Econometric Theory, **15**, 1999, 710–718. Printed in the United States of America.

OPTIMALITY FOR THE INTEGRATED CONDITIONAL MOMENT TEST

WM. BRENT BONING
The CNA Corporation

FALLAW SOWELL
Carnegie Mellon University

This paper proposes a version of the integrated conditional moment (ICM) test that is optimal for a class of composite alternatives. The ICM test is built on the fact that a random function based on a correctly specified model should have zero mean, whereas any misspecification in the conditional mean implies a divergent mean for the random function. We derive test statistics that are optimal for each basis element of an orthonormal decomposition of the function space for which the random function is an element. We then use a weighted summation of these test statistics to compose the single test statistic that is optimal for any pair of alternatives that are symmetric about zero. This test is equivalent to using a particular measure in the ICM test of Bierens and Ploberger (1997, *Econometrica* 65, 1129–1152).

1. INTRODUCTION

This paper proposes a version of the integrated conditional moment (ICM) test that is optimal for a class of composite alternatives. The ICM test is built on the fact that a random function based on a correctly specified model should have zero mean, whereas any misspecification in the conditional mean implies a divergent mean for the random function. We derive test statistics that are optimal for each basis element of an orthonormal decomposition of the function space for which the random function is an element. We then use a weighted summation of these test statistics to compose the single test statistic that is optimal for any pair of alternatives that are symmetric about zero. This test is equivalent to using a particular measure in the test of Bierens and Ploberger (1997).

Bierens (1982, 1990) proposes tests that are consistent for any misspecification in the conditional mean. Consistency is achieved by considering a continuum of moment conditions via a random function created from the residuals and the conditioning data. The tests in Bierens (1982, 1990) and Bierens and Ploberger (1997) are tests of whether this random function has a nonzero mean. Stinchcombe and White (1998) and de Jong (1996) provide generalizations of the random functions. This paper is most closely related to Bierens and Ploberger (1997).

We thank Marine Carrasco, two anonymous referees, and Donald Andrews for helpful suggestions and comments.
Address correspondence to: Fallaw Sowell, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

 © 1999 Cambridge University Press 0266-4666/99 $9.50

https://doi.org/10.1017/S026646669915504X Published online by Cambridge University Press

Bierens and Ploberger considered an orthonormal decomposition of the function space to characterize the asymptotic distribution of a single test statistic (a weighted L_2 norm of the random function) in terms of the eigenfunctions of the covariance of the random function. The functional theoretic techniques used are similar to those used in Carrasco and Florens (1997).

The next section introduces the model and the random function and summarizes the results of Bierens (1990) used in this paper. Section 3 presents the tests, and Section 4 discusses the optimality properties of the test.

2. MODELING AND THE SAMPLE RANDOM FUNCTION

Following Bierens (1990) the tests will be based on a random function constructed from the observed sample. The null hypothesis is a correctly specified parametric model

$$H_0 : y_j = f(x_j, \theta_0) + \epsilon_j$$

for a sample $(y_j, x_j')\ j = 1, \ldots, N$ where x_j is an m-dimensional vector of independent variables, $\theta_0 \in \Theta \subset R^k$ is a vector of unknown parameters, ϵ_j are independent[1] with mean 0 and variance σ^2 and are independent of all the x_j. The function $f(\cdot, \cdot)$ is assumed known. The null implies the moment condition

$$E_j[y_j - f(x_j, \theta_0)] = 0, \tag{1}$$

where E_j denotes the expectation conditional on x_j. The moment condition is used to determine the parameter estimates $\hat{\theta}_N$ by efficient nonlinear least squares. The alternative is that the model is incorrectly specified:

$$H_1 : \sup_{\theta \in \Theta} P[E[y_j | x_j] = f(x_j, \theta)] < 1.$$

The sample is used to construct the random function[2]:

$$X_N(\xi) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} (y_j - f(x_j, \hat{\theta}_N)) \exp(\xi' \Phi(x_j)),$$

where $\xi \in \Xi$, a compact subset of R^m not containing the origin. As in Bierens (1990), Φ is a bounded, continuous, one-to-one function. To reduce notation, we will use $w_j(\xi)$ to denote the weighting function $\exp(\xi' \Phi(x_j))$. The choice of $w_j(\xi)$ is not restrictive; Stinchcombe and White (1998) show that the results of Bierens (1990) extend to the class of analytic functions that are not polynomials. The results in this paper hold for other weighting functions. The moment condition (1) implies that the residuals should be orthogonal to every function of x_j. As ξ varies over Ξ, the weighting function becomes a continuum of functions of x_j, and the random function $X_N(\xi)$ considers a continuum of moment conditions implied by (1).

https://doi.org/10.1017/S026646669915504X Published online by Cambridge University Press

Under the alternative hypothesis the misspecified model is of the form $y_j = f(x_j,\theta) + g(x_j,\beta) + \epsilon_j$, where $g(x,\beta) = E[y|x] - f(x,\theta)$. With this alternative, the sample random function satisfies

$$X_N(\xi) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \epsilon_j \left(\exp(\xi'\Phi(x_j)) - b(\xi)'A^{-1} \frac{\partial f(x_j,\theta_0)}{\partial \theta'} \right) + \frac{1}{\sqrt{N}} \sum_{j=1}^{N} g(x_j,\beta) \left(\exp(\xi'\Phi(x_j)) - b(\xi)'A^{-1} \frac{\partial f(x_j,\theta_0)}{\partial \theta'} \right) + o_p(1), \tag{2}$$

where $b(\xi) = E[(\partial f(x_1,\theta_0)/\partial\theta') \exp(\xi'\Phi(x_1))]$ and $A = E[(\partial f(x_1,\theta_0)/\partial\theta')(\partial f(x_1,\theta_0)/\partial\theta)]$. The first term on the right hand side of (2) weakly converges to a random function that is the same under both the null and the local alternative. The second term converges to 0 under the null (because $g(x_j,\beta) = 0$) and diverges under the alternative.

We operate under assumptions A and B of Bierens (1990). Bierens's (1990) Theorem 3 shows that, under the null, the random function $X_N(\xi)$ converges to a Gaussian random function, $X(\xi) \in C(\Xi)$, with zero mean and covariance defined by

$$\Sigma(\xi_1,\xi_2) = E\left[\{y_j - f(x_j,\theta_0)\}^2 \left\{ w_j(\xi_1) - b(\xi_1)'A^{-1} \frac{\partial f(x_j,\theta_0)}{\partial \theta'} \right\} \times \left\{ w_j(\xi_2) - b(\xi_2)'A^{-1} \frac{\partial f(x_j,\theta_0)}{\partial \theta'} \right\} \right],$$

which we assume is continuous, positive definite, and symmetric. Under the alternative, the normalized sample random function $(1/\sqrt{N})X_N(\xi)$ converges to its mean function $\eta(\xi) \in C(\Xi)$.

3. THE TEST STATISTIC

The test we propose uses a complete orthonormal sequence of functions[3] $\{e_i(\xi)\}$ that forms an orthonormal basis in $L_2(\Xi)$, the inner product space of squared integrable functions on Ξ with the inner product

$$(\eta_1,\eta_2)_\Gamma \equiv \int_\Xi \int_\Xi \eta_1(\xi_1)\Gamma(\xi_1,\xi_2)\eta_2(\xi_2)\, d\xi_1\, d\xi_2,$$

where $\eta_1,\eta_2 \in L_2(\Xi)$ and $\Gamma : \Xi \cdot \Xi \to R$. To reduce notation, we drop the index variable ξ when there should be no confusion.

Because $\{e_i(\xi)\}$ $i = 1,2,\ldots$ is a complete sequence of orthonormal functions on $L_2(\Xi)$, every $\eta \in L_2(\Xi)$ can be written as

$$\eta(\xi) = \lim_{R\to\infty} \sum_{i=1}^{R} c_i e_i(\xi),$$

https://doi.org/10.1017/S026646669915504X Published online by Cambridge University Press

where $c_i \equiv (e_i, \eta)_\Gamma$ and lim indicates convergence in mean. A sequence of tests is created by squaring the inner products of the sample random function and each of the orthonormal elements[4]:

$$T_i = [(e_i, X)_{\Sigma^{-1}}]^2, \qquad i = 1, \ldots.$$

Under the null, $X(\xi)$ is a zero mean Gaussian random function with covariance $\Sigma(\xi_1, \xi_2)$; hence the inner product $(e_i, X)_{\Sigma^{-1}}$ will be a Gaussian random variable with zero mean and variance λ_i^{-1}, where λ_i are the eigenvalues of the covariance function that satisfy $\lambda_i = (e_i, e_i)_\Sigma$ (see Rozanov, 1982). This implies that under the null T_i normalized by λ_i has a chi-square distribution with one degree of freedom.[5]

This sequence of test statistics is used to construct the single test statistic:

$$CT = \sum_{i=1}^{\infty} \lambda_i^2 T_i, \tag{3}$$

$$= \sum_{i=1}^{\infty} \lambda_i^2 \left[\int_\Xi \int_\Xi X(\xi) \Sigma(\xi, \xi_1)^{-1} e_i(\xi_1)\, d\xi\, d\xi_1 \right]^2, \tag{4}$$

$$= \sum_{i=1}^{\infty} \lambda_i^2 \left[\int_\Xi \lambda_i^{-1} X(\xi) e_i(\xi)\, d\xi \right]^2, \tag{5}$$

$$= \int_\Xi X(\xi)^2\, d\xi, \tag{6}$$

where the last equality is an application of Parseval's theorem (see Rudin, 1976, p. 331, Theorem 11.45). The test statistic CT is a special case of the consistent ICM test of Bierens and Ploberger (1997) where $d\xi$ is the weighting measure. The distribution of CT is characterized in Bierens and Ploberger.

Subsequently we show that this test statistic is an optimal version of the integrated conditional moment test. In the following sections, we also discuss the test's power and its relationship to the tests in Bierens (1990) and Bierens and Ploberger (1997).

4. OPTIMALITY AND POWER

4.1. Generic Approach to Testing

Asymptotically, the null hypothesis implies a measure (say, μ_0) on $C(\Xi)$. The alternative implies a different measure (say, μ_1). The sample random function converges to a realization from one of these two measures. A test is implied by any functional $H: C(\Xi) \to R$. Applying the functional to the sample random function results in a test statistic. The continuous mapping theorem implies that if the functional is continuous then the asymptotic distribution under the null and alternative can be characterized by applying the functional to the random func-

https://doi.org/10.1017/S026646669915504X Published online by Cambridge University Press

tions drawn from μ_0 and μ_1, respectively. Following this approach, we need only to specify a functional and this determines the test statistic and its distributions under the null and under the local alternative.

Note that any functional can be used as a test statistic. Bierens (1990) suggests using $\max_{\Xi}|X(\xi)|$, and Bierens and Ploberger (1997) suggest using $\int_{\Xi} X^2(\xi)\, d\mu(\xi)$ for $\mu(\xi)$, a weighting function in $C(\Xi)$. Because a continuum of functionals can be used as the test statistic, a natural issue is to identify the best test. One candidate for the best test is the functional that has the highest power against a specific local alternative. The next subsection characterizes that functional.

4.2. Most Powerful Test

To address issues of power the Neyman–Pearson approach can be applied. The standard Neyman–Pearson lemma (Lehmann, 1959, p. 65) gives the form of the most powerful test when testing between two probability measures that possess densities. The Radon–Nikodym derivative generalizes this ratio when the probability measures do not have densities.

THEOREM 1. *Let $X(\xi)$ be a draw from the measure space $(C(\Xi),\mathcal{B})$. If the null hypothesis implies the measure μ_0 and the alternative hypothesis implies the measure μ_1, and if μ_0 is absolutely continuous with respect to μ_1, then the most powerful level α test will have a critical region $\mathcal{C}_*$, defined by*

(1) $(d\mu_1/d\mu_0)(X(\xi)) \geq k$ *for each* $X(\xi) \in \mathcal{C}_*$,
(2) $(d\mu_1/d\mu_0)(X(\xi)) < k$ *for each* $X(\xi) \notin \mathcal{C}_*$, *and*
(3) $\int_{\mathcal{C}_*} d\mu_0 = \alpha$,

where $(d\mu_1/d\mu_0)(X(\xi))$ is the Radon–Nikodym derivative of μ_1 with respect to μ_0 evaluated at $X(\xi)$.

Proof. See Appendix 3 in Sowell (1996). ■

This theorem shows that the Radon–Nikodym derivative is the functional that gives the most powerful test. Hence, optimal testing requires the Radon–Nikodym derivative of the measure implied by the local alternative with respect to the measure implied by the null. The general form of the Radon–Nikodym derivative is given in the following theorem.

THEOREM 2. *Let μ_0 and μ_1 denote Gaussian measures on $C(\Xi)$. Let the means and covariances of μ_0 and μ_1 be $(0,\Sigma(\xi_1,\xi_2))$ and $(\eta(\xi),\Sigma(\xi_1,\xi_2))$ respectively. If*

(1) *the mean function $\eta(\xi)$ is an element of $L_2(\Xi)$, the space of square integrable functions on Ξ, and*

https://doi.org/10.1017/S026646669915504X Published online by Cambridge University Press

(2) *the covariance function is continuous, positive definite, and symmetric then the Radon–Nikodym derivative of* μ_1 *with respect to* μ_0 *is*

$$\frac{d\mu_1}{d\mu_0}(X(\cdot)) = \exp\left\{ \int_{\Xi}\int_{\Xi} \eta(\xi_1)\Sigma(\xi_1,\xi_2)^{-1}X(\xi_2)\,d\xi_1\,d\xi_2 - \frac{1}{2}\int_{\Xi}\int_{\Xi} \eta(\xi_1)\Sigma(\xi_1,\xi_2)^{-1}\eta(\xi_2)\,d\xi_1\,d\xi_2 \right\}.$$

Proof. Because the covariance function is positive definite, its eigenfunctions form a complete orthonormal system for the space $L_2(\Xi)$. The theorem is an immediate application of Gihman and Skorohod (1974, Theorem 1, p. 490). ■

4.3. Optimality of T_i

The test statistic T_i has the greatest power against a specific local alternative. To show this optimality property, we derive an optimal test for the misspecified model with local alternatives studied in Bierens and Ploberger (1997):

$$y_j = f(x_j,\theta_0) + \frac{\gamma g(x_j)}{\sqrt{N}} + \epsilon_j,$$

where γ is a scalar parameter. First calculate the most powerful test for a specific value of γ. A weighting function can then be used to average these most powerful tests. The resulting test will have the greatest weighted average power for the local alternatives (see Andrews and Ploberger, 1994).

A simple generalization of Theorem 3 in Bierens (1990) shows that for this local alternative the sample random function weakly converges to a draw from $C(\Xi)$ with mean function $\eta(\xi)$ and covariance function $\Sigma(\xi_1,\xi_2)$. Note that $g(x_j) \neq 0$ implies that the mean function $\eta(\xi)$ will be nonzero, by (2). If there were no misspecification then the mean function would be zero and the covariance would remain the same. Asymptotically the two models imply different measures on the space $C(\Xi)$, and testing for this misspecification is equivalent to testing between these two measures.

Because the alternative's parameter enters linearly and using the definition of the inner product, we can write the Radon–Nikodym derivative as

$$\exp\left\{\gamma(\eta,X)_{\Sigma^{-1}} - \frac{\gamma^2}{2}(\eta,\eta)_{\Sigma^{-1}}\right\},$$

which is the functional that implies the most powerful test for the point local alternative $\gamma g(\cdot)/\sqrt{N}$, or equivalently,

$$\exp\{\gamma(\eta,X)_{\Sigma^{-1}}\}.$$

https://doi.org/10.1017/S026646669915504X Published online by Cambridge University Press

This can be used to construct a test that is optimal for the composite alternative of $\{-\gamma g(\cdot)/\sqrt{N}, \gamma g(\cdot)/\sqrt{N}\}$, where optimality is defined in terms of maximum average power over these two alternatives. The optimal test statistic is

$$\tfrac{1}{2}\exp\{-\gamma(\eta, X)_{\Sigma^{-1}}\} + \tfrac{1}{2}\exp\{\gamma(\eta, X)_{\Sigma^{-1}}\}.$$

Because this functional is monotonically increasing in $T = [(\eta, X)_{\Sigma^{-1}}]^2$, the test statistic T maximizes average power over these two alternatives for all $\gamma > 0$. Hence, the test statistic T_i, given in (3), is optimal for an alternative that implies the drift $e_i(\xi)$.

Under any local alternative $g(\cdot)/\sqrt{N}$, T_i is the square of a normal random variable that has mean $(e_i, \eta)_{\Sigma^{-1}}$ and variance λ_i^{-1}; hence $\lambda_i T_i$ has a noncentral chi-square distribution with one degree of freedom and noncentrality parameter $\lambda_i[(e_i, \eta)_{\Sigma^{-1}}]^2$. Because we assume that $\{e_i\}_{i=1}^{\infty}$ is a complete basis for $C(\Xi)$, the sequence of tests will be consistent against global alternatives that have nonzero projection into $C(\Xi)$. To see this suppose the actual misspecification were of the form $h(x_j)$ with implied drift of $\eta_h(\xi)$ (Bierens, 1990, Theorem 3). As long as $\eta_h(\xi)$ has a nonzero projection into $C(\Xi)$, at least one of the inner products $(e_i, X_N)_{\Sigma^{-1}}$ is nonzero, resulting in nontrivial power against the local alternative.

4.4. Optimality of *CT* and Relation to the ICM Test of Bierens and Ploberger (1997)

The *CT* test defined in equation (3) is optimal in the sense of having the greatest weighted average local power as defined in Andrews and Ploberger (1994). Consider a countable infinity of summable positive weights, $p_i > 0$ for $i = 1, 2, \ldots$. Let the basis elements be the eigenfunctions of the covariance function, let the weight p_i assigned to a basis element e_i be λ_i^2, the square of the associated eigenvalue, and consider the test statistic

$$\sum_{i=1}^{\infty} p_i T_i.$$

This statistic is an optimal test statistic in the sense that it has the greatest weighted average local power for the sequence of local alternatives that imply the basis elements $\{e_i(\xi)\}$ and the given weights p_i.

Because the sequence of basis elements is complete and orthonormal and the weights $p_i = \lambda_i^2$ are used, asymptotically the test reduces to

$$CT = \sum_{i=1}^{\infty} \lambda_i^2 T_i = \int_{\Xi} X(\xi)^2 \, d\xi$$

as shown previously. The last expression is the ICM test statistic presented in Bierens and Ploberger (1997) with weighting density

$$d\mu(\xi) = d\xi.$$

https://doi.org/10.1017/S026646669915504X Published online by Cambridge University Press

Because the *CT* test is an ICM test, the consistency arguments of Bierens (1990) hold. This shows the ICM test statistic can be interpreted as an optimal test for any set of alternatives that are a complete orthonormal sequence of alternatives $\{\phi_i(\xi)\}$ with respect to the weighting kernel $\Gamma(\xi_1,\xi_2)$, and the weights on the alternatives are w_i^2 where

$$w_i = (\phi_i,\phi_i)_\Gamma.$$

This is a strong argument for the weighting density in (6) when using the ICM test.

5. CONCLUSION

This paper presents a specific measure to be used in the integrated conditional moment test of Bierens and Ploberger (1997) that makes the test optimal in the sense of greatest weighted average local power.

Note that the optimality of the test is derived using an orthonormal basis of the function space for the mean of the random function implied by the alternative. Bierens and Ploberger (1997) use a *different* orthonormal basis to characterize the distribution of the ICM test. Thus, using the weighting measure for the ICM test provided in this paper and the results of Bierens and Ploberger, two important qualities of the ICM test have been established: the optimality of the test (the present work) and the asymptotic distribution of the test (Bierens and Ploberger).

NOTES

1. Independence of the ϵ_j is not essential; de Jong (1996) demonstrates that dependent structures for the error term are accommodated by an ICM-type test.

2. The random function $X_N(\xi)$ corresponds to $\sqrt{N}\hat{M}(t)$ in Bierens (1990).

3. Bierens and Ploberger (1997) used the eigenfunctions of the covariance function Σ. An alternative is to select a sequence of functional forms and use the Gram–Schmidt orthogonalization procedure to form an orthonormal basis from this sequence of functions.

4. The proposed test uses $\Sigma(\xi,\xi)^{-1}$ as a weighting function in the inner product and the Gram–Schmidt orthogonalization procedure to derive an orthonormal basis. The choice of $\Sigma(\xi,\xi)^{-1}$ leads to the optimality properties of the test. However, any weighting function in the inner product will result in a functional that can serve as a test, e.g., the inner product $(\eta_1,\eta_2) = \int_\Xi \int_\Xi \eta_1(\xi_1) r(\xi_1,\xi_2) \eta_2(\xi_2)\, d\xi_1\, d\xi_2$, where $r(\xi_1,\xi_2)$ is a nonnegative weighting function.

5. In general, the test statistics T_i will not be independent. We thank an anonymous referee for pointing this out.

REFERENCES

Andrews, D.W.K. & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.

Bierens, H.J. (1982) Consistent model specification tests. *Journal of Econometrics* 20, 105–134.

Bierens, H.J. (1990) A consistent conditional moment test of functional form. *Econometrica* 58, 1443–1458.

https://doi.org/10.1017/S026646669915504X Published online by Cambridge University Press

Bierens, H.J. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica* 65, 1129–1152.
Carrasco, M. & J.-P. Florens (1997) Generalization of GMM to a Continuum of Moment Conditions. Mimeo, Department of Economics, Ohio State University.
Gihman, I.I. & A.V. Skorohod (1974) *The Theory of Stochastic Processes I.* New York: Springer-Verlag.
de Jong, R.M. (1996) The Bierens test under data dependence. *Journal of Econometrics* 72, 1–32.
Lehmann, E.L. (1959) *Testing Statistical Hypotheses.* New York: Wiley.
Rozanov, Yu.A. (1982) *Markov Random Fields.* New York: Springer-Verlag.
Rudin, W. (1976) *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill.
Sowell, F. (1996) Optimal tests for parameter instability in the generalized method of moments framework. *Econometrica* 65, 1085–1107.
Stinchcombe, M.B. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–325.

https://doi.org/10.1017/S026646669915504X Published online by Cambridge University Press