t-j}X_t/E\epsilon_t^2\epsilon_{t-j})\epsilon_{t-j} \) is the optimal instruments in this case (See West (2002)).

Example 1 and 2 demonstrate the advantage of our general framework: robustness and efficiency are obtained without paying too much price. Note that in Example 2 the optimal instrument under heteroskedasticity is not explicitly defined-it is the case especially in time series, if instead the optimal estimator is pursued, some burdensome two-step procedure is required. Our framework evaluates the optimal instrument at some consistent estimation value, but it is asymptotically equivalent to the case where the optimal estimator is used.
5 Application to Testing of GARCH Models

5.1 The Null GARCH\((p,q)\) Model and The Testing Framework

Just for simplicity, we are considering the following conditional variance model:

\[
Y_t = \varepsilon_t h_t^{1/2}, \quad h_t = \omega_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \quad t \in \mathbb{Z}, \ p, q \in \mathbb{N},
\]

Denote \(I_{t-1} = (Y_{t-1}, Y_{t-2}, \cdots)\), \(\varepsilon_t\) is a sequence of random variables, satisfying

\[
E(\varepsilon_t^2 | I_{t-1}) = 1, \text{ a.s.}
\]

and \(\omega_0 > 0, \alpha_0i \geq 0, i = 1, \cdots, p, \beta_0j \geq 0, j = 1, \cdots, q\). Define the vector of parameters \(\theta = (\omega, \alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q)\), and the true parameter is denoted by \(\theta_0 = (\omega_0, \alpha_{01}, \cdots, \alpha_{0p}, \beta_{01}, \cdots, \beta_{0q})\).

Assumption A1. \(\varepsilon_t, t \in \mathbb{Z}\) is a strictly stationary and ergodic process satisfying \(E(\varepsilon_t^2 | I_{t-1}) = 1, \text{ a.s. and } \varepsilon_t^2\) has a nondegenerated distribution.

Assumption A2. \(Y_t, t \in \mathbb{Z}\) is a strictly stationary and ergodic process with \(E[|Y_t|^{2s}] < \infty\), for some \(s > 0\)

Assumption A3. \(\theta_0 \in \text{int}(\Theta)\), where \(\text{int}(\Theta)\) denotes the interior of \(\Theta\), and \(\Theta\) is compact.

Assumption A3. if \(q > 0\), \(A_{\theta_0}(z)\) and \(B_{\theta_0}(z)\) have no common root. For all \(\theta \in \Theta\), \(B_{\theta_0}(z)\) has its roots outside the unit circle. Moreover, \(A_{\theta_0}(1) \neq 0\) and \(\alpha_{0p} + \beta_{0q} \neq 0\), where \(A_{\theta_0}(z) = \sum_{i=1}^{p} \alpha_i z^i\) and \(B_{\theta_0}(z) = 1 - \sum_{i=1}^{q} \beta_i z^i, \ z \in \mathbb{C}\). By convention, \(A_{\theta_0}(z) \equiv 1\) if \(p = 0\) and \(B_{\theta_0}(z) \equiv 1\) if \(q = 0\).

Assumption A4. \(E|\varepsilon|^{4(1+\delta)} < 0\) for some \(\delta > 0\)

Assumption A1 to A4 make sure the consistency and asymptotically normality of QMLE estimator, for more details see Escanciano (2009).
Originally, Bollerslev (1986) presents a score type statistic for testing the GARCH model against a higher order GARCH model. Engle and Ng (1993) propose tests for asymmetry. Li and Mak (1994) construct the test for the adequacy of a GARCH model by considering the autocorrelation of the squared standardized error. Lundbergh and Teräsvirta (2002) present a unified framework for misspecification of GARCH model. Parametric Lagrange multiplier (LM) type tests of no ARCH in standardized errors, linearity and parameter constancy are proposed. Halunga and Orme (2009) propose tests on asymmetry and nonlinearity by considering the recursive nature of the GARCH model.

All specification tests mentioned above could be incorporated into the conditional moment testing framework, the conditional moment restriction is:

$$H_0 : E\left(\frac{Y_t^2}{h_t} - 1|I_{t-1}\right) = 0, \theta_0 \in \Theta, \quad (20)$$

The CM tests is to test the null:

$$H'_0 : E\left[\Lambda(I_{t-1}, \theta_0)\left(\frac{Y_t^2}{h_t} - 1\right)\right] = 0, \text{for some } \theta_0 \in \Theta$$

Suppose that we could find other unconditional moment restrictions such that

$$E\left[\Lambda[1](I_{t-1}, \theta_0)\left(\frac{Y_t^2}{h_t} - 1\right)\right] = 0, \text{for some } \theta_0 \in \Theta$$

We may follow Guo and Phillips (2001) to choose optimal instruments under semi-strong GARCH case. But the assumption $E(Y_t^8) < \infty$, which is required for the asymptotically normality of the estimator, is too strong for most practical situations.

Under Assumption A1 to A4, we may choose $\Lambda_t[1] = (\frac{Y_{t-1}^2}{h_{t-1}} - 1, \cdots, \frac{Y_{t-L}^2}{h_{t-L}} - 1)', \Lambda_t[1] = (\nabla_\theta_0(\frac{Y_{t-1}^2}{h_{t-1}}), \cdots, \nabla_\theta_0(\frac{Y_{t-L}^2}{h_{t-L}}))' \text{ or } \Lambda_t[1] = (Y_{t-1} - \sqrt{h_{t-1}}, \cdots, Y_{t-L} - \sqrt{h_{t-L}})'$. Just for illustration, in the following, We choose $\Lambda_t[1] = (Y_{t-1} - \sqrt{h_{t-1}}, \cdots, Y_{t-L} - \sqrt{h_{t-L}})'$. It
could been shown that Assumptions 1-7 are satisfied, when the null model follows the Assumptions A1 to A4. The transformed statistic has the form (9) with

$$\phi_t = \frac{Y^2_t}{h_t} - 1$$

Note that

$$\Xi[m] = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E[\Lambda_t[m] \nabla_\theta \left( \frac{Y^2_t}{h_t} - 1 \right)]$$

$$= E(-\frac{\Lambda_t[m]}{h_t} \nabla_\theta h_t), \text{ for } m = 0, 1.$$  

where $\nabla_\theta h_t = \nabla_\theta h(I_{t-1}, \theta_0)$. When $\beta_j \neq 0$ for $j = 1, \cdots, q$, $\nabla_\theta h_t$ has recursive characteristics,

$$\nabla_\theta h_t = s_{t-1} + \sum_{j=1}^{q} \beta_j \nabla_\theta h_{t-j}.$$  

where $s_{t-1} = (1, Y_{t-1}, \cdots, Y_{t-p}, h_{t-1}, \cdots, h_{t-q})'$.

### 5.2 Monte Carlo Study: Adequacy of ARCH/GARCH Model

In this section, we discuss the testing of the adequacy of ARCH/GARCH model and provide the monte Carlo results.

Lundbergh and Teräsvirta (2002) establish a parametric alternative to the GARCH model, assuming that

$$\varepsilon_t = z_t g_t^{1/2},$$

where $\{z_t\}$ is a sequence of independent, identically distributed random variables with zero mean, unit variance. $g_t = 1 + \pi' v_t$, where $v_t = (\varepsilon^2_{t-1}, \cdots, \varepsilon^2_{t-S})'$ and $\pi = (\pi_1, \cdots, \pi_s)$, $\pi_j \geq 0$, $j = 1, \cdots, S$. Lundbergh and Teräsvirta (2002) call the alternative model as "ARCH nested in GARCH" model. They test $H^*_0 : \pi = 0$ against $\pi \neq 0$, although the elements of $\pi$ are constrained to be non-negative. The statistic
used in Lundbergh and Teräsvirta (2002) has the form

\[ \hat{\xi}_T = T^{-1/2} \sum_{t=1}^{T} [\hat{\Lambda}_t \left( \frac{\hat{Y}_t^2}{h_t} - 1 \right)], \]

where \( \hat{\Lambda}_t = (\hat{\varepsilon}_{t-1}^2, \ldots, \hat{\varepsilon}_{t-S}^2)' \).

Li and Mak (1994) introduce a portmanteau statistic for testing the adequacy of the standard \( \text{GARCH}(p,q) \) model by testing the null hypothesis that the squared and standardized error process is not autocorrelated. The statistic still falls into the general framework, here \( \hat{\Lambda}_t = (\hat{\varepsilon}_{t-1}^2 - 1, \ldots, \hat{\varepsilon}_{t-S}^2 - 1)' \). Both Lundbergh and Teräsvirta (2002) statistic and Li and Mak (1994) statistic are equivalent asymptotically. In our Monte Carlo experiments, we follow Lundbergh and Teräsvirta (2002), choosing \( \hat{\Lambda}_t = (\hat{\varepsilon}_{t-1}^2, \ldots, \hat{\varepsilon}_{t-S}^2)' \).

The Monte Carlo experiment is conducted in MATLAB 7.6, using GARCH Toolbox to simulate strong GARCH (ARCH) models. The estimation is based on Gaussian maximum likelihood method.

The Monte Carlo experiment for assessing the size properties of the tests is based on an ARCH(1) model. namely

\[ Y_t = \sqrt{h_t} \varepsilon_t, \quad h_t = \alpha_0 + \alpha_1 Y_{t-1}^2, \]

where \( \varepsilon_t \sim N(0,1) \), or \( \varepsilon_t \sim t(d) \) (standardized Student t-distribution with degree of freedom \( d \)).

Following Li and Mak (1994), we choose \( \alpha_0 = 0.1, \alpha_1 = 0.3 \). Each model is replicated and estimated 10,000 times. For the \( t(d) \) distribution we choose \( d = 3, 5, 7 \) here. Note that when the degree of freedom of t distribution is \( d = 3 \), the asymptotically normality of QMLE estimator fails.

We report the empirical size 5% of testing no remaining ARCH in Figure 1 with sample size 250, and in Figure 2 with sample size 500, comparing our testing statistic
with the LM (Lagrange Multiplier) test and Wooldridge(1990)'s modified statistic. In both figures, we allow $L$ to change, while $S$ is fixed. The results of $S = 4$ is reported. We also report the results in Table 1, where $L$ is fixed.

Figure 1: Results of empirical size of testing no remaining ARCH effect, nominal size 5%, Sample size $T=250$, $S = 4$

Figure 1 and 2 and Table 1 show that our new testing statistic has good empirical size in all the cases, while the LM testing statistic tends to be oversized especially in the case of nonnomal distribution, Wooldridge (1990)'s modified statistic is downsized or oversized in some cases. Furthermore, given a sample size, the empirical size of the different value of $L$ is quite stable.

For the power checking, We consider an alternative ARCH(2) model.

\[ h_t = 0.1 + 0.2Y_{t-1}^2 + 0.2Y_{t-2}^2. \]
We report the results in Figure 3 and Table 2. Firstly, note that Wooldridge (1990)’s modified statistic has no power. The power of our testing statistic increases as $L$ increases, and is better than LM test in non-normal cases. In normal case, The power of our new transformed statistic is just a little bit lower than LM statistic. We conjecture that when the GARCH model is semi-strong or weak one, our new statistic will get more power. We can conclude that our testing statistic has very good power properties.
Table 1: Empirical size, Choosing $L = 20$

<table>
<thead>
<tr>
<th></th>
<th>$T=250$</th>
<th>$T=500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>1</td>
</tr>
<tr>
<td>$T_3$</td>
<td>N</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>LM</td>
<td>6.8</td>
</tr>
<tr>
<td>$T_5$</td>
<td>N</td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>LM</td>
<td>4.8</td>
</tr>
<tr>
<td>$T_7$</td>
<td>N</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>LM</td>
<td>4.4</td>
</tr>
<tr>
<td>$N(0, 1)$</td>
<td>N</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>LM</td>
<td>3.9</td>
</tr>
</tbody>
</table>


Table 2: Empirical power, Choosing $L = 20$

<table>
<thead>
<tr>
<th></th>
<th>$T=500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
</tr>
<tr>
<td>$T_3$</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>LM</td>
</tr>
<tr>
<td>$T_5$</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>LM</td>
</tr>
<tr>
<td>$T_7$</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>LM</td>
</tr>
<tr>
<td>$N(0, 1)$</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>LM</td>
</tr>
</tbody>
</table>

Figure 3: Results of empirical power of testing no remaining ARCH effect, nominal size 5%, Sample size T=500, S = 4

5.3 Testing for non-linearity

For the non-linearity testing, following Lundbergh and Teräsvirta (2002), the augmented version of model is

\[ Y_t = \varepsilon_t (h_t + g_t)^{1/2}, \]

where

\[ g_t = \sum_{j=1}^{q} \alpha_0 j \Pi_n (\varepsilon_{t-j}; \gamma, \mathbf{c}) + \sum_{j=1}^{q} \alpha_{1j} \Pi_n (\varepsilon_{t-j}; \gamma, \mathbf{c}) \varepsilon_{t-j}^2, \]
in which they consider a smooth transition alternative

\[ H_n(\epsilon_{t-j}; \gamma, c) = (1 + \exp(-\gamma \prod_{t=1}^{n}(x_t - c_t)))^{-1}, \gamma > 0, c_1 \leq \cdots \leq c_n, \]

where \( x_t \) is the transition variable at time \( t \), \( \gamma \) is a slope parameter, and \( c = (c_1, \ldots, c_n) \) a location vector.

By replacing the transition function \( H_n \) with a first-order Taylor approximation, the alternative hypothesis becomes

\[ h_t = \eta' s_{t-1} \]
\[ g_t = \beta_1' v_{1t} + \sum_{i=3}^{n+2} \beta_i' v_{it} + R_1, \]

where \( \beta_t = (\beta_{11}, \ldots, \beta_{ip})' \), \( v_{it} = (Y_{i,t-1}, \ldots, Y_{i,t-q})' \), \( i = 1, 3, \ldots, n + 2 \) and \( R_1 \) is the remainder. The new null hypothesis is \( H'_0 : \beta_1 = \beta_3 = \cdots = \beta_{n+2} = 0 \). One additional assumption is needed.

Under \( H_0 \), \( E(Y_{1(t-1)}^{2(n+2)}) < \infty \).

In this case, \( \hat{\Lambda}_t = (\hat{v}'_{1t}, \hat{v}'_{3t}, \ldots, \hat{v}'_{(n+2)t})' \). Under the GARCH(1,1) model, \( \hat{\Lambda}_t = (Y_{t-1}, Y_{t-1}^3)' \).

In the simulation, we we consider a GARCH(1,1) model under the null. In this case, \( \hat{\Lambda}_t = (Y_{t-1}, Y_{t-1}^3)' \), and compare our transformed statistic with Lagrange Multiplier statistic and Wooldridge (1990)’s modified statistic.

For the empirical size testing, we consider the following GARCH(1,1) model

\[ Y_t = \sqrt{h_t} \epsilon_t, \quad h_t = 0.1 + 0.1 Y_{t-1}^2 + 0.8 h_t. \]

We report the results in Figure 4. It shows that our transformed statistic has very good size properties. Both LM statistic and Wooldridge(1990)’s modified statistic are downsized, even in \( N(0,1) \) case. When degrees of freedom of t distribution are 3 and 5,
the assumption $E(Y_t^{2(n+2)}) < \infty$ fails, the LM statistic and Wooldridge (1990)'s sizes are downsized even further. Our transformed statistic is also downsized, but is much better than them.

For the power checking, we consider the GJR-GARCH model

$$h_t = 0.005 + 0.23[|Y_{t-1}| - 0.23Y_{t-1}]^2 + 0.7h_t.$$  

We report the results in Figure 5. It shows that the power of our new statistic is quite good. As $L$ increases, the power is quickly close to the LM or Wooldridge(1990)'s modified statistic. By consider both the size and power properties, we could conclude that our new transformed statistic has good size and power balance.
6 Conclusion

In this paper, we develop a new approach in the framework of conditional moment testing. Given a CM-test, additional moment restrictions are introduced to transform the statistic. It turns out the Wooldridge (1990)’s modified statistic is just a special case of our new methodology. When our framework applies to conditional variance model, the simulation results show that our new statistic has very good size property and nontrivial power against alternative, comparing with Lagrange Multiplier tests and Wooldridge (1990)’s modified statistic.
References


Appendix. In order to prove the theorem, We introduce a lemma

Lemma 1. Assume that the sequence of random functions \( \{Q_T(W_T, \theta) : \theta \in \Theta, T = 1, 2, \cdots\} \) where \( Q_T(W_T, \theta) \) is continuous on \( \Theta \) and \( \Theta \) is a compact subset of \( \mathbb{R}^P \), and the sequence of nonrandom functions \( \{\bar{Q}_T(W_T, \theta) : \theta \in \Theta, T = 1, 2, \cdots\} \), satisfy the following conditions:

1. \( \sup_{\theta \in \Theta} ||Q_T(W_T, \theta) - \bar{Q}_T(W_T, \theta)||_p \to 0 \)
2. \( \{\bar{Q}_T(W_T, \theta) : \theta \in \Theta, T = 1, 2, \cdots\} \) is continuous on \( \Theta \) uniformly in \( T \)

Let \( \hat{\theta}_T \) be a sequence of random vectors such that \( \hat{\theta}_T - \theta_0 \overset{p}{\to} 0 \), where \( \{\theta_0\} \subset \Theta \).

Then \( Q_T(W_T, \hat{\theta}_T) - \bar{Q}_T(W_T, \theta^0_T) \overset{p}{\to} 0 \)

Proof. See White (1994, Theorem 3.7).

Proof of Theorem 1. By Mean value theorem, we have

\[
T^{-1/2} \sum_{t=1}^{T} \hat{\Lambda}_t[m] \hat{\phi}_t = T^{-1/2} \sum_{t=1}^{T} \Lambda_t[m](I_{t-1}, \theta_0) \phi_t(Y_t, I_{t-1}, \theta_0) \\
+ T^{-1} \sum_{t=1}^{T} (\nabla_\theta \Lambda_t[m](I_{t-1}, \bar{\theta}_t) \otimes \phi_t(Y_t, I_{t-1}, \bar{\theta}_t) \sqrt{T}(\hat{\theta}_T - \theta_0) \\
+ T^{-1} \sum_{t=1}^{T} \Lambda_t[m](I_{t-1}, \bar{\theta}_t) \nabla_\theta \phi_t(Y_t, I_{t-1}, \bar{\theta}_t) \sqrt{T}(\hat{\theta}_T - \theta_0)
\]

where \( \bar{\theta}_t \) between \( \theta_0 \) and \( \hat{\theta}_T \), for \( t = 1, 2, \cdots, T \). Notice that Assumption 5 \( \sqrt{T}(\hat{\theta}_T - \theta_0) = O_p(1) \), so \( \sqrt{T}(\hat{\theta}_T - \theta_0) = O_p(1) \), for \( t = 1, 2, \cdots, T \). By Lemma 1 and Assumption 4, we have

\[
\frac{1}{T} \sum_{t=1}^{T} (\nabla_\theta \Lambda_t[m](I_{t-1}, \bar{\theta}_t) \otimes \phi_t(Y_t, I_{t-1}, \bar{\theta}_t) \overset{p}{\to} \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} (E[\nabla_\theta \Lambda_t[m]] \otimes \phi_t) \\
\frac{1}{T} \sum_{t=1}^{T} \Lambda_t[m](I_{t-1}, \bar{\theta}_t) \nabla_\theta \phi_t(Y_t, I_{t-1}, \bar{\theta}_t) \overset{p}{\to} \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E[\Lambda_t[m] \nabla_\theta \phi_t]
\]
Furthermore, by the law of iterated expectation

\[ T^{-1} \sum_{t=1}^{T} E[\nabla_{\theta} \Lambda_{t}[m] \otimes \phi_{t}] = T^{-1} \sum_{t=1}^{T} E[\nabla_{\theta} \Lambda_{t}[m] \otimes E(\phi_{t}|I_{t-1})] = 0. \]

Since under the null \( E(\phi_{t}|I_{t-1}) = 0 \). So we have

\[ \hat{\xi}[m] = \xi[m] + \Xi[m] \sqrt{T}(\hat{\theta}_T - \theta_0) + o_p(1) \]

and \( \Xi[m] = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E[\Lambda_{t}[m]\nabla_{\theta} \phi_{t}], \) for \( m = 0, 1 \)

Proof of Theorem 2. By Theorem 1, we have

\[
\begin{align*}
\hat{\xi}_T &= \hat{\xi}_T[0] - \hat{\Xi}_T[0](\hat{\Xi}_T[1]'\hat{\Xi}_T[1])^{-1}\hat{\Xi}_T[1]'\hat{\xi}_T[1] \\
&= \xi_T[0] + \Xi_T[0] \sqrt{T}(\hat{\theta}_T - \theta_0) + o_p(1) \\
&- \Xi[0](1\Xi_T[1]'\Xi_T[1])^{-1}1\Xi_T[1](\xi_T[1] + \Xi_T[1] \sqrt{T}(\hat{\theta}_T - \theta_0)) + o_p(1)) \\
&= \xi_T - \Xi_T[0](\Xi_T[1]'\Xi_T[1])^{-1}\Xi_T[1]'\xi_T[1] + o_p(1) \\
&= \Upsilon T^{-1/2} \sum_{t=1}^{T} (\Pi_t \phi_t) + o_p(1)
\end{align*}
\]

Where \( \Upsilon \) is defined in (15). By Assumption 7 and Slutsky’s Theorem, we can get (10) and (13), then (16) follows

Proof of Theorem 3. Based on Theorem 2, under \( H_{1T} \),

\[ T^{-1/2} \sum_{t=1}^{T} E(\Pi_t \phi_t) = \delta \]

\[
\begin{align*}
\frac{1}{T} \sum_{t=1}^{T} E[\nabla_{\theta} \Lambda_{t}[m](I_{t-1}, \theta_0) \phi_{t}(Y_t, I_{t-1}, \theta_0)] &= \frac{1}{T} \sum_{t=1}^{T} E[\nabla_{\theta} \Lambda_{t}[m](I_{t-1}, \theta_0) E(\phi_t(Y_t, I_{t-1}, \theta_0)|I_{t-1})] \\
&= T^{-1/2} \cdot T^{-1} \sum_{t=1}^{T} E[\nabla_{\theta} \Lambda_{t}[m](I_{t-1}, \theta_0) \delta[m] \to 0
\end{align*}
\]

38
for \( m = 0, 1 \). So we still have \( \Xi_T[m] = T^{-1} \sum_{t=1}^T E[\Lambda_t[m] \nabla \theta t] \). By similar argument in Proof of Theorem 1, we get \( \tilde{\xi}_T \tilde{\Sigma}^{-1} \tilde{\xi}_T \rightarrow_d \chi^2(S, \kappa \Sigma^{-1} \kappa) \), under \( H_1 \), where \( \kappa = \delta[0] - \Xi[0](\Xi[1]'\Xi[1])^{-1} \Xi[1]'\delta[1], \Sigma = \Upsilon \Gamma \Upsilon' \), \( \Upsilon \) is defined in (15).