AN IMPROVED CONSISTENT CONDITIONAL MOMENT TEST FOR REGRESSION MODELS IN THE PRESENCE OF HETEROSKEDASTICITY OF UNKNOWN FORM

Xuexin WANG*

This Version: March 2013

Abstract

The purpose of this paper is to propose a simple approach to consistent testing of specification of parametric regression models which is robust and efficient under 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. An innovative residual empirical process is proposed, employing a projection-based transformation. It is shown that the new residual empirical process is not affected by the uncertainty from the parameter estimation, only a preliminary $\sqrt{n}$-consistent estimator is needed. We establish its efficiency in the sense that the GMM estimator reaching the semiparametric efficiency bound under heteroskedasticity of unknown form is actually employed by the new empirical process under the null. Then a version of Bierens (1990) test based on the new empirical process 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.

JEL Classification: C12 C21

Keywords: Consistent Conditional Moment Test; Efficient tests; Estimation Effect Removal; Heteroskedasticity of Unknown Form.

*Address: The Wang Yanan Institute for Studies in Economics(WISE), Xiamen University, Xiamen, Fujian, China. E-mail: xuexinwang@xmu.edu.cn. I would like to thank my advisor Carlos Velasco for his helpful guidance and support. I am also grateful to Miguel A. Delgado, Juan Carlos Escanciano, Ignacio Lobato and Javier Hidalgo for their useful comments and suggestions.
1 Introduction

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, see, Härdle and Mammen (1993), Gozalo (1993), Hong and White (1995), Fan and Li (1996), and Zheng (1996), to mention but a few. The smoothing-based tests only require some consistent parameters estimator, and typically lead to asymptotic pivotal test statistics under the null. 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, see, for example, Bierens (1982, 1990), Bierens and Ploberger (1997), Stute (1997) and Escanciano (2006a). In contrast with smoothing-based tests, the non-smoothing-based tests have to handle the uncertainty from the parameters estimation, and lead to case-dependent limiting distributions. However it is well established that the non-smoothing-based tests are more powerful than smoothing-based tests against Pitman local alternatives.

Non-smoothing-based tests enjoy some optimal properties under conditional homoskedasticity and other assumptions. However, heteroskedasticity is a rule rather than an exception for most regression models. Even though non-smoothing-based test statistics proposed in previous literature are robust to heteroskedasticity of unknown form, they still suffer from the pitfalls of distorted size and deteriorated testing power. One of the reasons is that the estimator employed to handle the estimation effect becomes inefficient under heteroskedasticity of unknown form.

The purpose of the present paper is to propose a simple approach to consistent testing of specification of parametric regression models which is robust and efficient under 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 in the case of non-smoothing-based tests. Instead of following the “estimating-testing” paradigm of Newey (1985a,b) and Tauchen (1985), an innovative residual empirical process is proposed, employing a projection-based transformation. It is shown that the new residual empirical process is not affected by the uncertainty from the parameter estimation, only a preliminary $\sqrt{n}$-
consistent estimator is needed. We establish its efficiency in the sense that the new empirical process employs the GMM estimator that reaches the semiparametric efficiency bound under heteroskedasticity of unknown form. Then a version of Bierens (1990) test based on the new empirical process is proposed. Monte Carlo simulations show that Bierens (1990) test based on the new empirical process is more powerful for a large number of alternatives when heteroskedasticity of unknown form is presented.

The outline of the paper is as follows. In Section 2, we establish the testing framework. In Section 3, we define the class of weighting functions, the new residual empirical process and study its properties. Section 4 discusses the improved test statistic of Bierens (1990) based on the new empirical process. Section 5 conducts Monte Carlo simulations. Section 6 concludes.

2 Testing Framework

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 \(\mathcal{G} = \{f(X, \theta) : \mathbb{R}^d \to \mathbb{R} | \theta \in \Theta \subset \mathbb{R}^p\}\). To justify the correctness of the parametric model, we have to test the null hypothesis

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

against the alternative

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

The null hypothesis is equivalent to the conditional moment restriction

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

where \(e(\theta) = Y - f(X, \theta)\). The idea of non-smoothing-based tests is to convert the conditional moment restriction into an infinite number of unconditional moment restrictions, i.e,

\[
E[e(\theta_0)|X] = 0 \text{ a.s } \iff E[e(\theta_0)w(X,t)] = 0, \text{ for almost all } t \in T,
\]  

where \(w(X,t)\) are weighting functions.
where $T \subset \mathbb{R}^h$, $h \in \mathbb{N}$, and $w(X,t)$ is a proper weighting function such that the equivalence (3) holds. There are many weighting functions meeting the requirement of (3). One example is $w(X,t) = \exp(it'X)$ where $i = \sqrt{-1}$, $T = \mathbb{R}^d$, which is employed by Bierens (1982). Bierens (1990) proposes $w(X,t) = \exp(t'X)$, $T = \mathbb{R}^d$. Stute (1997) proposes the indicator function $w(X,t) = I(X < t)$, $T = \mathbb{R}^d$. Escanciano (2006a) introduces the weighting function $w(X,t) = I(\beta'X \leq u)$, $t = (\beta',u) \in T = S^d \times (-\infty,\infty)$, where $S^d = \{\beta \in \mathbb{R}^d : |\beta| = 1\}$.

Given a sample $(Y_j, X_j')$, $j = 1, \ldots, n$, and a $\sqrt{n}$-consistent estimator $\hat{\theta}$, the scaled sample analog of $E[e(\theta_0)w(X,t)]$, which forms a residual empirical process, is

$$\hat{M}(\hat{\theta},t) = n^{-1/2} \sum_{j=1}^{n} e_j(\hat{\theta})w(X_j,t), t \in \Pi \subset T,$$

where $e_j(\theta) = Y_j - f(X_j,\theta)$.

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

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

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

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

In contrast with smoothing-based tests, the non-smoothing-based tests have to handle the uncertainty from the parameter estimation. More specifically, it could be shown under some regularity conditions that

$$\hat{M}(\hat{\theta},t) = n^{-1/2} \sum_{j=1}^{n} e_j(\theta_0)w(X_j,t) - b(\theta_0,t)n^{1/2}(\hat{\theta} - \theta_0) + o_p(1),$$

where

$$b(\theta_0,t) = E\left[w(X,t) \frac{\partial f(X,\theta_0)}{\partial \theta'}\right].$$

In most cases, $b(\theta_0,t) \neq 0$. The existence of the so-called “estimation effect” $b(\theta_0,t)n^{1/2}(\hat{\theta} - \theta_0)$
makes the asymptotic theory more complicated, and limit results case-dependent.

Bierens (1982, 1990), and Bierens and Ploberger (1997) assume that the estimator $\hat{\theta}$ is based on the criteria function $\theta_0 = \arg\min_{\theta \in \Theta} E \{ [Y - f(X,\theta)]^2 \}$, which means the nonlinear least squared (NLS) estimator is employed. So we have

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

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

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 and conditional homoskedasticity, 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, and the testing power gets worse.1 In this case, it is natural to think about employing other more efficient estimators instead of the inefficient NLS one, Stute (1997) and Escanciano (2006a) assume an estimator $\hat{\theta}$ such that

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

where $h(\cdot)$ satisfies $E[h(Y, X, \theta_0)] = 0$, 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 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 is that while non-smoothing-based tests exploit the continuum of the unconditional moment restrictions, it is sufficient to employ a set

---

1For simulation evidence on the power deterioration of consistent tests when heteroskedasticity of unknown form is present, see Miles and Mora (2003).
of discrete countable unconditional moment restrictions in efficient estimation of parameters in the conditional moment restrictions models. 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 \( \{q_j(X)\}_{j=1}^{\infty} \) 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 \( \gamma_K \) such that

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

where \( q^K(X) = (q_1(X), \ldots, q_K(X))^\prime \) is a \( K \times 1 \) vector. Chamberlain (1987) firstly shows that an estimator obtained as the solution to

\[
\sum_{j=1}^{n} Q(X_j)e_j(\theta_0) = 0, \quad Q(X) = \frac{\partial f(X,\theta_0)}{\partial \theta} \sigma_0^{-2}(X),
\]

where \( \sigma_0^2(X) = E\left[ (Y - f(X,\theta_0))^2 | X \right] \), 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 (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 \( K \) for different choices of \( q^K(X) \), for example splines and power series, in a quite general framework. In next section, we will define a class of weighting function for non-moothing-based tests, and show that the weighting functions in this class also form the basis of efficient estimation of parameters of the regression models.

\footnote{Carrasco and Florens (2000) consider the continuum of unconditional moment restrictions in efficient estimation of the conditional moment restrictions models, however the singularity of the covariance matrix must be handled. Furthermore, the indexed parameter \( t \) has to be a scalar.}
3  A Class of Weighting Functions and A New Empirical Process

We focus on a class of weighting functions $\mathcal{W}$ such that

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

For any weighting function in this class, we have the following lemma.

**Lemma 1.** 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 $\mathcal{W}$, the equivalence in (3) holds.

**Proof.** See Stinchcombe and White (1998) Theorem 2.3. \hfill $\Box$

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

Examples of families satisfying this lemma are $w(t'\Phi(X)) = \exp(it't'\Phi(X))$ 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 residual empirical process such that

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

On the other hand, $w(t'\Phi(X))$ also forms a basis for efficient estimation. Note that we only need to consider a set of discrete countable unconditional moment restrictions in this case. 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

$$q^K(X) = (w(t'_1\Phi(X)), \cdots w(t'_K\Phi(X)))'.$$  \hfill (5)

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

**Corollary 1.** For any $\epsilon > 0$, $\varphi \in L_2$, and for each $K \times 1$ vector $q^K(X)$ defined in (5), there exist $K \times 1$ vectors $\gamma_K$ such that (4) holds.

**Proof.** See Appendix. \hfill $\Box$

It is possible to form a robust ICM or KS statistic based on the residual empirical process $\hat{M}(\theta, t)$ on some interval $\Pi$ 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 \( \hat{M}(\hat{\theta},t) \) to remove the estimation effect.

The idea is to form a new weighting functions by a proper linear combination of \( w(t'\Phi(X)) \) and \( q^K(X) \) such that there does not exist the estimation effect for the transformed empirical process. More specifically, given any \( \sqrt{n} \)-consistent estimator \( \hat{\theta} \), \( q^K(X) \) defined in (5), for any \( t \in \Pi \subset \mathbb{R}^d \), the new empirical process is

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

where the new weighting function depends on

\[
b(\theta,t) = \frac{1}{n} \sum_{i=1}^n w(t'\Phi(X_i)) \frac{\partial f(X_i,\theta)}{\partial \theta'},
\]

\[
\hat{A}(\theta,K) = \left[ \hat{\Lambda}(\theta,K)'\hat{\Omega}(\theta,K)^{-1} \hat{\Lambda}(\theta,K) \right]^{-1} \hat{\Lambda}(\theta,K)'\hat{\Omega}(\theta,K)^{-1},
\]

in which

\[
\hat{\Lambda}(\theta,K) = \frac{1}{n} \sum_{i=1}^n q^K(X_i) \frac{\partial f(X_i,\theta)}{\partial \theta'}
\]

\[
\hat{\Omega}(\theta,K) = \frac{1}{n} \sum_{i=1}^n e_i(\theta)^2 q^K(X_i)q^K(X_i)'.
\]

Remark: The intuition behind the removal of the estimation effect is that the transformed weighting function \( w(t'\Phi(X)) - b(\theta_0,t)A(\theta_0,K)q^K(X) \) is orthogonal to \( \frac{\partial f(X,\theta_0)}{\partial \theta'} \) now.

Remark: This transformation could be regarded as an application of the general methodology of projection-based transformation developed by Wang (2011), in which he proposes to use further moment conditions to remove the estimation effect in the test statistics. To see this point, we can rewrite (6) into

\[
\hat{M}(\hat{\theta},t,K) = n^{-1/2} \sum_{j=1}^n e_j(\hat{\theta}) w(t'\Phi(X_j)) - n^{-1/2}b(\hat{\theta},t)\hat{A}(\hat{\theta},K) \sum_{j=1}^n q^K(X_j)e_j(\hat{\theta}).
\]
\( \hat{A}(\hat{\theta}, K) = \frac{1}{n} \sum_{j=1}^{n} q^K(X_j) e_j(\hat{\theta}) \) is the generalized least squared (GLS) estimator of the regression of \( \hat{A}(\hat{\theta}, K) \) on \( \frac{1}{n} \sum_{j=1}^{n} q^K(X_j) e_j(\hat{\theta}) \). So \( M(\hat{\theta}, t, K) \) could be regarded as a scaled projection residual.

See Wang (2011) for more discussion on this.

Remark: To remove the estimation effect, the choice of the \( q^K(X) \) is not restricted to the one defined in (5). If \( q^K(X) = \frac{\partial f(X, \theta)}{\partial \theta} \) 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 \( q^K(X) = (w(t'_1 \Phi(X)), \ldots, w(t'_K \Phi(X)))' \) can be not only robust to heteroskedasticity but also efficient in the presence of heteroskedasticity of unknown form by using the optimal weighting \( \hat{\Omega}(\hat{\theta}, K) \) and increasing the dimension \( K \) as \( n \) grows, as we will show later.

Remark: The form of \( M(\hat{\theta}, t, K) \) also has a 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 class \( \mathcal{W} \). 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 1.** Let \( (Y_j, X_j)' \), \( j = 1, \ldots, n \), be a sample from a probability distribution \( F(Y, X) \) on \( \mathbb{R} \times \mathbb{R}^d \). Moreover, \( E(Y^2) < \infty \).

**Assumption 2.** The parameter space \( \Theta \) is a compact subset of \( \mathbb{R}^p \). \( \theta_0 \in \text{int}(\Theta) \).

**Assumption 3.** \( \sqrt{n}(\hat{\theta} - \theta_0) = O_p(1) \).

**Assumption 4.** \( E\left[ \sup_{\theta \in \Theta} (Y - f(X, \theta))^2 | X \right] < \infty \), \( \sigma^2(X) \) is bounded away from zero. There is \( \delta(Y, X) \) and \( \alpha > 0 \) such that for all \( \hat{\theta}, \theta \in \Theta \), \( |f(X, \hat{\theta}) - f(X, \theta)| \leq \delta(Y, X) \|\hat{\theta} - \theta\|^\alpha \) and \( E\left[ \delta(Y, X)^2 \right] < \infty \).

**Assumption 5.** \( f(X, \theta) \) is twice continuously differentiable in an open and convex neighborhood \( \Delta \) of \( \theta_0 \). \( E\left[ \left\| \frac{\partial^2 f(X, \theta)}{\partial \theta \partial \theta'} \right\| \right] \) is bounded, \( E\left[ \sup_{\theta \in \Delta} \left\| \frac{\partial f(X, \theta)}{\partial \theta} \right\|^2 \right] < \infty \), \( E\left[ \sup_{\theta \in \Delta} |Y - f(X, \theta)|^4 | X \right] < \infty \), and for all \( \theta \in \Delta \), \( |f(X, \theta) - f(X, \theta_0)| \leq \delta(Y, X) \|\theta - \theta_0\| \) and \( E\left[ \delta(Y, X)^2 | X \right] < \infty \).

**Assumption 6.** Denote \( Z \) as the support of \( X \), for each \( K \) there is a constant scalar \( \xi(K) \) and matrix \( B \) such that \( \hat{q}^K(X) = Bq^K(X) \) for every \( X \in Z \), \( \sup_{X \in Z} \|q^K(X)\| \leq \xi(K) \), \( \sqrt{K} \leq \xi(K) \), and \( E\left( \hat{q}^K(X) q^K(X)' \right) \) has smallest eigenvalue bounded away from zero uniformly in \( X \). There exists an integer \( D \), \( D \geq p \) such that when \( K \geq D \), \( E\left[ q^K(X) \frac{\partial f(X, \theta)}{\partial \theta} \right] \) 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 parameters estimation explicitly. To obtain a $\sqrt{n}$-consistent estimator, only an identification condition as we ask 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 4. The properties of $q^K(X)$ make sure that $\mathbb{E}\left[q^K(X) \frac{\partial f(X, \theta_0)}{\partial \theta}\right]$ 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 $\mathbb{E}\left[q^K(X) \frac{\partial f(X, \theta_0)}{\partial \theta}\right]$ when $K$ is large enough.

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

$$\tilde{M}(\hat{\theta}, t, K) = n^{-1/2} \sum_{j=1}^n e_j(\theta_0) \left[ w(t^\prime \Phi(X_j)) - b(\theta_0, t) A(\theta_0, K) q^K(X_j) \right] + o_p(1), \quad (9)$$

where $b(\theta_0, t) = p \lim \hat{b}(\theta_0, t)$, $A(\theta_0, K) = p \lim \hat{A}(\theta_0, K)$, and

$$\bar{M}(\hat{\theta}, t, K) \overset{d}{\rightarrow} N\left[0, s^2(\theta_0, t, K)\right], \quad (10)$$

where

$$s^2(\theta_0, t, K) = \mathbb{E}\left[e(\theta_0)^2 \left[ w(t^\prime \Phi(X)) - b(\theta_0, t) A(\theta_0, K) q^K(X) \right]^2\right].$$

**Proof.** See Appendix

Equation (9) shows that, unlike $\bar{M}(\hat{\theta}, t)$, $\bar{M}(\hat{\theta}, t, K)$ does not suffer from the estimation effect—the empirical process evaluated at any $\sqrt{n}$-consistent estimator is asymptotically the same as the empirical process evaluated at the true parameters.

Although the difference between $\bar{M}(\cdot, t)$ and $\bar{M}(\cdot, t, K)$ which are evaluated at the same es-
timator \( \hat{\theta} \) is not negligible, it turns out that \( \hat{M}(\hat{\theta}, t, K) \) is equivalent asymptotically to \( \hat{M}(\cdot, t) \) when is evaluated at a particular estimator. The following theorem establishes the relation between \( \hat{M}(\hat{\theta}, t, K) \) and \( \hat{M}(\cdot, t) \).

**Theorem 2.** When Assumptions 1 to 6 hold, \( K \geq D \), under \( H_0 \), for any \( t \in \Pi \subset \mathbb{R}^d \), \( \hat{M}(\hat{\theta}, t, K) \) is equivalent asymptotically to \( \hat{M}(\cdot, t) \) with

\[
\hat{\theta} = \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{j=1}^n q^K(X_j) \left( Y_j - f(X_j, \theta) \right) \Omega^{-1} \frac{1}{n} \sum_{j=1}^n q^K(X_j) \left( Y_j - f(X_j, \theta) \right),
\]

(11)

where \( \Omega \overset{p}{\rightarrow} \Omega(\theta_0, K) \).

**Proof.** See Appendix. \( \square \)

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

This theorem shows that, when \( K \) is large enough, \( \hat{M}(\cdot, t, K) \) 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 a 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 empirical process when \( K \) increases with sample size \( n \) in the following theorem.

**Theorem 3.** When Assumptions 1 to 6 hold, under \( H_0 \), for any \( t \in \Pi \subset \mathbb{R}^d \), when \( K \rightarrow \infty \) and \( \xi(K)^2K/n \rightarrow 0 \), \( \hat{M}(\hat{\theta}, t, K) \) is efficient in the sense that \( \hat{M}(\hat{\theta}, t, K) \) is equivalent to \( \hat{M}(\hat{\theta}, t) \) where \( \hat{\theta} \) reaches the semiparametric efficiency bound, such that

\[
\hat{M}(\hat{\theta}, t, K) = n^{-1/2} \sum_{j=1}^n e_j(\theta_0) \phi_*(X_j, t) + o_p(1)
\]

\[
s^2_\phi(\theta_0, t) = E \left[ e(\theta_0)^2 \phi_*(X, t)^2 \right]
\]

where

\[
\phi_*(X, t) = w(t'\Phi(X)) - b(\theta_0, t)A_*(\theta_0) \frac{\partial f(X, \theta_0)}{\partial \theta} \sigma_0^{-2}(X)
\]

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

This theorem establishes the efficiency in the sense that the GMM estimator that reaches the semiparametric efficiency bound under heteroskedasticity of unknown form is employed by the new empirical process.

Although we have established that \( \tilde{M}(\cdot, t, K) \) 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 \( E[q^K(X)e(\theta_0)] = 0 \), and established the increase rate of the dimension \( K \) for the efficiency, our approach is also more natural and much easier to compute, compared with \( \hat{M}(\cdot, t) \).

4 Bierens (1990) Test Based On The New Empirical Process

Based on \( \tilde{M}(\hat{\theta}, t, K) \), 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 time-consuming bootstrap procedures are 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 can give primitive conditions for Assumption 6.

Assumption 7. Choose \( (t'_1, \cdots, t'_K)' \) such that \( t'_j \in \mathbb{R}^d \setminus \Pi \) 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 nonparametric and semiparametric models. This assumption also sets the rules of how to choose \( (t'_1, \cdots, t'_K)' \) and the subset \( \Pi \), but they are hardly restrictive.

Lemma 2. Assumption 7 implies Assumption 6, further \( \xi(K) = CK^{3/2} \), where \( C > 0 \) is a constant. Finally, for any \( t \in \Pi \), \( s^2(\theta_0, t, K) > 0 \).

Proof. See Appendix.

Our approach conveniently avoids the extreme condition that \( s^2(\theta_0, t, K) = 0 \) for any proper \( K \). In Bierens (1990), an additional assumption has to be imposed, and it only could be estab-
lished that set \( S_{NLS}^\ast = \{ t \in \mathbb{R}^d : s_{NLS}^2(\theta_0,t) = 0 \} \), where

\[
s_{NLS}^2(\theta_0,t) = E \left\{ e(\theta_0)^2 \left[ \exp(t'\Phi(X)) - b(\theta_0,t)A_{NLS}(\theta_0) \frac{\partial f(X,\theta_0)}{\partial \theta} \right] \right\},
\]

\( A_{NLS}(\theta_0) = \left[ E \left( \frac{\partial f(X,\theta_0)}{\partial \theta} \frac{\partial f(X,\theta_0)}{\partial \theta'} \right) \right]^{-1} \), has Lebesgue measure zero and is not dense in \( \mathbb{R}^d \). In practice, given any proper \( K \), the function \( s^2(\theta_0,t,K) \) can be consistently estimated by

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

where \( \hat{b}(\hat{\theta},t) \) is defined by (7), and \( \hat{A}(\hat{\theta},K) \) by (8), and

\[
\tilde{W}(\hat{\theta},t,K) = \frac{\tilde{M}(\hat{\theta},t,K)^2}{\hat{s}^2(\hat{\theta},t,K)}
\]
is well defined for any \( t \in \Pi \) for sample size large enough.

Lemma 1 of Bierens (1990) shows that under \( H_1 \)

\[
S_{NLS} = \left\{ t \in \mathbb{R}^d : E \left[ (Y - f(X,\theta_0))\exp(t'\Phi(X)) - b(\theta_0,t)A_{NLS}(\theta_0) \frac{\partial f(X,\theta_0)}{\partial \theta} \right] = 0 \right\}
\]

has Lebesgue measure zero. Following the same logic, we can obtain that under \( H_1 \), the set

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

has Lebesgue measure zero.

In terms of the choice of \( \Pi \), Bierens (1990) has to assume that \( \Pi \subset \mathbb{R}^d \setminus S_{NLS} \cup S_{NLS}^\ast \), however we only need to assume \( \Pi \subset \mathbb{R}^d \setminus S \) here. This may have important impact on the testing power. Since the testing power negatively relies on the size of the set \( S_{NLS} \cup S_{NLS}^\ast \) or \( S \), 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.

Following Bierens (1990), we maximize \( \tilde{W}(\hat{\theta},t,K) \) over a subset \( \Pi \) of \( \mathbb{R}^d \). The results are summarized in the following theorem.

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

$$\Gamma_\ast(t_1,t_2) = E\left\{ (Y - f(X,\theta_0))^2 \phi_\ast(X,t_1)\phi_\ast(X,t_2)/\sqrt{s_\ast^2(\theta_0,t_1)s_\ast^2(\theta_0,t_2)} \right\}. \quad (12)$$

Moreover, \( \tilde{W}(\hat{\theta},\bar{t},K) \) with \( \bar{t} = \arg\max_{t \in \Pi} \tilde{W}(\hat{\theta},t,K) \) converges in distribution to \( \sup_{t \in \Pi} z_\ast^2(t) \). Furthermore under \( H_1 \), \( \tilde{W}(\hat{\theta},\bar{t},K)/n \to \eta_\ast(t) \) a.s. uniformly on \( \Pi \) and consequently \( \sup_{t \in \Pi} \tilde{W}(\hat{\theta},t,K)/n \to \sup_{t \in \Pi} \eta_\ast(t) \) a.s.

Proof. See Appendix \( \square \)

Note that the covariance function \( \Gamma_\ast(t_1,t_2) \) depends on the DGP of the model, so does the distribution of \( \sup_{t \in \Pi} z_\ast^2(t) \). 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 this by introducing some penalty functions. The alternative procedure similar to Bierens (1990) is the following. Choose independently of the data generating process real numbers \( \gamma > 0, \rho \in (0,1) \), and a point \( t_0 \in \Pi \). Let \( \bar{t} = \arg\max_{t \in \Pi} \tilde{W}(\hat{\theta},t,K) \) and let

$$\bar{t} = t_0, \text{ if } \tilde{W}(\hat{\theta},\bar{t},K) - \tilde{W}(\hat{\theta},t_0,K) \leq \gamma n^\rho; \bar{t} = \bar{t}, \text{ if } \tilde{W}(\hat{\theta},\bar{t},K) - \tilde{W}(\hat{\theta},t_0,K) \geq \gamma n^\rho.$$  

Then we have the following theorem.

**Theorem 5.** Let Assumptions 1-5, 7 hold. then under \( H_0 \), when \( K \to \infty \) and \( K^4/n \to 0 \), \( \tilde{W}(\hat{\theta},\bar{t},K) \to \chi^2_1 \) in distribution, whereas under \( H_1 \), \( \tilde{W}(\hat{\theta},\bar{t},K)/n \to \sup_{t \in \Pi} \eta_\ast(t) \) a.s.

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

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

**Theorem 6.** 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 \( \Pi \). Replace \( \bar{t} \) by \( \bar{t} = \arg\max_{t \in \{t_1,\cdots,t_L\}} \tilde{W}(\hat{\theta},t,K) \). Then the previous two theorems carry over for \( \bar{t} \).

Proof. Similar to the Proof of Bierens (1990) Theorem 5. \( \square \)
4.1 Local Alternative analysis

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

We consider the following local alternative:

\[ H^L_1: Y = f(X, \theta_0) + \frac{g(X)}{\sqrt{n}} + \epsilon(\theta_0), \]  

(13)

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

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

and the following theorem holds:

**Theorem 7.** Let Assumptions 1-5 and 7 hold, when \( K \to \infty \) and \( K^4/n \to 0 \), then \( \tilde{W}(\hat{\theta}, t, K) \) converges weakly to \( z^2_* \) under \( H^L_1 \), where \( z_* \) is a Gaussian element of \( C(\pi) \) with mean function

\[
\eta_*(t) = \frac{E^2\{g(X)\phi_*(X, t)\}}{s^2_*(\theta_0, t)}
\]

and covariance function \( \Gamma_*(t_1, t_2) \) defined in (12). Furthermore, under \( H^L_1 \), \( \tilde{W}(\hat{\theta}, t, K) \to \chi^2_1(\eta_*(t)) \).

**Proof.** See Appendix.

Note that under \( H^L_1 \), Bierens (1990) test has drift

\[
\eta_{NLS}(t) = \frac{E^2\{g(X)\exp(t'\Phi(X)) - b(\theta_0, t) A_{NLS}(\theta_0) \frac{\partial f(X, \theta_0)}{\partial \theta}\}}{s^2_{NLS}(\theta_0, t)}
\]

so we have the following corollary:

**Corollary 2.** When there exists conditional homoskedasticity, \( K \to \infty \) and \( K^4/n \to 0 \), our improved test is asymptotically equivalent to the Bierens (1990) test.

When there exists heteroskedasticity of unknown form, if \( E\{g(X)\frac{\partial f(X, \theta_0)}{\partial \theta}\} = 0, E\{g(X)\frac{\partial f(X, \theta_0)}{\partial \theta} \sigma^{-2}_0(X)\} = 0 \) and \( s^2_*(\theta_0, t) > s^2_{NLS}(\theta_0, t) \), then \( \eta_*(t) > \eta_{NLS}(t) \). We obtain a more powerful test.
5 Monte Carlo Simulations

We analyze 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 $z, v_1, v_2$, and $u$ follow the independent standard normal distribution, and let the regressors be $X_1 = z + v_1, X_2 = z + v_2$. The dependent variable is generated according to

$$Y = 1 + X_1 + X_2 + e.$$ 

Under the null, when the homoskedasticity is assumed, $e = u$, under heteroskedasticity, $e = (0.1 + 0.5x_1^2)^{1/2} u$. In both cases, OLS is employed to obtain the parameters estimator. Based on the OLS estimator and residuals, we calculate Bierens (1990) test and our improved Bierens (1990) test. Following Bierens (1990), we choose $L = \lfloor n/10 \rfloor - 1$ and $\Pi = [1, 5] \times [1, 5]$. $(t_1, \cdots, t_L)'$ have been drawn randomly from the uniform distribution on $\Pi$. $(t_1, \cdots, t_K)'$ 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

$$\{\gamma = 1, \rho = 0.5\} \quad \{\gamma = 0.5, \rho = 0.5\}$$

$$\{\gamma = 0.25, \rho = 0.5\} \quad \{\gamma = 0.25, \rho = 0.25\}.$$ 

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 1-4. 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: $Y = 1 + X_1 + X_2 + v_1 v_2 + u$. 

16
DGP 1.2: \( Y = 1 + X_1 + X_2 + v_1 v_2 + \left( 0.1 + 0.5x_1^2 \right)^{1/2} u. \)

DGP 2.1: \( Y = 1 + X_1 + X_2 + (1 + X_1 + X_2) \exp \left[ -0.01 (1 + X_1 + X_2)^2 \right] + u. \)

DGP 2.2: \( Y = 1 + X_1 + X_2 + (1 + X_1 + X_2) \exp \left[ -0.01 (1 + X_1 + X_2)^2 \right] + \left( 0.1 + 0.5x_1^2 \right)^{1/2} u. \)

DGP 3.1: \( Y = 1 + X_1 + X_2 + \sin (1 + X_1 + X_2) + u. \)

DGP 3.2: \( Y = 1 + X_1 + X_2 + \sin (1 + X_1 + X_2) + \left( 0.1 + 0.5x_1^2 \right)^{1/2} u. \)

DGP 4.1: \( Y_j = 1 + X_1 + X_2 + \cos (1 + X_1 + X_2) + u. \)

DGP 4.2: \( Y_j = 1 + X_1 + X_2 + \cos (1 + X_1 + X_2) + \left( 0.1 + 0.5x_1^2 \right)^{1/2} u. \)

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 5-12. To save space, results of sample size 200 and 400 are reported in one figure for each alternative. 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. Even though the model is misspecified, we still can obtain consistent estimator of parameters of \( X_1 \) and \( X_2 \) under this alternative, since \( E(v_1 v_2 X_i) = 0 \) 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 \( K \) we considered. 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 in the sense that the asymptotic variance of the GMM estimator decreases.
6 Conclusion

In this paper we propose a simple approach to consistent testing of functional form which is robust and efficient under heteroskedasticity of unknown form. By exploiting the duality property of one class of weighting functions for both consistent specification testing and efficient estimation of regression models, an innovative residual empirical process is proposed, employing a projection-based transformation. It has been shown that the new residual empirical process is not influenced by the parameter estimation effect, only a preliminary $\sqrt{n}$-consistent estimator is needed. We establish its efficiency in the sense that the GMM estimator reaching semiparametric efficiency bound under conditional heteroskedasticity of unknown form is employed in the new empirical process under the null. The simulation results of the new version of Bierens (1990) test based on the new empirical process demonstrate the good finite sample properties of the new test statistic.
Appendices

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 $w(t'_i X) = 1$. It is always possible to normalize $q^K(X)$ into this case when $w(t'_i X) \neq 1$. Firstly it is easy to check that $q(t_j X) \in L_2$ for $j = 1, 2, \cdots$. For $K = 2, 3, \cdots$, let

$$
\zeta_K(X) = \sum_{j=1}^K \alpha_{K,j} w(t'_j X),
$$

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

$$
E[\zeta_K(X)w(t'_j X)] = 0 \text{ if } j < K.
$$

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 $\varphi$ on the range of $X$ satisfying $E[\varphi(X)^2] < \infty$, with inner product $(\psi_K, \varphi) = E[\psi_K(X)\varphi(X)]$. Then by Theorem 2.4.2 of Brockwell and Davis (1991), for any $\varepsilon$, 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. To prove this theorem, instead of working on the $\bar{M}(\hat{\theta}, t, K)$ using mean value theorem directly, we rewrite $\bar{M}(\hat{\theta}, t, K)$ into

$$
\bar{M}(\hat{\theta}, t, K) = n^{-1/2} \sum_{j=1}^n e_j(\hat{\theta}) w(t'\Phi(X_j)) - n^{-1/2} \hat{b}(\hat{\theta}, t) \hat{A}(\hat{\theta}, K) \sum_{j=1}^n q^K(X_j)e_j(\hat{\theta}).
$$

We show that
\[ \hat{M}(\theta, t, K) = n^{-1/2} \sum_{j=1}^{n} e_j(\theta_0)w(t'\Phi(X_j)) - b(\theta_0, t)n^{1/2}(\hat{\theta} - \theta_0) + o_p(1) \]

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

\[ = n^{-1/2} \sum_{j=1}^{n} e_j(\theta_0)w(t'\Phi(X_j)) \]

\[ - b(\theta_0, t)A(\theta_0, K)n^{-1/2} \sum_{j=1}^{n} e_j(\theta_0)q^K(X_j) + o_p(1) \]

\[ = n^{-1/2} \sum_{j=1}^{n} e_j(\theta_0)[w(t'\Phi(X_j)) - b(\theta_0, t)A(\theta_0, K)q^K(X_j)] + o_p(1). \]

Since by the mean value theorem we have

\[ \hat{M}(\hat{\theta}, t) = n^{-1/2} \sum_{j=1}^{n} e(\theta_0)w(t'\Phi(X_j)) - \hat{b}(\theta, t)n^{1/2}(\hat{\theta} - \theta_0), \]

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

\[ E\left[ \sup_{\theta \in \Delta} \left\| w(t'\Phi(X)) \frac{\partial f(X, \theta)}{\partial \theta'} \right\|^2 \right] = E\left[ \left\| w(t'\Phi(X)) \right\| \left\| \frac{\partial f(X, \theta)}{\partial \theta'} \right\| \right] \]

\[ < \left[ E\left\| w(t'\Phi(X)) \right\|^2 \right]^{1/2} \left[ E\sup_{\theta \in \Delta} \left\| \frac{\partial f(X, \theta)}{\partial \theta'} \right\|^2 \right]^{1/2} \]

\[ < \infty. \]

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 \( \hat{b}(\theta, t) \xrightarrow{p} b(\theta_0, t), \hat{b}(\hat{\theta}, t) \xrightarrow{p} b(\theta_0, t) \). So

\[ n^{-1/2} \sum_{j=1}^{n} e(\hat{\theta})w(t'\Phi(X_j)) = n^{-1/2} \sum_{j=1}^{n} e(\theta_0)w(t'\Phi(X_j)) - b(\theta_0, t)n^{1/2}(\hat{\theta} - \theta_0) + o_p(1). \]
Similarly
\[
E \left[ \sup_{\theta \in \Delta} \left\| q^K(X) \frac{\partial f(X, \theta)}{\partial \theta'} \right\| \right] = E \left[ \left\| q^K(X) \sup_{\theta \in \Delta} \left\| \frac{\partial f(X, \theta)}{\partial \theta'} \right\| \right\] \\
< \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.
\]

Then
\[
\text{plim sup}_{\theta \in \Delta} \left\| \frac{1}{n} \sum_{j=1}^{n} q^K(X) \frac{\partial f(X, \theta)}{\partial \theta'} - E \left[ q^K(X) \frac{\partial f(X, \theta)}{\partial \theta'} \right] \right\| = 0,
\]

By similar argument, we have
\[
n^{-1/2} \sum_{j=1}^{n} e(\hat{\theta}) q^K(X_j) = n^{-1/2} \sum_{j=1}^{n} e(\theta_0) q^K(X_j) - \Lambda(\theta_0, K)n^{1/2}(\hat{\theta} - \theta_0) + o_p(1),
\]

and \( \hat{\Lambda}(\hat{\theta}, K) \overset{p}{\rightarrow} \Lambda(\theta_0, K) \).

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

Then by similar argument, \( \hat{\Omega}(\hat{\theta}, K) \overset{p}{\rightarrow} \Omega(\theta_0, K) \). Also by Assumptions 4 and 5, for any \( K > D, \Omega(\theta_0, K) \) is positive definite. By continuous mapping theorem, for \( K > D, \hat{\Lambda}(\hat{\theta}, K) \hat{\Omega}(\hat{\theta}, K)^{-1} \hat{\Lambda}(\hat{\theta}, K) \overset{p}{\rightarrow} \Lambda(\theta_0, K)'\Omega(\theta_0, K)^{-1}\Lambda(\theta_0, K), \hat{\Lambda}(\hat{\theta}, K)' \hat{\Omega}(\hat{\theta}, K)^{-1} \overset{p}{\rightarrow} \Lambda(\theta_0, K)'\Omega(\theta_0, K)^{-1}. \) Note also that \( A(\theta_0, K) \Lambda(\theta_0, K) = I \).

To prove (10), we rewrite \( \hat{M}(\hat{\theta}, t, K) \) as
\[
\hat{M}(\hat{\theta}, t, K) = [1, -b((\theta_0, t)A(\theta_0, K))]n^{-1/2} \sum_{j=1}^{n} e_j(\theta_0) \left( w(t' \Phi(X_j)) q^K(X_j) \right)' + o_p(1). \quad (14)
\]

By Lindberg-Feller central limit theory and Slutsky theorem, we have
\[
\hat{M}(\hat{\theta}, t, K) \overset{d}{\rightarrow} N \left[ 0, s^2(\theta_0, t, K) \right].
\]
Proof of Theorem 2. By the mean value theorem, we have

\[
\hat{M}(\hat{\theta}, t) = n^{-1/2} \sum_{j=1}^{n} [Y_j - f(X_j, \theta_0)]w(t'\Phi(X_j)) - b(\theta_0, t)n^{1/2}(\hat{\theta} - \theta_0) + o_p(1). \tag{15}
\]

From (11), we have

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

Plug in (15), so we have

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

\[
- b(\theta_0, t) \left[\left[\Lambda(\theta_0, K)'\Omega(\theta_0, K)^{-1}\Lambda(\theta_0, K)\right]^{-1} \left[\Lambda(\theta_0, K)'\Omega(\theta_0, K)^{-1} n^{-1/2} \sum_{j=1}^{n} q^K(X_j)[Y_j - f(X_j, \theta_0)]\right]
\]

Since we also have

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

then

\[
\hat{M}(\hat{\theta}, t, K) = \hat{M}(\hat{\theta}, t) + o_p(1).
\]

Proof of Theorem 3. Based on Theorem 2, we only need establish that GMM estimator \( \hat{\theta} \) 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 \( \tilde{\theta} \) satisfies

\[
n^{1/2}(\tilde{\theta} - \theta)^{d} \to N(0, V)
\]

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 \( \hat{\theta} \) reaches semiparametric efficiency bound, It is straightforward to obtain \( s^2_\epsilon(\theta_0, t) \). Furthermore, the result of \( A_\epsilon(\theta_0) \) comes directly from Lemma A.4 in Donald et al., (2003). Note that \( \left[\Lambda(\theta_0, K)'\Omega(\theta_0, K)^{-1}\Lambda(\theta_0, K)\right]^{-1} p \to...
To prove which means that the condition of nonsingularity is satisfied.

Proof of Theorem 4. We still assume that that $X$ is bounded itself, so that we may choose $\Phi(X) = X$. We set $\exp(t'X) = 1$. 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 $[E\zeta_K(X)^2] > 0$ almost surely. So for $K = 2, 3, \cdots$, 

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

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 $E\left(q^K(X)\tilde{q}^K(X)'ight) = I_K$, which means that the condition of nonsingularity is satisfied.

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

\[
\sup_{X \in \mathbb{Z}}||\tilde{q}^K(X)|| \leq C\left[\sum_{k=1}^{K} k^2\right]^{1/2} \\
\leq CK^{3/2}.
\]

To prove $s^2(\theta_0, t, K) > 0$, note that for any $t \in \Pi$, $t \neq t_j$ for $j = 1, \cdots, K$. Denote $q^{K+1}(X) = (\exp(t'_1\Phi(X)), \cdots, \exp(t'_K\Phi(X)), \exp(t'\Phi(X)))'$. Then based on Lemma 2 we can obtain that $E\left(q^{K+1}(X) q^{K+1}(X)'ight)$ has smallest eigenvalue bounded away from zero. Note that $E[(Y - f(X, \theta_0))|X]^2 > 0$, then $E\left((Y_j - f(X_j, \theta_0))^2 q^{K+1}(X) q^{K+1}(X)'ight)$ is positive definite. From (14) in the proof of Theorem 1 it is easy to obtain that $s^2(\theta_0, t, K) > 0$.

Proof of Theorem 4. The result under $H_1$ follows straightforwardly from the uniform law of large numbers. Under $H_0$, Define 

$z_n(\theta_0, t, K) = n^{-1/2} \sum_{j=1}^{n} [Y_j - f(X_j, \theta_0)][\exp(t'\Phi(X_j)) - b(\theta_0, t)A(\theta_0, K) q^K(X_j)]/\sqrt{s^2(\theta_0, t, K)}.$
Following the Proof of (9) in Theorem 1, we have under $H_0$

$$p \lim_{n \to \infty} \sup_{t \in \Pi} |\tilde{W}(\hat{\theta}, t, K) - z_n^2(\theta_0, t, K)| = 0.$$  

Following the Proof of Lemma 4 in Bierens (1990), we can obtain under $H_0$, $z_n(\theta_0, t, K)$ is tight. Then We allow $K \to \infty$, $\frac{K^4}{n} \to 0$, the following result holds

$$p \lim_{n \to \infty} \sup_{t \in \Pi} |z_n^2(\theta_0, t, K) - z^2(\theta_0, t)| = 0.$$  

It is also easy to prove that for arbitrary $t_1, \ldots, t_m$ in $\Pi$, $(z_n(\theta_0, t_1, K), \ldots, z_n(\theta_0, t_m, K))'$ is asymptotically distributed as $(z(\theta_0, t_1), \ldots, z(\theta_0, t_m))'$. Then $z_n$ converges weakly to $z$. Following the functional limit theory of Billingsley (1968 p. 47), we have the results.

Proof of Theorem 7. Similar to the proof of Theorem 4.
References


Figure 1: Size of testing at 5% level, Sample size 200, \( u_j = e_j \)

\[
\begin{align*}
\gamma &= 1, \rho = 0.5 \\
\gamma &= 0.5, \rho = 0.5 \\
\gamma &= 0.25, \rho = 0.5 \\
\gamma &= 0.25, \rho = 0.25
\end{align*}
\]

Figure 2: Size of testing at 5% level, Sample size 400, \( u_j = e_j \)

\[
\begin{align*}
\gamma &= 1, \rho = 0.5 \\
\gamma &= 0.5, \rho = 0.5 \\
\gamma &= 0.25, \rho = 0.5 \\
\gamma &= 0.25, \rho = 0.25
\end{align*}
\]
Figure 3: Size of testing at 5% level, Sample size 200, $u_j = (0.1 + 0.5x_i^2)^{1/2} e_j$

$\gamma = 1, \rho = 0.5$

$\gamma = 0.5, \rho = 0.5$

$\gamma = 0.25, \rho = 0.5$

$\gamma = 0.25, \rho = 0.25$

Figure 4: Size of testing at 5% level, Sample size 400, $u_j = (0.1 + 0.5x_i^2)^{1/2} e_j$

$\gamma = 1, \rho = 0.5$

$\gamma = 0.5, \rho = 0.5$

$\gamma = 0.25, \rho = 0.5$

$\gamma = 0.25, \rho = 0.25$
Figure 5: Power of testing at 5% level, DGP 1.1

- \(\gamma = 1, \rho = 0.5\)
- \(\gamma = 0.5, \rho = 0.5\)
- \(\gamma = 0, \rho = 0.5\)
- \(\gamma = 0.25, \rho = 0.5\)
- \(\gamma = 0.25, \rho = 0.25\)

The New Test, Sample Size 200
Bierens (1990) Test, Sample size 200
The New Test, Sample Size 400
Bierens (1990) Test, Sample Size 400
Figure 6: Power of testing at 5% level, DGP 1.2

\[ \gamma = 1, \rho = 0.5 \]

\[ \gamma = 0.5, \rho = 0.5 \]

\[ \gamma = 0.25, \rho = 0.5 \]

\[ \gamma = 0.25, \rho = 0.25 \]

---

Bierens (1990) Test, Sample Size 200

The New Test, Sample Size 200

The New Test, Sample Size 400

Bierens (1990) Test, Sample Size 400
Figure 7: Power of testing at 5% level, DGP 2.1

- For $\gamma = 1, \rho = 0.5$, the New Test (red) and Bierens (1990) Test (black) show similar power at sample sizes of 200 and 400.
- For $\gamma = 0.5, \rho = 0.5$, the New Test (red) has higher power compared to Bierens (1990) Test (black) at sample sizes of 200 and 400.
- For $\gamma = 0, \rho = 0.5$, the New Test (red) and Bierens (1990) Test (black) show similar power at sample sizes of 200 and 400.
- For $\gamma = 0.25, \rho = 0.5$, the New Test (red) has higher power compared to Bierens (1990) Test (black) at sample sizes of 200 and 400.
- For $\gamma = 0.25, \rho = 0.25$, the New Test (red) has higher power compared to Bierens (1990) Test (black) at sample sizes of 200 and 400.
Figure 8: Power of testing at 5% level, DGP 2.2

\[ \gamma = 1, \rho = 0.5 \]
\[ \gamma = 0.5, \rho = 0.5 \]
\[ \gamma = 0.25, \rho = 0.5 \]
\[ \gamma = 0.25, \rho = 0.25 \]

The New Test, Sample Size 200
The New Test, Sample Size 400
Bierens (1990) Test, Sample size 200
Bierens (1990) Test, Sample Size 400
Figure 9: Power of testing at 5% level, DGP 3.1

\[ \gamma = 1, \rho = 0.5 \]
\[ \gamma = 0.5, \rho = 0.5 \]
\[ \gamma = 0.25, \rho = 0.5 \]
\[ \gamma = 0.25, \rho = 0.25 \]

- **Bierens (1990) Test, Sample Size 200**
- **Bierens (1990) Test, Sample Size 400**
- **The New Test, Sample Size 200**
- **The New Test, Sample Size 400**
Figure 10: Power of testing at 5% level, DGP 3.2

The New Test, Sample Size 200
Bierens (1990) Test, Sample size 200

The New Test, Sample Size 400
Bierens (1990) Test, Sample Size 400
Figure 11: Power of testing at 5% level, DGP 4.1
Figure 12: Power of testing at 5% level, DGP 4.2

\[ \gamma = 1, \rho = 0.5 \]

\[ \gamma = 0.5, \rho = 0.5 \]

\[ \gamma = 0.25, \rho = 0.5 \]

\[ \gamma = 0.25, \rho = 0.25 \]

The New Test, Sample Size 200

The New Test, Sample Size 400

Bierens (1990) Test, Sample size 200

Bierens (1990) Test, Sample Size 400