UNIVERSIDAD CARLOS III DE MADRID

TESIS DOCTORAL

Three Essays on Specification Testing

Autor:

Xuexin Wang

Director:

Carlos Velasco Gómez

DEPARTAMENTO DE ECONOMÍA

Getafe, Julio 2012

TESIS DOCTORAL

Three Essays on Specification Testing

Autor: Xuexin Wang

Director: Carlos Velasco Gómez

Presidente: Dr. D. Miguel Ángel Delgado, Catedrático de la Universidad Carlos III de Madrid

Vocal: Dr. D. Ignaco N. Lobato, Catedrático de la ITAM

Secretario: Dr. D. Juan Carlos Escanciano Reyero, Profesor Titular de la Indiana University, Bloomington

Calificación:

Getafe, de de

Acknowledgements

The Writing of my Ph.D dissertation has been a challenging journey in my academic life. I could not have embarked on this expedition and traveled this far without the help and support of advisors, professors, colleagues, friends and my family.

First of all, I would like to express my deepest gratitude to Professor Carlos Velasco, my advisor. It has been a great honor to be his Ph.D. student. Over the past four years, he has continuously provided the support, encouragement and invaluable guidance to my academic research. I appreciate all his contributions of time, patience and insight, which make my Ph.D. dissertation stand on the right foot and hold it to a high research standard.

I would like to thank my committee members, Professor Miguel Delgado, Professor Ignacio N. Lobato and Professor Abderrahim Taamouti, for their insightful comments and constructive criticisms on my research.

I am also indebted to participants at seminars of Universidad Carlos III, ESEM 2011, 2011 North American Summer Meeting of the Econometric Society for their useful comments. The financial support from the Spanish Ministerio de Educación y Ciencia, Ref. no. SEJ2007-62908/ECON, is gratefully acknowledged.

Special thanks go to Professor Seongman Moon. As my mentor, he has steered me through some rough terrains during my formative years as a Master student and has generously extended his counsel and friendship in my doctoral period.

It is my pleasure to attend different courses on econometrics taught by Professor Jesus Gozalo, Juanjo Dolado, Miguel Delgado, Winfried Stute at UC3M. Their excellent tutoring stimulated my interest in econometrics and paved the way for my academic research. I thank them for all their effort.

My heartfelt gratitude goes to my family. This dissertation is simply impossible without them. I am indebted to my father, Wang Yongxiang, and my mother, Sun Yiying for their unflagging love and support. Living in a poor rural area of china, they work industriously to earn the family's living and spare no effort to provide the best possible environment for me and my elder sisters to grow up. They always stand by my side, giving my the freedom to chase my dream. My two elder sisters, Wang Xiaoyu and Wang

Xiaoli, also play an important role in my pursuit. I rely on them to take care of my parents when I am abroad. They are also reliable counselors, when I encounter difficulties in life.

Resumen en Castellano

En el primer capítulo, **"A General Approach to Conditional Moment Specification Testing with Projections"**, se desarrolla un enfoque general para el análisis de especificación del modelo dentro del marco de pruebas de especificación de los momentos condicionales. A diferencia del paradigma de cálculos aproximados y pruebas desarrollado por Newey (1985a, b) y Tauchen (1985), la nueva metodología de estimación, teniendo como objetivo explorar la naturaleza de prueba de especificación de los momentos condicionales, elimina el efecto de estimación no despreciable a través de una transformación del estadístico de prueba basada en la proyección. Es decir, las restricciones de los momentos condicionales no sólo implican las limitaciones de los momentos no-condicionales que estamos examinando, sino también muchas otras más. Nuestro procedimiento incluye como un caso especial el estadístico modificado de Wooldridge (1990). Este enfoque es robusto a las desviaciones de las hipótesis de distribución que no hemos llegado a examinar, por otra parte sólo se necesita un preliminar estimador $\sqrt{T}$-consistente, y la transformación es asintóticamente libre d distribución. Además, el estadístico transformado alcanza la eficiencia asintótica en términos de la estimación GMM. Ponemos como un ejemplo, la aplicación de nuestra metodología para examinar la idoneidad y la no linealidad del modelo GARCH. En comparación con el tipo de pruebas de multiplicador Lagrange (LM) y el estadístico modificado de Wooldridge (1990), el resultado de la simulación muestra que nuestro estadístico nuevo tiene unas propiedades de tamaño y una potencia no-trivial muy satisfactorias. Por último, destacamos los méritos de dicho enfoque con una aplicación a los datos diarios de S&P 500.

En el segundo capítulo, **"An Improved Consistent Conditional Moment Test for Regression Models in The Presence of Heteroskedasticity of Unknown Form"**, exploramos la propiedad de dualidad de una clase de funciones de ponderación para prueba de especificación consistente de los modelos de regresión, asıcomo para estimación eficiente de los mismos. Empleando una proyección basada en transformación, proponemos un nuevo y consistente estadístico de prueba de forma funcional. Se muestra que el estadístico de prueba explora la estimación de parámetro asintótica y eficiente en una forma de heterocedasticidad desconocida. Además, es muy fácil de calcular con un solo estimador $\sqrt{n}$-consistente preliminar. Mientras tanto, planteamos una nueva versión de la prueba de Bierens (1990), y analizamos sus propiedades asintóticas. Realizamos simulaciones de Monte Carlo para demostrar las buenas propiedades de muestra finita del nuevo estadístico.

En el tercer capítulo, **"A Joint Portmanteau Test for Conditional Mean and Variance Time Series Models"(con Carlos Velasco)**, se propone una nueva prueba de conjunto para Portmanteau paramétrica medias condicionales y varianzas de los modelos de series de tiempo lineales y no lineales. El uso de la prueba de articulación está motivada por el hecho de que las pruebas marginales para la varianza condicional puede conducir a conclusiones erróneas cuando el media condicional queda mal. La nueva prueba se basa en una distribución libre de transformación en las autocorrelaciones de muestra de ambos residuales normalizados y cuadrados residuales normalizados, que se extienden y Delgado Velasco (2011). Las versiones sólidas de la prueba son adecuadamente en cuenta una mayor dependencia momento el orden. El rendimiento finito-muestra de la nueva prueba se compararon con los de ensayos bien conocidos a través de simulaciones.

Dissertation Abstract

In the first chapter, **"A General Approach to Conditional Moment Specification Testing with Projections"**, we develop a general approach for model specification analysis within the conditional moment specification testing framework. Unlike the estimating-testing paradigm developed by Newey (1985a,b) and Tauchen (1985), the new methodology removes the non-negligible estimation effect via a projection-based transformation of the test statistic, exploiting the nature of conditional moment specification testing. That is, the conditional moment restrictions not only imply the unconditional moment restrictions we are testing, but also many other unconditional moment restrictions. Our testing procedure includes Wooldridge (1990)'s modified statistic as a special case. This approach is robust to departures from the distributional assumptions that are not being tested, moreover only a preliminary $\sqrt{T}$-consistent estimator is needed, and the transformation is asymptotically distribution free. Furthermore, the transformed statistic reaches asymptotic efficiency in the sense of GMM estimation. As examples, we apply our methodology to test the adequacy and nonlinearity of the 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. Finally, an application to the S&P 500 daily data highlights the merits of our approach.

In the second chapter, **"An Improved Consistent Conditional Moment Test for Regression Models in The Presence of Heteroskedasticity of Unknown Form"**, we exploit the duality property of one class of weighting functions for both consistent specification testing and efficient estimation of regression models. A new consistent test statistic of functional form is proposed employing a transformation-based projection. It is shown that the new test statistic exploits asymptotic efficient parameter estimation under heteroskedasticity of unknown form. Further, it is quite easy to compute, only a preliminary $\sqrt{n}$-consistent estimator is needed. Then a new version of Bierens (1990) test is proposed, and its asymptotic properties are analyzed. Monte Carlo simulations are conducted to demonstrate the good finite sample properties of the new test statistic.

In the third chapter, **"A Joint Portmanteau Test for Conditional Mean and Variance Time Series Models"**, we propose a new joint Portmanteau Test for parametric conditional means and variances of linear and nonlinear time series models. The use of the joint test is motivated from the fact that marginal tests for the conditional variance may lead to misleading conclusions when the conditional mean is misspecified. The new test is based on a distribution-free transformation on the sample autocorrelations of both normalized residuals and squared normalized residuals, extending Delgado and Velasco (2011). The robust versions of the test are properly account for higher order moment dependence. The finite-sample performance of the new test is compared with those of well known tests through simulations.

Contents

Chapter 1

A General Approach to Conditional Moment Specification Testing with Projections

1.1 Introduction

Models based on conditional moment restrictions are very important in econometrics. For example in macroeconomics and finance, rational expectations and dynamic asset pricing models give rise to conditional moment restrictions in the form of stochastic Euler equations. In the context of maximum likelihood models, conditional moment restrictions appear in the score functions when exogenous variables are present. So it becomes crucial to check the validity of the conditional moment restrictions.

Conditional moment tests, which are firstly proposed by Newey (1985a), aim at testing the "directional" validity of the conditional moment restrictions via a finite number of unconditional moment conditions implied by the conditional moment restrictions. This framework includes many specification testing procedures as special cases, examples are Lagrange Multiplier (LM) test for nested hypothesis, Hausman's statistic, White's statistic, Portmanteu tests, and so on. Although conditional moment tests are not able to detect some misspecifications, they are still useful when the econometrician has a specific alternative in mind, and they may be optimal in the direction of the precisely specified alternatives. On the other hand, "Omnibus" specification tests have been developed to consistently test these conditional moment restrictions, see, for example, Bierens (1982, 1990), Zheng (1996). These consistent tests can be interpreted as conditional moment tests with an infinite number of moment restrictions.

In this paper we focus on the conditional moment tests. This is a classical research topic in econometrics. In a maximum likelihood setting with independent observations, Newey (1985a) and Tauchen (1985) derive the asymptotic distribution of conditional moment test statistic. Newey (1985b) considers the statistic in the GMM framework. White (1994) incorporates both the MLE and GMM approaches into a general framework. We label the previously mentioned approaches as an “estimating-testing” paradigm, since they all rely on some specific parameter estimator to handle the uncertainty of the parameter estimation. On the other hand, Wooldridge (1990) proposes a transformed statistic in conditional moment tests framework to remove the estimation effect. The martingale transformation is first introduced by Khmaladze (1981) as a method of removing the estimation effect of parametric empirical processes. Later on, Stute, Thies, and Zhu (1998) and Bai (2003) apply the idea of martingale transformation to parametric empirical processes for regression models and the conditional distributions of dynamic models respectively. Delgado and Velasco (2011) 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 but in a discrete context.

This paper also proposes removing the estimation effect by transforming the test statistic. By exploiting the nature of the conditional moment tests that the conditional moment restrictions not only imply the unconditional moment restrictions we are testing, but also many other unconditional moment restrictions, we develop a general approach for model specification analysis. This approach is robust to departures from the distributional assumptions that are not being tested, moreover only a preliminary $\sqrt{T}$-consistent estimator is needed, and the transformation is asymptotically distribution free. Our testing procedure includes Wooldridge (1990) modified statistic as a special case. 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 instruments 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.

The outline of the paper is the following. In Section 2, we propose the new test statistics. In Section 3, we study its properties. Section 4 discusses its relation with other tests. Section 5 discusses the efficient tests. Section 6 applies the new methodology to specification testing of GARCH models. Section 7 concludes.

1.2 Test Statistics

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. For time series applications, let $I_{t-1} = \{X_t, X_{t-1}, \cdots; Y_{t-1}, Y_{t-2}, \cdots\}$ represents the information set at time t. For cross section applications, we set $I_{t-1} = X_t$ and assume that the observations are independently distributed. In econometrics the interest lies in explaining Y_t in terms of the information set I_{t-1}. Frequently this implies that certain conditional moment restrictions are satisfied. That is, there is a $J \times 1$ vector of functions $\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta})$ defined on a parameter set $\boldsymbol{\Theta} \subset \mathbb{R}^P$ such that

$$E(\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}_0)|I_{t-1}) = 0, \text{ for some } \boldsymbol{\theta}_0 \in \boldsymbol{\Theta},\ t = 1, 2, \cdots. \tag{1.1}$$

The function $\phi_t(\cdot)$ can be derived from residuals of the structural models or other circumstances. Wooldridge (1990) calls $\phi_t(.)$ as a "generalized residual vector". For example, in dynamic model analysis, one typically considers parametric models such that $E(Y_t|I_{t-1}) = f_t(I_{t-1}, \boldsymbol{\theta}_0)$, for some $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, $t = 1, 2, \cdots$. By definition, we have $E(Y_t - f_t(I_{t-1}, \boldsymbol{\theta}_0)|I_{t-1}) = 0$, for some $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, $t = 1, 2, \cdots$, which means that $\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}_0) = Y_t - f_t(I_{t-1}, \boldsymbol{\theta}_0)$. In the framework of conditional maximum likelihood estimation, suppose that the parameter conditional likelihood function is $L(Y_t|I_{t-1}, \boldsymbol{\theta})$. Under some regularity conditions, by differentiating the identity $\int L(Y_t|I_{t-1}, \boldsymbol{\theta}_0)dy = 1$, we can obtain $E[\partial l(Y_t|I_{t-1}, \boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}|I_{t-1}] = 0$, $t = 1, 2, \cdots$, where $l(Y_t|I_{t-1}, \boldsymbol{\theta}_0) = \ln[L(Y_t|I_{t-1}, \boldsymbol{\theta}_0)]$. In this case $\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}_0) = \partial l(Y_t|I_{t-1}, \boldsymbol{\theta}_0)$, and $J = P$. For simplicity, we set $J = 1$, assuming that $\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta})$ is a scalar random function in this paper.[1]

The idea of the conditional moment tests is that, instead of testing conditional moment restriction (1.1) directly, its validity is tested by choosing some functions of the information set I_{t-1} and checking whether the sample covariance function between these functions and $\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta})$ are significantly different from zero. That is, we test the "directional" validity of the conditional moment restriction (1.1) by testing

$$H_0 : E(\Lambda_t(I_{t-1}, \boldsymbol{\theta}_0)\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}_0)) = 0, \text{ for some } \boldsymbol{\theta}_0 \in \boldsymbol{\Theta},\ t = 1, 2, \cdots, \tag{1.2}$$

[1] Although the setting of $J > 1$ is more general in some degree, it complicates the presentation of the idea.

where $\Lambda_t(I_{t-1}, \boldsymbol{\theta})$ is an $S \times 1$ vector function of I_{t-1} and $\boldsymbol{\theta}$, and the alternative is the negation of the null. Note that we allow the function $\Lambda_t(\cdot,\cdot)$ to depend on $\boldsymbol{\theta}$ in order to cover a wide range circumstances that are of interest to economists.[2] 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):

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

where $\hat{\rho}(j)$ is the sample autocorrelation function of $\{\hat{e}_t\}_{t=1}^{T}$, $\hat{e}_t = e(Y_t, I_{t-1}, \hat{\boldsymbol{\theta}}_{\boldsymbol{T}}) = Y_t - f(I_{t-1}, \hat{\boldsymbol{\theta}}_T)$, and $f(I_{t-1}, \hat{\boldsymbol{\theta}}_T)$ 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

$$\Lambda_t(I_{t-1}, \boldsymbol{\theta}) = (e(Y_{t-1}, I_{t-2}, \boldsymbol{\theta}), \cdots, e(Y_{t-S}, I_{t-S-1}, \boldsymbol{\theta}))',$$

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

For the conditional moment test (1.2), when $\boldsymbol{\theta}_0$ is known, the test statistic is based on the $S \times 1$ vector $\hat{\xi}_T(\boldsymbol{\theta}_0)$, where

$$\hat{\xi}_T(\boldsymbol{\theta}) = T^{-1/2}\sum_{t=1}^{T}\Lambda_t(I_{t-1}, \boldsymbol{\theta})\phi_t(I_{t-1}, \boldsymbol{\theta}).$$

In this case, some central limit theory usually could be applied directly, which makes it quite straightforward to obtain that a quadratic form of $\hat{\xi}_T(\boldsymbol{\theta}_0)$ follows an asymptotic chi-square distribution.

But typically $\boldsymbol{\theta}_0$ is unknown, and has to be estimated firstly. Assume that there is an estimator $\hat{\boldsymbol{\theta}}_T$ such that $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) = O_p(1)$, then a computable test statistic of the CM test is the $S \times 1$ vector $\hat{\xi}_T(\hat{\boldsymbol{\theta}}_T)$. Now it becomes more difficult to obtain the asymptotic properties of the quadratic form of $\hat{\xi}_T$, since there exists an "estimation effect", when

[2] Wooldridge (1990) considers a more complicated unconditional moment restrictions such that $E(\Lambda_t(I_{t-1}, \boldsymbol{\theta}_0, \pi_0)C_t(I_{t-1}, \boldsymbol{\theta}_0, \pi_0)\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}_0)) = 0$, $t = 1, 2, \cdots$, where π denotes an $N \times 1$ vector of nuisance parameters, and $C_t(I_{t-1}, \boldsymbol{\theta}, \pi)$ is an $S \times S$ symmetric and positive semidefinite weighting matrix. Since the estimator of nuisance parameters π does not affect the asymptotic theory of the test statistic, it is not worthwhile incorporating them in the function Λ_t. As for the weighting matrix, it is not necessary to seperate it from function Λ_t.

model parameters have to be estimated. To see this point, denote[3]

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

Under some regularity conditions, it could be shown that under H_0

$$\hat{\xi}_T(\hat{\boldsymbol{\theta}}_T) = \hat{\xi}_T(\boldsymbol{\theta}_0) + \Xi(\boldsymbol{\theta}_0)\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + o_p(1), \tag{1.3}$$

where $\Xi(\boldsymbol{\theta}_0) = p\lim_{T\to\infty}\hat{\Xi}_T(\boldsymbol{\theta}_0)$, $\hat{\Xi}_T(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T}\Lambda_t(I_{t-1},\boldsymbol{\theta})\Phi_t(I_{t-1},\boldsymbol{\theta})$. There are rare cases when $\Xi(\boldsymbol{\theta}_0) = 0$ holds. When $\Xi(\boldsymbol{\theta}_0) \neq 0$, the existence of the term $\Xi(\boldsymbol{\theta}_0)\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$, which is called "estimation effect", makes the asymptotic inference more complicated. In order to derive the covariance matrix correctly, the asymptotic joint distribution of $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$ and $\hat{\xi}_T(\boldsymbol{\theta}_0)$ has to be considered, which depends on the model and DGP characteristics, the method of estimating $\hat{\boldsymbol{\theta}}_T$ and even the unknown parameter $\boldsymbol{\theta}_0$.

Instead of following the "estimating-testing" paradigm, we transform $\hat{\xi}_T(\hat{\boldsymbol{\theta}}_T)$ so as to remove the estimation effect. The basic idea is that even though we are testing (1.2), it is the conditional moment restriction (1.1) that holds firstly. Under (1.1) there exist many other unconditional moment restrictions in addition to the "directional" unconditional moment restrictions (1.2). It is possible to find a vector of unconditional moment restrictions with $L \times 1$ dimension under (1.1) such that

$$E(\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta}_0)\phi_t(Y_t,I_{t-1},\boldsymbol{\theta}_0)) = 0, \text{ for some } \boldsymbol{\theta}_0 \in \boldsymbol{\Theta},\ t = 1,2,\cdots, \tag{1.4}$$

where $\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta})$ is $L \times 1$ vector function of I_{t-1} and $\boldsymbol{\theta}$, for $t = 1,2,\cdots$. 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 residuals of some ARMA model.

[3] Following convention, if function $a(\boldsymbol{\theta})$ is a $S \times 1$ function with $\boldsymbol{\theta} \in \mathbb{R}^P$, $\nabla_{\theta}\, a(\boldsymbol{\theta})$ denotes the $S \times P$ Jacobian matrix.

The sample analog of (1.4), when $\boldsymbol{\theta}_0$ is known, is then $\hat{\xi}_{1,T}(\boldsymbol{\theta}_0)$, where

$$\hat{\xi}_{1,T}(\boldsymbol{\theta}) = T^{-1/2}\sum_{t=1}^{T}\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta})\phi_t(Y_t,I_{t-1},\boldsymbol{\theta}).$$

The computable statistic, when we just have a $\sqrt{T}$-consistent estimator $\hat{\boldsymbol{\theta}}_T$, is $\hat{\xi}_{1,T}(\hat{\boldsymbol{\theta}}_T)$.

Note that under (1.1), $\hat{\xi}_{1,T}$ is also affected by estimation effect, i.e.,

$$\hat{\xi}_{1,T}(\hat{\boldsymbol{\theta}}_T) = \hat{\xi}_{1,T}(\boldsymbol{\theta}_0) + \Xi_1(\boldsymbol{\theta}_0)\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + o_p(1), \tag{1.5}$$

where $\Xi_1(\boldsymbol{\theta}_0) = p\lim_{T\to\infty}\hat{\Xi}_{1,T}(\boldsymbol{\theta}_0)$, $\hat{\Xi}_{1,T}(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T}\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta})\Phi_t(I_{t-1},\boldsymbol{\theta})$. The linear expansion (1.5) of extra statistic $\hat{\xi}_{1,T}(\hat{\boldsymbol{\theta}}_T)$ forms our bricks for the removal of the estimation effect of the statistic $\hat{\xi}_T(\hat{\boldsymbol{\theta}}_T)$.

Then the transformed statistic we propose is

$$\tilde{\xi}_T(\hat{\boldsymbol{\theta}}_T) = \hat{\xi}_T(\hat{\boldsymbol{\theta}}_T) - \hat{\Xi}_T(\hat{\boldsymbol{\theta}}_T)\left(\hat{\Xi}_{1,T}(\hat{\boldsymbol{\theta}}_T)'\hat{\Xi}_{1,T}(\hat{\boldsymbol{\theta}}_T)\right)^{-1}\hat{\Xi}_{1,T}(\hat{\boldsymbol{\theta}}_T)'\hat{\xi}_{1,T}(\hat{\boldsymbol{\theta}}_T). \tag{1.6}$$

This transformation could be regarded as a detrending operation on $\hat{\xi}_T(\hat{\boldsymbol{\theta}}_T)$. We first regress $\hat{\xi}_{1,T}(\hat{\boldsymbol{\theta}}_T)$ on $\hat{\Xi}_{1,T}(\hat{\boldsymbol{\theta}}_T)$, $\left(\hat{\Xi}_{1,T}(\hat{\boldsymbol{\theta}}_T)'\hat{\Xi}_{1,T}(\hat{\boldsymbol{\theta}}_T)\right)^{-1}\hat{\Xi}_{1,T}(\hat{\boldsymbol{\theta}}_T)'\hat{\xi}_{1,T}(\hat{\boldsymbol{\theta}}_T)$ is the least squares estimator, which requires $\Xi_1(\boldsymbol{\theta}_0)'\Xi_1(\boldsymbol{\theta}_0)$ to be of full rank. Notice that $\hat{\xi}_{1,T}(\hat{\boldsymbol{\theta}}_T)$ is a $L\times 1$ vector, $\hat{\Xi}_{1,T}(\hat{\boldsymbol{\theta}}_T)$ is a $L\times P$ matrix, so at least $L\geq P$ is required. Finally, multiplying this estimator by $\hat{\Xi}_T(\hat{\boldsymbol{\theta}}_T)$, we obtain the predicted value of $\hat{\xi}_T(\hat{\boldsymbol{\theta}}_T)$, then $\tilde{\xi}_T(\hat{\boldsymbol{\theta}}_T)$ is the residual.

It will be shown that under H_0

$$\tilde{\xi}_T(\hat{\boldsymbol{\theta}}_T) = \hat{\xi}_T(\boldsymbol{\theta}_0) - \Xi(\boldsymbol{\theta}_0)\left(\Xi_1(\boldsymbol{\theta}_0)'\Xi_1(\boldsymbol{\theta}_0)\right)^{-1}\Xi_1(\boldsymbol{\theta}_0)'\hat{\xi}_{1,T}(\boldsymbol{\theta}_0) + o_p(1), \tag{1.7}$$

i.e., $\tilde{\xi}_T(\hat{\boldsymbol{\theta}}_T)$ has not estimation effect: the asymptotic distribution of $\tilde{\xi}_T(\cdot)$, which is evaluated at any $\sqrt{T}$-consistent estimator $\hat{\boldsymbol{\theta}}_T$, is unchanged when the estimator is replaced by the true parameter for $\hat{\xi}_T(\cdot)$, $\hat{\xi}_{1,T}(\cdot)$, and both $\hat{\Xi}_T(\cdot)$ and $\hat{\Xi}_{1,T}(\cdot)$ are replaced by their probability limits. Note that the original statistic $\hat{\xi}_T(\hat{\boldsymbol{\theta}}_T)$ does not generally have this asymptotic property. Based on $\tilde{\xi}_T(\hat{\boldsymbol{\theta}}_T)$, we can take a quadric form as test statistic.

On the face of it, the transformation (1.6) could be seen as a discrete version of the martingale transformation of Khmaladze (1981), or the idea of "recursive residuals" used by Delgado and Velasco (2011), 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 (2011) 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 (2011), in their simplest case, exploit the asymptotic i.i.d property of the residual autocorrelation under the true parameters. Our framework instead assumes some generic central limit theorem holds, as Assumption 7 will show below. The price of that 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.

1.3 Asymptotic Properties

In the following we present the assumptions for our analysis. In order to incorporate circumstances as general as possible, similar to Wooldridge (1990), high-level assumptions are employed.

Assumption 1. $\boldsymbol{\Theta} \subset \mathbb{R}^P$ *is compact parameter space, and* $\boldsymbol{\theta_0} \in Int(\boldsymbol{\Theta})$. $E(\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}_0)|I_{t-1}) = 0$, *for some* $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, $t = 1, 2, \cdots$.

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

Assumption 2. *Both* $\{\Lambda_t(I_{t-1}, \boldsymbol{\theta})\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ *and* $\{\Lambda_{1,t}(I_{t-1}, \boldsymbol{\theta})\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$

are sequences of vector random functions such that $\Lambda_t(\cdot,\boldsymbol{\theta})\phi_t(\cdot,\cdot,\boldsymbol{\theta})$ *and* $\Lambda_{1,t}(\cdot,\boldsymbol{\theta})\phi_t(\cdot,\cdot,\boldsymbol{\theta})$ *are Borel measurable for each* $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, *and both* $\Lambda_t(I_{t-1},\cdot)\phi_t(Y_t,I_{t-1},\cdot)$ *and* $\Lambda_{1,t}(I_{t-1},\cdot)\phi_t(Y_t,I_{t-1},\cdot)$ *are continuously differentiable on the interior of* $\boldsymbol{\Theta}$ *for all* Y_t, $\sigma(I_{t-1})$, $t=1,2,\cdots$.

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

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\frac{1}{T}\sum_{t=1}^{T}\{\Lambda_t(I_{t-1},\boldsymbol{\theta})\phi_t(Y_t,I_{t-1},\boldsymbol{\theta})-E[\Lambda_t(I_{t-1},\boldsymbol{\theta})\phi_t(Y_t,I_{t-1},\boldsymbol{\theta})]\}\|\xrightarrow{p}0,$$

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\frac{1}{T}\sum_{t=1}^{T}\{\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta})\phi_t(Y_t,I_{t-1},\boldsymbol{\theta})-E[\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta})\phi_t(Y_t,I_{t-1},\boldsymbol{\theta})]\}\|\xrightarrow{p}0,$$

where $\|\cdot\|$ *denotes Euclidean norm.*

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

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\frac{1}{T}\sum_{t=1}^{T}\{\nabla_{\theta}[\Lambda_t(I_{t-1},\boldsymbol{\theta})\phi_t(Y_t,I_{t-1},\boldsymbol{\theta})]-E\nabla_{\boldsymbol{\theta}}[\Lambda_t(I_{t-1},\boldsymbol{\theta})\phi_t(Y_t,I_{t-1},\boldsymbol{\theta})]\}\|\xrightarrow{p}0,$$

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\frac{1}{T}\sum_{t=1}^{T}\{\nabla_{\theta}[\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta})\phi_t(Y_t,I_{t-1},\boldsymbol{\theta})]-E\nabla_{\boldsymbol{\theta}}[\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta})\phi_t(Y_t,I_{t-1},\boldsymbol{\theta})]\}\|\xrightarrow{p}0.$$

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. $\sqrt{T}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)=O_p(1)$.

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 the $L \times P$ matrix $\Xi_1(\boldsymbol{\theta}_0)$ is of full rank, and $L \geq P$.*

Assumption 6 levies some conditions on the choice of $\Lambda_{1,t}(\cdot)$ for $t = 1, 2, \cdots$, and the dimension L.

Assumption 7. *Define the $(S+L) \times 1$ vector*

$$\Pi_t(I_{t-1}, \boldsymbol{\theta}) = (\Lambda'_t(I_{t-1}, \boldsymbol{\theta}), \Lambda'_{1,t}(I_{t-1}, \boldsymbol{\theta}))', \tag{1.8}$$

and assume that under H_0

$$T^{-1/2} \sum_{t=1}^{T} \Pi_t(I_{t-1}, \boldsymbol{\theta}_0) \phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\Gamma}(\boldsymbol{\theta}_0)), \tag{1.9}$$

where $\boldsymbol{\Gamma}(\boldsymbol{\theta}) = AVar(T^{-1/2} \sum_{t=1}^{T} (\Pi_t(I_{t-1}, \boldsymbol{\theta}) \phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}))$, and $\boldsymbol{\Gamma}(\boldsymbol{\theta}_0) > 0$.

Note that $\sum_{t=1}^{T} \Pi_t(I_{t-1}, \boldsymbol{\theta}_0) \phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}_0)$ 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 central limit theorems in the case of ergodic stationary martingale difference process and α- or ϕ-mixing process.

In the following Theorem, we justify the asymptotic properties of the new test statistic $\tilde{\xi}_T(\hat{\boldsymbol{\theta}}_T)$ we proposed in (1.6).

Theorem 1. *When Assumptions 1-6 hold, under H_0, the asymptotic expansion (1.7) holds.*

Proof. See Appendix □

With Theorem 1, define

$$\underset{S\times(S+L)}{\hat{\boldsymbol{\Upsilon}}_T(\boldsymbol{\theta})} = \left[\mathbf{I}_s, -\hat{\Xi}_T(\boldsymbol{\theta}) \left(\hat{\Xi}_{1,T}(\boldsymbol{\theta})' \hat{\Xi}_{1,T}(\boldsymbol{\theta})\right)^{-1} \hat{\Xi}_{1,T}(\boldsymbol{\theta})'\right],$$

where $\mathbf{I}_s$ represents the $S \times S$ identity matrix, and $\hat{\Sigma}_T(\boldsymbol{\theta}) = \hat{\boldsymbol{\Upsilon}}_T(\boldsymbol{\theta}) \hat{\boldsymbol{\Gamma}}_T(\boldsymbol{\theta}) \hat{\boldsymbol{\Upsilon}}_T(\boldsymbol{\theta})'$, $\Sigma(\boldsymbol{\theta}) = \text{plim}_{T\to\infty} \hat{\Sigma}_T(\boldsymbol{\theta})$. Under some regularity conditions, $\hat{\Sigma}_T(\hat{\boldsymbol{\theta}}_T)$ is a positive definite matrix with probability 1, we propose a test based on a quadratic form of $\tilde{\xi}_T$ such that

$$\hat{N}_T(\hat{\boldsymbol{\theta}}_T) = \tilde{\xi}_T(\hat{\boldsymbol{\theta}}_T)'\hat{\Sigma}_T(\hat{\boldsymbol{\theta}}_T)^{-1}\tilde{\xi}_T(\hat{\boldsymbol{\theta}}_T).$$

The following Theorem establish its asymptotic distribution under the null, which is parameter free.

Theorem 2. *When Assumptions 1-7 hold, under H_0,*

$$\hat{N}_T(\hat{\boldsymbol{\theta}}_T) \xrightarrow{d} \chi^2(S), \tag{1.10}$$

where $\chi^2(S)$ represents Chi-square distribution with S degrees of freedom.

Proof. See Appendix ☐

1.3.1 Local Alternatives

We consider the following class of local alternative hypothesis H_{1T} : for some $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, as given in Assumption 5

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} E(\Lambda_t(I_{t-1},\boldsymbol{\theta}_0)\phi_t(Y_t,I_{t-1},\boldsymbol{\theta}_0)) &= \frac{\delta}{\sqrt{T}}(1+o(1)) \\ T^{-1}\sum_{t=1}^{T} E(\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta}_0)\phi_t(Y_t,I_{t-1},\boldsymbol{\theta}_0)) &= \frac{\delta_1}{\sqrt{T}}(1+o(1)), \end{aligned}$$

where δ and δ_1 are $S \times 1$ and $L \times 1$ nonrandom constant vectors, respectively. Now Assumption 7 is replaced by the following Assumption.

Assumption 8. *Assume that under H_{1T}*

$$T^{-1/2}\sum_{t=1}^{T} \Pi_t(I_{t-1},\boldsymbol{\theta}_0)\phi_t(Y_t,I_{t-1},\boldsymbol{\theta}_0) \xrightarrow{d} N(\gamma,\boldsymbol{\Gamma}(\boldsymbol{\theta}_0)),$$

where $\gamma = (\delta',\delta_1')'$.

The following theorem provides the limiting distribution of $\hat{N}_T(\hat{\boldsymbol{\theta}}_T)$ under the local alternative.

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

$$\hat{N}_T(\hat{\boldsymbol{\theta}}_T) \xrightarrow{d} \chi^2(S, \kappa'\Sigma(\boldsymbol{\theta}_0)^{-1}\kappa),$$

where $\kappa = \delta - \Xi(\boldsymbol{\theta}_0)\left(\Xi_1(\boldsymbol{\theta}_0)'\Xi_1(\boldsymbol{\theta}_0)\right)^{-1}\Xi_1(\boldsymbol{\theta}_0)'\delta_1$.

Proof. See Appendix. □

This theorem shows that the test has power against some alternatives when $\kappa \neq 0$.

1.4 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. Remember that $\Phi_t(I_{t-1}, \boldsymbol{\theta}) = E[\nabla_{\boldsymbol{\theta}}\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta})|I_{t-1}]$. The modified statistic of Wooldridge (1990) is

$$\bar{\xi}_T\left(\hat{\boldsymbol{\theta}}_T\right) = T^{-1/2}\sum_{t=1}^{T}(\Lambda_t(I_{t-1}, \hat{\boldsymbol{\theta}}_T) - (\Phi_t(I_{t-1}, \hat{\boldsymbol{\theta}}_T)\hat{B}_T\left(\hat{\boldsymbol{\theta}}_T\right))')\phi_t(Y_t, I_{t-1}, \hat{\boldsymbol{\theta}}_T),$$

where

$$\hat{B}_T\left(\hat{\boldsymbol{\theta}}_T\right) = \left(\sum_{t=1}^{T}\Phi_t(I_{t-1}, \hat{\boldsymbol{\theta}}_T)'\Phi_t(I_{t-1}, \hat{\boldsymbol{\theta}}_T)\right)^{-1}\sum_{t=1}^{T}\Phi_t(I_{t-1}, \hat{\boldsymbol{\theta}}_T)'\Lambda_t(I_{t-1}, \hat{\boldsymbol{\theta}}_T)'.$$

Denoting $\hat{\Xi}_{W,T}\left(\hat{\boldsymbol{\theta}}_T\right) = \frac{1}{T}\sum_{t=1}^{T}\Phi_t(I_{t-1}, \hat{\boldsymbol{\theta}}_T)'\Phi_t(I_{t-1}, \hat{\boldsymbol{\theta}}_T)$, we rewrite Wooldridge's transformation as

$$\begin{aligned}
\bar{\xi}_T\left(\hat{\boldsymbol{\theta}}_T\right) &= \hat{\xi}_T\left(\hat{\boldsymbol{\theta}}_T\right) - \hat{\Xi}_T(\hat{\boldsymbol{\theta}}_T)\left(\hat{\Xi}_{W,T}(\hat{\boldsymbol{\theta}}_T)\right)^{-1} T^{-1/2}\sum_{t=1}^{T}\Phi_t(I_{t-1}, \hat{\boldsymbol{\theta}}_T)'\phi_t(Y_t, I_{t-1}, \hat{\boldsymbol{\theta}}_T) \\
&= \hat{\xi}_T\left(\hat{\boldsymbol{\theta}}_T\right) - \hat{\Xi}_T\left(\hat{\boldsymbol{\theta}}_T\right)\left(\hat{\Xi}_{W,T}(\hat{\boldsymbol{\theta}}_T)'\hat{\Xi}_{W,T}(\hat{\boldsymbol{\theta}}_T)\right)^{-1}\hat{\Xi}_{W,T}(\hat{\boldsymbol{\theta}}_T)'T^{-1/2}\sum_{t=1}^{T}\Phi_t(I_{t-1}, \hat{\boldsymbol{\theta}}_T)'\phi_t(Y_t, I_{t-1}, \hat{\boldsymbol{\theta}}_T).
\end{aligned}$$

Note that the second equation holds because of the symmetry of $\hat{\Xi}_{W,T}(\hat{\boldsymbol{\theta}}_T)$. So Wooldridge (1990)'s modified statistic $\bar{\xi}_T\left(\hat{\boldsymbol{\theta}}_T\right)$ turns out to be equivalent to $\tilde{\xi}_T\left(\hat{\boldsymbol{\theta}}_T\right)$, where $\hat{\xi}_{1,T}\left(\hat{\boldsymbol{\theta}}_T\right) = T^{-1/2}\sum_{t=1}^{T}\Phi_t(I_{t-1}, \hat{\boldsymbol{\theta}}_T)'\phi_t(Y_t, I_{t-1}, \hat{\boldsymbol{\theta}}_T)$. While Wooldridge (1990) claims its robustness and the 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 turns out that $\tilde{\xi}_T\left(\hat{\boldsymbol{\theta}}_T\right)$ is equivalent asymptotically to an untransformed statistic $\hat{\xi}_T(\cdot)$ when is evaluated at a particular estimator. The following theorem establishes the relation between $\tilde{\xi}_T\left(\hat{\boldsymbol{\theta}}_T\right)$ and $\hat{\xi}_T(\cdot)$.

Theorem 4. *When Assumption 1-6 hold, under H_0, $\tilde{\xi}_T\left(\hat{\boldsymbol{\theta}}_T\right)$ is equivalent to $\hat{\xi}_T\left(\check{\boldsymbol{\theta}}_T\right)$ asymptotically, where*

$$\check{\boldsymbol{\theta}}_T = \underset{\boldsymbol{\theta}}{argmin}\frac{1}{T}\hat{\xi}_{1,T}(\boldsymbol{\theta})'\hat{\xi}_{1,T}(\boldsymbol{\theta}).$$

Note that $\frac{1}{T}\hat{\xi}_{1,T}(\boldsymbol{\theta})'\hat{\xi}_{1,T}(\boldsymbol{\theta})$ is an GMM estimation objective function with identity weighting matrix I_L. This means that our new statistic is asymptotically equivalent to the testing statistic which is evaluated at the estimator that is obtained from using the unconditional moment restrictions $E(\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta})\phi_t(Y_t,I_{t-1},\boldsymbol{\theta}))$.

1.5 Efficient 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, which is implicit in our transformation by Theorem 4. As we know, in the 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.

Denote

$$\boldsymbol{\Gamma}^{11}(\boldsymbol{\theta}) = AVar\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta})\phi_t(Y_{t-1},I_{t-1},\boldsymbol{\theta})\right]. \tag{1.11}$$

By considering the optimal weighting matrix $\hat{\Gamma}_T^{11}\left(\hat{\boldsymbol{\theta}}_T\right)$, our projected statistic becomes

$$\tilde{\xi}_T\left(\hat{\boldsymbol{\theta}}_T\right) = \hat{\xi}_T\left(\hat{\boldsymbol{\theta}}_T\right) - \hat{\Xi}_T\left(\hat{\boldsymbol{\theta}}_T\right)\left(\hat{\Xi}_{1,T}\left(\hat{\boldsymbol{\theta}}_T\right)'\hat{\Gamma}_T^{11}\left(\hat{\boldsymbol{\theta}}_T\right)^{-1}\hat{\Xi}_{1,T}\left(\hat{\boldsymbol{\theta}}_T\right)\right)^{-1}\hat{\Xi}_{1,T}\left(\hat{\boldsymbol{\theta}}_T\right)'\hat{\Gamma}_T^{11}\left(\hat{\boldsymbol{\theta}}_T\right)^{-1}\hat{\xi}_{1,T}\left(\hat{\boldsymbol{\theta}}_T\right).$$

Note firstly that $\hat{\Gamma}_T^{11}\left(\hat{\boldsymbol{\theta}}_T\right)$ is a postive definite matrix with probability 1, given Assumption 7. Secondly even though a weighting function is introduced, $\tilde{\xi}_T\left(\hat{\boldsymbol{\theta}}_T\right)$ still does not depend on $\hat{\boldsymbol{\theta}}_T$ asymptotically, so in this case, similar results to both Theorem 2 and 3 all hold. Also notice that now $\hat{\boldsymbol{\Upsilon}}_T(\boldsymbol{\theta})$ becomes

$$\underset{S\times(S+L)}{\hat{\boldsymbol{\Upsilon}}_T(\boldsymbol{\theta})} = \left[\mathbf{I}_s, -\hat{\Xi}_T(\boldsymbol{\theta})\left(\hat{\Xi}_{1,T}(\boldsymbol{\theta})'\hat{\Gamma}_T^{11}(\boldsymbol{\theta})^{-1}\hat{\Xi}_{1,T}(\boldsymbol{\theta})\right)^{-1}\hat{\Xi}_{1,T}(\boldsymbol{\theta})'\hat{\Gamma}_T^{11}(\boldsymbol{\theta})^{-1}\right].$$

It is straightforward to show that $\tilde{\xi}_T\left(\hat{\boldsymbol{\theta}}_T\right)$ is asymptotically equivalent to $\hat{\xi}_T\left(\check{\boldsymbol{\theta}}_T\right)$, where $\check{\boldsymbol{\theta}}_T$ satisfies

$$\check{\boldsymbol{\theta}}_T = \underset{\boldsymbol{\theta}}{argmin}\frac{1}{T}\hat{\xi}_{1,T}(\boldsymbol{\theta})'\hat{\Gamma}_T^{11}\left(\bar{\boldsymbol{\theta}}_T\right)^{-1}\hat{\xi}_{1,T}(\boldsymbol{\theta}),$$

where $\bar{\boldsymbol{\theta}}_T$ is a $\sqrt{T}$-consistent estimator. This means that $\tilde{\xi}_T(\cdot)$ evaluated at any $\sqrt{n}$ consistent estimator is equivalent to $\hat{\xi}_T(\cdot)$ evaluated at the two-step GMM estimator based on moment conditions $E\left[\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta}_0)\phi_t(Y_{t-1},I_{t-1},\boldsymbol{\theta}_0)\right] = 0$. In this sense, our approach could be regarded as a one-step procedure in testing scenario.

This argument sheds some light on the limitations of the Wooldridge (1990)'s modified statistic. 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 $\Phi_t(I_{t-1},\boldsymbol{\theta})'$ of the conditional moment restrictions). The possible choices of $\Lambda_t(I_{t-1},\boldsymbol{\theta})$ will be restrained by the form of $\Phi_t(I_{t-1},\boldsymbol{\theta})'$. Furthermore, as an instrument, $\Phi_t(I_{t-1},\boldsymbol{\theta})'$ is not necessarily the optimal one in most cases. In our framework, it is convenient to incorporate $\Phi_t(I_{t-1},\boldsymbol{\theta})'$ into $\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta})$, which could bring about potentially more powerful test than Wooldridge (1990)'s modified statistic from using more efficient estimate by the interpretation of Theorem 4.

Delgado and Velasco (2011) consider general cases that residuals of time series models exhibit higher-order serial dependence. They introduce some proper weight matrix to standardize residuals sample autocorrelations, then employ the idea of recursive residuals to remove the estimation effect in the standardized residuals sample autocorrelation. However the weighting matrix we introduced in this paper has the interpretation of optimal weighting matrix of GMM estimation.

Our general approach deviates from the Estimating-testing paradigm of Newey (1985a,b), Tauchen (1985) and White (1994). Newey (1985b) considers specification testing in the framework of GMM. In the notation of this paper, he wants to test the validity of some linear combination of the moment restrictions $E[(\Lambda_t(I_{t-1},\boldsymbol{\theta}_0)',\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta}_0)')'\phi_t(Y_{t-1},I_{t-1},\boldsymbol{\theta}_0)] = 0$, and the estimator of $\boldsymbol{\theta}_0$ is based on GMM estimation employing all the given moment restrictions under some weighting function. But without any specification of the form of the linear combination and further assumptions, the test statistic has to use a generalized inverse to standardize the sample moment conditions evaluated at the parameter estimators. Moreover, Newey (1985b) derives the optimal GMM test in the case that the distribution information is assumed to be known up to some unknown parameters, and the optimality is only about testing for this particular linear combination of the unconditional moment restrictions. On the other hand our framework assumes conditional moment restrictions hold, and we explicitly focus on testing the validity of moment conditions $E[(\Lambda_t(I_{t-1},\boldsymbol{\theta}_0)\phi_t(Y_{t-1},I_{t-1},\boldsymbol{\theta}_0)] = 0$. Therefore it is impossible to obtain the same optimality results as Newey (1985b) in general. However, in some special cases, we can establish the asymptotically equivalence between our new framework and Newey (1985b)'s optimal test. Generally, the optimal test in the sense of Newey (1985b) has the interpretation of some form of a score test. To derive the optimal test, knowledge of data generating process is required. However in the following alternative, the optimal tests can be calculated without further knowledge of the data generating process. More specifically, Suppose

$$T^{-1}\sum_{t=1}^{T} E[(\Lambda_t(I_{t-1},\boldsymbol{\theta}_0)',\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta}_0)')'\phi_t(Y_{t-1},I_{t-1},\boldsymbol{\theta}_0)] = \left(\frac{\delta'}{\sqrt{T}}(1+o(1)),0\right)', \qquad (1.12)$$

which means that under this misspecification $E[\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta}_0)'\phi_t(Y_{t-1},I_{t-1},\boldsymbol{\theta}_0)] = 0$, for $t = 1,\cdots T$, remains satisfied.

Separate $\boldsymbol{\Gamma}(\theta)$, which is defined in Assumption 7, into

$$\begin{pmatrix} \boldsymbol{\Gamma}^{00}(\theta) & \boldsymbol{\Gamma}^{01}(\theta) \\ \boldsymbol{\Gamma}^{10}(\theta) & \boldsymbol{\Gamma}^{11}(\theta) \end{pmatrix},$$

where $\boldsymbol{\Gamma}^{00}(\theta)$ is an $S\times S$ matrix, and $\boldsymbol{\Gamma}^{11}(\theta)$ is the $L\times L$ matrix defined in (1.11). Define $\Lambda_t^*(I_{t-1},\boldsymbol{\theta}) = \Lambda_t(I_{t-1},\boldsymbol{\theta}) - \hat{\boldsymbol{\Gamma}}_T^{01}(\theta)\hat{\boldsymbol{\Gamma}}_T^{11}(\theta)^{-1}\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta})$. In Newey (1985b) the optimal test is to test moment restrictions $E[\Lambda_t^*(I_{t-1},\boldsymbol{\theta})\phi_t(Y_t,I_{t-1},\boldsymbol{\theta})] = 0$, for $t = 1,2,\cdots$, and the optimal GMM estimator comes from the GMM estimation based on moment restrictions $E[\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta}_0)\phi_t(Y_{t-1},I_{t-1},\boldsymbol{\theta}_0)] = 0$, for $t = 1,2,\cdots$. So based on Theorem 4 it could be shown that one version of our test statistic given by

$$\tilde{\xi}_T^*(\hat{\boldsymbol{\theta}}_T) = \hat{\xi}_T^*\left(\hat{\boldsymbol{\theta}}_T\right) - \hat{\Xi}_T^*(\hat{\boldsymbol{\theta}}_T)\left(\hat{\Xi}_{1,T}\left(\hat{\boldsymbol{\theta}}_T\right)'\hat{\Gamma}_T^{11}\left(\hat{\boldsymbol{\theta}}_T\right)^{-1}\hat{\Xi}_{1,T}\left(\hat{\boldsymbol{\theta}}_T\right)\right)^{-1}\hat{\Xi}_{1,T}\left(\hat{\boldsymbol{\theta}}_T\right)'\hat{\Gamma}_T^{11}\left(\hat{\boldsymbol{\theta}}_T\right)^{-1}\hat{\xi}_{1,T}\left(\hat{\boldsymbol{\theta}}_T\right)$$

is asymptotically equivalent to Newey (1985b)'s optimal GMM test (Proposition 3) under the alternative (1.12) , where $\hat{\xi}_T^*(\boldsymbol{\theta}) = T^{-1/2}\sum_{t=1}^T \Lambda_t^*(I_{t-1},\boldsymbol{\theta})\phi_t(Y_t,I_{t-1},\boldsymbol{\theta})$ and $\hat{\Xi}_T^*(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^T \Lambda_t^*(I_{t-1},\boldsymbol{\theta})\Phi_t(I_{t-1},\boldsymbol{\theta})$, in the sense that the asymptotic variance is the same and the drift of the chi-square statistics is the same. Note that it is easy to obtain the asymptotic variance of $\tilde{\xi}_T^*(\hat{\boldsymbol{\theta}}_T)$ is

$$\Sigma^*(\boldsymbol{\theta}_0) = \boldsymbol{\Gamma}^{00}(\boldsymbol{\theta}_0) - \boldsymbol{\Gamma}^{01}(\boldsymbol{\theta}_0)\boldsymbol{\Gamma}^{11}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Gamma}^{10}(\boldsymbol{\theta}_0) + \Xi^*(\boldsymbol{\theta}_0)\left(\Xi_1(\boldsymbol{\theta}_0)'\Gamma^{11}(\boldsymbol{\theta}_0)^{-1}\Xi_1(\boldsymbol{\theta}_0)\right)^{-1}\Xi^*(\boldsymbol{\theta}_0)',$$

where $\Xi^*(\boldsymbol{\theta}) = p\lim_{T\to\infty}\hat{\Xi}_T^*(\boldsymbol{\theta})$, which is the same as (31b) in Newey (1985b).

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

Example 1. *Consider the conditional moment restrictions*

$$E(\phi(z,\boldsymbol{\theta}_0)|x) = 0, \textit{ for some } \boldsymbol{\theta}_0 \in \boldsymbol{\Theta},\ t = 1,2,\cdots, \tag{1.13}$$

where z denotes a single observation, $\boldsymbol{\theta}$ a $p\times 1$ vector of parameter, $\phi(z,\boldsymbol{\theta})$ 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 γ_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 $\Lambda_1(x) = q^K(x)$ to get optimal testing statistic under some conditions, since the estimator reaches the semiparametric efficiency bound, for more details see Donald et al (2003).

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

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

where $X_t = (y_{t-1}, \cdots, y_{t-p})'$, and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)'$ satisfy the condition that the roots of the associated lag polynomial lie outside of the unit circle; and ϵ_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 using the notation of our testing framework, we have $\phi_t = \epsilon_t$, $\Phi_t = X_t'$.

If we assume that conditional homoskedasticity $E(\epsilon_t^2 | I_{t-1}) = E(\epsilon_t^2)$ holds, the LM test is optimal, since the OLS or QMLE estimators reach efficiency. But the assumption of conditional homoskedasticity is too strong, and when it does not hold, Then the OLS or QMLE estimators are not optimal. There exists more efficient or most efficient GMM estimators. In this case, if we choose instrument variable $\Lambda_1 = (y_{t-1}, \cdots, y_{t-L})'$, where $L > p$, we will obtain a 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

$$\hat{\Xi}_{1,T}(\boldsymbol{\theta})' \hat{\Gamma}_T^{11}(\boldsymbol{\theta})^{-1} \hat{\xi}_{1,T}(\boldsymbol{\theta}) = T^{-1/2} \sum_{j=1}^{\infty} (E\left(\epsilon_{t-j} X_t\right) / E\left(\epsilon_t^2 \epsilon_{t-j}^2\right)) \epsilon_{t-j} \epsilon_t$$

to get the optimal test, since $\sum_{j=1}^{\infty} \frac{E\left(\epsilon_{t-j} X_t\right)}{E\left(\epsilon_t^2 \epsilon_{t-j}^2\right)} \epsilon_{t-j}$ is the optimal instrument in this case (See West (2002)).

Examples 1 and 2 demonstrate the potential advantages of our general framework:

robustness and efficiency are obtained without paying too much price. 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 actual optimal estimator is used.

1.6 Application to Testing of GARCH Models

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

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

$$Y_t = \varepsilon_t h_t^{1/2},\, h_t = \omega_0 + \sum_{i=1}^{p} \alpha_{0i} Y_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} h_{t-j},\, t \in \mathbb{Z},\, p, q \in \mathbb{N}.$$

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

$$E(\varepsilon_t(\boldsymbol{\theta}_0)) = 0,\, E(\varepsilon_t^2(\boldsymbol{\theta}_0)|I_{t-1}) = 1, \text{ a.s},$$

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 $\boldsymbol{\theta} = (\omega, \alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q)'$, and the true parameter is denoted by $\boldsymbol{\theta}_0 = (\omega_0, \alpha_{01}, \cdots, \alpha_{0p}, \beta_{01}, \cdots, \beta_{0q})'$.

Assumption A.1. *ε_t, $t \in \mathbb{Z}$ is a strictly stationary and ergodic process satisfying $E(\varepsilon_t^2|I_{t-1}) = 1$, a.s, and ε_t^2 has a nondegenerated distribution. Y_t, $t \in \mathbb{Z}$ is a strictly stationary and ergodic process with $E[|Y_t|^{2s}] < \infty$, for some $s > 0$.*

Assumption A.2. *$\boldsymbol{\theta}_0 \in int(\boldsymbol{\Theta})$, where $int(\boldsymbol{\Theta})$ denotes the interior of $\boldsymbol{\Theta}$, and $\boldsymbol{\Theta}$ is compact.*

Assumption A.3. *if $q > 0$, $A_{\theta_0}(z)$ and $B_{\theta_0}(z)$ have no common root. For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $B_\theta(z)$ has its roots outside the unit circle. Moreover, $A_{\theta_0}(1) \neq 0$ and $\alpha_{0p} + \beta_{0q} \neq 0$, where $A_{\theta_0}(z) = \sum_{i=1}^{p} \alpha_{0i} z^i$ and $B_{\theta_0}(z) = 1 - \sum_{i=1}^{q} \beta_{0i} z^i$, $z \in \mathbb{C}$. By convention, $A_{\theta_0}(z) \equiv 1$ if $p = 0$ and $B_{\theta_0}(z) \equiv 1$ if $q = 0$.*

Assumption A.4. *$E|\varepsilon_t|^{4(1+\delta)} < 0$ 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 errors. 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:

$$E\left(\frac{Y_t^2}{h_t(I_{t-1},\boldsymbol{\theta}_0)}-1|I_{t-1}\right)=0,\ \boldsymbol{\theta}_0\in\boldsymbol{\Theta}.$$

The CM tests is to test the implication such that

$$H_0:E\left[\Lambda_t(I_{t-1},\boldsymbol{\theta}_0)\left(\frac{Y_t^2}{h_t(I_{t-1},\boldsymbol{\theta}_0)}-1\right)\right]=0,\text{ for some }\boldsymbol{\theta}_0\in\boldsymbol{\Theta}.$$

Suppose that we could find other unconditional moment restrictions such that

$$E\left[\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta}_0)\left(\frac{Y_t^2}{h_t(I_{t-1},\boldsymbol{\theta}_0)}-1\right)\right]=0,\text{for some }\boldsymbol{\theta}_0\in\boldsymbol{\Theta}.$$

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 choose $\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta})=(\frac{Y_{t-1}^2}{h_{t-1}(I_{t-1},\boldsymbol{\theta})},\cdots,\frac{Y_{t-L}^2}{h_{t-L}(I_{t-1},\boldsymbol{\theta})})'$. 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 (1.6) with $\phi_t((Y_t,I_{t-1},\boldsymbol{\theta}))=\frac{Y_t^2}{h_t(I_{t-1},\boldsymbol{\theta})}-1$. Note that

$$\begin{aligned}\Xi(\boldsymbol{\theta}_0)&=\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}E\left[\Lambda_t(I_{t-1},\boldsymbol{\theta}_0)\nabla_\theta\left(\frac{Y_t^2}{h_t(I_{t-1},\boldsymbol{\theta}_0)}-1\right)\right]\\&=E\left[-\frac{\Lambda_t(I_{t-1},\boldsymbol{\theta}_0)}{h_t(I_{t-1},\boldsymbol{\theta}_0)}\nabla_\theta h_t(I_{t-1},\boldsymbol{\theta}_0)\right]\end{aligned}$$

$$
\begin{aligned}
\Xi_1(\boldsymbol{\theta}_0) &= \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\left[\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta}_0)\nabla_\theta\left(\frac{Y_t^2}{h_t(I_{t-1},\boldsymbol{\theta}_0)}-1\right)\right] \\
&= E\left[-\frac{\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta}_0)}{h_t(I_{t-1},\boldsymbol{\theta}_0)}\nabla_\theta h_t(I_{t-1},\boldsymbol{\theta}_0)\right]
\end{aligned}
$$

When $\beta_{0j} \neq 0$ for $j = 1,\cdots q$, $\nabla_\theta h_t$ has recursive characteristics,

$$
\nabla_\theta h_t(\boldsymbol{\theta}_0) = \mathbf{s}_{t-1}(\boldsymbol{\theta}_0)' + \sum_{j=1}^{q} \beta_{0j}\nabla_\theta h_{t-j}(\boldsymbol{\theta}_0),
$$

where $\mathbf{s}_{t-1}(\boldsymbol{\theta}) = (1, Y_{t-1},\cdots,Y_{t-p}, h_{t-1}(\boldsymbol{\theta}),\cdots,h_{t-q}(\boldsymbol{\theta}))'$.

1.6.2 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. $g_t = 1 + \boldsymbol{\pi}'\mathbf{v}_t$, where $\mathbf{v}_t = (\varepsilon_{t-1}^2,\cdots,\varepsilon_{t-S}^2)'$ and $\boldsymbol{\pi} = (\pi_1,\cdots,\pi_s)$, $\pi_j \geq 0$, $j = 1,\cdots,S$. Lundbergh and Teräsvirta (2002) call the alternative model as "ARCH nested in GARCH" model. They test $H_0' : \boldsymbol{\pi} = \mathbf{0}$ against $\boldsymbol{\pi} \neq \mathbf{0}$, although the elements of $\boldsymbol{\pi}$ are constrained to be non-negative. The statistic used in Lundbergh and Teräsvirta (2002) has the form

$$
\hat{\xi}_T(I_{t-1},\boldsymbol{\theta}) = T^{-1/2}\sum_{t=1}^{T}\Lambda_t(I_{t-1},\boldsymbol{\theta})\left(\frac{Y_t^2}{h_t(I_{t-1},\boldsymbol{\theta})}-1\right),
$$

where $\Lambda_t(I_{t-1},\boldsymbol{\theta}) = (\varepsilon_{t-1}^2(I_{t-2},\boldsymbol{\theta}),\cdots,\varepsilon_{t-S}^2(I_{t-S-1},\boldsymbol{\theta}))'$.

Li and Mak (1994) introduce a portmanteau statistic for testing the adequacy of the standard 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 $\Lambda_t(I_{t-1},\boldsymbol{\theta}) = (\varepsilon_{t-1}^2(I_{t-2},\boldsymbol{\theta})-1,\cdots,\varepsilon_{t-S}^2(I_{t-S-1},\boldsymbol{\theta})-1)'$. 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 $\Lambda_t(I_{t-1},\boldsymbol{\theta}) = (\varepsilon_{t-1}^2(I_{t-2},\boldsymbol{\theta}),\cdots,\varepsilon_{t-S}^2(I_{t-S-1},\boldsymbol{\theta}))'$. Note that in this case, we have to choose $\Lambda_{1,t}(I_{t-1},\boldsymbol{\theta}) = (\varepsilon_{t-S-2}^2(I_{t-S-3},\boldsymbol{\theta}),\cdots,\varepsilon_{t-S}^2(I_{t-S-L-1},\boldsymbol{\theta}))'$.

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

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

where $\varepsilon_t \sim N(0,1)$, or $\varepsilon_t \sim t(d)$ (standardized Student t-distribution with degree of freedom d). We choose $\alpha_0 = 0.1, \alpha_1 = 0.8$. 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 1.1, where L is fixed at $L = 20$.

Figure 1.6.2 and 1.6.2 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.

$$h_t = 0.2 + 0.2Y_{t-1}^2 + 0.2Y_{t-2}^2.$$

We report the results in Figure 1.6.2 and Table 1.2. Firstly, note that Wooldridge (1990)'s

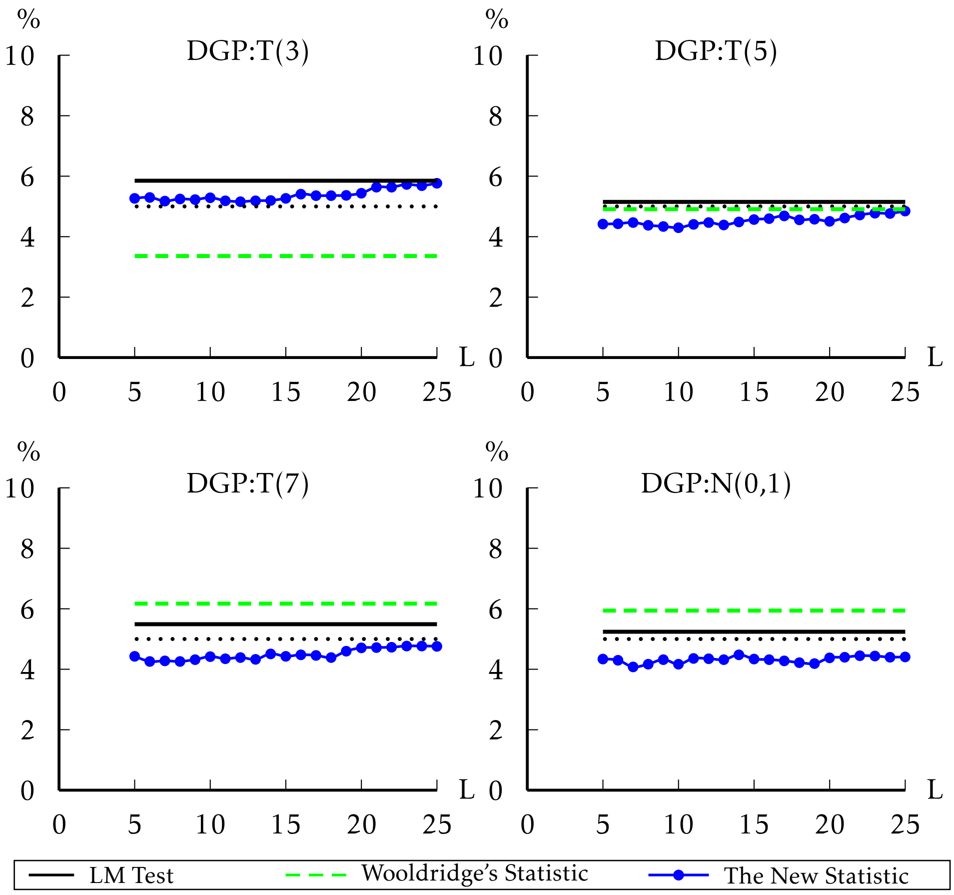


Figure 1.1: Results of empirical size of testing no remaining ARCH effect, nominal size 5%, Sample size T=250, $S = 4$.

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.

1.6.3 Testing for non-linearity

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

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

Table 1.1: Empirical size, Choosing $L = 20$.

		T=250				T=500			
	S	1	3	5	7	1	3	5	7
T_3	N_T	3.7	4.9	5.8	6.6	3.5	4.3	5.4	6.3
	W	2.0	3.9	2.7	1.7	2.8	5.8	5.3	4.3
	LM	3.5	5.3	6.2	6.5	3.3	4.9	5.8	6.7
T_5	N_T	4.1	4.6	5.0	5.4	3.3	3.9	5.0	5.7
	W	3.5	5.3	4.5	3.2	2.4	7.7	8.1	7.5
	LM	4.3	4.9	5.2	5.3	3.7	5.2	5.8	6.3
T_7	N_T	4.1	4.6	4.7	5.0	4.5	5.1	5.3	5.5
	W	4.1	6.1	5.3	3.8	5.0	8.0	7.9	7.3
	LM	4.4	5.4	5.4	5.4	4.4	5.5	5.9	6.1
$N(0,1)$	N_T	4.8	4.3	4.3	4.7	4.9	4.7	4.8	4.7
	W	4.5	5.0	5.7	5.1	5.2	6.4	6.8	6.4
	LM	4.9	4.9	5.2	5.4	4.7	4.9	5.0	5.2

Empirical size of the 10000 Monte-Carlo experiments, Choosing $L = 20$. H_0 :ARCH(1). $\hat{N}_T$ represents the new statistic, W, Wooldridge's modified statistic, LM, Lagrange Multiplier statistic.

Table 1.2: Empirical power, Choosing $L = 20$.

		T=500						
	S	1	2	3	4	5	6	7
T_3	$\hat{N}_T$	8.6	35.1	33.0	31.5	30.4	28.9	27.8
	W	2.9	5.6	6.3	6.0	6.6	5.5	5.4
	LM	4.4	34.5	32.0	31.6	30.4	28.6	28.5
T_5	$\hat{N}_T$	11.5	59.0	57.2	54.0	51.9	51.1	49.1
	W	10.5	19.9	14.2	11.2	11.9	12.6	11.2
	LM	5.0	57.7	58.1	54.2	51.8	49.9	48.5
T_7	$\hat{N}_T$	10.5	70.2	67.0	64.5	62.8	59.4	57.1
	W	16.0	31.0	22.5	17.6	14.4	13.7	12.0
	LM	9.4	69.5	67.0	64.7	61.6	59.4	57.2
$N(0,1)$	$\hat{N}_T$	11.0	83.7	81.0	76.9	75.0	73.2	70.8
	W	23.5	56.6	44.8	36.6	31.4	27.8	25.3
	LM	23.0	83.6	82.0	78.3	76.0	73.6	72.6

Empirical power of the 1000 Monte-Carlo experiments, Choosing $L = 20$. H_1 :ARCH(2). $\hat{N}_T$ represents the new statistic, W, Wooldridge's modified statistic, LM, Lagrange Multiplier statistic.

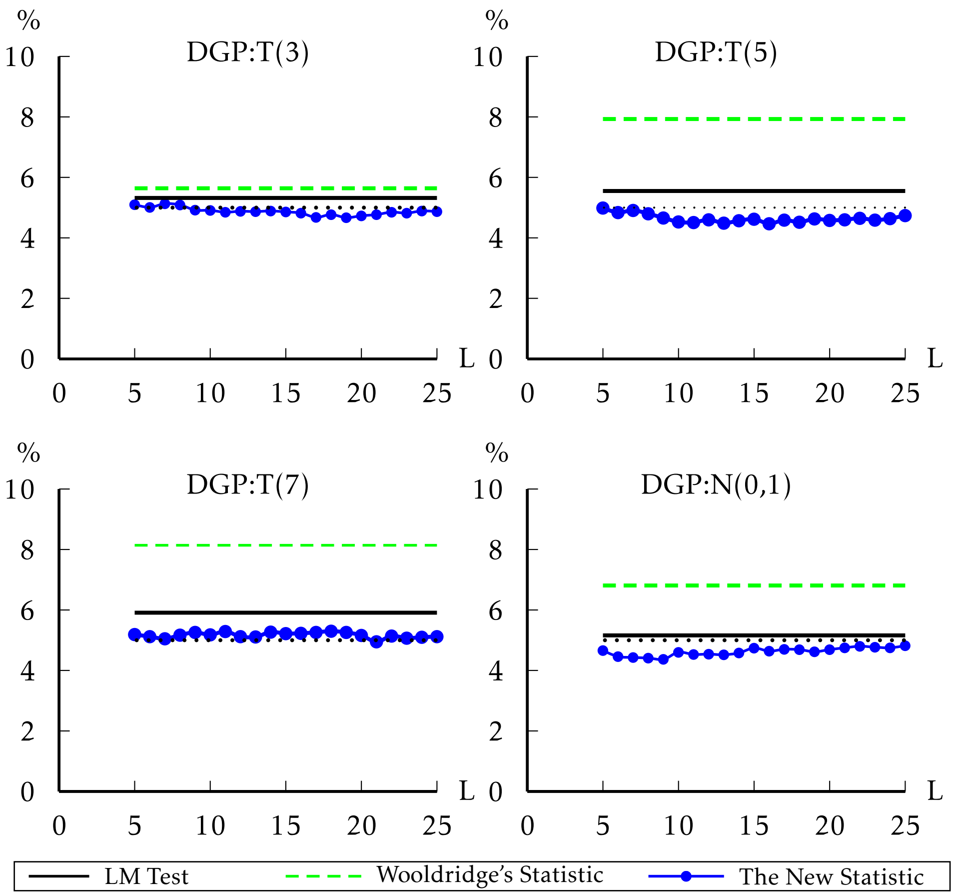


Figure 1.2: Results of empirical size of testing no remaining ARCH effect, nominal size 5%, Sample size T=500, $S = 4$.

where

$$g_t = \sum_{j=1}^{q} \alpha_{0j} \overline{H}_n(\epsilon_{t-j};\gamma,\mathbf{c}) + \sum_{j=1}^{q} \alpha_{1j} \overline{H}_n(\epsilon_{t-j};\gamma,\mathbf{c})\varepsilon_{t-j}^2,$$

in which they consider a smooth transition alternative

$$\overline{H}_n(\epsilon_{t-j};\gamma,\mathbf{c}) = (1 + exp(-\gamma \prod_{l=1}^{n}(x_t - c_l)))^{-1}, \gamma > 0, c_1 \leq \cdots \leq c_n,$$

where x_t is the transition variable at time t,γ is a slope parameter, and $\mathbf{c} = (c_1, ..., c_n)$ a location vector.

By replacing the transition function $\overline{H}_n$ with a first-order Taylor approximation, the

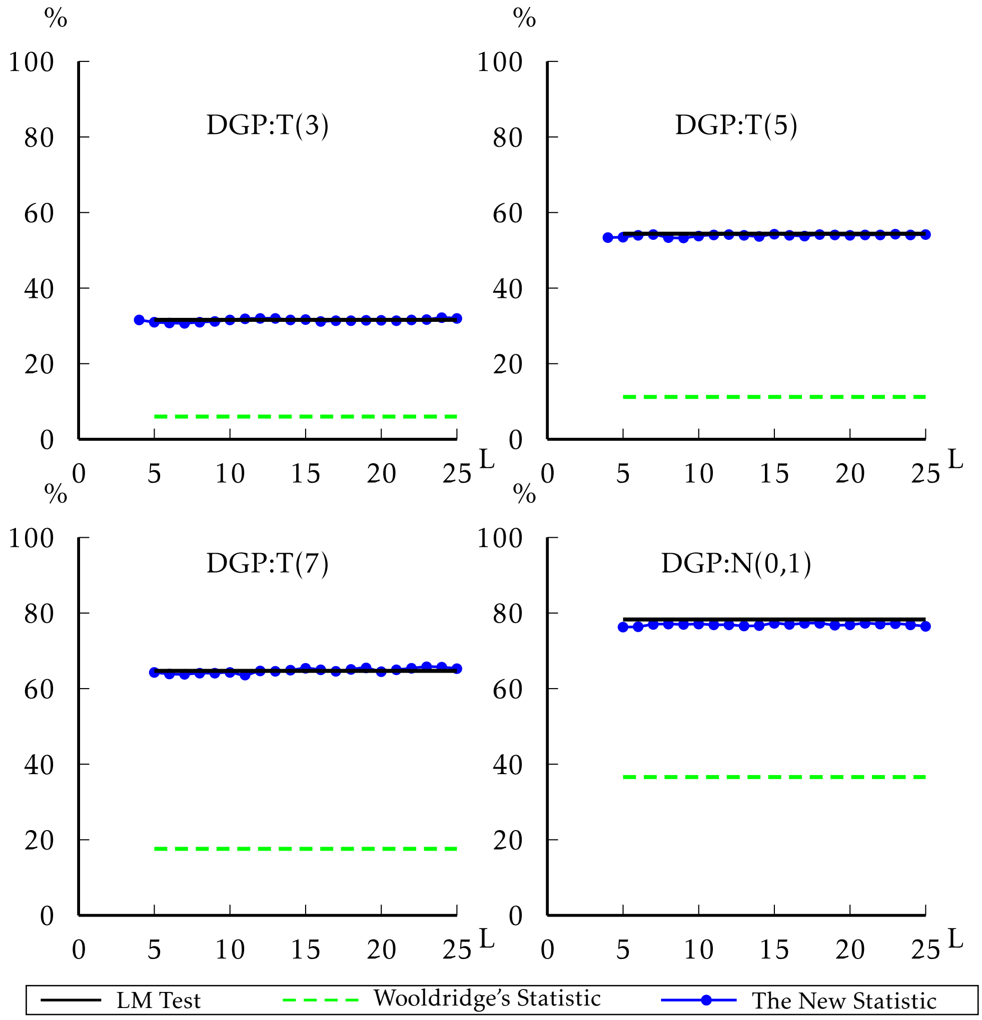


Figure 1.3: Results of empirical power of testing no remaining ARCH effect, nominal size 5%, Sample size T=500, $S = 4$.

alternative hypothesis becomes

$$h_t = \eta' \mathbf{s}_{t-1}$$

$$g_t = \boldsymbol{\beta}_1' \mathbf{v}_{1t} + \sum_{i=3}^{n+2} \boldsymbol{\beta}_i' \mathbf{v}_{it} + R_1,$$

where $\boldsymbol{\beta}_i = (\beta_{i1}, \cdots, \beta_{ip})'$, $\mathbf{v}_{it} = (Y_{t-1}^i, \cdots, Y_{t-q}^i)'$, $i = 1, 3, \cdots, n+2$ and R_1 is the remainder. The new null hypothesis is $H_0' : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_3 = \cdots = \boldsymbol{\beta}_{n+2} = 0$. One additional assumption is needed.

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

In this case, $\hat{\Lambda}_t = (\hat{\mathbf{v}}'_{1t}, \hat{\mathbf{v}}'_{3t}, \cdots, \hat{\mathbf{v}}'_{(n+2)t})'$. Under the GARCH(1,1) model, $\Lambda_t(I_{t-1}, \boldsymbol{\theta}) = (Y_{t-1}, Y_{t-1}^3)'$.

In the simulation, we consider a GARCH(1,1) model under the null. In this case, $\Lambda_t(I_{t-1}, \boldsymbol{\theta}) = (Y_{t-1}, Y_{t-1}^3)'$, and $\Lambda_{1,t}(I_{t-1}, \boldsymbol{\theta}) = (\varepsilon_{t-1}^2(I_{t-1}, \boldsymbol{\theta}), \cdots, \varepsilon_{t-L}^2(I_{t-1}, \boldsymbol{\theta}))'$. Note that the new statistic N_T, LM test, and Wooldridge (1990)'s modified statistic follow χ^2 distribution with degree of freedom of 2 asymptotically.

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

$$Y_t = \sqrt{h_t}\varepsilon_t,\, h_t = 0.1 + 0.1Y_{t-1}^2 + 0.8h_t.$$

We report the results in Figure 1.6.3. 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 $E(Y_{t-1}^{2(n+2)}) < \infty$ 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 1.6.3. 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.

1.6.4 An Empirical Application: S&P 500 Daily return Data

In this subsection, we apply our new methodology to the S&P 500 daily return. The range of the return data spans from May-23-2003 to Apr-29-2011, as shown in Figure 1.6.4. We start with testing for the adequacy of the GARCH/ARCH model, applying the new statistic $\hat{N}_T$ defined in subsection 5.2, the LM test and Wooldridge (1990)'s modifed statistic. The null of ARCH model is rejected, even for very large p. Table 1.3 reports the p-values of testing for adequacy of the null ARCH(5) model. All three tests reject the null

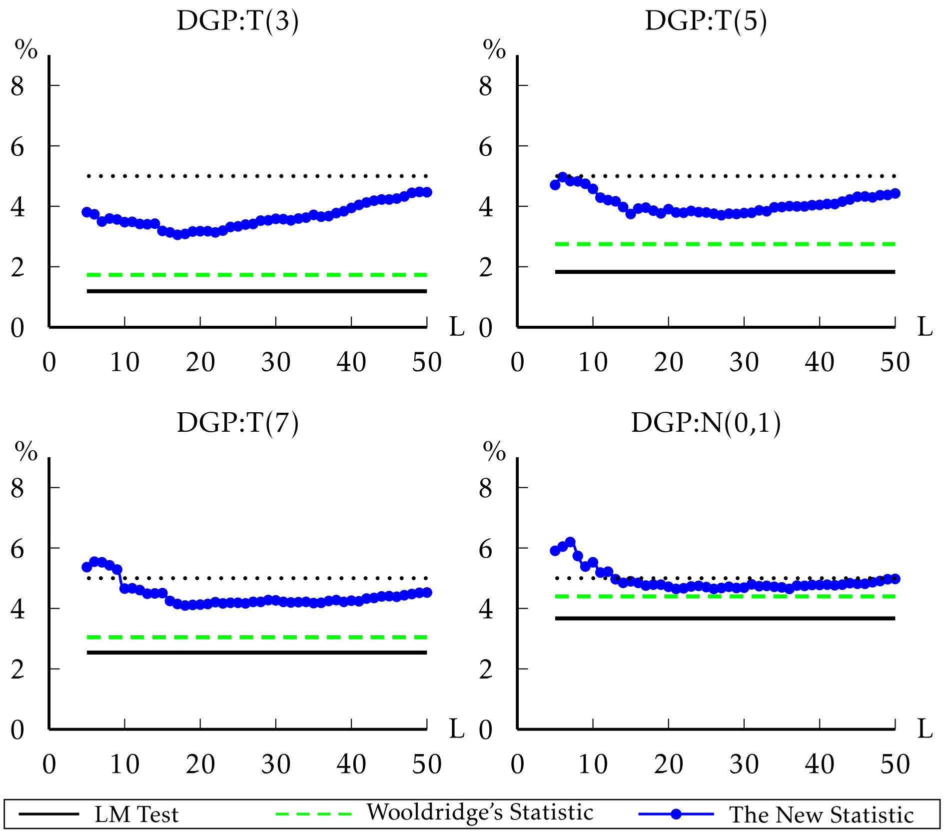


Figure 1.4: Results of empirical size of testing the nonlinearity, nominal size 5%, Sample size T=1000.

ARCH(5) at 1%. Then we estimate the data with a GARCH(1,1) model, employing the Gaussian maximum likelihood method. We test the adequacy of the null GARCH(1,1) model, reporting the results in Table 1.4, in which p-values are presented. For LM test, the results are mixed: it rejects the null GARCH(1,1) Model at 5% for S from 1 to 4, however for S from 5 to 9, it can not reject the null GARCH(1,1). One the other hand, Wooldridge (1990)'s modified statistic and the new statistic reject the null at 5% for all S from 1 to 10. Given the robustness of the new statistic and Wooldridge (1990)'s modified statistic, we conclude that GARCH(1,1) model is not sufficient to capture the variance properties of the S&P 500 daily return data.

For testing the nonlinearity, the p-value of LM test is 0.078, the null of GARCH(1,1) can not be rejected at 5%. On the other hand, p-value of Wooldridge (1990)'s modified statistic is 0.007, the p-value of the new statistic, for $L = 50$, is 0.000. Both Wooldridge (1990)'s modified statistic and the new statistic reject the null of GARCH(1,1) at 5%.

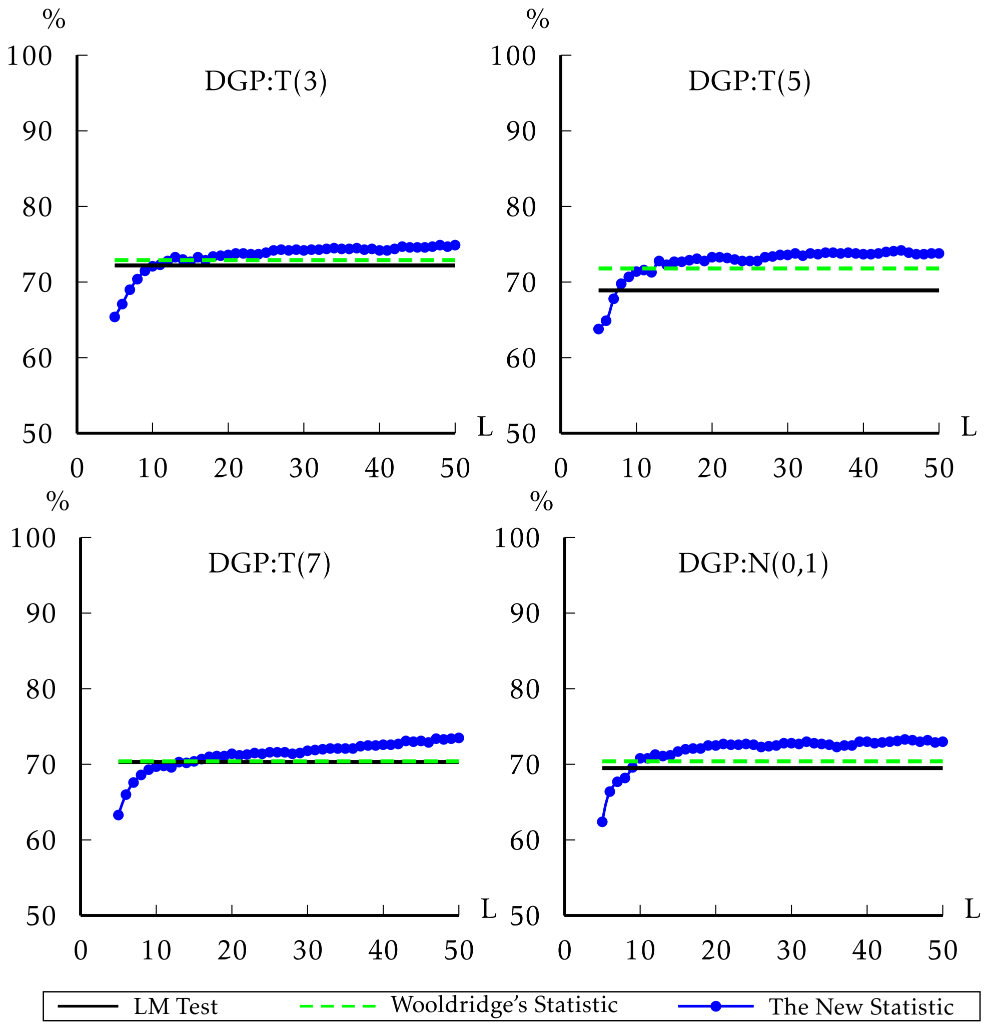


Figure 1.5: Results of empirical power of testing the nonlinearity, nominal size 5%, Sample size T=1000.

Given the fact that the new statistic has good size and power properties, we conclude that we need model the nonlinearity of the S&P 500 daily return data.

1.7 Conclusion

In this paper, we develop a new approach in the framework of conditional moment testing. Given a conditional moment test, additional unconditional 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. This new approach gives rise to the robustness and efficiency, however without paying too much price: only a $\sqrt{T}$-consistent

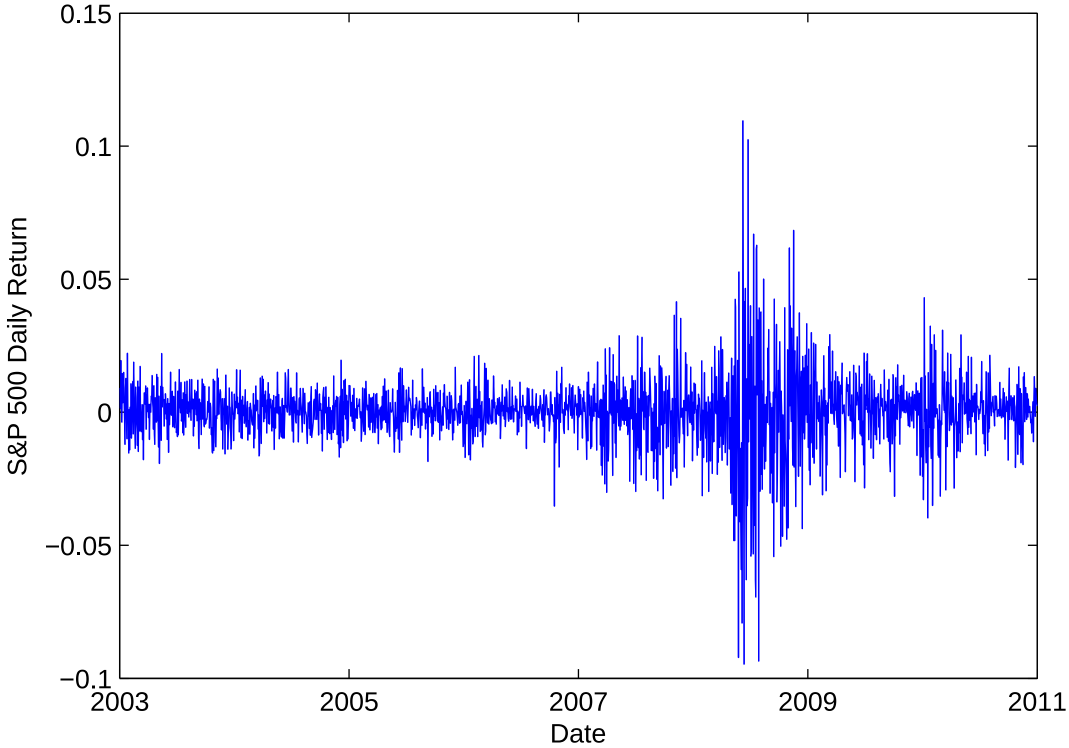


Figure 1.6: S&P 500 Daily Return Data: May-23-2003 to Apr-29-2011

estimator is needed. When our framework applies to conditional variance models, the simulation results show that our new statistic has very good size properties and nontrivial power against alternative, comparing with Lagrange Multiplier tests and Wooldridge (1990)'s modified statistic.

Table 1.3: p-values of Testing the Adequacy of ARCH(5) model for S&P 500 daily return data. $\hat{N}_T$ represents the new statistic, $L = 50$. W, Wooldridge's modified statistic. LM, Lagrange Multiplier statistic.

S		1	2	3	4	5	6	7	8	9	10
$\hat{N}_T$	L=50	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.005	0.002
W		0.084	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LM		0.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 1.4: p-values of Testing the Adequacy of GARCH(1,1) model for S&P 500 daily return data. N_T represents the new statistic, $L = 50$. W, Wooldridge's modified statistic. LM, Lagrange Multiplier statistic.

S		1	2	3	4	5	6	7	8	9	10
$\hat{N}_T$	L=50	0.002	0.005	0.003	0.004	0.005	0.000	0.001	0.004	0.010	0.002
W		0.000	0.000	0.000	0.001	0.003	0.005	0.009	0.013	0.012	0.010
LM		0.004	0.015	0.002	0.026	0.075	0.114	0.114	0.093	0.078	0.000

1.8 Appendix

In order to prove the theorems, we introduce a lemma

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

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

Let $\bar{\boldsymbol{\theta}}_T$ *be a sequence of random vectors such that* $\bar{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_T^0 \xrightarrow{p} 0$*, where* $\{\boldsymbol{\theta}_T^0\} \subset \boldsymbol{\Theta}$*. Then* $Q_T(W_T, \bar{\boldsymbol{\theta}}_T) - \bar{Q}_T(W_T, \boldsymbol{\theta}_T^0) \xrightarrow{p} 0.$

Proof. *See White (1994, Theorem 3.7).*

Proof of Theorem 1. *we firstly prove (1.3). By Mean value theorem, we have*

$$\begin{aligned}T^{-1/2}\sum_{t=1}^{T}\Lambda_t(I_{t-1},\hat{\boldsymbol{\theta}}_T)\phi_t(Y_t,I_{t-1},\hat{\boldsymbol{\theta}}_T) &= T^{-1/2}\sum_{t=1}^{T}\Lambda_t(I_{t-1},\boldsymbol{\theta}_0)\phi_t(Y_t,I_{t-1},\boldsymbol{\theta}_0)\\ &+T^{-1}\sum_{t=1}^{T}(\nabla_\theta\Lambda_t(I_{t-1},\bar{\boldsymbol{\theta}}_T)\otimes\phi_t(Y_t,I_{t-1},\bar{\boldsymbol{\theta}}_T)\sqrt{T}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)\\ &+T^{-1}\sum_{t=1}^{T}\Lambda_t(I_{t-1},\bar{\boldsymbol{\theta}}_T)\nabla_\theta\phi_t(Y_t,I_{t-1},\bar{\boldsymbol{\theta}}_T)\sqrt{T}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0),\end{aligned}$$

where $\bar{\boldsymbol{\theta}}_T$ between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_T$, for $t = 1,2,\cdots,T$. Notice that Assumption 5 $\sqrt{T}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0) = O_p(1)$, so $\sqrt{T}(\bar{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0) = O_p(1)$, for $t = 1,2,\cdots,T$. By Lemma 1 and Assumption 4, we have

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}(\nabla_\theta\Lambda_t(I_{t-1},\bar{\boldsymbol{\theta}}_T)\otimes\phi_t(Y_t,I_{t-1},\bar{\boldsymbol{\theta}}_t)\xrightarrow{p}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(E[\nabla_\theta\Lambda_t(I_{t-1},\boldsymbol{\theta}_0))\otimes\phi_t(Y_t,I_{t-1},\boldsymbol{\theta}_0)]\\ &\frac{1}{T}\sum_{t=1}^{T}\Lambda_t(I_{t-1},\bar{\boldsymbol{\theta}}_T)\nabla_\theta\phi_t(Y_t,I_{t-1},\bar{\boldsymbol{\theta}}_t)\xrightarrow{p}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E[\Lambda_t(I_{t-1},\boldsymbol{\theta}_0)\Phi_t(I_{t-1},\boldsymbol{\theta}_0)].\end{aligned}$$

Furthermore, by the law of iterated expectation

$$T^{-1}\sum_{t=1}^{T}E[\nabla_\theta\Lambda_t(I_{t-1},\boldsymbol{\theta}_0))\otimes\phi_t(Y_t,I_{t-1},\boldsymbol{\theta}_0)=0.$$

Since under the null $E(\phi_t(Y_t,I_{t-1},\boldsymbol{\theta}_0)|I_{t-1}) = 0$. So we prove that (1.3) holds. By similar argument we can prove that (1.5) holds. Based on (1.3) and (1.5), we have

$$\begin{aligned}\tilde{\xi}_T(\hat{\boldsymbol{\theta}}_T) &= \hat{\xi}_T(\boldsymbol{\theta}_0)+\Xi(\boldsymbol{\theta}_0)\sqrt{T}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)+o_p(1)\\ &-\Xi(\boldsymbol{\theta}_0)(\Xi_1(\boldsymbol{\theta}_0)'\Xi_1(\boldsymbol{\theta}_0))^{-1}\Xi_1(\boldsymbol{\theta}_0)'(\hat{\xi}_{1,T}(\boldsymbol{\theta}_0)+\Xi_1(\boldsymbol{\theta}_0)\sqrt{T}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0))+o_p(1))\\ &= \hat{\xi}_T(\boldsymbol{\theta}_0)-\Xi(\boldsymbol{\theta}_0)\left(\Xi_1(\boldsymbol{\theta}_0)'\Xi_1(\boldsymbol{\theta}_0)\right)^{-1}\Xi_1(\boldsymbol{\theta}_0)'\hat{\xi}_{1,T}(\boldsymbol{\theta}_0)+o_p(1)\end{aligned}$$

Proof of Theorem 2. *Based on Theorem 1, under H_0 and (1.1),*

$$\tilde{\xi}_T(\hat{\boldsymbol{\theta}}_T) = \hat{\Upsilon}_T(\hat{\boldsymbol{\theta}}_T)T^{-1/2}\sum_{t=1}^{T}(\Pi_t(I_{t-1},\boldsymbol{\theta}_0)\phi_t(Y_t,I_{t-1},\boldsymbol{\theta}_0))+o_p(1).$$

By Assumption 7 and Slutzky's Theorem, it is easy to obtain that

$$\tilde{\xi}_T(\hat{\boldsymbol{\theta}}_T) \xrightarrow{d} N\left(0, \boldsymbol{\Upsilon}(\boldsymbol{\theta}_0)\boldsymbol{\Gamma}(\boldsymbol{\theta}_0)\boldsymbol{\Upsilon}(\boldsymbol{\theta}_0)'\right).$$

Also note that by assumption 7, $\boldsymbol{\Upsilon}(\boldsymbol{\theta}_0)\boldsymbol{\Gamma}(\boldsymbol{\theta}_0)\boldsymbol{\Upsilon}(\boldsymbol{\theta}_0)' > \mathbf{0}$*, then we have*

$$\hat{N}_T(\hat{\boldsymbol{\theta}}_T) \xrightarrow{d} \chi^2(S).$$

Proof of Theorem 3.

$$\lim_{T\to\infty} T^{-1/2} \sum_{t=1}^{T} (\Pi_t(I_{t-1}, \boldsymbol{\theta}_0)\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}_0)) = \gamma.$$

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} E[\nabla_\theta \Lambda_t(I_{t-1}, \boldsymbol{\theta}_0)\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}_0)] &= \frac{1}{T}\sum_{t=1}^{T} E[\nabla_\theta \Lambda_t(I_{t-1}, \boldsymbol{\theta}_0)E(\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}_0)|I_{t-1})] \\ &= T^{-1/2}\cdot T^{-1}\sum_{t=1}^{T} E[\nabla_\theta \Lambda_t(I_{t-1}, \boldsymbol{\theta}_0)\delta \to 0,\end{aligned}$$

Similarly

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} E[\nabla_\theta \Lambda_{1,t}(I_{t-1}, \boldsymbol{\theta}_0)\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}_0)] &= \frac{1}{T}\sum_{t=1}^{T} E[\nabla_\theta \Lambda_{1,t}(I_{t-1}, \boldsymbol{\theta}_0)E(\phi_t(Y_t, I_{t-1}, \boldsymbol{\theta}_0)|I_{t-1})] \\ &= T^{-1/2}\cdot T^{-1}\sum_{t=1}^{T} E[\nabla_\theta \Lambda_{1,t}(I_{t-1}, \boldsymbol{\theta}_0)\delta_1 \to 0,\end{aligned}$$

So we still have $\hat{\Xi}_T(\boldsymbol{\theta}_0) = T^{-1}\sum_{t=1}^{T} E[\Lambda_t(I_{t-1}, \boldsymbol{\theta}_0)\Phi_t(I_{t-1}, \boldsymbol{\theta}_0)]$, $\hat{\Xi}_{1,T}(\boldsymbol{\theta}_0) = T^{-1}\sum_{t=1}^{T} E[\Lambda_{1,t}(I_{t-1}, \boldsymbol{\theta}_0)\Phi_t(I_{t-1}, \boldsymbol{\theta}_0)]$. *By similar argument in Proof of Theorem 2, we get* $\tilde{\xi}_T'\widehat{\boldsymbol{\Sigma}}^{-1}\tilde{\xi}_T \to_d \chi^2(S, \kappa'\boldsymbol{\Sigma}^{-1}\kappa)$ *under* H_{1T}.

Chapter 2

An Improved Consistent Conditional Moment Test for Regression Models in The Presence of Heteroskedasticity of Unknown Form

2.1 Introduction

The purpose of the present paper is to propose a simple approach to consistent testing of functional form which is robust and efficient under heteroskedasticity of unknown form. More precisely, let $(Y, X')'$ be a random vector in a $(1+d)$-dimensional Euclidean space, where X is a $d \times 1$ vector and Y is a scalar. When $E(|Y|) < \infty$, there exists a Borel measurable function f such that $E(Y|X) = f(X)$. In parametric modeling, $f(X)$ is assumed to belong to a parametric family $\mathcal{G} = \left\{ f(X, \theta) : \mathbb{R}^d \to \mathbb{R} | \theta \in \Theta \subset \mathbb{R}^p \right\}$. To justify the correctness of the parametric model, we have to test the null hypothesis

$$H_0 : \Pr\left[E(Y|X) = f(X, \theta_0)\right] = 1 \text{ for some } \theta_0 \in \Theta \tag{2.1}$$

against the alternative

$$H_1 : \Pr\left[E(Y|X) = f(X, \theta)\right] < 1 \text{ for all } \theta \in \Theta.$$

The null hypothesis is equivalent to

$$E\left[e(\theta_0)|X\right] = 0 \text{ a.s., for some } \theta_0 \in \Theta, \tag{2.2}$$

where $e(\theta) = Y - f(X, \theta)$.

There is a vast amount of literature on consistently testing the correct specification of a parametric regression model. Generally, these tests can be classified into two groups. The first class is based on smoothing methods, comparing the fitted parametric regression function with a nonparametric function estimator, for example Härdle and Mammen (1993), Gozalo (1993), Hong and White (1995), Fan and Li (1996), and Zheng (1996), to mention but a few. Under the null hypothesis, the smoothing-based tests only require some consistent estimate of parameter $\hat{\theta}$, and typically lead to asymptotic pivotal tests statistics. However, they depend on a smoothing parameter, and there has been much concern over their small sample properties.

The second class of tests are based on the integrated nonparametric curves, avoiding smoothing estimation by means of converting the conditional moment restriction into an infinite number of unconditional moment restrictions, i.e.,

$$E\left[e(\theta_0)|X\right] = 0 \text{ a.s} \Leftrightarrow E\left[e(\theta_0)\, w(X,t)\right] = 0, \text{ for almost all } t \in \mathcal{T}, \tag{2.3}$$

where $\mathcal{T} \subset \mathbb{R}^h$, $h \in \mathbb{N}$, and $w(X,t)$ is a proper weighting function such that the equivalence (2.3) holds. There are many weighting functions meeting the requirement of (2.3). One example is $w(X,t) = \exp(it'X)$ where $i = \sqrt{-1}$, $\mathcal{T} = \mathbb{R}^d$, which is employed by Bierens (1982). Bierens (1990) proposes $w(X,t) = \exp(t'X)$, $\mathcal{T} = \mathbb{R}^d$. Stute (1997) proposes the indicator function $w(X,t) = I(X < t)$. Escanciano (2006a) introduces weighting function $w(X,t) = I(\beta'X \le u)$, with $t = (\beta', u)' \in \mathcal{T} = S^d \times (-\infty, \infty)$, where $S^d = \left\{\beta \in \mathbb{R}^d : \left|\beta\right| = 1\right\}$. Escanciano (2006b) summarizes different weighting functions into one general class.

Given a sample $\left(Y_j, X_j'\right)'$, $j = 1, \cdots, n$, a $\sqrt{n}$-consistent estimator $\hat{\theta}$, for any t in some subset $\Pi \in \mathcal{T}$, the scaled sample analog of $E\left[e(\theta_0)\, w(X,t)\right]$ is

$$\hat{M}\left(\hat{\theta}, t\right) = n^{-1/2} \sum_{j=1}^{n} e_j\left(\hat{\theta}\right) w\left(X_j, t\right).$$

Stinchcombe and White (1998) coin this kind of specification tests as the one with nuisance parameters present only under the alternative, since there is the nuisance parameter t in test statistic $\hat{M}\left(\hat{\theta}, t\right)$. Bierens (1982) proposes to integrate the nuisance parameter out. The so-called integrated conditional moment (ICM) test statistic has the form

$$ICM = \int_{\Pi} \left|\hat{M}\left(\hat{\theta}, t\right)\right|^2 d\mu(t),$$

where $\mu(t)$ is a probability measure on Π that is absolutely continuous with respect to Lebesgue measure on $\Pi \subset \mathcal{T}$. Or we can maximize $\hat{M}\left(\hat{\theta}, t\right)$ over Π, which is a Kolmogorov-Smirnov type statistic

$$KS = \sup_{t \in \Pi} \left|\hat{M}\left(\hat{\theta}, t\right)\right|^2.$$

In contrast with smoothing-based tests, the non-smoothing-based tests have to handle the so-called "estimation effect". More specifically, it could be shown that

$$\hat{M}\left(\hat{\theta},t\right)=n^{-1/2}\sum_{j=1}^{n}e_j\left(\theta_0\right)w\left(X_j,t\right)-b(t)'n^{1/2}\left(\hat{\theta}-\theta_0\right)+o_p\left(1\right),$$

where

$$b(t)=E\left[\frac{\partial f\left(X,\theta_0\right)}{\partial\theta}w\left(X,t\right)\right].$$

The asymptotic variance of $\hat{M}\left(\hat{\theta},t\right)$ depends on the model and the estimation approach applied to obtain the estimate $\hat{\theta}$. Bierens (1982, 1990), and Bierens and Ploberger (1997) assume that the estimator $\hat{\theta}$ is based on the criteria function $\theta_0=\arg\min_{\theta\in\Theta}E\left\{\left[Y-f\left(X,\theta\right)\right]^2\right\}$, which means that the nonlinear least squared (NLS) estimator is employed. So we have

$$n^{1/2}\left(\hat{\theta}-\theta_0\right)=E\left[\frac{\partial f\left(X,\theta_0\right)}{\partial\theta}\frac{\partial f\left(X,\theta_0\right)}{\partial\theta'}\right]^{-1}n^{-1/2}\sum_{j=1}^{n}e_j\left(\theta_0\right)\frac{\partial f\left(X_j,\theta_0\right)}{\partial\theta}+o_p\left(1\right).$$

Following the "estimating-testing" paradigm of Newey (1985a,b) and Tauchen (1985), it is quite straightforward to obtain the asymptotic theory of $\hat{M}\left(\hat{\theta},t\right)$ for any $t\in\Pi$. Then the asymptotic distribution of ICM or KS statistic could be established on Π by employing weakly convergence theory. Both ICM and KS statistics follow non-standard and model-dependent distributions asymptotically, and normally bootstrap procedures have to be applied though. Bierens (1990) develops a procedure which could be easily implemented without employing bootstrap techniques.

Under conditional homoskedasticity, it is well established that the non-smoothing-based tests are more powerful than smoothing-based tests against Pitman local alternatives. Furthermore, under the assumption of normal errors the ICM test is asymptotically admissible, in the sense that there does not exist a test that is uniformly more powerful; see Bierens and Ploberger (1997). But when there exists unknown heteroskedasticity, the NLS estimator becomes inefficient, the optimality of ICM test breaks down, the testing power gets worse.[1] In order to robustify the testing statistic under the possible het-

[1] For simulation evidence on the power deterioration of consistent tests when heteroskedasticity of unknown form is present, see Miles and Mora (2003).

eroskedasticity, Stute (1997) and Escanciano (2006a) assume an estimator such that

$$n^{1/2}\left(\hat{\theta}-\theta_0\right)=n^{-1/2}\sum_{j=1}^{n}h\left(Y_j,X_j,\theta_0\right)+o_p(1),$$

where $h(\cdot)$ is such that $E[h(Y,X,\theta_0)]=0$, and $H(\theta_0)=E[h(Y,X,\theta_0)h'(Y,X,\theta_0)]$ exists and is positive definite. This form includes the NLS estimator as a special case, however it does not provide any useful clue of how to choose more efficient or the most efficient estimator in practice.

On the other hand, the efficient estimation of parameters of conditional moment restriction models can be pursued by employing exactly the same idea as the non-smoothing based consistent specification testing. The only difference between non-smoothing consistent tests and efficient estimation is that while non-smoothing consistent tests exploit the continuum of the unconditional moment restrictions, it is sufficient to employ discrete countable unconditional moment restrictions in efficient estimation of the conditional moment restrictions model.[2] More specifically, let Z be the support of distribution of X, define L_2 be the space of measurable functions $g:Z\to\mathbb{R}$ with $E[g^2(X)]<\infty$. We say a sequence of $\{q_j\}_{j=1}^{\infty}$ in L_2 is L_2-complete if for any $\epsilon>0$, and any $\varphi\in L_2$, there exists a positive integer K and a $K\times 1$ vector γ_K such that

$$\left\{E\left[\left\{\varphi(X)-q^K(X)'\gamma_K\right\}^2\right]\right\}^{1/2}<\epsilon, \tag{2.4}$$

where $q^K(X)=(q_1(X),\cdots,q_K(X))'$ is a $K\times 1$ vector. There are a lot of choices of $q^K(X)$ satisfying (2.4), examples are splines, power series, and Fourier series. Chamberlain (1987) firstly shows that an estimator obtained as the solution to

$$\sum_{j=1}^{n}Q\left(X_j\right)e(\theta_0)=0,\ Q(X)=\frac{\partial f(X,\theta_0)}{\partial\theta}\sigma_0^{-2}(X),$$

where

$$\sigma_0^2(X)=E\left[(Y-f(X,\theta_0))^2|X\right],$$

[2]Carrasco and Florens (2000) consider the continuum of unconditional moment restrictions in efficient estimation of the conditional moment restrictions model, however the singularity of the covariance matrix has to be handled. Furthermore, the indexed parameter t has to be a scalar.

achieves the semiparametric efficiency bound. Chamberlain (1987, 1992) show that the asymptotic variance of the GMM estimator based on the unconditional moment restrictions $E\left[q^K(X)e(\theta_0)\right] = 0$, where $q^K(X)$ satisfies (2.4), comes arbitrarily close to the semiparametric efficiency bound as $K \to \infty$. Intuitively, since the conditional moment restriction is equivalent to a sequence of unconditional moment restrictions, as K grows with the sample size, all of the information of the conditional moment restriction is eventually accounted for. One special advantage of this approach, as Newey (1993) points out, is that the linear combination of $q^K(X)$ can approximate $Q(X)$ very well with only a few terms. Hahn (1997) and Donald, Imbens and Newey (2003) establish the rate of increase of the number of instruments for different choices of $q^K(X)$, for example splines and power series, in a quite general framework. However this rate of growth seems to have very little practical relevance. No methodology has been established to find the number of instruments for a given sample size which would guarantee that the resultant sequence of GMM estimators would achieve the semiparametric efficiency bound.

In this paper we exploit the duality property of one class of weighting functions, which will be defined in next section, for both consistent specification testing and efficient estimation. Instead of following the "estimating-testing" paradigm of Newey (1985a,b) and Tauchen (1985), we propose a new test statistic, employing a transformation-based projection. It is shown that the new test statistic exploits asymptotic efficient parameter estimator under heteroskedasticity of unknown form, which will bring about potentially improved tests. Further, it is quite easy to compute, only an initial $\sqrt{n}$-consistent estimator is needed. Monte Carlo simulations show that Bierens (1990) test based on the new test statistic is more powerful for a large number of alternatives when heteroskedasticity of unknown form is presented.

The outline of the paper is as following. In Section 2, we define the class of weighting functions and the new test statistic, study its properties. Section 3 discusses the improved test statistic of Bierens (1990). Section 4 conducts Monte Carlo simulations. Section 5 concludes.

2.2 The Class of Weighting Functions and The New Test Statistic

From the perspective of consistent specification testing, it is possible to consider a class of weighting functions as general as the one defined in Escanciano (2006b), however we find that it is convenient to focus on a class of weighting functions $\mathcal{W}$ such that

$$\mathcal{W} = \left\{ w(t'X), t \in \mathbb{R}^d, w \text{ is an analytic function that is nonpolynomial} \right\},$$

where X is a bounded $d \times 1$ random vector.

Lemma 2. *Let X be a random vector in $\mathbb{R}^d$, $\Phi(\cdot)$ a bounded one-to-one mapping from $\mathbb{R}^d$ into $\mathbb{R}^d$, for any weighting function $w(t'\Phi(X))$ in the class $\mathcal{W}$, the equivalence in (2.3) holds.*

Proof. See Stinchcombe and White (1998) Theorem 2.3. □

Remark: Bierens and Ploberger (1997) give an alternative version of conditions of the equivalence.

Examples of families satisfying this lemma are $w(t'\Phi(X)) = \exp(it'\Phi(X))$ and $w(t'\Phi(X)) = \exp(t'\Phi(X))$.

For $w \in \mathcal{W}$ and any $t \in \Pi \subset \mathbb{R}^d$, we have a statistic such that

$$\hat{M}\left(\hat{\theta}, t\right) = n^{-1/2} \sum_{j=1}^{n} e_j\left(\hat{\theta}\right) w\left(t'\Phi(X_j)\right).$$

Then we can form the ICM or KS statistic on some interval Π. Note that, to reach the testing consistency, it is not necessary to integrate or maximize over the whole interval $\mathbb{R}^d$, only a subset is needed.

On the other hand, $w(t'\Phi(X))$ also forms a basis for efficient estimation. In this case, we only need to consider discrete countable unconditional moment restrictions. For any fixed sequence $\{t_j\}_{j=1}^{\infty}$, which is dense in some subset of $\mathbb{R}^d$, $q_j(X) = w(t_j'\Phi(X))$, $j = 1, 2, \cdots$, and for each positive integer K, define the $K \times 1$ vector

$$q^K(X) = \left(w(t_1'\Phi(X)), \cdots w(t_K'\Phi(X))\right)'. \tag{2.5}$$

Note that we omit in the notation $q^K(X)$ the dependence of this $\{t_j\}_{j=1}^K$ sequence. We have the following Corollary:

Corollary 1. *For any $\varphi \in L_2$ and for each $K \times 1$ vector $q^K(X)$ defined in (2.5), there are $K \times 1$ vectors γ_K such that (2.4) holds.*

Proof. See Appendix. □

It is possible to form a robust ICM or KS statistic based on $\hat{M}\left(\hat{\theta}, t\right)$ by employing the "Estimating-testing" paradigm of Newey (1985a,b) and Tauchen (1985), using a GMM estimator based on the unconditional moment conditions $E[q^K(X)e(\theta_0)] = 0$. But it is unclear to choose the dimension K in finite samples, and it is tedious to compute the efficient estimator in the first place. More importantly, there may exist global identification problems for nonlinear models, in which the parameters are not identified by the unconditional moment restrictions $E[q^K(X)e(\theta_0)] = 0$. To overcome these two problems, we propose an innovative transformation on $\hat{M}\left(\hat{\theta}, t\right)$ to *remove* the estimation effect. The idea is to form a new weights by a proper linear combination of $w(t'\Phi(X))$ and $q^K(X)$. More specifically, given any $\sqrt{n}$-consistent estimator $\hat{\theta}$, $q^K(X)$ defined in (2.5), for any $t \in \Pi \subset \mathbb{R}^d$, the new test statistic is

$$\tilde{M}(\hat{\theta}, t) = n^{-1/2} \sum_{j=1}^{n} [Y_j - f(X_j, \hat{\theta})][w(t'\Phi(X_j)) - \hat{A}\left(X_j\right)' \hat{b}(\hat{\theta}, t)],$$

where the new weights depend on

$$\hat{b}(\hat{\theta}, t) = \frac{1}{n} \sum_{j=1}^{n} w\left(t'\Phi(X_j)\right) \frac{\partial f(X_j, \hat{\theta})}{\partial \theta} \tag{2.6}$$

$$\hat{A}\left(X_j\right) = \left[\hat{\Lambda}\left(\hat{\theta}\right)' \hat{\Omega}\left(\hat{\theta}\right)^{-1} \hat{\Lambda}\left(\hat{\theta}\right)\right]^{-1} \hat{\Lambda}\left(\hat{\theta}\right)' \hat{\Omega}\left(\hat{\theta}\right)^{-1} q^K\left(X_j\right) \tag{2.7}$$

$$\hat{\Lambda}\left(\hat{\theta}\right) = \frac{1}{n} \sum_{j=1}^{n} q^K\left(X_j\right) \frac{\partial f(X_j, \hat{\theta})}{\partial \theta'}$$

$$\hat{\Omega}\left(\hat{\theta}\right) = \frac{1}{n} \sum_{j=1}^{n} (Y_j - f(X_j, \hat{\theta}))^2 q^K\left(X_j\right) q^K\left(X_j\right)'.$$

Remark: This transformation to remove the estimation effect could be regarded as an application of the general methodology in the conditional moment specification testing developed by Wang (2011), in which he proposes to use further moment conditions to remove the estimation effect in the test statistics. In Theorem 1, we will show that we will show that in our modified statistics exploits the moment conditions $E[q^K(X)e(\theta_0)] = 0$. The intuition behind the removal of the estimation effect is that the transformed weighting function $w(t'\Phi(X)) - \hat{A}(X)'\hat{b}(\hat{\theta},t)$ in $\tilde{M}(\hat{\theta},t)$ is orthogonal to $\frac{\partial f(X,\hat{\theta})}{\partial\theta}$. So $\tilde{M}(\hat{\theta},t)$ is not affected by the estimation effect asymptotically.

Remark: To remove the estimation effect, the choice of the $q^K(X)$ is not restricted to q^K defined in (2.5). If $q^K(X) = \frac{\partial f(X,\theta_0)}{\partial\theta}$ is chosen, we obtain Wooldridge (1990)'s modified statistic. While Wooldridge (1990)'s modified statistic is only robust to heteroskedasticity in the sense that White's heteroskedasticity-robust variance estimate is used, the statistic choosing $q^K(X) = \left(w\left(t_1'\Phi(X)\right), \cdots, w\left(t_K'\Phi(X)\right)\right)'$ can be not only robust to heteroskedasticity but also efficient in the presence of heteroskedasticity of unknown form by choosing the optimal weighting $\hat{\Omega}\left(\hat{\theta}\right)$ and increasing the dimension K as n grows, as we will show later.

Remark: The form of $\tilde{M}(\hat{\theta},t)$ also has connection with the martingale transformation approach employed by Stute et al. (1998). In their case, $w(X,t) = I(X < t)$, but this function does not fall into the function class defined by Lemma 1. While martingale transformation approach focuses on obtaining asymptotic distribution-free statistics, our transformation focuses on obtaining efficient statistics under heteroskedasticity of unknown form.

Now we present the assumptions:

Assumption 9. *Let* $\left(Y_j, X_j'\right)'$, $j = 1, \cdots, n$, *be a sample from a probability distribution* $F(Y,X)$ *on* $\mathbb{R} \times \mathbb{R}^d$. *Moreover,* $E\left(Y^2\right) < \infty$.

Assumption 10. *The parameter space* Θ *is a compact subset of* $\mathbb{R}^p$. $\theta_0 \in int(\Theta)$.

Assumption 11. $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$.

Assumption 12. $E\left[\sup_{\theta\in\Theta}(Y - f(X,\theta))^2 | X\right] < \infty$, $\sigma_0^2(X)$ *is bounded away from zero. There is* $\delta(Y,X)$ *and* $\alpha > 0$ *such that for all* $\bar{\theta}$, $\theta \in \Theta$, $\left|f\left(X,\bar{\theta}\right) - f(X,\theta)\right| \leq \delta(Y,X)\left\|\bar{\theta} - \theta\right\|^{\alpha}$ *and* $E\left[\delta(Y,X)^2\right] < \infty$.

Assumption 13. *$f(X,\theta)$ is twice continuously differentiable in a open and convex neighborhood Δ of θ_0. $E\left[\left\|\frac{\partial^2 f(X,\theta_0)}{\partial\theta\partial\theta'}\right\|\right]$ is bounded, $E\left[\frac{\partial f(X,\theta_0)}{\partial\theta}\frac{\partial f(X,\theta_0)}{\partial\theta'}\right]$ is nonsingular. $E\left[\sup_{\theta\in\Delta}\left\|\frac{\partial f(X,\theta)}{\partial\theta}\right\|^2\right] < \infty$, $E\left[\sup_{\theta\in\Delta}|Y-f(X,\theta)|^4|X\right] < \infty$, and for all $\theta\in\Delta$, $|f(X,\theta)-f(X,\theta_0)| \le \delta(Y,X)\|\theta-\theta_0\|$ and $E\left[\delta(Y,X)^2|X\right] < \infty$.*

Assumption 14. *Denote Z as the support of X, for each K there is a constant scalar $\xi(K)$ and matrix B such that $\tilde{q}^K(X) = Bq^K(X)$ for every $X\in Z$, $sup_{X\in Z}\left\|\tilde{q}^K(X)\right\| \le \xi(K)$, $\sqrt{K}\le\xi(K)$, and $E\left(\tilde{q}^K(X)\tilde{q}^K(X)'\right)$ has smallest eigenvalue bounded away from zero uniformly in X. There exists an integer D, $D\ge p$ such that when $K\ge D$, $E\left[q^K(X)\frac{\partial f(X,\theta_0)}{\partial\theta'}\right]$ is of full rank.*

Assumptions 1 and 2 are standard regularity conditions. Assumption 1 restricts our analysis to an i.i.d context. It is possible to extend it to dependent data following De Jong (1996). Assumption 3 shows that we only need a $\sqrt{n}$-consistent estimator. Since we are dealing with a testing problem, we do not present the identification conditions of parameter estimation explicitly, just assuming some $\sqrt{n}$-consistent estimator is obtainable. To obtain a $\sqrt{n}$-consistent estimator, only an identification condition as weak as Dominguez and Lobato's (2004) is needed. Assumption 4 imposes some restrictions on second moment condition of the error term and the smoothness of the function $f(X,\theta)$. Assumption 5 is essential for asymptotic normality when the number of moment conditions is growing with the sample size. Assumption 6 imposes a normalization on the approximate function, bounds the second moment restriction away from singularity and restricts the magnitude of the series terms. The magnitude of the series terms is important, playing a crucial role in the asymptotic theory of GMM estimation when K increases with sample size n. Primitive conditions for this assumption are given in the case of $w(\cdot)=\exp(\cdot)$ when we discuss the improved Bierens (1990) statistic in Section 3. The properties of $q^K(X)$ make sure that $E\left[q^K(X)\frac{\partial f(X,\theta_0)}{\partial\theta'}\right]$ is of full rank as K goes to infinity. However, in some cases, K has to be regarded as a fixed number. So it is necessary to explicitly assume the nonsingularity of $E\left[q^K(X)\frac{\partial f(X,\theta_0)}{\partial\theta'}\right]$ when K is large enough.

Theorem 5. *When Assumptions 1 to 6 hold, $K\ge D$, under H_0, for any $t\in\Pi\subset\mathbb{R}^d$,*

$$\tilde{M}(\hat{\theta},t) = n^{-1/2}\sum_{j=1}^{n}[Y_j - f(X_j,\theta_0)][w(t'\Phi(X_j)) - A\left(X_j\right)' b(t)] + o_p(1), \tag{2.8}$$

and

$$\tilde{M}(\hat{\theta},t) \xrightarrow{d} N\left[0, s^2(t)\right], \tag{2.9}$$

where

$$A\left(X_j\right) = \left[\Lambda'\Omega^{-1}\Lambda\right]^{-1}\Lambda'\Omega^{-1}q^K\left(X_j\right)$$

$$\Lambda = E\left[q^K(X)\frac{\partial f(X,\theta_0)}{\partial\theta'}\right]$$

$$\Omega = E\{[Y - f(X,\theta_0)]^2 q^K(X)q^K(X)'\}$$

$$s^2(t) = E\left\{[Y - f(X,\theta_0)]^2\left[w(t'\Phi(X)) - A(X)'b(t)\right]^2\right\}.$$

Proof. See Appendix □

Equation (2.8) shows that, unlike the statistic $\hat{M}\left(\hat{\theta},t\right)$, the new statistic $\tilde{M}(\hat{\theta},t)$ does not suffer from the estimation effect: the statistic evaluated at any $\sqrt{n}$-consistent parameter estimator is asymptotically the same as the statistic evaluated at the true parameter.

Although the difference between the two statistics evaluated at the same estimator $\hat{\theta}$ is not negligible, it turns out that $\tilde{M}\left(\hat{\theta},t\right)$ is equivalent asymptotically to $\hat{M}(\cdot,t)$ which is evaluated at a particular estimator. The following theorem establishes the relation between $\tilde{M}(\cdot,t)$ and $\hat{M}(\cdot,t)$.

Theorem 6. *When Assumptions 1 to 6 hold, $K \geq D$, under H_0, for any $t \in \Pi \subset \mathbb{R}^d$, $\tilde{M}(\hat{\theta},t)$ is equivalent asymptotically to $\hat{M}\left(\check{\theta},t\right)$ with*

$$\check{\theta} = \arg\min_{\theta\in\Theta}\frac{1}{n}\sum_{j=1}^{n}q^K\left(X_j\right)'\left(Y_j - f(X_j,\theta)\right)\bar{\Omega}^{-1}\frac{1}{n}\sum_{j=1}^{n}q^K\left(X_j\right)\left(Y_j - f(X_j,\theta)\right) \tag{2.10}$$

where $\bar{\Omega} \xrightarrow{p} \Omega$.

Proof. See Appendix. □

Note that $\check{\theta}$ satisfies Assumption 3. Actually under Assumption 1 to 6, $\hat{M}\left(\check{\theta},t\right)$ could be regarded as a special case of $\tilde{M}(\cdot,t)$: when $\hat{\theta} = \check{\theta}$, $\tilde{M}(\cdot,t)$ retreats to $\hat{M}\left(\check{\theta},t\right)$.

This theorem shows that, when K is large enough, $\tilde{M}(\cdot,t)$ evaluated at any $\sqrt{n}$ consistent estimator is equivalent to $\hat{M}(\cdot,t)$ evaluated at the two-step GMM estimator based on moment conditions $E\left[q^K(X)e(\theta_0)\right]=0$. In this sense, our approach could be regarded as a one-step procedure in testing scenario. This theorem does not levy any restriction on K. To reach the efficiency, we have to control the increase of K as n increases. We establish the asymptotic efficiency of the new test statistic when K increases with sample size n in the following Theorem.

Theorem 7. *When Assumptions 1 to 6 hold, under H_0, for any $t\in\Pi\subset\mathbb{R}^d$, $\tilde{M}(\hat{\theta},t)$ is an efficient statistic in the sense that $\tilde{M}(\hat{\theta},t)$ is equivalent to $\hat{M}\left(\check{\theta},t\right)$ where $\check{\theta}$ reaches the semiparametric efficiency bound, such that*

$$\tilde{M}(\hat{\theta},t)=n^{-1/2}\sum_{j=1}^{n}[Y_j-f(X_j,\theta_0)]\left[w(t'\Phi(X_j))-A_*\left(X_j\right)'b(t)\right]+o_p(1),$$

$$s_*^2(t)=E\left\{[Y-f(X,\theta_0)]^2\left[w(t'\Phi(X_j))-A_*\left(X_j\right)'b(t)\right]^2\right\},$$

where

$$A_*\left(X_j\right)=E\left[\frac{\partial f(X,\theta_0)}{\partial\theta}\sigma_0^{-2}(X)\frac{\partial f(X,\theta_0)}{\partial\theta'}\right]^{-1}\frac{\partial f(X_j,\theta_0)}{\partial\theta}\sigma_0^{-2}\left(X_j\right),$$

when $K\to\infty$ and $\xi(K)^2K/n\to 0$.

Proof. See Appendix. □

This theorem establishes that $s_*^2(t)$ is the lowest variance we can obtain in the semiparametric framework for $\tilde{M}(\hat{\theta},t)$. In Section 3, we will demonstrate the efficiency of the KS test statistic when there exists heteroskedasticity of unknown form.

Although we have established that $\tilde{M}(\cdot,t)$ evaluated at any $\sqrt{n}$ consistent estimator is equivalent to $\hat{M}(\cdot,t)$ evaluated at the two-step GMM estimator based on moment conditions $E\left[q^K(X)e(\theta_0)\right]=0$, and established the increase rate of the dimension K for the efficiency. Our approach is more natural and much easier to compute, comparing with $\hat{M}(\cdot,t)$, since in practice, we normally have a fixed sample size. It is still unclear to choose the dimension K. Furthermore, it is tedious to compute the efficient GMM estimator.

2.3 Bierens (1990) Test Based On The New Statistic

Based on $\tilde{M}(\hat{\theta},t)$, we can form ICM or KS tests. In the case of KS tests, Bierens (1990)'s procedure is attractive, since its null asymptotic distribution is tractable, and the time-consuming bootstrap procedure is avoided.

In Bierens (1990) $w(\cdot) = \exp(\cdot)$, so $q^K(X) = (\exp(t_1'\Phi(X)),\cdots,\exp(t_K'\Phi(X)))'$. In this case, we could give primitive conditions for Assumption 6.

Assumption 15. *Choose $(t_1,\cdots,t_K)'$ such that $t_j \in \mathbb{R}^d \backslash \Pi$ for $j = 1,\cdots,K$, and $t_j \neq t_i$ for any $j,i = 1,\cdots,K$. For $\Phi(X)$, the Borel measurable bounded one-to-one mapping from $\mathbb{R}^d$ into $\mathbb{R}^d$, has a probability density function that is bounded away from zero. There exists an integer D, $D \geq p$ such that when $K \geq D$, $E\left[q^K(X)\frac{\partial f(X,\theta_0)}{\partial \theta'}\right]$ is of full rank.*

This assumption imposes restrictions on the probability density function of X. Similar assumption has been used by Newey (1997) in the case of series estimation of non-parametric and semiparametric models. This assumption also sets the rules of how to choose $(t_1,\cdots,t_K)'$ and the subset Π, they are hardly restrictive.

Lemma 3. *Assumption 7 implies Assumption 6, further $\xi(K) = CK^{3/2}$, where $C > 0$ is a constant. Finally, for any $t \in \Pi$, $s^2(t) > 0$.*

Proof. See Appendix. □

In practice, Given any proper K, the function $s^2(t)$ can be consistently estimated by

$$\hat{s}^2(t) = \frac{1}{n}\sum_{j=1}^{n}(Y_j - f(X_j,\hat{\theta}))^2[\exp(t'\Phi(X_j)) - \hat{A}\left(X_j\right)'\hat{b}(\hat{\theta},t)]^2,$$

where $\hat{b}(\hat{\theta},t)$ is defined by (2.6), and $\hat{A}\left(X_j\right)$ by (2.7). From this lemma

$$\tilde{W}(t) = \frac{\left[\tilde{M}(\hat{\theta},t)\right]^2}{\hat{s}^2(t)}$$

is well defined for any $t \in \Pi$ for sample size large enough.

From Lemma 1 of Bierens (1990), it is straightforward to obtain that under H_1, the set

$$S = \left\{t \in \mathbb{R}^d : E\left[(Y - f(X,\theta_0))\left(\exp(t'\Phi(X)) - A(X)'b(t)\right)\right] = 0\right\}$$

has Lebesgue measure zero.

Our approach conveniently avoids the extreme condition that $s^2(t) = 0$. In Bierens (1990), an additional assumption has to be imposed, and it only could be established that set $S^* = \left\{t \in \mathbb{R}^d : s^2(t) = 0\right\}$ has Lebesgue measure zero and is not dense in $\mathbb{R}^d$. So in Bierens' case it has to be assumed that $\Pi \subset \mathbb{R}^d \backslash S \cup S^*$, on the other hand, we only need assume $\Pi \subset \mathbb{R}^d \backslash S$. This may have important impact on the testing power. Since the testing power negatively relies on the size of the set $S \cup S^*$. The new test statistic may have more power than the Bierens (1990) test, where the NLS estimator is employed, even in the case of conditional homoskedasticity.

We summarize our results in the following theorem.

Theorem 8. *Under Assumptions 1-5,7, there exists a nondense subset S of* $\mathbb{R}^d$ *with Lebesgue measure zero such that for every* $t \in \Pi \subset \mathbb{R}^d \backslash S$, $\tilde{W}(t) \to \chi_1^2$ *in distribution under* H_0. *Whereas under* H_1, $\tilde{W}(t)/n \to \eta(t)$, *where* $\eta(t) > 0$.

Following Bierens (1990), we maximize $\tilde{W}(t)$ over a subset Π of $\mathbb{R}^d$.

Theorem 9. *Let Assumptions 1-5 and 7 hold, when* $K \to \infty$ *and* $K^4/n \to 0$ *then* $\tilde{W}$ *converges weakly to* z^2 *under* H_0, *where z is a Gaussian element of* $C(\Pi)$ *with covariance function*

$$\Gamma(t_1,t_2) = E\left\{\begin{array}{c} [Y - f(X,\theta_0)]^2 [\exp(t_1'\Phi(X)) - A_*(X)'b(t_1)] \\ \times[\exp(t_2'\Phi(X)) - A_*(X)'b(t_2)]/\sqrt{s_*^2(t_1)}\sqrt{s_*^2(t_2)} \end{array}\right\}. \tag{2.11}$$

Moreover, $\tilde{W}(\tilde{t})$ *with* $\tilde{t} = \arg\max_{t\in\Pi} \tilde{W}(t)$ *converges in distribution to* $\sup_{t\in\Pi} z^2(t)$. *Furthermore under* H_1, $\tilde{W}(t)/n \to \eta(t)$ *a.s. uniformly on* Π *and consequently* $\sup_{t\in\Pi} \tilde{W}(t)/n \to \sup_{t\in\Pi} \eta(t)$ *a.s.*

Proof. See Appendix □

Note that the covariance function $\Gamma(t_1, t_2)$ depends on the DGP of the model, so does the distribution of $sup_{t\in\Pi} z^2(t)$. Then critical values should be tabulated for each model and each DGP. Normally some bootstrap procedure should be applied to overcome this problem. Bierens (1990) circumvents the bootstrap procedure by introducing some penalty function. The alternative procedure similar to Bierens (1990) is the following.

Theorem 10. *Let Assumptions 1-5, 7 hold. Choose independently of the data generating process real numbers* $\gamma > 0$, $\rho \in (0,1)$*, and a point* $t_0 \in \Pi$*. Let* $\tilde{t} = \arg\max_{t\in\Pi} \tilde{W}(t)$ *and let*

$$\bar{t} = t_0, \text{ if } \tilde{W}(\tilde{t}) - \tilde{W}(t_0) \le \gamma n^{\rho}; \ \bar{t} = \tilde{t}, \text{ if } \tilde{W}(\tilde{t}) - \tilde{W}(t_0) \ge \gamma n^{\rho},$$

then under H_0*,* $\tilde{W}(\bar{t}) \to \chi_1^2$ *in distribution, whereas under* H_1*,* $\tilde{W}(\bar{t})/n \to \sup_{t\in\Pi} \eta(t)$ *a.s.*

Proof. Similar to the Proof of Bierens (1990) Theorem 4. □

In practice, it may be quite laborious to determine $\tilde{t} = \arg\min_{t\in\Pi} \tilde{W}(t)$ on the continuum set Π. We can simplify this problem by discretizing the maximum problem by the following theorem.

Theorem 11. *Choose a sequence of positive integers L converging to infinity with n, and choose a sequence* (t_i) *such that* $\{t_1, t_2, t_3, \cdots\}$ *is dense in* Π*. Replace* $\tilde{t}$ *by* $t = argmax_{t\in\{t_1,\cdots,t_L\}} \tilde{W}(t)$*. Then the previous two theorems carry over.*

Proof. Similar to the Proof of Bierens (1990) Theorem 5. □

2.3.1 Local Alternative Analysis

In this Subsection, we will compare the local alternative properties of the improved Bierens (1990) test proposed in this paper with the Bierens (1990) test where a NLS estimator is employed.

We consider the following local alternative:

$$H_1^L : Y = f(X, \theta_0) + \frac{g(X)}{\sqrt{n}} + e(\theta_0), \tag{2.12}$$

where the error $e(\theta_0)$ is the same as under the null hypothesis. Under this local alternative,

$$\tilde{M}(\hat{\theta},t) = n^{-1/2}\sum_{j=1}^{n} e_j(\theta_0)\left[\exp(t'\Phi(X_j)) - A\left(X_j\right)' b(t)\right] + \frac{1}{n}\sum_{j=1}^{n} g\left(X_j\right)\left[\exp(t'\Phi(X_j)) - A\left(X_j\right)' b(t)\right] + o_p(1).$$

Theorem 12. *Let Assumptions 1-5 and 7 hold, when $K \to \infty$ and $K^4/n \to 0$, then $\tilde{W}$ converges weakly to z^2 under H_1^L, where z is a Gaussian element of $C(\Pi)$ with mean function*

$$\eta_*(t) = \frac{E^2\left\{g(X)\left[\exp(t'\Phi(X)) - A_*(X)' b(t)\right]\right\}}{s_*^2(t)}$$

and covariance function $\Gamma(t_1,t_2)$ defined in (2.11). Furthermore, under H_1^L, $\tilde{W}(\bar{t}) \to \chi_1^2(\eta_*(\bar{t}))$.

Proof. See Appendix. □

Note that under H_1^L, Bierens (1990) test has drift

$$\eta_{NLS}(t) = \frac{E^2\left\{g(X)\left[\exp(t'\Phi(X)) - A_{NLS}(X)' b(t)\right]\right\}}{s_{NLS}^2(t)},$$

where

$$A_{NLS}(X) = \left[E\left(\frac{\partial f(X,\theta_0)}{\partial\theta}\frac{\partial f(X,\theta_0)}{\partial\theta'}\right)\right]^{-1}\frac{\partial f(X,\theta_0)}{\partial\theta},$$

$$s_{NLS}^2(t) = E\left\{\left[Y - f(X,\theta_0)\right]^2\left[\exp(t'\Phi(X)) - A_{NLS}(X)' b(t)\right]^2\right\}.$$

Also note that

$$A_*(X) = \left\{E\left[\frac{\partial f(X,\theta_0)}{\partial\theta}\sigma_0^{-2}\left(X_j\right)\frac{\partial f(X,\theta_0)}{\partial\theta'}\right]\right\}^{-1}\frac{\partial f(X,\theta_0)}{\partial\theta}\sigma_0^{-2}\left(X_j\right).$$

When there exists homoskedasticity, our improved test is asymptotically equivalent to the Bierens (1990) test. When there exists conditional heteroskedasticity, as long as $\eta_*(t) > \eta_{NLS}(t)$ for any $t \in \Pi$ we will obtain a more powerful test. The fact that $s_*^2(t) \leq s_{NLS}^2(t)$ provides the possible improvement in the power. However $\eta_*(t)$ and $\eta_{NLS}(t)$ also depend on $E[g(X)A_*(X)]$ and $E[g(X)A_{NLS}(X)]$ respectively. It is possible that there are some alternatives that we could not obtain improved tests.

2.4 Monte Carlo Simulations

We show in the following Monte Carlo simulations the finite sample properties of the improved test, comparing with the Bierens test where the NLS estimator is employed.

Let z_j, v_{1j}, v_{2j}, and u_j be independent random drawings from the standard normal distribution, and let the regressors be $X_{1j} = z_j + v_{1j}$, $X_{2j} = z_j + v_{2j}$. The dependent variable is generated according to

$$Y_j = 1 + X_{1j} + X_{2j} + e_j$$

Under the null, when the homoskedasticity is assumed, $e_j = u_j$, under heteroskedasticity, $e_j = \left(0.1 + 0.5x_{1j}^2\right)^{1/2} u_j$. In both cases, OLS is employed to obtain the parameter estimator. Based on the OLS estimator and residuals, we calculate Bierens (1990) test and our improved Bierens (1990) test. Following Bierens (1990), we choose $L = [n/10] - 1$ and $\Pi = [1,5] \times [1,5]$. $(t_1, \cdots, t_L)'$ have been drawn randomly from the uniform distribution on Π. $(t_1, \cdots, t_K)'$ have been drawn randomly from the uniform distribution on subset $[-1,1] \times [-1,1]$. We use the weighting function with $\Phi(x_1, x_2) = \left(\tan^{-1}(x_1/2), \tan^{-1}(x_2/2)\right)'$. The Monte Carlo simulations have been conducted for sample size 200 and 400 with four sets of values of the penalty parameters

$$\gamma = 1, \rho = 0.5 \qquad \gamma = 0.5, \rho = 0.5$$
$$\gamma = 0.25, \rho = 0.5 \quad \gamma = 0.25, \rho = 0.25.$$

For both sample sizes, we report the results of choosing K starting from 3 to 20. Note that $K = 3$ is minimum dimension requirement for the model.

For the empirical size check, 10,000 replications are used. We report the results in Figures 2.6-2.6. Firstly note that in both homoskedasticity and heteroskedasticity cases, the empirical size of the new statistic is quite stable or becomes stable quickly as K increases. In the homoskedasticity case, its empirical size properties are comparable to Bierens (1990)'s statistic, even when $K = 3$. In the heteroskedasticity case, under reasonable penalty parameters situations, while Bierens (1990) statistic is undersized, the empirical size of the new statistic is a little bit undersized, when K is a small number; it becomes very close to the nominal size, when K increases. Note that when penalty parameters are too small ($\gamma = 0.25, \rho = 0.25$), both statistics are all heavily oversized.

For the power check, 1000 replications are used. We consider the following alternatives

DGP 1.1: $Y_j = 1 + X_{1j} + X_{2j} + v_{1j}v_{2j} + u_j$.

DGP 1.2: $Y_j = 1 + X_{1j} + X_{2j} + v_{1j}v_{2j} + \left(0.1 + 0.5x_{1j}^2\right)^{1/2} u_j$.

DGP 2.1: $Y_j = 1 + X_{1j} + X_{2j} + \left(1 + X_{1j} + X_{2j}\right)\exp\left[-0.01\left(1 + X_{1j} + X_{2j}\right)^2\right] + u_j$.

DGP 2.2: $Y_j = 1 + X_{1j} + X_{2j} + \left(1 + X_{1j} + X_{2j}\right)\exp\left[-0.01\left(1 + X_{1j} + X_{2j}\right)^2\right] + \left(0.1 + 0.5x_{1j}^2\right)^{1/2} u_j$.

DGP 3.1: $Y_j = 1 + X_{1j} + X_{2j} + \sin\left(1 + X_{1j} + X_{2j}\right) + u_j$.

DGP 3.2: $Y_j = 1 + X_{1j} + X_{2j} + \sin\left(1 + X_{1j} + X_{2j}\right) + \left(0.1 + 0.5x_{1j}^2\right)^{1/2} u_j$.

DGP 4.1: $Y_j = 1 + X_{1j} + X_{2j} + \cos\left(1 + X_{1j} + X_{2j}\right) + u_j$.

DGP 4.2: $Y_j = 1 + X_{1j} + X_{2j} + \cos\left(1 + X_{1j} + X_{2j}\right) + \left(0.1 + 0.5x_{1j}^2\right)^{1/2} u_j$.

Remark: The first alternative is considered by Bierens (1990); the second is same as the alternative 3 in Escanciano (2006a); The third is similar to the alternative 4 in Escanciano (2006a). The fourth just changes the sin function in the third alternative into a cos function.

We report the results in Figures 2.5-2.12. To save space, results of sample size 200 and 400 are reported in one figure. For the first alternative, our new statistic is worse than Bierens (1990)'s in both homoskedasticity and heteroskedasticity cases when K is large. But it is comparable to Bierens (1990)'s test when K is small. Note that this is a quite special alternative. We still can obtain consistent estimators of X_{1j} and X_{2j} under this alternative, since $E(v_{1j}v_{2j}X_{ij}) = 0$ for $i = 1, 2$.

For the alternative 2 and 3, under homoskedasticity, the power of the new statistic is quite close to Bierens (1990)'s test for all the K. In heteroskedasticity case, the new statistic has very good power properties even when K is small; as K increases, the difference of the power between the new statistic and Bierens (1990)'s test reaches as much as 20%. For the alternative 4, surprisingly, the new statistic is even more powerful than Bierens (1990)'s test in the homoskedasticity case. In the heteroskedasticity case, the power improvement is even more dramatic.

All in all, the new statistic has good size properties and improves the power significantly when there exists heteroskedasticity of unknown form. The choice of K is not restrictive. For a large number of alternatives, the general pattern of the test results is that when K increases, we obtain better power properties. This is in accordance with the intuition behind the GMM estimation: adding more moments can not hurt, asymptotically, in the sense that the asymptotic variance of the GMM estimator decreases.

2.5 Conclusion

In this paper, we propose a new testing statistic in consistent conditional moment testing framework, exploiting the duality property of one class of functions for both consistent specification testing and efficient estimation of regression. It has be shown that the new statistic is robust to heteroskedasticity and can show some efficiency gain under heteroskedasticity of unknown form. Further, it is quite easy to compute, only a $\sqrt{n}$-consistent estimator needed. Based on our new testing statistic, a new version of Bierens (1990) test is then proposed. Monte Carlo simulations show that our Bierens (1990) testing method employing our new test statistic has good finite sample properties in the presence of heteroskedasticiy of unknown form for a large number of alternatives.

2.6 Appendix

Proof of Corollary 1. Without loss of generality we may assume that X is bounded itself, so that we may choose $\Phi(X) = X$. We set $w(t_1'X) = 1$. It is always possible to normalize $q^K(X)$ into this case when $w(t_1'X) \neq 1$. Firstly it is easy to check that $q(t_jX) \in L_2$, for $j = 1, 2, \cdots$. For $K = 2, 3, \cdots$, let

$$\zeta_K(X) = \sum_{j=1}^{K} \alpha_{K,j} w\left(t_j'X\right),$$

where $\alpha_{K,K} = 1$, and the other $\alpha_{K,j}$ are chosen such that

$$E\left[\zeta_K(X) w\left(t_j'X\right)\right] = 0 \text{ if } j < K.$$

For $K = 1, 2, \cdots$, define function $\psi_K(X)$ on the range of X such that

$$\psi_1(X) = 1,$$

$$\psi_K(X) = \begin{cases} \zeta_K(X)/\left[E\zeta_K(X)^2\right]^{1/2}, & \text{if } \left[E\zeta_K(X)^2\right] > 0 \\ 0, & \text{if } \left[E\zeta_K(X)^2\right] = 0 \end{cases}$$

for $K > 1$. Then $\psi_K(X)$, $K = 1, 2, \cdots$ form an orthonormal system of the Hilbert space of H of Borel measurable functions φ on the range of X satisfying $E\left[\varphi(X)^2\right] < \infty$, with inner product $(\psi_K, \varphi) = E\left[\psi_K(X)\varphi(X)\right]$. Then by Theorem 2.4.2 of Brockwell and Davis (1991), for any ε, there exists a positive integer K and constant $c_1, \cdots, c_K$ such that

$$\left[E\left(\varphi(X) - \sum_{j=1}^{K} c_j \psi_j(X)\right)^2\right]^{1/2} < \varepsilon,$$

then the conclusion follows. □

Proof of Theorem 1.

$$
\begin{aligned}
\tilde{M}(\hat{\theta},t) &= n^{-1/2}\sum_{j=1}^{n}[Y_j - f(X_j,\theta_0)]w(t'\Phi(X_j)) - b(t)'n^{1/2}\left(\hat{\theta}-\theta\right) + o_p(1) \\
&- \left\{ n^{-1/2}\sum_{j=1}^{n}[Y_j - f(X_j,\theta_0)]q^K\left(X_j\right) - \Lambda n^{1/2}\left(\hat{\theta}-\theta\right) + o_p(1) \right\}' \\
&\times \left\{ \left[\left(\Lambda'\Omega^{-1}\Lambda\right)^{-1}\Lambda'\Omega^{-1} \right]' b(t) + o_p(1) \right\} \\
&= n^{-1/2}\sum_{j=1}^{n}[Y_j - f(X_j,\theta_0)]w(t'\Phi(X_j)) \\
&- \left[\left(\Lambda'\Omega^{-1}\Lambda\right)^{-1}\Lambda'\Omega^{-1} n^{-1/2}\sum_{j=1}^{n}[Y_j - f(X_j,\theta_0)]q^K\left(X_j\right) \right]' b(t) + o_p(1) \\
&= n^{-1/2}\sum_{j=1}^{n}[Y_j - f(X_j,\theta_0)][w(t'\Phi(X_j)) - A\left(X_j\right)' b(t)] + o_p(1).
\end{aligned}
$$

Since by the mean value theorem we have

$$
n^{-1/2}\sum_{j=1}^{n}[Y_j - f(X_j,\hat{\theta})]w(t'\Phi(X_j)) = n^{-1/2}\sum_{j=1}^{n}[Y_j - f(X_j,\theta_0)]w(t'\Phi(X_j)) - \hat{b}(\dot{\theta},t)n^{1/2}\left(\hat{\theta}-\theta\right)
$$

where $\dot{\theta}$ lies on the line joining $\hat{\theta}$ and θ_0, $\dot{\theta}$, $\hat{\theta} \in \Delta$, an open convex neighborhood of θ_0, with $\dot{\theta} \xrightarrow{p} \theta_0$. By Assumption 4 and the fact that $E[w^2(t'\Phi(X))]$ is finite, the dominance condition holds by Cauchy-Schwarz inequality

$$
\begin{aligned}
E\left[\sup_{\theta\in\Delta}\left\| w(t'X)\frac{\partial f(X,\theta)}{\partial\theta'}\right\|\right] &= E\left[\left\| w(t'\Phi(X))\right\| \sup_{\theta\in\Delta}\left\|\frac{\partial f(X,\theta)}{\partial\theta'}\right\|\right] \\
&< \left[E\left\| w(t'\Phi(X))\right\|^2\right]^{1/2}\left[E\sup_{\theta\in\Delta}\left\|\frac{\partial f(X,\theta)}{\partial\theta'}\right\|^2\right]^{1/2} \\
&< \infty.
\end{aligned}
$$

So we have weakly uniformly convergence of

$$
p\lim\sup_{\theta\in\Delta}\left\|\frac{1}{n}\sum_{j=1}^{n} w(t'\Phi(X))\frac{\partial f(X_j,\theta)}{\partial\theta'} - E\left[w(t'\Phi(X))\frac{\partial f(X,\theta)}{\partial\theta'}\right]\right\| = 0,
$$

then $\hat{b}(\dot{\theta},t) \xrightarrow{p} b(t)$, $\hat{b}(\hat{\theta},t) \xrightarrow{p} b(t)$. So

$$n^{-1/2}\sum_{j=1}^{n}[Y_j - f(X_j,\hat{\theta})]w(t'\Phi(X_j)) = n^{-1/2}\sum_{j=1}^{n}[Y_j - f(X_j,\theta_0)]w(t'\Phi(X_j)) - b(t)'n^{1/2}\left(\hat{\theta}-\theta\right) + o_p(1).$$

Similarly

$$\begin{aligned} E\left[\sup_{\theta\in\Delta}\left\|q^K(X)\frac{\partial f(X,\theta)}{\partial\theta'}\right\|\right] &= E\left[\left\|q^K(X)\right\|\sup_{\theta\in\Delta}\left\|\frac{\partial f(X,\theta)}{\partial\theta'}\right\|\right] \\ &< \left[E\left\|q^K(X)\right\|^2\right]^{1/2}\left[E\sup_{\theta\in\Delta}\left\|\frac{\partial f(X,\theta)}{\partial\theta'}\right\|^2\right]^{1/2} \\ &< \infty. \end{aligned}$$

Then

$$p\lim\sup_{\theta\in\Delta}\left\|\frac{1}{n}\sum_{j=1}^{n}q^K(X)\frac{\partial f(X,\theta)}{\partial\theta'} - E\left[q^K(X)\frac{\partial f(X,\theta)}{\partial\theta'}\right]\right\| = 0,$$

By similar argument, we have

$$n^{-1/2}\sum_{j=1}^{n}[Y_j - f(X_j,\hat{\theta})]q^K\left(X_j\right) = n^{-1/2}\sum_{j=1}^{n}[Y_j - f(X_j,\theta_0)]q^K\left(X_j\right) - \Lambda n^{1/2}\left(\hat{\theta}-\theta\right) + o_p(1),$$

and $\hat{\Lambda}\left(\hat{\theta}\right) \xrightarrow{p} \Lambda$.

Similarly

$$\begin{aligned} E\left[\sup_{\theta\in\Delta}\left\|q^K(X)q^K(X)'[Y-f(X,\theta)]^2\right\|\right] &= E\left[\left\|q^K(X)q^K(X)'\right\|\sup_{\theta\in\Delta}\left\|[Y_j-f(X,\theta)]^2\right\|\right] \\ &< \left[E\left\|q^K(X)q^K(X)'\right\|^2\right]^{1/2}\left[E\sup_{\theta\in\Delta}[Y-f(X,\theta)]^4\right]^{1/2} \\ &< \infty. \end{aligned}$$

Then by similar argument, $\hat{\Omega}\left(\hat{\theta}\right) \xrightarrow{p} \Omega$. Also by Assumptions 4 and 5, for any $K > D$, Ω is positive definite. By continuous mapping theorem, for $K > D$, $\hat{\Lambda}\left(\hat{\theta}\right)'\hat{\Omega}\left(\hat{\theta}\right)^{-1}\hat{\Lambda}\left(\hat{\theta}\right) \xrightarrow{p} \Lambda'\Omega^{-1}\Lambda$, $\hat{\Lambda}\left(\hat{\theta}\right)'\hat{\Omega}\left(\hat{\theta}\right)^{-1} \xrightarrow{p} \Lambda'\Omega^{-1}$.

To prove (2.9), we rewrite $\tilde{M}(\hat{\theta},t)$ as

$$\tilde{M}(\hat{\theta},t)=[1,-\left[\left(\Lambda'\Omega^{-1}\Lambda\right)^{-1}\Lambda'\Omega^{-1}\right]'b(t)]n^{-1/2}\sum_{j=1}^{n}[Y_j-f(X_j,\theta_0)]\left(w(t'\Phi(X_j)),q^K\left(X_j\right)'\right)'+o_p(1). \tag{2.13}$$

By Lindberg-Feller central limit theory and slutsky theorem, we have

$$\tilde{M}(\hat{\theta},t)\xrightarrow{d}N\left[0,s^2(t)\right]$$

□

Proof of Theorem 2. By the mean value theorem, we have

$$\hat{M}\left(\check{\theta},t\right)=n^{-1/2}\sum_{j=1}^{n}[Y_j-f(X_j,\theta_0)]w\left(t'X_j\right)-b(t)'n^{1/2}\left(\check{\theta}-\theta_0\right)+o_p(1). \tag{2.14}$$

From (2.10), we have

$$n^{1/2}\left(\check{\theta}-\theta_0\right)=\left[\Lambda'\Omega^{-1}\Lambda\right]^{-1}\Lambda'\Omega^{-1}n^{-1/2}\sum_{j=1}^{n}q^K\left(X_j\right)[Y_j-f(X_j,\theta_0)]+o_p(1).$$

Plug in (2.14), so we have

$$\begin{aligned}\hat{M}\left(\check{\theta},t\right)=&\,n^{-1/2}\sum_{j=1}^{n}[Y_j-f(X_j,\theta_0)]w\left(t'X_j\right)\\&-\left[\left(\Lambda'\Omega^{-1}\Lambda\right)^{-1}\Lambda'\Omega^{-1}\right]'b(t)n^{-1/2}\sum_{j=1}^{n}q^K\left(X_j\right)[Y_j-f(X_j,\theta_0)]+o_p(1).\end{aligned}$$

Since we also have

$$\tilde{M}(\hat{\theta},t)=n^{-1/2}\sum_{j=1}^{n}[Y_j-f(X_j,\theta_0)][w(t'\Phi(X_j))-A\left(X_j\right)'b(t)]+o_p(1),$$

then

$$\tilde{M}(\hat{\theta},t)=\hat{M}\left(\check{\theta},t\right)+o_p(1).$$

□

Proof of Theorem 3. Based on Theorem 2, we only need establish that GMM estimator $\check{\theta}$ reaches the semiparametric efficiency bound. It is easy to check Assumptions 1-5 in Donald et al. (2003) are all satisfied. Note that Assumption 1 in Donald et al. (2003) corresponds to Corollary 1 in this paper, by Theorem 5.4 of Donald et al. (2003), we have $K \to \infty$ and $\xi(K)^2 K/n \to 0$, the GMM estimator $\check{\theta}$ satisfies

$$n^{1/2}\left(\check{\theta}-\theta\right) \xrightarrow{d} N(0,V)$$

where $V = \left\{E\left[E\left[(Y-f(X,\theta_0))^2|X\right]^{-1}\frac{\partial f(X,\theta_0)}{\partial\theta}\frac{\partial f(X,\theta_0)}{\partial\theta'}\right]\right\}^{-1}$. Since $\check{\theta}$ reaches semiparametric efficiency bound, It is straightforward to obtain s_*^2. Furthermore, the result of $A_*(X)$ comes directly from Lemma A4 in Donald et al. (2004). □

Proof of Lemma 2. We still assume that that X is bounded itself, so that we may choose $\Phi(X) = X$. We set $\exp(t_1'X) = 1$. Note that we can always normalize $q^K(X)$ into $q^K(X) = (1, \exp((t_2-t_1)'X), \cdots, \exp((t_K-t_1)'X))'$. For $K = 1,2,\cdots$, since the probability density function of X is bounded away from zero, then the second moment of $\zeta_K(X)$ defined in the proof of Lemma 1 is larger than zero, that is $\left[E\zeta_K(X)^2\right] > 0$ almost surely. So for $K = 2,3,\cdots$,

$$\psi_K(X) = \zeta_K(X)/\left[E\zeta_K(X)^2\right]^{1/2}.$$

For any K, define $\tilde{q}^K(X) = (\psi_1(X),\cdots,\psi_K(X))'$. When $t_{jK} \neq t_{iK}$ for $j,i = 1,\cdots,K$, $\tilde{q}^K(X)$ is linear transformation of $q^K(X)$: $q^K(X) = B\tilde{q}^K(X)$, where B is a nonsingular lower triangular matrix. So $\tilde{q}^K(X) = B^{-1}q^K(X)$. Since $(\psi_1(X),\cdots,\psi_K(X))'$ is an orthonormal set, so $E\left(\tilde{q}^K(X)\tilde{q}^K(X)'\right) = I_K$, which means that the condition of nonsingularity is satisfied.

Note that $\|\tilde{q}^K(X)\| = [\sum_{j=1}^K \psi_j(X)^2]^{1/2}$, $\zeta_K(X) = \sum_{j=1}^K \alpha_{K,j}\exp\left(t_j'X\right)$. So we have

$$\begin{aligned}\sup_{X\in Z}\|\tilde{q}^K(X)\| &\leq C[\sum_{k=1}^K k^2]^{1/2}\\ &\leq CK^{3/2}.\end{aligned}$$

To prove $s^2(t) > 0$, note that for any $t \in \Pi$, $t \neq t_j$ for $j = 1,\cdots,K$. Denote $q^{K+1}(X) = \left(\exp(t_1'\Phi(X)),\cdots,\exp(t_K'\Phi(X)),\exp(t'\Phi(X))\right)'$. Then based on Lemma 2 we can obtain that

$E\left(q^{K+1}(X)q^{K+1}(X)'\right)$ has smallest eigenvalue bounded away from zero. Note that $E[(Y - f(X,\theta_0))|X]^2 > 0$, then $E\left(\left(Y_j - f(X_j,\theta_0)\right)^2 q^{K+1}(X)q^{K+1}(X)'\right)$ is positive definite. From (2.13) in the proof of Theorem 1 it is easy to obtain that $s^2(t) > 0$. □

Proof of Theorem 5. The result under H_1 follows straightforwardly from the uniform law of large numbers. Under H_0, Define

$$z_n(t) = n^{-1/2}\sum_{j=1}^{n}[Y_j - f(X_j,\theta_0)][\exp(t'\Phi(X_j)) - A\left(X_j\right)' b(t)]/\sqrt{(}s^2(t)),$$

where $b(t)$, $A\left(X_j\right)$ and $s^2(t)$ are defined in Theorem 1. Following the Proof of (2.8) in Theorem 1, we have under H_0

$$p\lim_{n\to\infty}\sup_{t\in\Pi}|\tilde{W}(t) - z_n^2(t)| = 0.$$

Further, following the Proof of Lemma 4 in Bierens (1990), we can obtain under H_0, z_n is tight. It is also easy to prove that for arbitrary $t_1,\cdots,t_m$ in Π, $(z_n(t_1),\cdots,z_n(t_m))'$ is asymptotically distributed as $(z(t_1),\cdots,z(t_m))'$. Then z_n converges weakly to z. Following the functional limit theory of Billingsley (1968 p. 47), we have the results. □

Proof of Theorem 8. Similar to the proof of Theorem 5. □

Figure 2.1: Size of testing at 5% level, Sample size 200, $u_j = e_j$

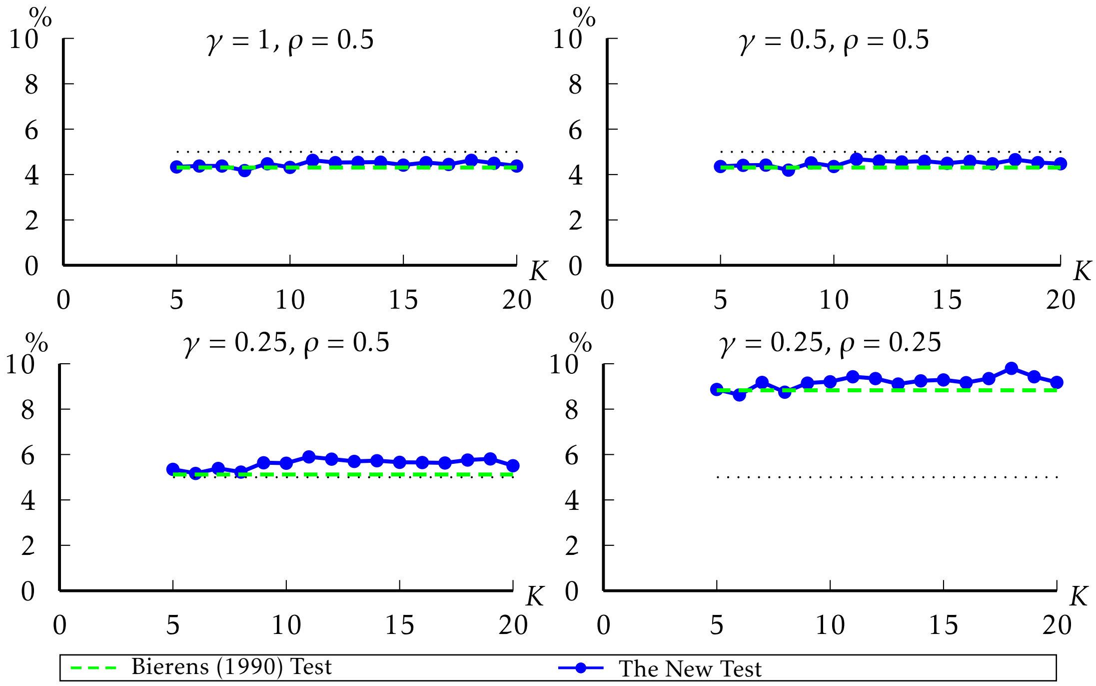


Figure 2.2: Size of testing at 5% level, Sample size 400, $u_j = e_j$

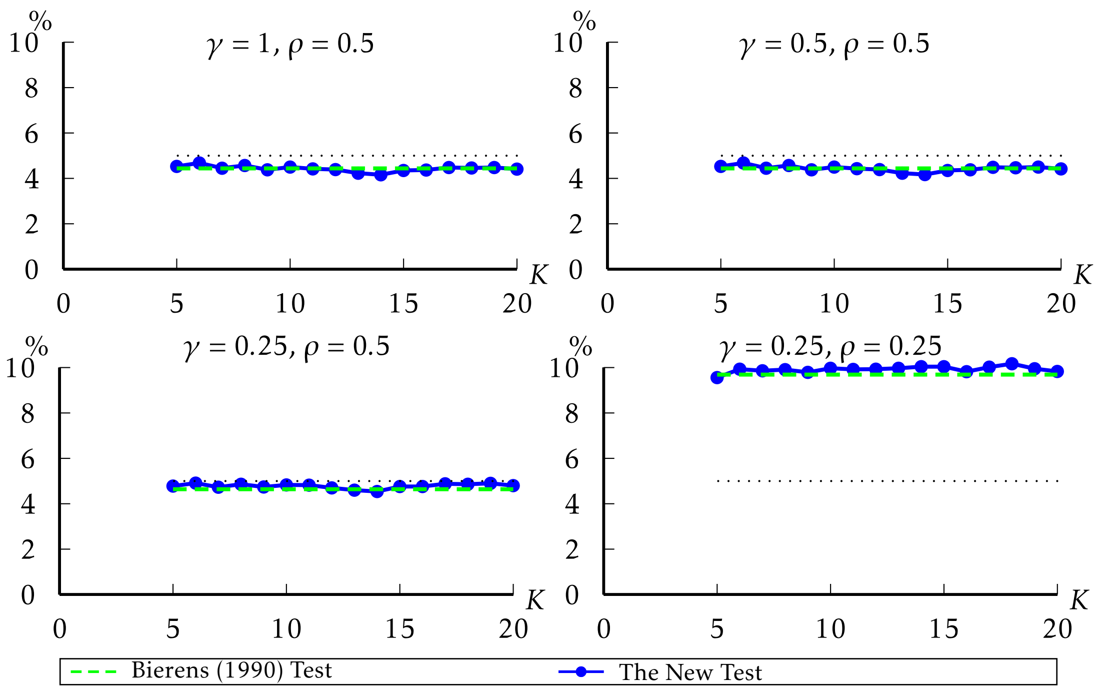


Figure 2.3: Size of testing at 5% level, Sample size 200, $u_j = \left(0.1 + 0.5x_{1j}^2\right)^{1/2} e_j$

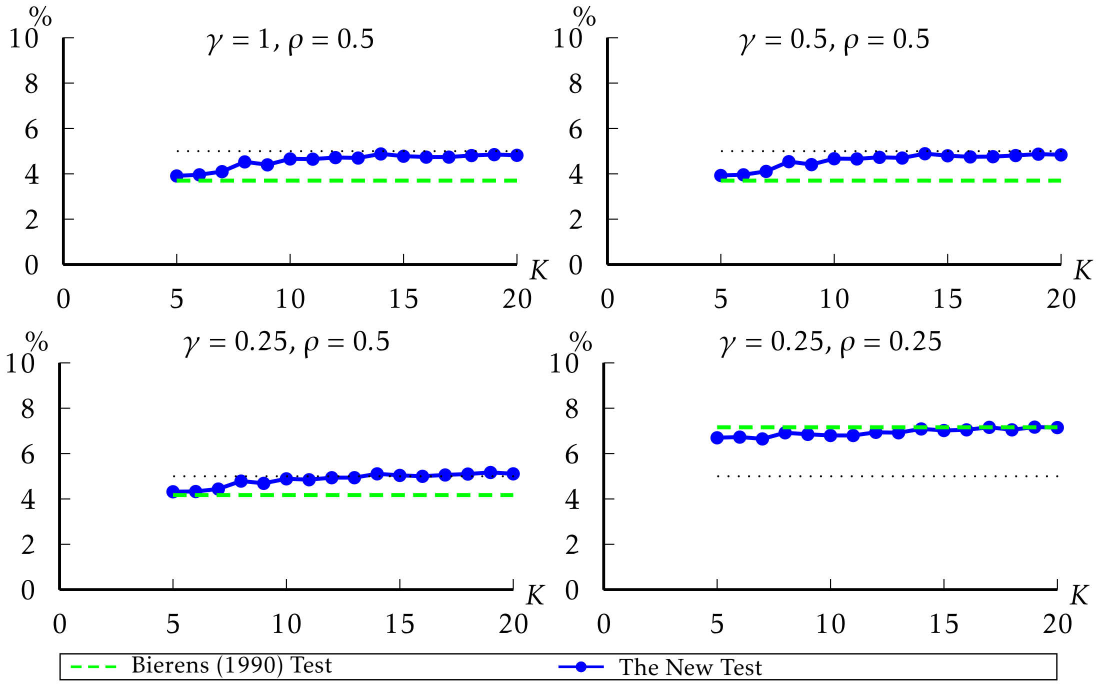


Figure 2.4: Size of testing at 5% level, Sample size 400, $u_j = \left(0.1 + 0.5x_{1j}^2\right)^{1/2} e_j$

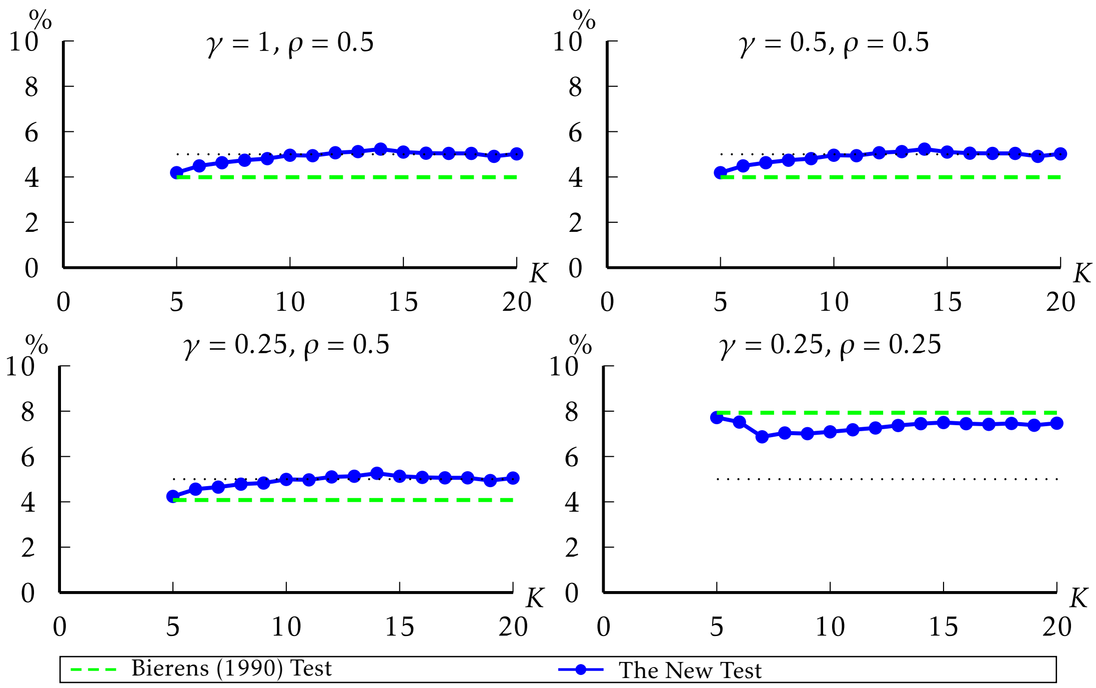


Figure 2.5: Power of testing at 5% level, DGP 1.1

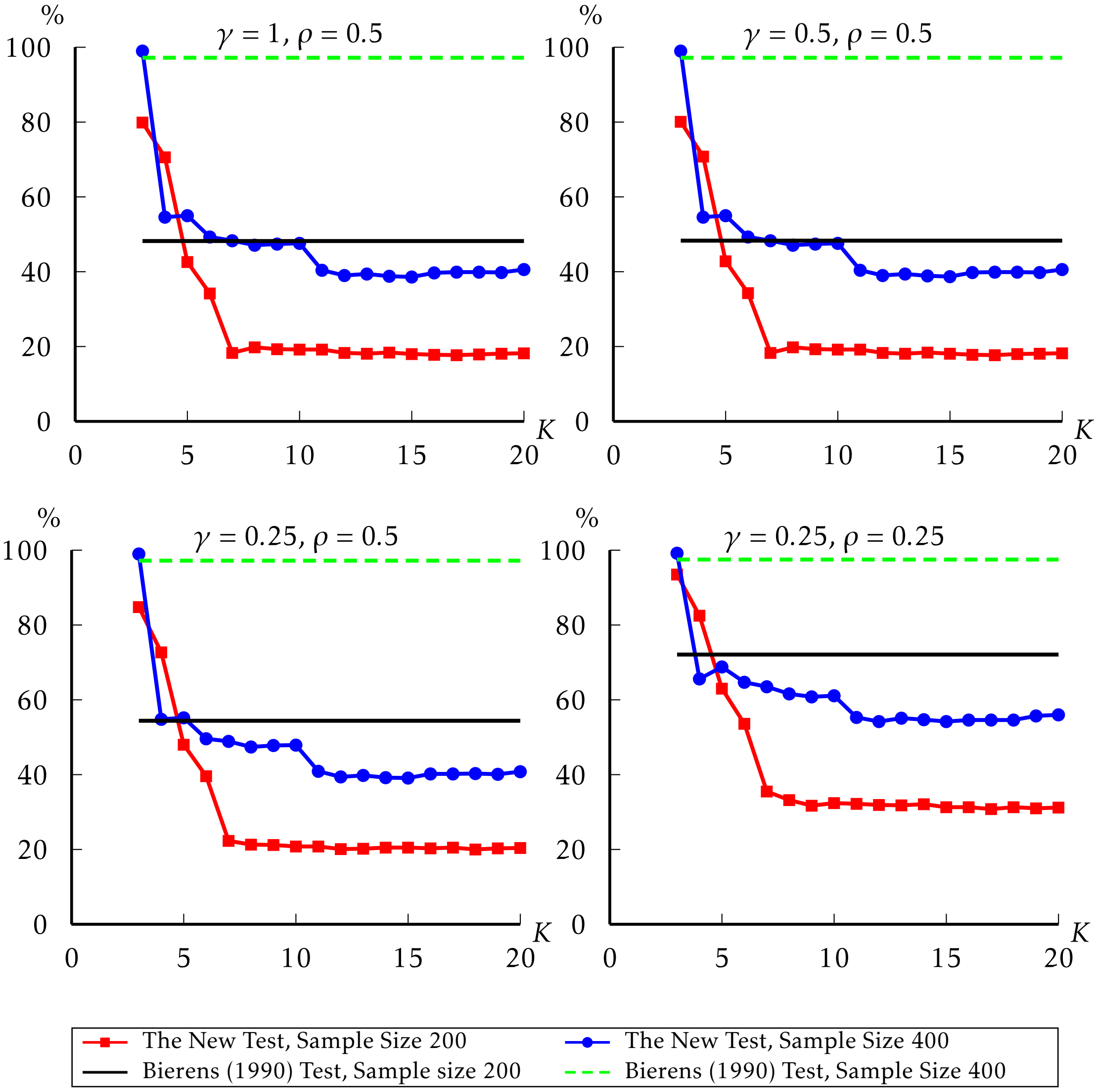


Figure 2.6: Power of testing at 5% level, DGP 1.2

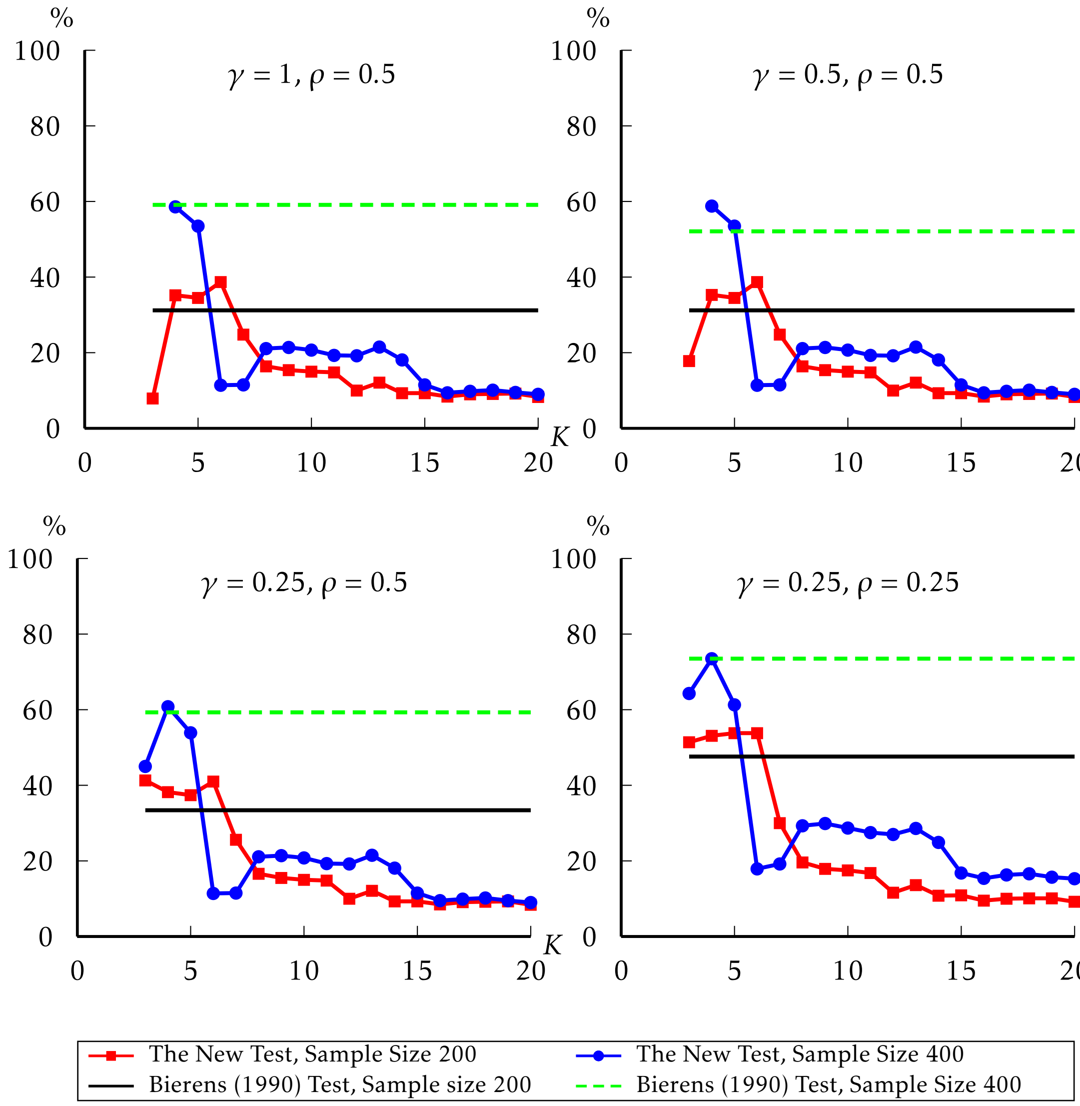


Figure 2.7: Power of testing at 5% level, DGP 2.1

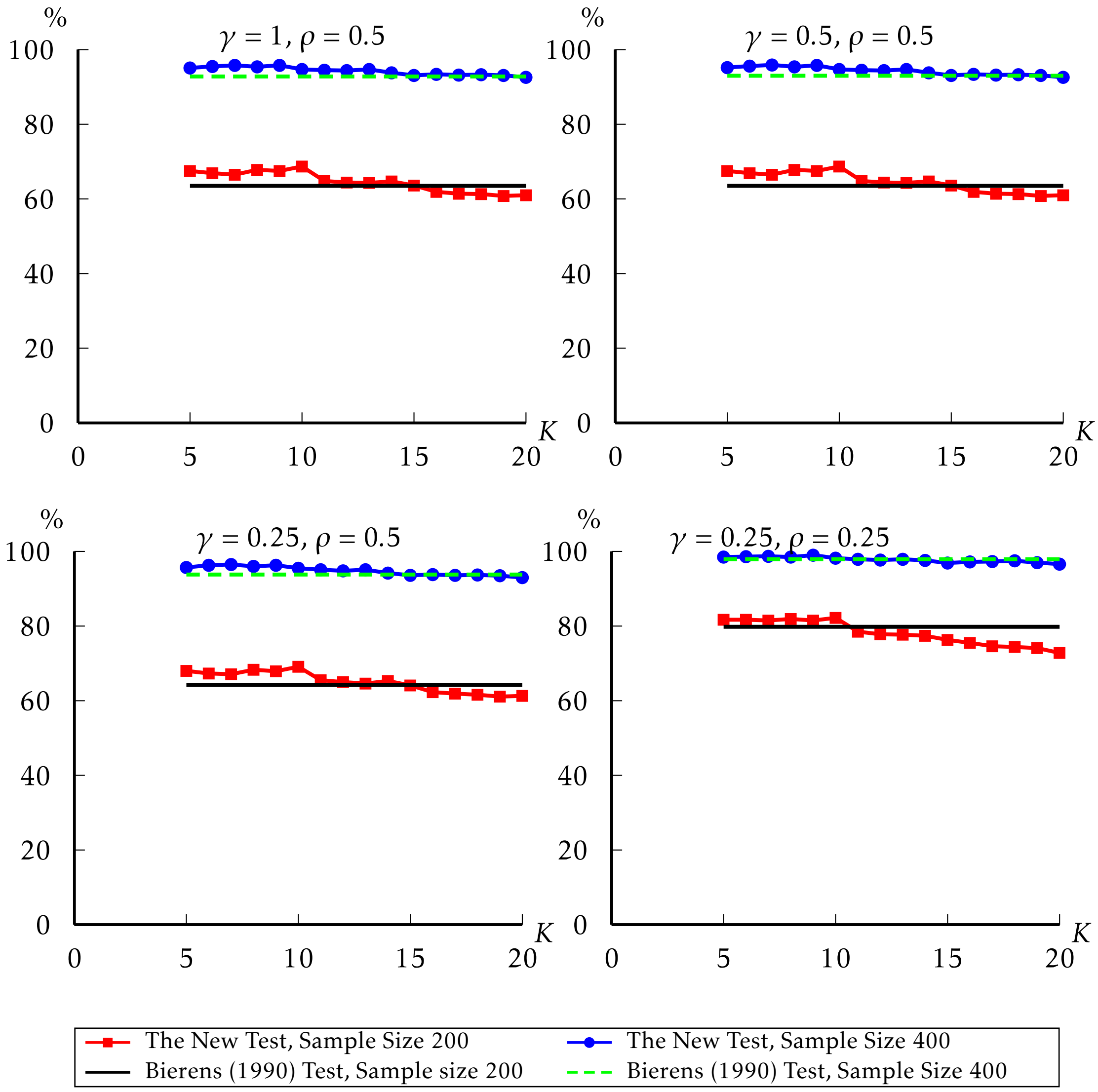


Figure 2.8: Power of testing at 5% level, DGP 2.2

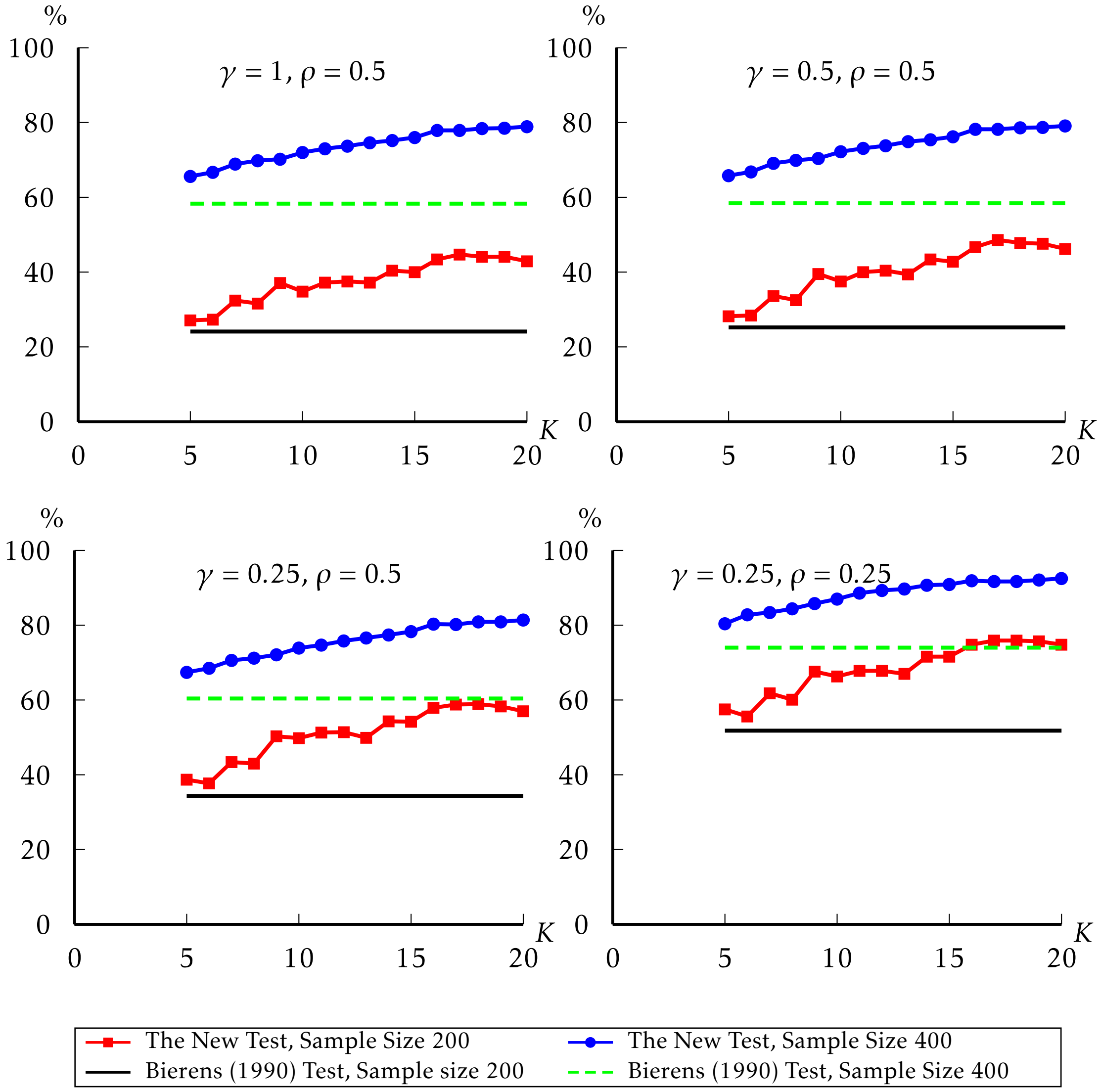


Figure 2.9: Power of testing at 5% level, DGP 3.1

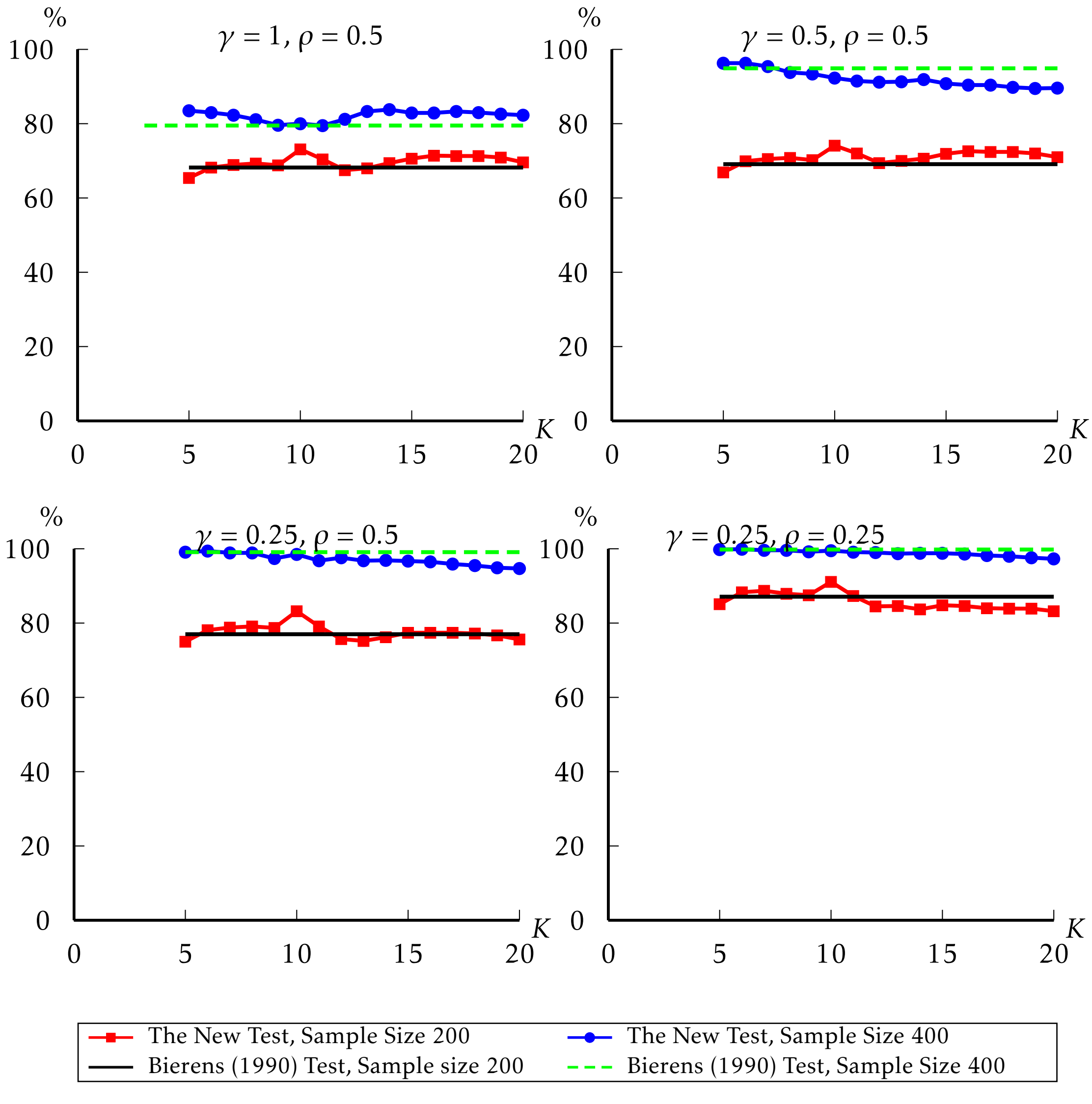


Figure 2.10: Power of testing at 5% level, DGP 3.2

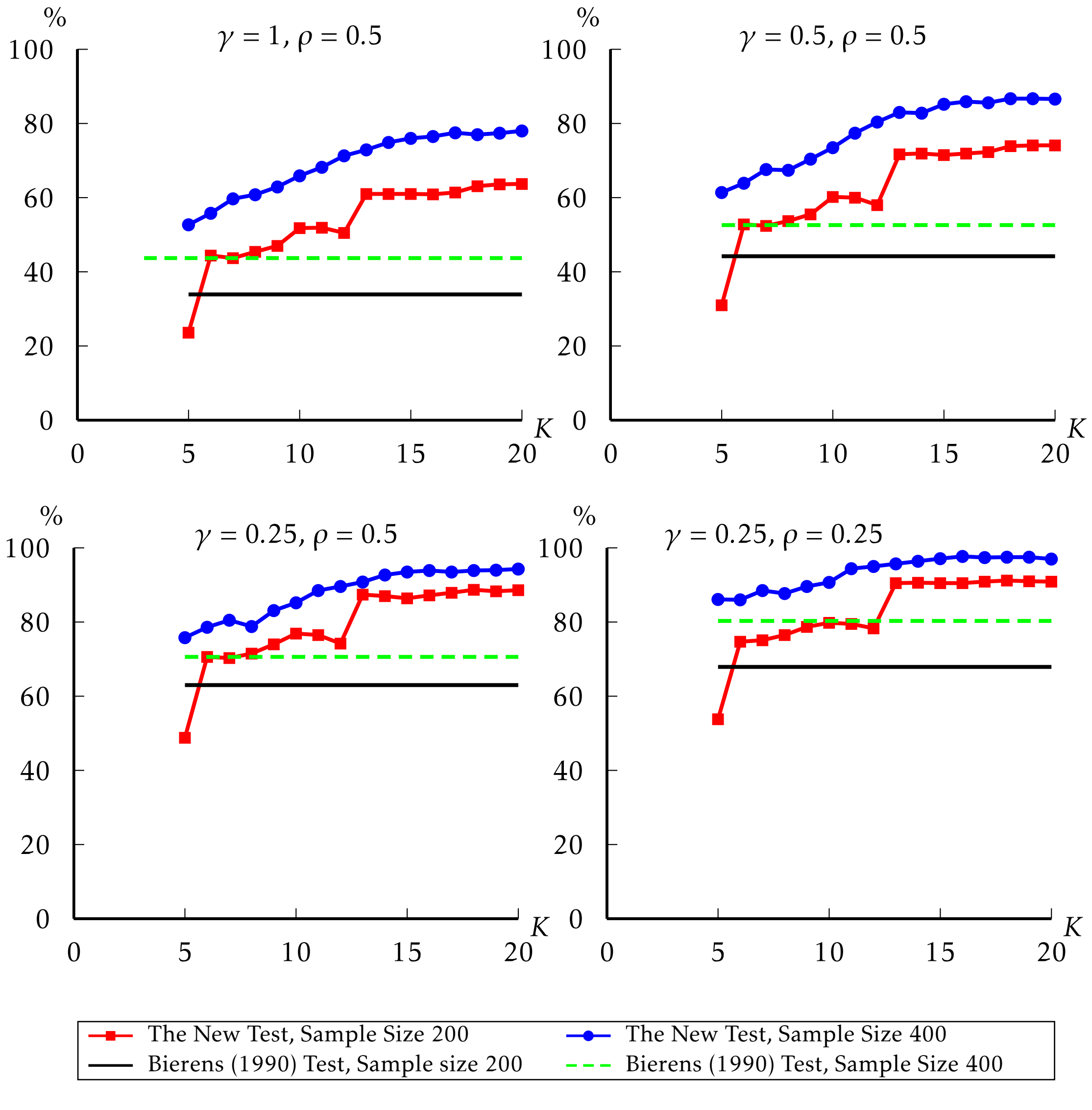


Figure 2.11: Power of testing at 5% level, DGP 4.1

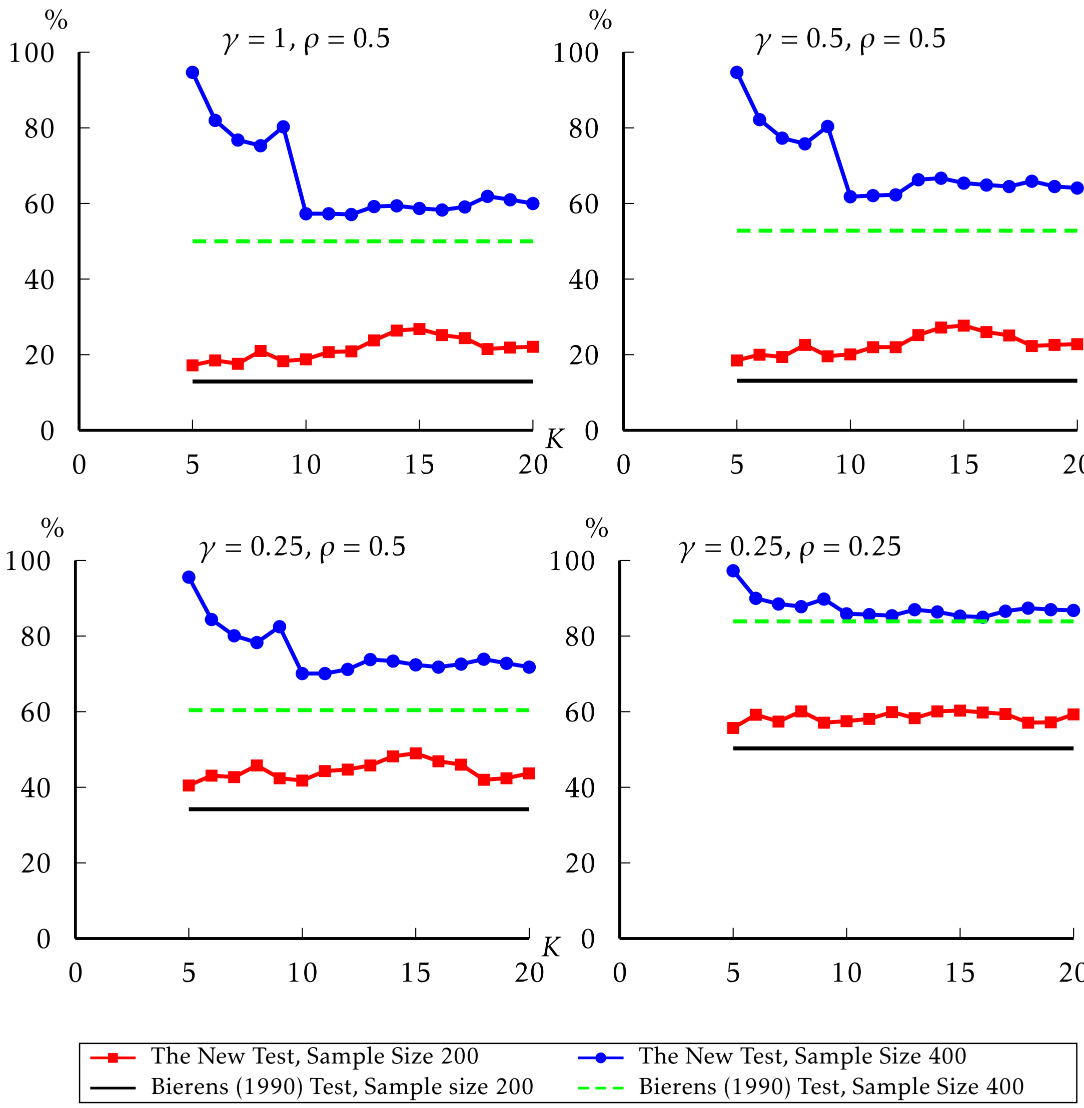
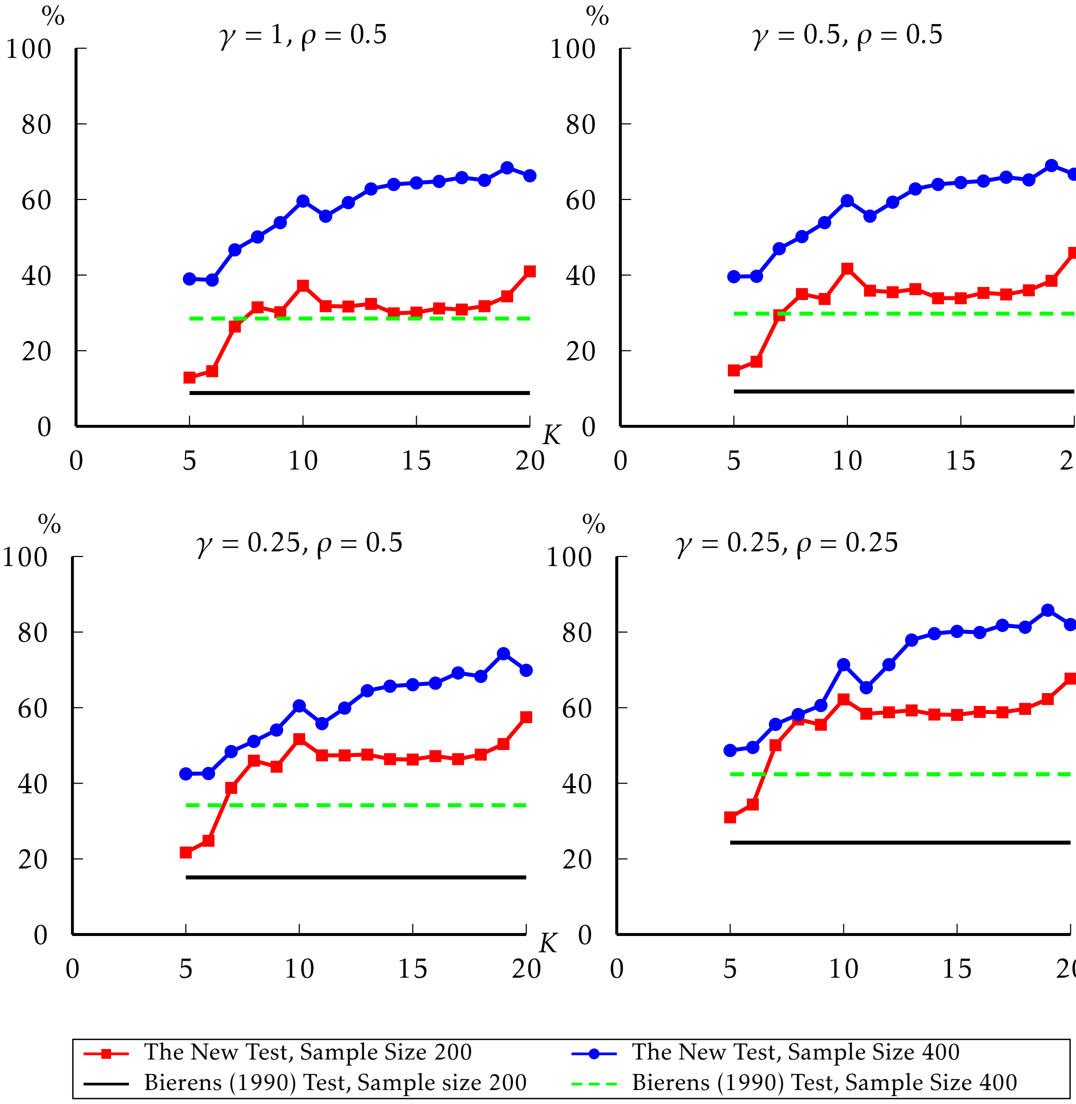


Figure 2.12: Power of testing at 5% level, DGP 4.2



Chapter 3

A Joint Portmanteau Test for Conditional Mean and Variance Time Series Models

3.1 Introduction

During the time series model building processes, it is important to check whether the residuals of a time series model are approximately uncorrelated, since a good model should be able to describe the dependence structure of the data adequately, and one important measurement of dependence is via the autocorrelation functions of residuals. But these only attend to linear dependence, so when modelling other dynamic espects, such as the conditional variance, or using nonlinear specification, further dependence measures of residuals have to be considered.

It has been a long history of studying the distribution of residual autocorrelations in linear time series models. With the populization of the Box-Jenkins modelling approach in 1970s, in ARMA modelling, Box and Pierce (1970) and Ljung and Box (1978) propose the famous portmanteau test to check the adequacy of an ARMA model. Box and Pierce (1970) and Durbin (1970) show that, although that the sample autocorrelations of ARMA model residuals under the true parameters are asymptotic independently normal distributed, this does not hold when genuine innovations are substituted by estimated residuals. More specifically, consider the ARMA (p_1, p_2) model

$$e_{\theta t} = \varphi_\theta (L) Y_t, \; t \in \mathbb{Z},$$

where $\varphi_\theta (z) = A_\theta (z) B_\theta^{-1} (z)$, $A_\theta (z) = 1 - \sum_{j=1}^{p_1} a_j z^j$, $B_\theta (z) = 1 - \sum_{j=1}^{p_2} b_j z^j$, in which $A_\theta (z)$ and $B_\theta (z)$ have no roots in common.

Consider the residual $\{e_{\theta t}\}_{t \in \mathbb{Z}}$, define the residual sample autocorrelation function

$$\rho_\theta (j) = \frac{\gamma_\theta (j)}{\gamma_\theta (0)}, \; j \in 1, 2, \cdots,$$

where $\gamma_\theta (j) = Cov\left(e_{\theta t}, e_{\theta t-j}\right)$, $j \in \mathbb{Z}$, is the corresponding autocorrelation function.

Given observations $\{Y_t\}_{t=1}^T$, ρ_θ is estimated by the sample autocorrelation function

$$\hat{\rho}_\theta (j) = \frac{\hat{\gamma}_\theta (j)}{\hat{\gamma}_\theta (0)}, \; j \in 1, \cdots, T-1,$$

where

$$\hat{\gamma}_\theta(j) = \frac{1}{T}\sum_{t=j+1}^{T}(e_{\theta t} - \bar{e}_\theta)\left(e_{\theta t-j} - \bar{e}_\theta\right)$$

is the sample autocovariance function and $\bar{e}_\theta = T^{-1}\sum_{t=1}^{T} e_{\theta t}$ is the residual sample mean.

The null hypothesis is

$$H_0 : \rho_{\theta_0}(j) = 0 \text{ for all } j \in 1, 2, \cdots \text{ and some } \theta_0 \in \Theta.$$

When $\left\{e_{\theta_0 t}\right\}_{t\in\mathbb{Z}}$ are i.i.d for some $\theta_0 \in \Theta$ or martingale difference sequence with some restrictions on higher order powers, it is well known that $\left\{\sqrt{T}\hat{\rho}_\theta(j)\right\}_{j=1}^{s}$ are asymptotically independent distributed as standard normals, So the portmanteau test

$$BPL(s) = T(T+2)\sum_{j=1}^{s}(T-j)^{-1}\hat{\rho}^2_{\theta_0}(j) \tag{3.1}$$

follows a χ^2 distribution with s degrees of freedom under the null. When we only have a maximum likelihood estimator $\hat{\theta}_T$,

$$BPL(s) = T(T+2)\sum_{j=1}^{s}(T-j)^{-1}\hat{\rho}^2_{\hat{\theta}_T}(j)$$

is approximated by Box and Pierce (1970) as χ^2 distribution with $s-(p+q)$ degrees of freedom. Note that the degrees of freedom of the Box-Pierce-Ljung test depend on the number of the estimated parameters due to the impact of the parameters estimation uncertainty.

When it comes to nonlinear time series models, the simple form of Box-Pierce-Ljung test for ARMA models breaks down. Normally Lagrange multiplier type approach is applied to derive the asymptotic theory of the quadratic form of the residual sample autocorrelations, which depends on the model and the estimator considered.

Quite recently Delgado and Velasco (2011) develop an asymptotically distribution-free transform of the sample autocorrelations of residuals in general parametric time series models. It has been shown that the proposed Box-Pierce type test statistic based on the transformed autocorrelation is not affected by the estimation effect.

For financial time series, where dynamic conditional heteroskedasticity is the norm, the ARMA model with constant variance is inadequate to describe the data. Engle (1982) proposes that the conditional variance of $e_{\theta t}$ can be modeled as

$$e_{\theta t} = h(I_{t-1}, \theta)\, \varepsilon_{\theta t},\ \varepsilon_{\theta t} \sim i.i.d\left(0, \sigma^2\right),$$

where I_{t-1}denotes the information set at t. There are many possible spefications of the function $h(I_{t-1}, \theta)$. Engle (1982) proposes the autoregressive conditional heteroskedasticity (ARCH) model. Bolleslev (1986) proposes the GARCH models. Since then the GARCH (ARCH) models become more and more popular and sucessful in economics and finance. In this case the autocorrelations of squared normalized residuals derived from these models should be useful in checking the model adequacy of $h(I_{t-1}, \theta)$. In this regard, the Box-Pierce statistic on the first s autocorrelations of squared normalized residuals is proposed by Higgins and Bera (1992) for checking of the model adequacy of the ARCH model specifications. However, a χ^2 distribtution with s degrees of freedom is used as the asymptotic distibution for the statistic, it turns out to be incorrect. Li and Mak (1994) propose the portmanteau statistic based on the correct asymptotic distribution of the autocorrelations of squared normalized residuals. Lundbergh and Tirasvirta (2002) establish the asymptotic equivalence between Li and Mak's statistic and the LM statistic.

Nowadays, dynamic econometric models that jointly parameterize conditional means and conditional variances are becoming increasingly popular in the analysis of economic time series. This kind of models appears in several dynamic contexts, such as asset pricing, portfolio choices, and market risk management. While there exist portmanteau tests for conditional mean models or for conditional variance models, the literature on joint model checking for the conditional mean and variance functions is rather scarce. The joint portmanteau test is motivated by the fact that portmanteau test only for conditional variance may lead to misleading conclusion when the conditional mean is misspecified. Wong and Ling (2005) mixed the Box-Pierce-Ljung test statistic and Li-Mak test statistic to jointly test the model adequacy of the conditional mean and variance models. In this paper, instead, we propose an asymptotic distribution-free transform of sample autocorrelation of standardized residuals and square of the standardized residuals at the same time, extending Delgado and Velasco (2011) approach to the conditional mean and variance models scenarios. We then consider the Box-Pierce type tests based on these transformation.

The Outline of the paper is as following. In Section 2, we establish the transform. Section 3 studies its asymptotic properties, and propose the Box-Pierce test statistic. Section 4 is a Monte Carlo study of the joint portmanteau test.

3.2 Transformed Residual autocorrelations

We consider the following data generating process:

$$Y_t = f(I_{t-1}, \theta) + h(I_{t-1}, \theta)\varepsilon_{\theta t},\ t \in \mathbb{Z} \tag{3.2}$$

where $f(I_{t-1},\theta)$ and $h(I_{t-1},\theta)$ are the parametric specifications for $f(I_{t-1})$ and $h(I_{t-1})$, respectively; I_t is the information set generated by $\{Y_t, Y_{t-1}, \cdots\}$; θ is a finite-dimensional unknown parameter vector such that $\theta \in \Theta \subset \mathbb{R}^K$. There exists some $\theta_0 \in \Theta$ such that

$$E(\varepsilon_{\theta_0 t}|I_{t-1}) = 0,\ E(\varepsilon^2_{\theta_0 t}|I_{t-1}) = 1. \tag{3.3}$$

By normalizing $\varepsilon_{\theta_0 t}$ in this way, we have $f(I_{t-1},\theta_0) = E(Y_t|I_{t-1})$, $h^2(I_{t-1},\theta_0) = Var(Y_t|I_{t-1})$. This asumption is weaker than assuming $\left\{\varepsilon_{\theta_0 t}\right\}_{t\in\mathbb{Z}}$ i.i.d with mean 0 and variance 1, which is normally assumed in the literature. This makes sense, since there is a growing econometrics and finance literature documenting time-varying conditional skewness and kurtosis in economic and financial time series, see, e.g. Gallant et al. (1991), Hansen (1994), Harvey and Siddique (1999) and Jondeau and Rockinger (2003). This specification covers most commonly used linear and nonlinear dynamic time series models. Examples include the autoregressive conditional heteroskedascity (ARCH), autoregresive moving average (ARMA), bilinear, nonlinear moving average, Markov regime-switching, smooth transition, exponential, and threshold autoregressive models.

Consider $\{\varepsilon_{\theta t}\}_{t\in\mathbb{Z}}$ and $\left\{\varepsilon_{\theta t}^{2}\right\}_{t\in\mathbb{Z}}$, define the residual sample autocorrelation functions

$$\rho_{\theta}(j) = \tfrac{\gamma_{\theta}(j)}{\gamma_{\theta}(0)},\ j \in 1,2,\cdots,$$
$$\delta_{\theta}(j) = \tfrac{\eta_{\theta}(j)}{\eta_{\theta}(0)},\ j \in 1,2,\cdots,$$

where $\gamma_{\theta}(j) = Cov\left(\varepsilon_{\theta t}, \varepsilon_{\theta t-j}\right)$, $\eta_{\theta}(j) = Cov\left(\varepsilon_{\theta t}^{2}, \varepsilon_{\theta t-j}^{2}\right)$, $j \in 1,2,\cdots$, respectively, are the corresponding autocovariance functions of the standardized residuals and the square of the residuals.

If the model (3.2) is correctly specified the null hypothsis

$$\begin{aligned} H_0: \quad & \rho_{\theta_0}(1) = \rho_{\theta_0}(2) = \cdots = 0 \\ & \delta_{\theta_0}(1) = \delta_{\theta_0}(2) = \cdots = 0 \end{aligned}$$

is satisfied.

Given observations $\{Y_t\}_{t=1}^{T}$, $\rho_{\theta}(j)$ and $\delta_{\theta}(j)$ are estimated by the sample autocorrelation functions

$$\hat{\rho}_{\theta}(j) = \frac{\hat{\gamma}_{\theta}(j)}{\hat{\gamma}_{\theta}(0)},\ j \in 1,2,\cdots,$$
$$\hat{\delta}_{\theta}(j) = \frac{\hat{\eta}_{\theta}(j)}{\hat{\eta}_{\theta}(0)},\ j \in 1,2,\cdots,$$

where

$$\hat{\gamma}_{\theta}(j) = \frac{1}{T}\sum_{t=j+1}^{T}\left(\varepsilon_{\theta t} - \bar{\varepsilon}_{\theta}\right)\left(\varepsilon_{\theta t-j} - \bar{\varepsilon}_{\theta}\right),\ j \in \mathbb{Z}$$
$$\hat{\eta}_{\theta}(j) = \frac{1}{T}\sum_{t=j+1}^{T}\left(\varepsilon_{\theta t}^{2} - \bar{\varepsilon}_{\theta}^{2}\right)\left(\varepsilon_{\theta t-j}^{2} - \bar{\varepsilon}_{\theta}^{2}\right),\ j \in \mathbb{Z}$$

are the sample autocovariance functions and $\bar{\varepsilon}_{\theta} = T^{-1}\sum_{t=1}^{T}\varepsilon_{\theta t}$, $\bar{\varepsilon}_{\theta}^{2} = T^{-1}\sum_{t=1}^{T}\varepsilon_{\theta t}^{2}$.

Define $\hat{\rho}_{\theta_0}^{(m)} = \left(\hat{\rho}_{\theta_0}(1),\cdots,\hat{\rho}_{\theta_0}(m)\right)'$, $\hat{\delta}_{\theta_0}^{(m)} = \left(\hat{\delta}_{\theta_0}(m),\cdots,\hat{\delta}_{\theta_0}(m)\right)'$ for a fixed m. When $\left\{\varepsilon_{\theta_0 t}\right\}$ is a sequence of independently and identically distributed (i.i.d) innovations with mean 0 and variance 1, we have

$$\sqrt{T}\begin{bmatrix} \hat{\rho}_{\theta_0}^{(m)} \\ \hat{\delta}_{\theta_0}^{(m)} \end{bmatrix} \to_d N\left(0, I^{(2m)}\right).$$

If martingale difference restriction is satisfied, $\{\varepsilon_{\theta t}\}$ may exhibit higher-order serial dependence conditions. In this case,

$$\sqrt{T}\hat{\delta}_{\theta_0}^{(m)} \to_d N\left(0, \Sigma_{\theta_0}^{(m)}\right),\ \Sigma_{\theta}^{(m)} = \left[\frac{\sigma_{\theta}^{(i,j)}}{\eta_{\theta}(0)^2}\right]_{i,j=1}^{m},$$

where $\sigma_{\theta_0}^{(i,j)} = E\left[\left(\varepsilon_{\theta_0 t}^2 - 1\right)^2\left(\varepsilon_{\theta_0 t-i}^2 - 1\right)\left(\varepsilon_{\theta_0 t-j}^2 - 1\right)\right]$.

Furthermore, when conditional symmetry does not hold ($E\left(\varepsilon_{\theta_0 t}^3 | I_{t-1}\right) \neq 0$), under H_0 we have

$$\sqrt{T}\begin{bmatrix} \hat{\rho}_{\theta_0}^{(m)} \\ \hat{\delta}_{\theta_0}^{(m)} \end{bmatrix} \to_d N\left(0, \Omega_{\theta_0}^{(2m)}\right).$$

where

$$\Omega_{\theta_0}^{(2m)} = \begin{pmatrix} I^{(m)} & \Pi_{\theta_0}^{(m)} \\ \Pi_{\theta_0}^{(m)\prime} & \Sigma_{\theta_0}^{(m)} \end{pmatrix}.$$

and

$$\Pi_{\theta_0}^{(m)} = \left[\frac{\upsilon_{\theta_0}^{(i,j)}}{\eta_{\theta_0}(0)}\right]_{i,j=1}^{m}$$

and $\upsilon_{\theta_0}^{(i,j)} = E\left[\varepsilon_{\theta_0 t}^3 \varepsilon_{\theta_0 t-i}\left(\varepsilon_{\theta_0 t-j}^2 - 1\right)\right]$. If $E\left(\varepsilon_{\theta_0 t}^3 | I_{t-1}\right)$ is a constant, then we have

$$\Omega_{\theta_0}^{(2m)} = \begin{pmatrix} I^{(m)} & 0 \\ 0 & \Sigma_{\theta_0}^{(m)} \end{pmatrix}.$$

In practice, we do not know the true values of the parameters, they have to be estimated in the first place. Assume that there exists an estimator $\hat{\theta}_T$ such that

$$\hat{\theta}_T = \theta_0 + O_p\left(T^{-1/2}\right). \tag{3.4}$$

then we have

$$\hat{\rho}_{\hat{\theta}_T}^{(m)} = \left(\hat{\rho}_{\hat{\theta}_T}(1), \cdots, \hat{\rho}_{\hat{\theta}_T}(m)\right)'$$
$$\hat{\delta}_{\hat{\theta}_T}^{(m)} = \left(\hat{\delta}_{\hat{\theta}_T}(m), \cdots, \hat{\delta}_{\hat{\theta}_T}(m)\right)'.$$

In the following Proposition, we show that the residual autocorrelations suffer from estimation effects.

Proposition 1. *Under H_0, (3.4) and Assumption A.1 to A.3 in the Appendix,*

$$\sqrt{T}\hat{\rho}_{\hat{\theta}_T}^{(m)} = \sqrt{T}\hat{\rho}_{\theta_0}^{(m)} + \nabla\rho_{\theta_0}^{(m)}\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_p(1)$$

$$\sqrt{T}\hat{\delta}_{\hat{\theta}_T}^{(m)} = \sqrt{T}\hat{\delta}_{\theta_0}^{(m)} + \nabla\delta_{\theta_0}^{(m)}\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_p(1),$$

where $\nabla\rho_{\theta_0}^{(m)} = p\lim \frac{\partial}{\partial\theta'}\hat{\rho}_{\theta_0}^{(m)}, \nabla\delta_{\theta_0}^{(m)} = p\lim \frac{\partial}{\partial\theta'}\hat{\delta}_{\theta_0}^{(m)}$.

Now things become more complicated. In order to derive the covariance matrix of $\hat{\rho}_{\hat{\theta}_T}^{(m)}$ and $\hat{\delta}_{\hat{\theta}_T}^{(m)}$ correctly, the asymptotic joint distribution of $\sqrt{T}\left(\hat{\theta}_T - \theta_0\right)$ and $\hat{\rho}_{\theta_0}^{(m)}$, $\hat{\delta}_{\theta_0}^{(m)}$ has to be considered, which depends on the model and characteristics, the method of estimating $\hat{\theta}_T$ and the unknown parameter value θ_0.

In this article, we propose an asymptotically distribution-free transform of the sample autocorrelations of residuals. Consider a positive definite matrix of statistics $\hat{\Omega}_{\theta}^{(2m)}$ such that

$$\hat{\Omega}_{\hat{\theta}_T}^{(2m)} = \Omega_{\theta_0}^{(2m)} + o_p(1) \tag{3.5}$$

under H_0.

We first normalize $\hat{\rho}_{\hat{\theta}_T}^{(m)}$ and $\hat{\delta}_{\hat{\theta}_T}^{(m)}$ into

$$\sqrt{T}\begin{bmatrix} \tilde{\rho}_{\hat{\theta}_T}^{(m)} \\ \tilde{\delta}_{\hat{\theta}_T}^{(m)} \end{bmatrix} = \sqrt{T}\hat{\Omega}_{\hat{\theta}_T}^{(2m)-1/2}\begin{bmatrix} \hat{\rho}_{\hat{\theta}_T}^{(m)} \\ \hat{\delta}_{\hat{\theta}_T}^{(m)} \end{bmatrix}.$$

Based on Proposition 1, it is easy to obtain that

$$\sqrt{T}\tilde{\rho}_{\hat{\theta}_T}^{(m)} = \sqrt{T}\tilde{\rho}_{\theta_0}^{(m)} + \nabla\tilde{\rho}_{\theta_0}^{(m)}\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_p(1)$$

$$\sqrt{T}\tilde{\delta}_{\hat{\theta}_T}^{(m)} = \sqrt{T}\tilde{\delta}_{\theta_0}^{(m)} + \nabla\tilde{\delta}_{\theta_0}^{(m)}\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_p(1),$$

where

$$\begin{bmatrix} \nabla\tilde{\rho}_{\theta_0}^{(m)} \\ \nabla\tilde{\delta}_{\theta_0}^{(m)} \end{bmatrix} = \Omega_{\theta_0}^{(2m)-1/2} \begin{bmatrix} \nabla\rho_{\theta_0}^{(m)} \\ \nabla\delta_{\theta_0}^{(m)} \end{bmatrix}.$$

Group $\tilde{\rho}_{\hat{\theta}_T}^{(m)}$ and $\tilde{\delta}_{\hat{\theta}_T}^{(m)}$ into $\tilde{\lambda}_{\hat{\theta}_T}(1) = \left(\tilde{\rho}_{\hat{\theta}_T}^{(m)}(1), \tilde{\delta}_{\hat{\theta}_T}^{(m)}(1)\right)', \cdots, \tilde{\lambda}_{\hat{\theta}_T}(m) = \left(\tilde{\rho}_{\hat{\theta}_T}^{(m)}(m), \tilde{\delta}_{\hat{\theta}_T}^{(m)}(m)\right)'$, and define

$$\tilde{\Lambda}_{\hat{\theta}_T}(i) = \begin{pmatrix} \nabla\tilde{\rho}_{\hat{\theta}_T}^{(m)}(i) \\ \nabla\tilde{\delta}_{\hat{\theta}_T}^{(m)}(i) \end{pmatrix},$$

for $i = 1, \cdots, m$.

For $i = 1, \cdots, m-k$, the transformation of the pairs of residual autocorrelations is

$$\begin{aligned} \bar{\lambda}_{\hat{\theta}_T}(i) = & \left[I_2 + \tilde{\Lambda}_{\hat{\theta}_T}(i) \left(\sum_{j=i+1}^{m} \tilde{\Lambda}_{\hat{\theta}_T}(j)' \tilde{\Lambda}_{T\hat{\theta}_T}(j) \right)^{-1} \tilde{\Lambda}_{\hat{\theta}_T}(i)' \right]^{-1/2} \\ & \times \left[\tilde{\lambda}_{\hat{\theta}_T}(i) - \tilde{\Lambda}_{\hat{\theta}_T}(i) \left(\sum_{j=i+1}^{m} \tilde{\Lambda}_{\hat{\theta}_T}(j)' \tilde{\Lambda}_{\hat{\theta}_T}(j) \right)^{-1} \sum_{j=i+1}^{m} \tilde{\Lambda}_{\hat{\theta}_T}(j)' \tilde{\lambda}_{\hat{\theta}_T}(j) \right], \end{aligned}$$

where we make a recursive projection on $\tilde{\Lambda}_{\hat{\theta}_T}(i)$, employing $\tilde{\Lambda}_{\hat{\theta}_T}(j)$, $j = i+1, \cdots, m$.

3.3 Main Results

In this section we show that our transformed (squared) residual autocorrelations are asymptotically distribution free and propose new specification tests based on them.

We prove in the following theorem that, under H_0, $\bar{\lambda}_{\hat{\theta}_T}^{(m-k)} = \left(\bar{\lambda}_{\hat{\theta}_T}(1)', \cdots, \bar{\lambda}_{\hat{\theta}_T}(m-k)'\right)'$ and $\bar{\lambda}_{\theta_0}^{(m-k)}$ are asymptotically equivalent, and $\sqrt{T}\bar{\lambda}_{\theta_0}^{(m-k)}$ is asymptotically distributed as a vector of independent standard normals.

Theorem 13. *Under H_0, $m > k$, Assumption A.1 to A.4 in the Appendix and with $\hat{\theta}_T$ satisfying (3.4) and (3.5).*

$$\bar{\lambda}_{\hat{\theta}_T}^{(m-k)} = \bar{\lambda}_{\theta_0}^{(m-k)} + o_p\left(T^{-1/2}\right)$$

and

$$\sqrt{T}\bar{\lambda}_{\theta_0}^{(m-k)} \to_d N(0, I_{2m-2k}).$$

3.3.1 The Local Alternative

We consider the following local alternative,

$$H_{1T}: \rho_{\theta_0}(j) = \frac{r_{\theta_0\rho}(j)}{\sqrt{T}}, \delta_{\theta_0}(j) = \frac{r_{\theta_0\delta}(j)}{\sqrt{T}}, \text{ for all } j = 1, 2, ...$$

In order to describe the asymptotic distribution of $\bar{\lambda}_{\hat{\theta}_T}^{(m-k)}$ under H_{1T}, define firstly the vector $\bar{\tau}_\theta^{(m-k)} = \left(\bar{\tau}_\theta(1)', \cdots, \bar{\tau}_\theta(m-k)'\right)'$ as the projected and standardized drift, where

$$\bar{\tau}_\theta(i)' = \tau_\theta(i) - \Lambda_\theta(i)\left(\sum_{j=i+1}^{m} \Lambda_\theta(j)' \Lambda_\theta(j)\right)^{-1} \sum_{j=i+1}^{m} \Lambda_\theta(j)' \tau_\theta(j)$$

for $i = 1, 2, \cdots, m-k$, and

$$\tau_\theta^m = \Omega_\theta^{(2m)-1/2} r_\theta^m,$$

where $r_\theta(i) = \left(r_{\theta\rho}(i), r_{\theta\delta}(i)\right)'$ and $r_\theta^m = \left(r_\theta(1)', \cdots, r_\theta(m)'\right)'$.

Theorem 14. *Under H_{1T}, $m > k$, Assumptions A.1 to A.4 in the Appendix and with $\hat{\theta}_T$ and $\hat{\Omega}_{\hat{\theta}_T}^{(2m)}$ satisfying (3.4) and (3.5), respectively,*

$$\bar{\lambda}_{\hat{\theta}_T}^{(m-k)} = \bar{\lambda}_{\theta_0}^{(m-k)} + o_p\left(T^{-1/2}\right)$$

and

$$\sqrt{T}\bar{\lambda}_{\theta_0}^{(m-k)} \to_d N\left(\bar{\tau}_{\theta_0}^{(m-k)}, I_{2m-2k}\right).$$

3.3.2 Box-Pierce Type Tests

Based on Theorems 1 and 2, we can establish asymptotic properties of a Box-Pierce type test statistic.

For $1 \leq s \leq m-k$, the Box-Pierce type test statistic is defined as

$$\bar{B}_{\hat{\theta}_T}^{(2m)}(s) = T\sum_{j=1}^{s} \bar{\lambda}_{\hat{\theta}_T}(j)' \bar{\lambda}_{\hat{\theta}_T}(j)$$

Proposition 2. *Under* H_0, $\bar{B}_{\hat{\theta}_T}^{(2m)}(s) \to_d \chi^2_{2s}$. *Under* H_{1T}, $\bar{B}_{\hat{\theta}_T}^{(2m)}(s) \to_d \chi^2_{2s}\left(\sum_{j=1}^{s} \bar{\tau}_{\theta_0}^{(m-k)}(j)' \bar{\tau}_{\theta_0}^{(m-k)}(j)\right)$.

The proof is obvious, so it is omitted.

Proposition 2 shows that Box-Pierce type test statistic has power as long as the projected drift is non-zero, so that H_{1T} effectively represent local alternative from the model.

3.3.3 ARMA-GARCH Model

We consider in this subsection an ARMA(P, Q)-GARCH(p,q) model

$$\begin{aligned} Y_t &= \sum_{j=1}^{P} a_{0j} Y_{t-j} + e_t - \sum_{j=1}^{Q} b_{0j} e_{t-j}, \\ e_t &= h_t \varepsilon_t, \\ h_t^2 &= \varpi_0 + \sum_{j=1}^{q} \alpha_{0j} e_{t-j}^2 + \sum_{j=1}^{p} \beta_{0j} h_{t-j}^2, \end{aligned}$$

and show how to compute our new test statistics.

The parameter vector is denoted by $\vartheta = \left(a_1, \cdots, a_P, b_1, \cdots, b_Q\right)'$ for the conditional mean part of the model, $\nu = \left(\varpi, \alpha_1, \cdots, \alpha_q, \beta_1, \cdots, \beta_p\right)'$ for the conditional variance part of the model, $\theta = (\vartheta', \nu')'$. The true parameter vector is $\theta_0 = \left(\vartheta_0', \nu_0'\right)'$, where $\vartheta_0 = \left(a_{01}, \cdots, a_{0P}, b_{01}, \cdots, b_{0Q}\right)'$, $\nu_0 = \left(\varpi_0, \alpha_{01}, \cdots, \alpha_{0q}, \beta_{01}, \cdots, \beta_{0p}\right)'$. The usual identification conditions are assumed. In the following, for any generic function g_θ indexed by parameters $\theta \in \Theta_0$,

$$\dot{g}_\theta = \frac{\partial g_\theta}{\partial \theta'}.$$

We can write the model as

$$\varepsilon_t(\vartheta, \nu) = \frac{e_t(\vartheta)}{h_{\theta t}} = \frac{\varphi_\vartheta(L) Y_t}{h_{\theta t}} = \frac{Y_t - \{1 - \varphi_\vartheta(L)\} Y_t}{h_{\theta t}},$$

where $\varphi_\vartheta(z) = A_\vartheta(z) B_\vartheta^{-1}(z)$, $A_\vartheta(z) = 1 - \sum_{j=1}^{P} a_j z^j$, $B_\vartheta(z) = 1 - \sum_{j=1}^{Q} b_j z^j$.

So that

$$\begin{aligned}
\dot{\varepsilon}_{\vartheta t}(\vartheta,\nu) &= \frac{\partial}{\partial\vartheta}\varepsilon_t(\vartheta,\nu) = \frac{\dot{\varphi}_{\vartheta}(L)\,Y_t}{h_{\theta t}} - \frac{\dot{h}_{\vartheta t}}{h_{\theta t}}\frac{\varphi_{\vartheta}(L)\,Y_t}{h_{\theta t}} \\
&= \left\{\frac{\dot{\varphi}_{\vartheta}(L)}{\varphi_{\vartheta}(L)} - \frac{\dot{h}_{\vartheta t}}{h_{\theta t}}\right\}\frac{e_t(\vartheta)}{h_{\theta t}} \\
\dot{\varepsilon}_{\nu t}(\vartheta,\nu) &= \frac{\partial}{\partial\nu}\varepsilon_t(\vartheta,\nu) = -\frac{\dot{h}_{\nu t}}{h_{\theta t}}\frac{\varphi_{\theta}(L)\,Y_t}{h_{\theta t}} \\
&= -\frac{\dot{h}_{\nu t}}{h_{\theta t}}\frac{e_t(\vartheta)}{h_{\theta t}}.
\end{aligned}$$

It is easy to obtain that

$$\hat{\rho}^{(m)}_{\hat{\theta}_T}(j) = \rho_{\theta_0}(j) - (\vartheta-\vartheta_0)'E\left[\frac{\dot{\varphi}_{\vartheta_0}(L)}{\varphi_{\vartheta_0}(L)}\varepsilon_{\theta_0 t}\varepsilon_{\theta_0 t-j}\right] + o_p\left(T^{-1/2}\right),$$

$$\delta^{(m)}_{\hat{\theta}_T}(j) = \delta^{(m)}_{\theta_0}(j) - 2(\theta-\theta_0)'E\left[\frac{\dot{h}_{\theta_0 t}}{h_{\theta_0 t}}\varepsilon^2_{\theta_0 t}\left(\varepsilon^2_{\theta_0 t-j}-1\right)\right] + o_p\left(n^{-1/2}\right)$$

Example ARMA(1,1)-**GARCH**(1,1) **model:**

$$\begin{aligned}
\frac{\dot{h}_{\vartheta t}}{h_t} &= \frac{1}{2}h_t^{-2}\left(2\alpha\sum_{j=0}^{\infty}\beta^j e_{\vartheta t-1-j}\dot{e}_{\vartheta t-1-j}\right) \\
&= \alpha\sum_{j=0}^{\infty}\beta^j\frac{e_{\vartheta t-1-j}}{h_t}\frac{\dot{e}_{\vartheta t-1-j}}{h_t} \\
&= \alpha\frac{e_{\vartheta t-1}}{h_{\theta t}}\frac{\dot{e}_{\vartheta t-1}}{h_{\theta t}} \quad \text{if ARCH(1).}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\dot{h}_{\nu t}}{h_{\theta t}} &= \frac{1}{2}\begin{pmatrix} h_{\theta t}^{-2}/(1-\beta) \\ \sum_{j=0}^{\infty}\beta^j\frac{e^2_{\vartheta t-1-j}}{h^2_{\theta t}} \\ -h_{\theta t}^{-2}/(1-\beta)^2 + \alpha\sum_{j=1}^{\infty}j\beta^{j-1}\frac{e^2_{\vartheta t-1-j}}{h^2_{\theta t}} \end{pmatrix} \\
&= \frac{1}{2}\begin{pmatrix} h_{\theta t}^{-2} \\ \frac{e^2_{\vartheta t-1}}{h^2_{\theta t}} \end{pmatrix} \quad \text{if ARCH(1).}
\end{aligned}$$

3.4 Monte Carlo Simulations

We carry out some Monte Carlo Simulations to compare the finite-sample performance of the new test statistic with those of the Li-Mak test. The null model is the AR(1)-ARCH(1) model,

$$Y_t = 0.5Y_{t-1} + e_t,$$
$$e_t = h_t \varepsilon_t,$$
$$h_t^2 = 0.01 + 0.4e_{t-1}^2.$$

We consider sample sizes T=200 and 500 and 10,000 replication in each experiment. Parameters are estimated by Quasi maximum likelihood estimation. Nominal value of all tests is 5%. For the sake of comparison, we use values for s from 1 up to 12 and set $m = 15$ for both sample sizes. We use two different estimates of $\Omega_{\theta_0}^{(2m)}$. They are $\hat{\Omega}_{\hat{\theta}_T}^{(2m)} = I^{(2m)}$ and

$$\hat{\Omega}_{\hat{\theta}_T}^{(2m)} = \begin{pmatrix} I^{(m)} & 0 \\ 0 & \hat{\Sigma}_{\hat{\theta}_T}^{(m)} \end{pmatrix}.$$

The first estimate exploits a possible asymptotic i.i.d property of the sample autocorrelations. We consider ε_t ∼i.i.d N(0,1) or i.i.d standardized student t distribution with 10 degrees of freedom, we also consider the semistrong version of the AR(1)-ARCH(1) model.[1] We compare the New Box-Pierce type statistic with the Li-Mak statistic.

Figue 3.4 and 3.4 report the simulated size. When the innovations are (a) i.i.d normal distributed; (b) i.i.d standardized student t distributed with 10 degrees of freedom; (c) semistrong ARCH with normal distribution; (d) semistrong ARCH with standardized Student t distribution with 10 degrees of freedom. We can observe from Figure 1 and 2 that when the innovations follow an i.i.d normal distribution, Li-Mak statistic has good size properties. The new transformed Box-Pierce statistics with $\hat{\Omega}_{\hat{\theta}_T}^{(2m)} = I^{(2m)}$ underreject for $T = 200$, but have good size levels for $T = 500$. On the other hand, the new trans-

[1] To obtain the semistrong ARCH model, first generate the ARCH (1) model $h_t = 0.01/(1+\sqrt{0.4}) + \sqrt{0.4}e_t^2$ with sample size $2T$, then choose the even-number observations. It could be shown that these observations follow semistrong ARCH (1) model $h_t = 0.01 + 0.4e_t^2$, see Franq and Zakoian (2010) Chapter 4.1.1 for more details.

formed Box-Pierce statistics with $\hat{\Omega}^{(2m)}_{\hat{\theta}_T} = \begin{pmatrix} I^{(m)} & 0 \\ 0 & \hat{\Sigma}^{(m)}_{\hat{\theta}_T} \end{pmatrix}$ overreject for $T = 200$, but have good size levels very close to nominal size for $T = 500$. When the innovations follow i.i.d standardized Student t distribution with 10 degrees of freedom, Li-Mak statistics heavily overreject for both $T = 200$ and $T = 500$. Similar results are obtained for Li-Mak statistics when the innovations follow a semistrong ARCH model with normal distribution. However the new transformed Box-Pierce statistics have good size properties in both cases. When it comes to the case that the innovations follow the semistrong ARCH model with standardized Student t distribution, Li-Mak statistics overreject for $T = 200$, but become closer to the nominal size for $T = 500$.

To study the power properties of the new test, we consider the following alternative models

$$Y_t = 0.5Y_{t-1} + 0.2Y_{t-2} + e_t,\ h_t = 0.01 + 0.4e^2_{t-1}.$$

$$Y_t = 0.5Y_{t-1} + 0.2Y_{t-2} + e_t,\ h_t = 0.01 + 0.4e^2_{t-1} + 0.2e^2_{t-2}.$$

$$Y_t = 0.5Y_{t-1} + 0.2Y_{t-2} + e_t,\ h_t = 0.01 + 0.4e^2_{t-1} + 0.5h_{t-2}.$$

For the second and third alternatives, note that there are misspecification in both the conditional mean and conditional variance functions. We report the percentage of rejections under these alternative hypotheses in Figures 3.4-3.4 respectively. For the first alternative, which is a AR(2)-ARCH(1) model, it is clear that Li-Mak tests have no power.

3.5 Appendix

In this appendix we present the sufficient assumptions for the proofs of our results. First we introduce some notations. For any generic function g_θ indexed by parameters $\theta \in \Theta_0$,

$$\dot{g}_\theta = \frac{\partial g_\theta}{\partial \theta'}.$$

Assumption A1. $\{Y_t\}$ is a strictly stationary and ergodic process.

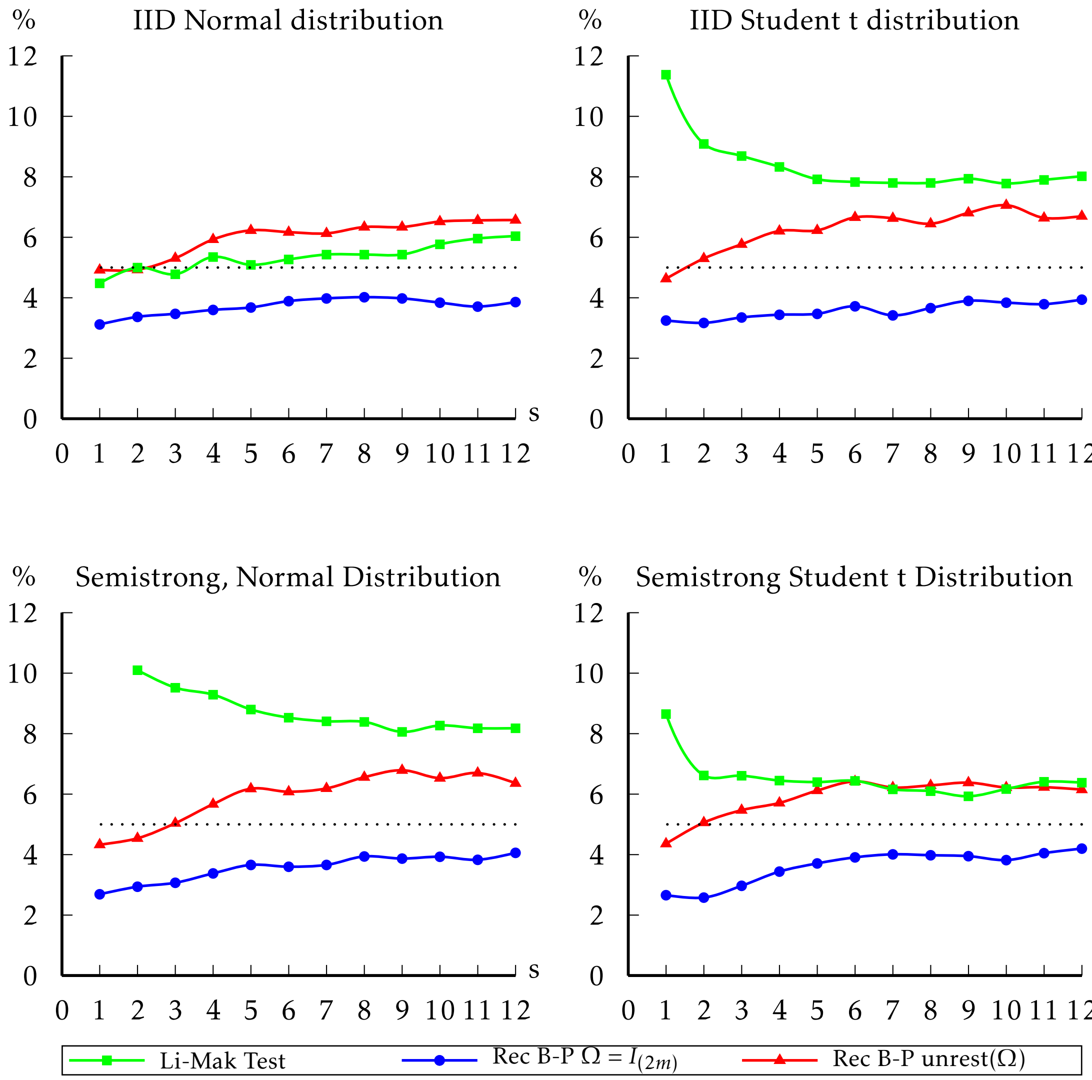


Figure 3.1: Percentage of rejections of Portmanteau tests in terms of the lag s. Nominal level is 5%. $T = 200$. Li-Mak tests compare with a χ^2_s critical value, Rec B-P are tests $\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$ based on recursive projected autocorrelations compared to χ^2_{2s}. Models are AR(1)-ARCH(1) with i.i.d normal distribution, i.i.d Student t distribution with degrees of freedom of 10, semistrong ARCH with normal distribution and semistrong ARCH with Student t distribution with 10 degrees of freedom .

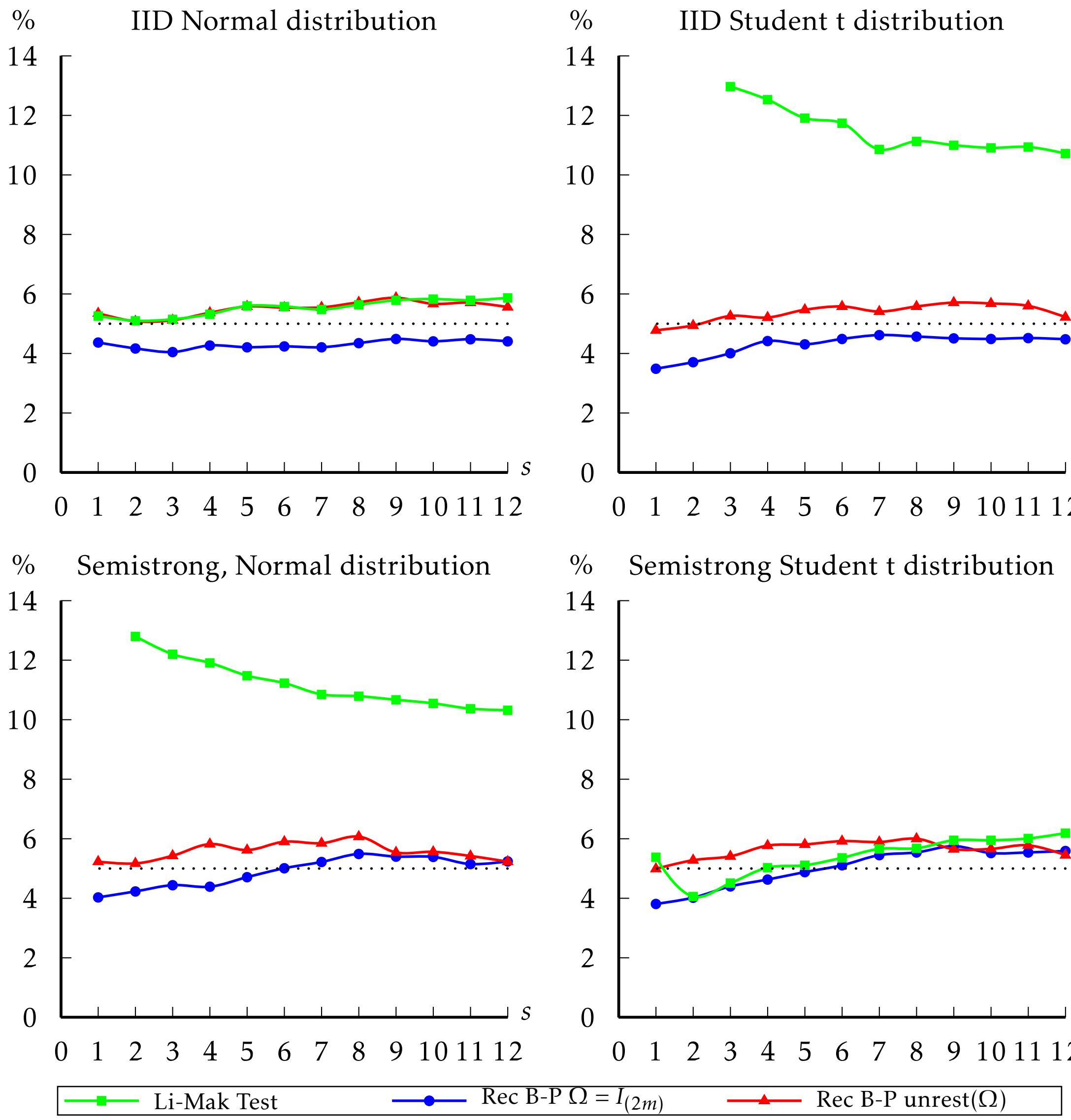


Figure 3.2: Percentage of rejections of Portmanteau tests in terms of the lag s. Nominal level is 5%. $T = 500$. Li-Mak tests compare with a χ^2_s critical value, Rec B-P are tests $\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$ based on recursive projected autocorrelations compared to χ^2_{2s}. Models are AR(1)-ARCH(1) with i.i.d normal distribution, i.i.d Student t distribution with degrees of freedom of 10, semistrong ARCH with normal distribution and semistrong ARCH with Student t distribution with 10 degrees of freedom.

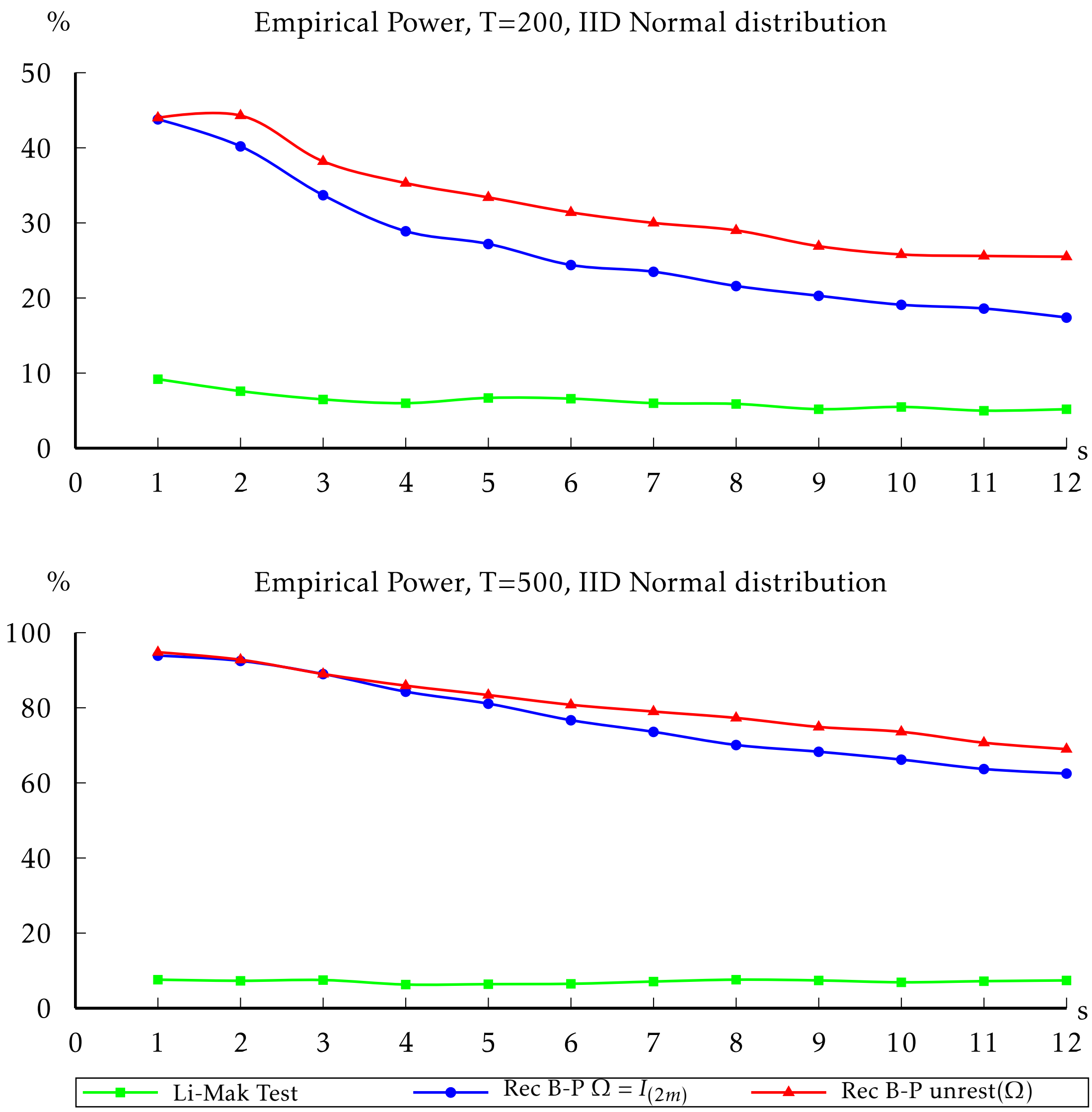


Figure 3.3: Percentage of rejections of Portmanteau tests in terms of the lag s. Nominal level is 5%. $T = 200$ and $T = 500$. Li-Mak tests compare with a χ^2_s critical value, Rec B-P are tests $\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$ based on recursive projected autocorrelations compared to χ^2_{2s}. The null is AR(1)-ARCH(1) model. the alternative is AR(2)-ARCH(1) $Y_t = 0.5Y_{t-1} + 0.2Y_{t-2} + e_t$, $h_t = 0.01 + 0.4e^2_{t-1}$ with i.i.d normal distribution

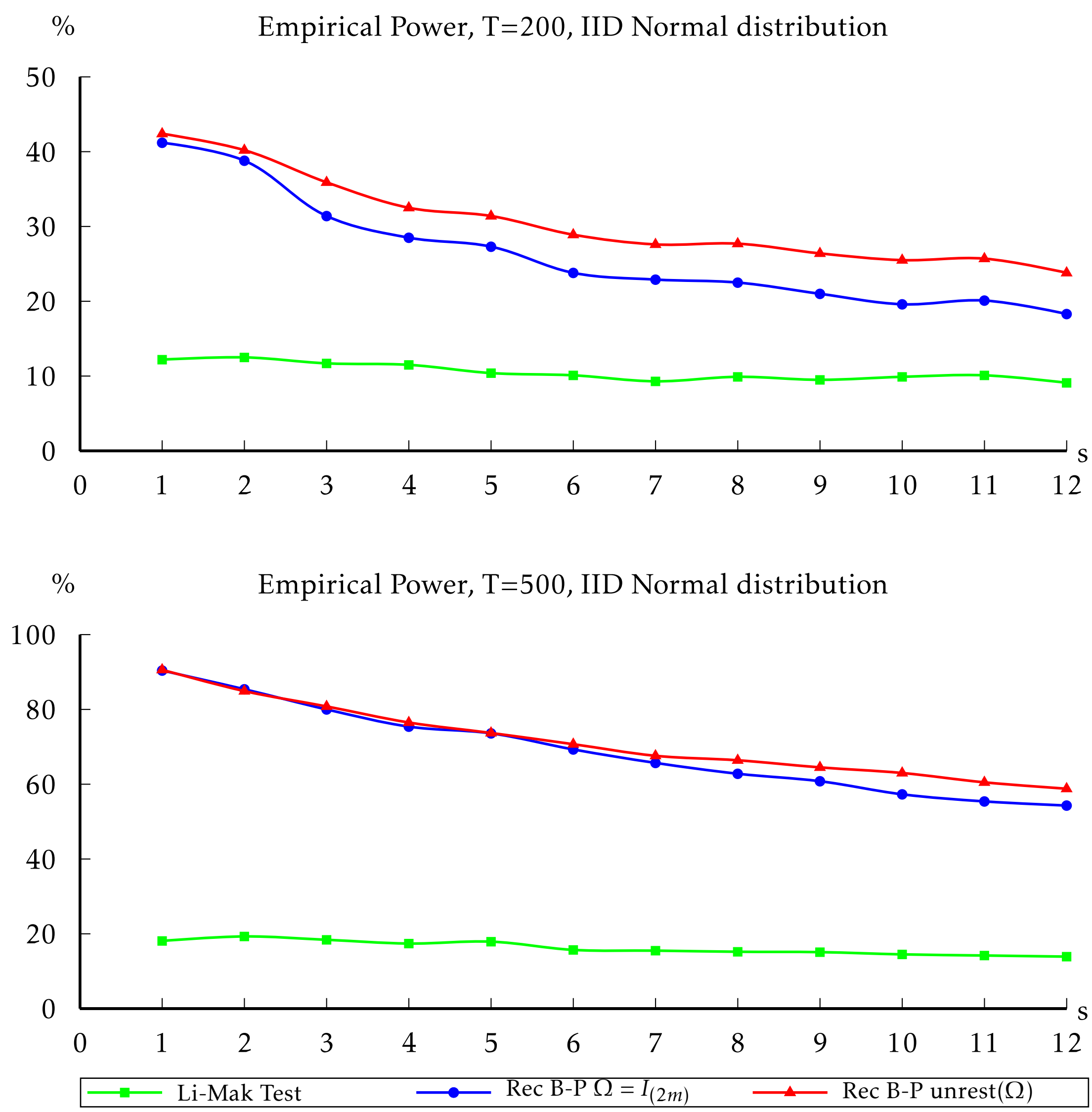


Figure 3.4: Percentage of rejections of Portmanteau tests in terms of the lag s. Nominal level is 5%. $T = 200$ and $T = 500$. Li-Mak tests compare with a χ^2_s critical value, Rec B-P are tests $\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$ based on recursive projected autocorrelations compared to χ^2_{2s}. The null is AR(1)-ARCH(1) model. the alternative is AR(2)-ARCH(2) $Y_t = 0.5Y_{t-1} + 0.2Y_{t-2} + e_t$, $h_t = 0.01 + 0.4e^2_{t-1} + 0.2e^2_{t-2}$ with i.i.d normal distribution

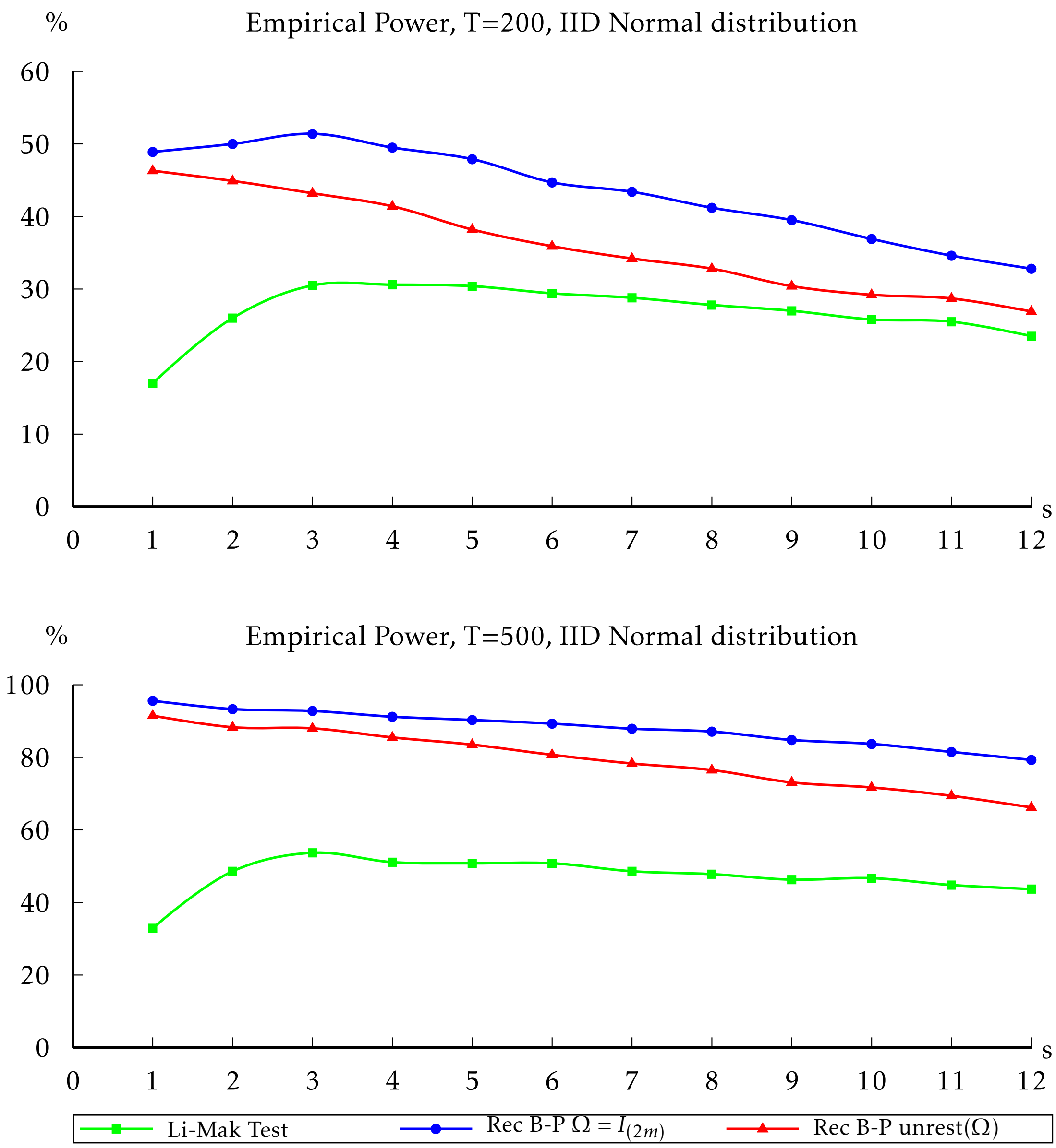


Figure 3.5: Percentage of rejections of Portmanteau tests in terms of the lag s. Nominal level is 5%. $T = 200$ and $T = 500$. Li-Mak tests compare with a χ^2_s critical value, Rec B-P are tests $\bar{B}^{(2m)}_{T\hat{\theta}_T}(s)$ based on recursive projected autocorrelations compared to χ^2_{2s}. The null is AR(1)-ARCH(1) model. the alternative is AR(2)-GARCH(1,1) $Y_t = 0.5Y_{t-1} + 0.2Y_{t-2} + e_t$, $h_t = 0.01 + 0.4e^2_{t-1} + 0.5h_{t-2}$ with i.i.d normal distribution

Assumption A2. $\{\varepsilon_{\theta t}\}$ is a strictly stationary and ergodic process satisfying (3.3) for $\theta \in \Theta$.

Assumption A3. Let Θ_0 be a small convex neighborhood of θ_0. The functions $f(I_{t-1}, \cdot)$ and $h(I_{t-1}, \cdot)$ are twice continuously differentiable with respect to $\theta \in \Theta_0$ a.s. $E \sup_{\theta \in \Theta_0} \left\| h^{-1}(I_{t-1}, \theta)(\partial/\partial\theta) f(I_{t-1}, \theta) \right\|^4 < \infty$. $E \sup_{\theta \in \Theta_0} \left\| h^{-1}(I_{t-1}, \theta)(\partial/\partial\theta) h(I_{t-1}, \theta) \right\|^4 < \infty$. $E \sup_{\theta \in \Theta_0} \left[\varepsilon_{\theta t}^8 \right] < \infty$.

Assumption A4. For $m > k$,

$$\sum_{j=m-k+1}^{m} \tilde{\Lambda}_{\theta_0}(j)' \tilde{\Lambda}_{\theta_0}(j)$$

is positive definite.

Assumptions A1 and A2 are standard conditions. The 8th moment of standard error $\varepsilon_{\theta t}$ assumption in Assumption A3 seems strong in the general framework of the conditional mean and conditional variance models, but it could be relaxed for example in ARMA-GARCH models following Francq and Zakoian (2004) and Berkes et al. (2003). Assumption A4 is similar to Delagdo and Velasco (2011).

Proof of Proposition 1: First for $\hat{\rho}_{\hat{\theta}_T}^{(m)}$ note that under H_0, Since $E \sup_{\theta \in \Theta_0} \left[\varepsilon_{\theta t}^8 \right] < \infty$, $\hat{\theta}_T = \theta_0 + O_p\left(T^{-1/2}\right)$, it is easy to obtain that $\hat{\gamma}_{T\hat{\theta}_T}(0) = 1 + o_p(1)$. Furthermore $\bar{\varepsilon}_{\hat{\theta}_T} = o_p(1)$. We only need to consider $\frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \varepsilon_{\hat{\theta}_T t} \varepsilon_{\hat{\theta}_T t-j}$, for $j = 1, \cdots, m$. By first-order Taylor expansion, we have $\frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \varepsilon_{\hat{\theta}_T t} \varepsilon_{\hat{\theta}_T t-j} = \frac{1}{\sqrt{T}} \sum_{t=j+1}^{T} \varepsilon_{\theta_0 t} \varepsilon_{\theta_0 t-j} + \frac{1}{T} \sum_{t=j+1}^{T} (\dot{\varepsilon}_{\check{\theta} t} \varepsilon_{\check{\theta} t-j} + \varepsilon_{\check{\theta} t} \dot{\varepsilon}_{\check{\theta} t-j}) \sqrt{T} \left(\hat{\theta}_T - \theta_0\right)$, where $\left\|\check{\theta} - \theta_0\right\| \leq \left\|\hat{\theta}_T - \theta_0\right\|$. Note that

$\varepsilon_{\theta t} = \frac{Y_t - f(I_{t-1}, \theta)}{h(I_{t-1}, \theta)}$,

$$\dot{\varepsilon}_{\theta t} = -\varepsilon_{\theta t} h(I_{t-1}, \theta)^{-1} \frac{\partial h(I_{t-1}, \theta)}{\partial \theta'} - h(I_{t-1}, \theta)^{-1} \frac{\partial f(I_{t-1}, \theta)}{\partial \theta'}$$

$$
\begin{aligned}
E_{\sup\theta\in\Theta_0}\left(\left\|\dot{\varepsilon}_{\theta t}\varepsilon_{\theta t-j}\right\|\right) &\leq E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t}\varepsilon_{\theta t-j}h\left(I_{t-1},\theta\right)^{-1}\frac{\partial h\left(I_{t-1},\theta\right)}{\partial\theta'}\right\| \\
&+ E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t-j}h\left(I_{t-1},\theta\right)^{-1}\frac{\partial f\left(I_{t-1},\theta\right)}{\partial\theta'}\right\| \\
&\leq \left[E_{\sup\theta\in\Theta_0}\|\varepsilon_{\theta t}\|^2\right]^{1/2}\left[E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t-j}h\left(I_{t-1},\theta\right)^{-1}\frac{\partial h\left(I_{t-1},\theta\right)}{\partial\theta'}\right\|^2\right]^{1/2} \\
&+ \left[E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t-j}\right\|^2\right]^{1/2}\left[E_{\sup\theta\in\Theta_0}\left\|h\left(I_{t-1},\theta\right)^{-1}\frac{\partial f\left(I_{t-1},\theta\right)}{\partial\theta'}\right\|^2\right]^{1/2} \\
&\leq \left[E_{\sup\theta\in\Theta_0}\|\varepsilon_{\theta t}\|^2\right]^{1/2} \\
&\times\left[\left[E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t-j}\right\|^4\right]^{1/2}\left[E_{\sup\theta\in\Theta_0}\left\|h\left(I_{t-1},\theta\right)^{-1}\frac{\partial h\left(I_{t-1},\theta\right)}{\partial\theta'}\right\|^4\right]^{1/2}\right]^{1/2} \\
&+ \left[E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t-j}\right\|^2\right]^{1/2}\left[E_{\sup\theta\in\Theta_0}\left\|h\left(I_{t-1},\theta\right)^{-1}\frac{\partial f\left(I_{t-1},\theta\right)}{\partial\theta'}\right\|^2\right]^{1/2} \\
&< \infty
\end{aligned}
$$

Similarly we could obtain $E_{\sup\theta\in\Theta_0}\left(\left\|\varepsilon_{\theta t}\dot{\varepsilon}_{\theta t-j}\right\|\right) < \infty$. So the dominance condition holds, then we have the weakly uniformly convergence of

$$
p\lim\sup_{\theta\in\Theta_0}\left\|\frac{1}{T}\sum_{t=j+1}^{T}\left(\dot{\varepsilon}_{\theta t}\varepsilon_{\theta t-j}+\varepsilon_{\theta t}\dot{\varepsilon}_{\theta t-j}\right)-E(\dot{\varepsilon}_{\theta t}\varepsilon_{\theta t-j}+\varepsilon_{\theta t}\dot{\varepsilon}_{\theta t-j})\right\| = 0.
$$

Then we have $\frac{1}{T}\sum_{t=j+1}^{T}(\dot{\varepsilon}_{\check{\theta}t}\varepsilon_{\check{\theta}t-j}+\varepsilon_{\check{\theta}t}\dot{\varepsilon}_{\check{\theta}t-j}) \rightarrow_p E(\dot{\varepsilon}_{\theta_0 t}\varepsilon_{\theta_0 t-j})$, $j = 1,\cdots,m$. So we prove that

$$
\sqrt{T}\hat{\rho}_{\hat{\theta}_T}^{(m)} = \sqrt{T}\hat{\rho}_{\theta_0}^{(m)} + \nabla\rho_{\theta_0}^{(m)}\sqrt{T}\left(\hat{\theta}_T-\theta_0\right)+o_p(1),
$$

where $\nabla\rho_{\theta_0}^{(m)}(j) = E(\dot{\varepsilon}_{\theta_0 t}\varepsilon_{\theta_0 t-j})$.

For $\hat{\delta}_{\hat{\theta}_T}^{(m)}$, Note that $\hat{\eta}_{\hat{\theta}_T}(0) = \eta_{\theta_0}(0)+o_p(1)$.We only need focus on $\hat{\eta}_{\hat{\theta}_T}(j)$, $j = 1,\cdots,m$.

$$
\begin{aligned}
\hat{\eta}_{\hat{\theta}_T}(j) &= \frac{1}{T}\sum_{t=j+1}^{T}\left(\varepsilon_{\hat{\theta}_T t}^2-\bar{\varepsilon}_{T\hat{\theta}_T}^2\right)\left(\varepsilon_{\hat{\theta}_T t-j}^2-\bar{\varepsilon}_{\hat{\theta}_T}^2\right) \\
&= \frac{1}{T}\sum_{t=j+1}^{T}\left(\varepsilon_{\hat{\theta}_T t}^2-1\right)\left(\varepsilon_{\hat{\theta}_T t-j}^2-1\right)+o_p(1).
\end{aligned}
$$

By first-order Taylor expansion

$$\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\left(\varepsilon^2_{\hat{\theta}_T t}-1\right)\left(\varepsilon^2_{\hat{\theta}_T t-j}-1\right)=\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\left(\varepsilon^2_{\theta_0 t}-1\right)\left(\varepsilon^2_{\theta_0 t-j}-1\right)$$
$$+\frac{2}{T}\sum_{t=j+1}^{T}(\dot{\varepsilon}_{\breve{\theta}t}\varepsilon_{\breve{\theta}t}\left(\varepsilon^2_{\breve{\theta}t-j}-1\right)$$
$$+\left(\varepsilon^2_{\breve{\theta}t}-1\right)\varepsilon_{\breve{\theta}t-j}\dot{\varepsilon}_{\breve{\theta}t-j})\sqrt{T}\left(\hat{\theta}_T-\theta_0\right),$$

where $\left\|\breve{\theta}-\theta_0\right\|\leq\left\|\hat{\theta}_T-\theta_0\right\|$. Next,

$$E_{\sup\theta\in\Theta_0}\left(\left\|\dot{\varepsilon}_{\theta t}\varepsilon_{\theta t}\varepsilon^2_{\theta t-j}\right\|\right)\leq E_{\sup\theta\in\Theta_0}\left\|\varepsilon^2_{\theta t}\varepsilon^2_{\theta t-j}h(I_{t-1},\theta)^{-1}\frac{\partial h(I_{t-1},\theta)}{\partial\theta'}\right\|$$
$$+E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t}\varepsilon^2_{\theta t-j}h(I_{t-1},\theta)^{-1}\frac{\partial f(I_{t-1},\theta)}{\partial\theta'}\right\|,$$

since

$$\begin{aligned}E_{\sup\theta\in\Theta_0}\left\|\varepsilon^2_{\theta t}\varepsilon^2_{\theta t-j}h(I_{t-1},\theta)^{-1}\frac{\partial h(I_{t-1},\theta)}{\partial\theta'}\right\| &\leq \left[E_{\sup\theta\in\Theta_0}\left\|\varepsilon^2_{\theta t-j}h(I_{t-1},\theta)^{-1}\frac{\partial h(I_{t-1},\theta)}{\partial\theta'}\right\|^2\right]^{1/2}\\ &\quad\times\left[E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t-j}\right\|^4\right]^{1/2}\\ &\leq \left[E_{\sup\theta\in\Theta_0}\left\|\varepsilon^8_{\theta t-j}\right\|E_{\sup\theta\in\Theta_0}\left\|h(I_{t-1},\theta)^{-1}\frac{\partial h(I_{t-1},\theta)}{\partial\theta'}\right\|^4\right]^{1/4}\\ &\quad\times\left[E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t-j}\right\|^4\right]^{1/2}\\ &\leq \infty.\end{aligned}$$

$$\begin{aligned}E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t}\varepsilon^2_{\theta t-j}h(I_{t-1},\theta)^{-1}\frac{\partial f(I_{t-1},\theta)}{\partial\theta'}\right\| &\leq \left[E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t}h(I_{t-1},\theta)^{-1}\frac{\partial f(I_{t-1},\theta)}{\partial\theta'}\right\|^2\right]^{1/2}\\ &\quad\times\left[E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t-j}\right\|^4\right]^{1/2}\\ &\leq \left[E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t}\right\|^4E_{\sup\theta\in\Theta_0}\left\|h(I_{t-1},\theta)^{-1}\frac{\partial f(I_{t-1},\theta)}{\partial\theta'}\right\|^4\right]^{1/4}\\ &\quad\times\left[E_{\sup\theta\in\Theta_0}\left\|\varepsilon_{\theta t-j}\right\|^4\right]^{1/2}\\ &<\infty.\end{aligned}$$

So the dominance condition holds,

$$p \lim \sup_{\theta \in \Theta_0} \left\| \frac{1}{T} \sum_{t=j+1}^{T} \dot{\varepsilon}_{\theta t} \varepsilon_{\theta t} \varepsilon_{\theta t-j}^2 - E(\dot{\varepsilon}_{\theta t} \varepsilon_{\theta t} \varepsilon_{\theta t-j}^2) \right\| = 0,$$

and similarly we can obtain that

$$p \lim \sup_{\theta \in \Theta_0} \left\| \frac{1}{T} \sum_{t=j+1}^{T} \dot{\varepsilon}_{\theta t} \varepsilon_{\theta t} - E(\dot{\varepsilon}_{\theta t} \varepsilon_{\theta t}) \right\| = 0.$$

Then $\frac{1}{T} \sum_{t=j+1}^{T} \dot{\varepsilon}_{\breve{\theta} t} \varepsilon_{\breve{\theta} t} \left(\varepsilon_{\breve{\theta} t-j}^2 - 1 \right) \to_p E\left[\dot{\varepsilon}_{\theta t} \varepsilon_{\theta t} \left(\varepsilon_{\theta t-j}^2 - 1 \right) \right] = E\left[\left(\varepsilon_{\theta t-j}^2 - 1 \right) h\left(I_{t-1}, \theta\right)^{-1} \frac{\partial h(I_{t-1}, \theta)}{\partial \theta'} \right]$. Similarly we can obtain $\frac{1}{T} \sum_{t=j+1}^{T} \left(\varepsilon_{\breve{\theta} t}^2 - 1 \right) \varepsilon_{\breve{\theta} t-j} \dot{\varepsilon}_{\breve{\theta} t-j} \to_p E\left[\left(\varepsilon_{\breve{\theta} t}^2 - 1 \right) \varepsilon_{\breve{\theta} t-j} \dot{\varepsilon}_{\breve{\theta} t-j} \right] = 0$.

So we prove that

$$\sqrt{T} \tilde{\delta}_{\hat{\theta}_T}^{(m)} = \sqrt{T} \tilde{\delta}_{\theta_0}^{(m)} + \nabla \tilde{\delta}_{\theta_0}^{(m)} \sqrt{T} \left(\hat{\theta}_T - \theta_0 \right) + o_p(1),$$

where $\nabla \tilde{\delta}_{\theta_0}^{(m)}(j) = 2E\left[\left(\varepsilon_{\theta t-j}^2 - 1 \right) h\left(I_{t-1}, \theta\right)^{-1} \frac{\partial h(I_{t-1}, \theta)}{\partial \theta'} \right]$.

Proof of Theorem 1: The proof is similar to the reasoning in Brown, Durbin and Evans (1975) and Delgado and Velasco (2011) under Assumption A4, and it is omitted.

Bibliography

[1] Andrews, W K, (1987). Consistency in nonlinear econometric models: A generic uniform law of large numbers. Econometrica 55 (1987), 1465–1471.

[2] Bai, J. (2003). Testing Parametric Conditional Distributions of Dynamic Models. Review of Economics and Statistics, 85, 531-549.

[3] Berkes, I., Horväth, Lajos and Kokoszka, Piotr, (2003). Asymptotics for GARCH Squared Residual Correlations. Econometric Theory, 19, issue 04, p. 515-540.

[4] Bierens H. (1982). Consistent Model Specification Test. Journal of Econometrics, 20, 105-34.

[5] Bierens H.J., 1990. A Consistent Conditional Moment Test of Functional Form. Econometrica 58, 1443-1458.

[6] Bierens, H.J., Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests. Econometrica 65, 1129-1152.

[7] Billingsley, P., 1968. Convergence of Probability Measures. New York: Wiley.

[8] Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods (2nd Edition). Springer.

[9] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307-327.

[10] Brown, R.L., Durbin, J. and Evans, J.M. (1975). Techniques for testing constancy of regression relationships over time. J. Roy. Statist. Soc. Ser. B. 37, 149-192.

[11] Box, G. E. P. and Pierce, D. A. (1970) Distribution of the residual autocorrelations in autoregressive integrated moving average time series models. Journal of American Statistical Association 65, 1509–26.

[12] Chamberlain, G., 1987. Asymptotic efficiency in estimation with conditional moment restrictions. Journal of Econometrics 34, 305–334.

[13] Chamberlain, G., 1992. Comment: sequential moment restrictions in panel data. Journal of Business and Economic Statistics 10, 20-26.

[14] Cragg, J. G. (1983). More efficient estimation in the presence of heteroskedasticity of unknown form. Econometrica 51, 751-763.

[15] Carrasco, M., Florens J.P. 2000. Generalization of GMM to a continuum of moment conditions. Econometric Theory 16, 797-834.

[16] Davidson, R., and MacKinnon, J. G. (1993). Estimation and Inference in Econometrics. New York: Oxford University Press.

[17] De Jong, R.M. & Bierens, H.J., 1994. On the Limit Behavior of a Chi-Square Type Test if the Number of Conditional Moments Tested Approaches Infinity. Econometric Theory, vol. 10, 70-90.

[18] De Jong, R.M., 1996. The Bierens test under data dependence. Journal of Econometrics 72, 1-32.

[19] Delgado, M.A., Velasco, C., (2011). An Asymptotically Pivotal Transform of the Residuals Sample Autocorrelations With Application to Model Checking. Journal of the American Statistical Association. September 1, 2011, 106(495): 946-958.

[20] Domínguez, M.A. and I.N. Lobato, 2004. Consistent Estimation of Models Defined by Conditional Moment Restrictions. Econometrica, 1601-1615.

[21] Donald, G. Imbens, W. Newey, Whitney K., 2003. Empirical likelihood estimation and consistent tests with conditional moment restrictions. Journal of Econometrics, Elsevier, vol. 117(1), 55-93.

[22] Engle, R.F., Ng, V.K. (1993). Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749-1777.

[23] Escanciano, J.C., 2006a. A consistent diagnostic test for regression models using projections. Econometric Theory 22, 1030-1051.

[24] Escanciano, J.C., 2006b. Goodness-of-fit statistics for linear and nonlinear time series models. Journal of the American Statistical Association 101, 531-541.

[25] Escanciano, J.C., (2008). Joint and marginal specification tests for conditional mean and variance models. Journal of Econometrics, Elsevier, vol. 143(1), pages 74-87, March.

[26] Escanciano, Juan Carlos, 2009. Quasi-Maximum Likelihood Estimation Of Semi-Strong Garch Models. Econometric Theory, vol. 25(02), 561-570.

[27] Fan, Y., Li, Q., 1996. Consistent model specification tests: omitted variables and semi-parametric functional forms. Econometrica 64, 865-890.

[28] Fan, J.Q., and Yao, Q.W., (2003). Nonlinear Time Series: Nonparametric and Parametric Methods. Springer-Verlag, New York.

[29] Francq Christian, Zakoïan Jean-Michel (2004). Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes. Bernoulli Vol. 10, pp. 605-637

[30] Gallant A.R., Hsieh D.A., Tauchen G.,(1991). On fitting a recalcitrant series: the Pound/Dollar exchange rate, 1974–1983. W.A. Barnett, J. Powell, G. Tauchen (Eds.), Nonparametric and Semiparametric Methods in Econometrics and Statistics, Cambridge University Press, Cambridge, UK, pp. 199–240

[31] Gozalo, P.L., 1993. A consistent model specification test for nonparametric estimation of regression function models. Econometric Theory 9, 451-477.

[32] Gourieroux, C. (1997). ARCH Models and Financial Applications. New York: Spring-Verlag.

[33] Guo, Binbin and Phillips, Peter C. B. Efficient Estimation of Second Moment Parameters in ARCH Models. UC Santa Cruz Economics Working Paper No. 541.

[34] Hahn, J., 1997. Efficient estimation of panel data models with sequential moment restrictions. Journal of Econometrics 79, 1-21.

[35] Halunga, Andreea G., Orme, Chris D., (2009). First-Order Asymptotic Theory For Parametric Misspecification Tests Of Garch Models. Econometric Theory, vol. 25(02), 364-410.

[36] Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. The Annals of Statistics 21, 1926-1947.

[37] Harvey, C., Siddique A. (1999). Autoregressive conditional skewness. Journal of Financial and Quantitative Analysis, 34, pp. 465–487.

[38] Hayashi, F. (2000). Econometrics, Princeton. NJ: Princeton University Press.

[39] Hong, Yongmiao and Lee, Tae-Hwy, (2003). Diagnostic Checking for The Adequacy of Nonlinear Time Series Models. Econometric Theory, 19, issue 06, p. 1065-1121.

[40] Hong, Y., White, H., 1995. Consistent specification testing via nonparametric series regression. Econometrica 63, 1133-1159.

[41] Khmaladze, E. V. (1981). Martingale Approach in the Theory of Goodness-of-Fit Tests. Theory of Probability and Its Applications, 26, 240-257.

[42] Kuersteiner, Guido M., (2001). Optimal instrumental variables estimation for ARMA models. Journal of Econometrics, 104, issue 2, 359-405

[43] Kuan, C.M., Lee W.M., (2006). Robust M Tests Without Consistent Estimation of the Asymptotic Covariance Matrix. Journal of the American Statistical Association, Vol. 101, No. 475, 1264.

[44] Li, W.K., Mak, T.K. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. Journal of Time Series Analysis 15, 627-636.

[45] Ljung, G. M., and Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. Biometrika, 65, 297–303.

[46] Lundbergh, S., Teräsvirta, T. (2002). Evaluating GARCH Models. Journal of Econometrics 110, 417-435.

[47] Miles, D., J. Mora., 2003 On the performance of nonparametric specification tests in regression models. Computational Statistics & Data Analysis 42, 477– 490.

[48] Newey, W. K. (1985a). Maximum Likelihood Specification Testing and Conditional Moment Tests. Econometrica, 53, 1047-1070.

[49] Newey, W. K. (1985b). Generalized method of moments specification testing. Journal of Econometrics 29, 229-256.

[50] Newey, W.K., 1993. Efficient estimation of models with conditional moment restrictions. In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), Handbook of Statistics, Vol. 11: Econometrics, North-Holland, Amsterdam.

[51] Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147–168.

[52] Newey, W.K., McFadden, D., 1994. Large Sample Estimation and Hypothesis Testing. In: Engle, McFadden, D. (Eds.), Handbook of Econometrics, Vol. 4. North-Holland Amsterdam, pp. 2113–2245.

[53] Pötscher B.M. and Prucha I.R.,(1989). A uniform law of large numbers for dependent and heterogeneous data processes. Econometrica, 57, 675–683.

[54] Stinchcombe, M., White, H., 1998. Consistent specification testing with nuisance parameters present only under the alternative. Econometric Theory 14, 295-325.

[55] Stute, W., 1997. Nonparametric model checks for regression. The Annals of Statistics 25, 613-641.

[56] Stute,W., Thies, S., and Zhu, L.-X. (1998). Model Checks for Regression: An Innovation Process Approach. The Annals of Statistics, 26, 1916-1934.

[57] Tauchen, G. (1985). Diagnostic Testing and Evaluation of Maximum Likelihood Model. Journal of Econometrics, 30, 415-443.

[58] Wang, X., 2011. A General Approach to Conditional Moment Specification Testing with Projections. Working Paper, UC3M

[59] White, H. (1994). Estimation, Inference and Specification Analysis. New York: Cambridge University Press.

[60] Wong H, Ling S. 2005. Mixed portmanteau tests for time-series models. Journal of Time Series Analysis 26: 569-579.

[61] Wooldridge, J. (1990). A unified approach to robust, regression-based specification tests. Econometric Theory, 6, 17-43.

[62] Wooldridge, J. (1991). On the application of robust, regression-based diag- nostics to models of conditional means and conditional variances. Journal of Econometrics 47, 5-46.

[63] Wooldridge, J. M. (1994). Estimation and Inference for Dependent Processes. In Handbook of Econometrics, Volume 4, ed. R. F. Engle and D. L. McFadden. Amsterdam: North-Holland, 2639-2738.

[64] Zheng, J.X., 1996. A consistent test of functional form via nonparametric estimation techniques. Journal of Econometrics 75, 263-289.