ANNALS OF ECONOMICS AND FINANCE **2**, 31–58 (2001)

Consistent Specification Tests for Regression Models

Chunrong Ai*

Department of Economics, University of Florida
Gainesville, FL 32611, USA

This paper presents three procedures for testing specifications of regression models. The first procedure is parametric and it tests the specification of an unconditional moment restriction model against a nonnested unconditional moment restriction model. The second procedure is nonparametric and it tests the specification of a conditional moment restriction model against a nonnested conditional moment restriction model. The third procedure also is nonparametric but it tests the specification of a conditional moment restriction model against all alternative specifications. All procedures permit heteroskedasticity of unknown form and are shown to be consistent. The test statistics of the first two procedures are simply the t-ratios of a 2SLS estimator while that of the third procedure is computable from LS output via auxiliary LS regressions.

Key Words: Nonparametric; Hypothesis Testing; Moment Restrictions; Series Estimator.

JEL Classification Numbers: C20

1. INTRODUCTION

Economic theory relates economic variables but rarely provides a functional form. In practice, linear, loglinear, semi-loglinear, and trans-loglinear forms are commonly used. However, if these functional forms misspecify the true relationship, parameter estimates will generally be inconsistent. Thus, it is important to subject the model to evaluation. Several procedures for testing functional forms have been proposed.[1] Most of those

* I would like to thank Dave Denslow, Charlie Hadlock, and Jeffrey Wooldridge for useful comments.

[1]Examples include Godfrey and Wickens (1981), Utts (1982), MacKinnon, White, and Davidson (1983), Davidson and Mackinnon (1981, 1985), Wooldrige (1991, 1992, 1994), De Jong and Bierens (1994), Hong and White (1996), and Fan and Li (1996). For a list of other references, see Godfrey, McAleer, and McKenzie (1988).

1529-7373/2001

procedures compare the regression model $y = f(x, \beta_0) + u$ with another regression model $\phi(y) = q(x, \alpha_0) + \varepsilon$, while the other procedures test the specification of the regression model $y = f(x, \beta_0) + u$ against all alternative specifications. All those procedures focus on conditional mean specifications and require the regressors be exogenous. However, in empirical studies, regressors often are endogenous and models often are specified as conditional moment restrictions (see Dadkhah and Valbuena (1985) and Hansen and Singleton (1982) for examples). The main objective of this paper is to extend the existing test procedures to general regression models. Specifically, we present test procedures for testing a conditional moment restriction model:

$$E\{g(z, \beta_0)|x\} = 0 \quad \text{for some } \beta_0 \in B \text{ with probability one,} \tag{1}$$

against a nonnested conditional moment restriction model:

$$E\{h(z, \alpha_0)|x\} = 0 \quad \text{for some } \alpha_0 \in A \text{ with probability one,} \tag{2}$$

where $g(.)$ and $h(.)$ are two (possibly nonlinear) functions of known form and α_0 and β_0 are the true values of vector-valued unknown parameters.[2] The vector-valued z and x denote observed data and x may or may not overlap with z.[3] The variables x can be interpreted as instrumental variables so that the variables z are permitted to correlate with the residuals.

We present three testing procedures. The first procedure is for testing a weaker version of (1) against (2). Let w denote a vector-valued function of x. (1) and (2) imply

$$E\{g(z, \beta_0)w\} = 0 \text{ for some } \beta_0 \in B, \tag{1'}$$

$$E\{h(z, \alpha_0)w\} = 0 \text{ for some } \alpha_0 \in A. \tag{2'}$$

The first procedure is for testing (1′) against (2′). This procedure is parametric when the dimension of w is fixed and independent of sample sizes. The second procedure is the same as the first one but allows the dimension of w to grow with sample size. This procedure is nonparametric and it is for testing (1) against (2). The third procedure also is nonparametric but it tests the validity of (1) against all alternative specifications.

The first two tests are closely related to those of Davidson and Mackinnon (1981, 1985) and Mackinnon, White, and Davidson (1993), and Wooldridge (1994) and can be viewed as conditional moment tests of Newey (1985) and Tauchen (1985). But, unlike those tests, the proposed tests: are applicable

[2]Extensions to the case where $g(.)$ and $h(.)$ are vector-valued functions are described in the discussion following Theorem 1.

[3]All vectors in this paper refer to column vectors.

to more general conditional moment models; permit heteroskedasticity of unknown form; and allow the number of moment restrictions to grow with sample size. By allowing the dimension of w to grow with sample size, we obtain a consistent test for (1) against (2). The existing tests, on the other hand, are not always consistent. Our third test is closely related to those of De Jong and Bierens (1994) and Hong and White (1996). In both studies, conditional expectations are estimated by series estimators and the number of terms in the series approximations are then allowed to grow with sample size to assure that the approximation errors are zero in the limit. Neither studies address the heteroskedasticity issue explicitly, though the test of De Jong and Bierens is heteroskedasticity robust. Our test, on the other hand, addresses the heteroskedasticity issue explicitly, thereby is more powerful.

As indicated earlier, the framework (1) - (2) encompasses many of the testing problems studied in the literature. To see this specifically, consider testing a linear specification $y = x_1'\beta_0 + u$ against a loglinear one $\ln(y) = x_2'\alpha_0 + \varepsilon$. With x as the union of x_1 and x_2 and with z as the union of y and x, (1) and (2) are satisfied with $g(z, \beta_0) = y - x_1'\beta_0$ and $h(z, \alpha_0) = \ln(y) - x_2'\alpha_0$. As a second example, consider testing the linear specification against the more general model of Box and Cox (1964). (1) and (2) are satisfied with $g(z, \beta_0) = y - x_1'\beta_0$ and $h(z, \alpha_0) = y^{\lambda_0} - x_2'\alpha_0$, where λ_0 is a coefficient to be chosen by researchers or estimated from data. As a third example, consider testing the model $y = x_1'\beta_0 + u$ against the nonparametric alternative $y = E\{y|x\} + \varepsilon$. Then (1) is satisfied with $g(z, \beta_0) = y - x_1'\beta_0$ if and only if $x_1'\beta_0$ is the correct specification of $E\{y|x\}$.

The paper is organized as follows: Section 2 introduces and discusses the first test procedure, Section 3 introduces and discusses the second procedure, Section 4 introduces and discusses the third procedure, and Section 5 concludes the paper. Omitted proofs are relegated to Appendix B.

2. PARAMETRIC PROCEDURE

To motivate the first two testing procedures, we apply the well-known artificial nesting approach (Atkinson (1970) and Quandt (1974)). Nest (1) and (2) in a larger model:

$$E\{(1 - \delta_0)g(z, \beta_0) + \delta_0 h(z, \alpha_0)|x\} = 0 \text{ with probability one,}$$

which writes

$$g(z, \beta_0) = \delta_0(h(z, \alpha_0) - g(z, \beta_0)) + \eta, \tag{3}$$

where the disturbance η satisfies $E\{\eta|x\} = 0$ under both (1) and (2) but the true value δ_0 satisfies $\delta_0 = 0$ under (1) and $\delta_0 = 1$ under (2) respectively. The models (1) and (2) and the models (1$'$) and (2$'$) can be tested by checking whether $\delta_0 = 0$ or $\delta_0 = 1$. We will estimate δ_0 from (3) and check its significance. However, as noted in the literature, (3) alone cannot identify all parameters.[4] To resolve this identification problem, we follow Davidson and MacKinnon (1981) by estimating β_0 from model (1) and estimating α_0 from model (2) using some sensible estimation methods. For example, $\hat{\beta}$ and $\hat{\alpha}$ could be the GMM estimators of models (1) and (2) respectively.[5] It then follows from Hansen (1982) that, under some regularity conditions, $\hat{\alpha}$ and $\hat{\beta}$ converge to some limits α^* and β^* respectively at rate $\sqrt{N}$, where the limits satisfy: $\beta^* = \beta_0$ under (1) and $\alpha^* = \alpha_0$ under (2). The estimators $\hat{\alpha}$ and $\hat{\beta}$ do not have to be the GMM estimators but they must satisfy:

Assumption 2.1. *$\sqrt{N}(\hat{\alpha} - \alpha^*) = O_p(1)$ and $\sqrt{N}(\hat{\beta} - \beta^*) = O_p(1)$.*

Replacing α_0 and β_0 with the estimators, we obtain:

$$g(z, \hat{\beta}) = \delta_0(h(z, \hat{\alpha}) - g(z, \hat{\beta})) + \hat{\eta}, \tag{4}$$

where $\hat{\eta}$ is the sum of η and the errors resulted from substituting $\hat{\alpha}$ and $\hat{\beta}$ for α_0 and β_0. Our test procedures are based on the model:

$$g(z, \hat{\beta}) = g_\beta(z, \hat{\beta})'\lambda_0 + \delta_0(h(z, \hat{\alpha}) - g(z, \hat{\beta})) + \hat{\eta}, \tag{5}$$

where $g_\beta(.)$ denotes the first derivatives of $g(.)$ with respect to β and the true value $\lambda_0 = 0$ holds under both (1) and (2). The first derivative term is used to purge the effects of $\hat{\alpha}$ and $\hat{\beta}$.

Since the right-hand side regressors in (5) may be correlated with the residuals, we estimate λ_0 and δ_0 by 2SLS using w as instruments. We then test $\delta_0 = 0$ using the t-ratio of the 2SLS estimator of δ_0. Let (z_i, x_i) for $i = 1, 2, \cdots, N$ denote sample realizations of (z, x) satisfying:

Assumption 2.2. *(z_i, x_i) for $i = 1, \cdots, N$ are i.i.d..*

Denote sample realizations of w by w_i. Let $\hat{\Gamma}$ and $\hat{\gamma}$ denote the OLS estimators from regressing $g_\beta(z_i, \hat{\beta})$ and $h(z_i, \hat{\alpha}) - g(z_i, \hat{\beta})$ on w_i respectively. Let $h_\alpha(z, \alpha)$ denote the first derivative with respect to α and let $g_{\beta\beta}(z, \beta)$ denote the second derivatives with respect to β. Let β_j denote the j-th element of β. Suppose that the following condition holds:

Assumption 2.3. *(i) $N^{-1}\sum_{i=1}^{N} w_i w_i'$ is nonsingular and $E\{ww'\}$ is positive definite;*

[4] See Gourieroux and Monfort (1994, p. 2611).

[5] The GMM estimators can be obtained with little programming effort in TSP. All you need to do is to provide the moment function $g(z, \beta)$ or $h(z, \alpha)$ and the instruments w. TSP then computes the estimates and the standard errors.

(ii) $E\{\sup_{\beta\in B}|g_\beta(z,\beta)w'|\}, E\{\sup_{\alpha\in A}|h_\alpha(z,\alpha)w'\}$, *and* $E\{\sup_{\beta\in B}|g_{\beta\beta j}(z,\beta)w'|\}$ *for all* j *are finite.*

Then $\hat{\Gamma}$ and $\hat{\gamma}$ converge to $\Gamma^* = E\{g_\beta(z,\beta^*)w'\}E\{ww'\}^{-1}$ and $\gamma^* = E\{ww'\}^{-1}E\{w(h(z,\alpha^*)-g(z,\beta^*))\}$ respectively in probability at rate $\sqrt{N}$ (see Lemma B.1 of Appendix B).

Let $\hat{\theta}$ denote the OLS estimator from regressing $w_i'\hat{\gamma}$ on $\hat{\Gamma}w_i$. $\hat{\theta}$ converges to $\theta^* = (\Gamma^*E\{ww'\}\Gamma^{*\prime})^{-1}\Gamma^*E\{ww'\}\gamma^*$ in probability at rate $\sqrt{N}$ (Lemma B.1) if the following condition holds:

Assumption 2.4. $N^{-1}\sum_{i=1}^{N}\hat{\Gamma}w_iw_i'\hat{\Gamma}'$ *is nonsingular and* $\Gamma^*E\{ww'\}\Gamma^{*\prime}$ *is positive definite.*

The 2SLS estimator $\hat{\delta}$ is given by

$$\hat{\delta} = \left\{\sum_{i=1}^{N}(w_1'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})^2\right\}^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})g(z_i,\hat{\beta}).$$

Clearly, $N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})^2$ must be nonzero for all N and converge to a positive constant $E\{(w'\gamma^*-(\Gamma^*w)'\theta^*)^2\}$ in probability. The constant $E\{(w'\gamma^*-(\Gamma^*w)'\theta^*)^2\}$ is positive if and only if the following condition holds:

Assumption 2.5. $\gamma^* \neq \Gamma^{*\prime}\theta^*$.

We now state other conditions. Define $u = g(z,\beta^*)$ and $\varepsilon = h(z,\alpha^*)$. Denote sample realizations of u and ε by u_i and ε_i respectively.

Assumption 2.6. $E\{u^2|x\}$ *and* $E\{u^4|x\}$ *are bounded and bounded away from zero uniformly over* x.

Assumption 2.7. *For given* z, $g(z,\beta)$ *and* $h(z,\alpha)$ *are continuous and twice continuously differentiable with respect to* α *and* β.

Assumption 2.2 rules out dependent data but it can be weakened to permit time series data (see the discussion following Theorem 1). Assumption 2.3(i) rules out multicollinearity. This condition is always imposed in the regression literature. Assumption 2.3(ii) is a stochastic dominance condition. This condition is often imposed in the nonlinear econometric literature to assure uniform convergence. Assumption 2.6 is needed for proving consistency of the White heteroskedasticity covariance matrix while Assumption 2.7 is needed for linearizing $g(z,\hat{\beta})$ and $h(z,\hat{\alpha})$. Assumptions 2.4

and 2.5 assure that the 2SLS estimator $\hat{\delta}$ is well-defined. Assumption 2.5 requires that (1′) and (2′) be intrinsically nonnested in the sense that they do not hold simultaneously.[6] Though difficult to verify, this condition can be tested using the estimator $\hat{\gamma} - \hat{\Gamma}'\hat{\theta}$ which is asymptotically normal with mean zero if Assumption 2.5 is not satisfied. For instance, a Chi-square type of test can be constructed. [7]

Under these conditions, we show that $\sqrt{N}\hat{\delta}$ is asymptotically normal with mean zero and variance $\sigma_\delta^2 = (E\{(w'\gamma - w'\Gamma'\theta)^2\})^{-2}(E\{(w'\gamma - w'\Gamma'\theta)^2 u\})$ under (1′). Following White (1980), a consistent estimator for σ_δ^2 is given by

$$\hat{\sigma}_\delta^2 = \left\{N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma} - (\hat{\Gamma}w_i)'\hat{\theta})^2\right\}^{-2}\left\{N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma} - (\hat{\Gamma}w_i)'\hat{\theta})^2\hat{u}_i^2\right\}, \quad (6)$$

where $\hat{u}_i = g(z_i, \hat{\beta})$ or $\hat{u}_i = g(z, \hat{\beta}) - g_\beta(z, \hat{\beta})'\hat{\lambda} - \hat{\delta}(h(z, \hat{\alpha}) - g(z, \hat{\beta}))$. Both estimated residuals are consistent for the true residual u_i, but the latter is used in computing the variance when the TSP 2SLS procedure is applied to (5) and the Robust option is requested. The t-ratio $\hat{t} = \sqrt{N}\hat{\delta}/\hat{\sigma}_\delta$ is then asymptotically standard normal under (1′). Under (2′), we show that $\hat{t}$ converges to $+\infty$. Thus, a one-sided test is consistent for testing (1′) against (2′). We summarize these results in a theorem.

THEOREM 2.1. *Under Assumptions 2.1 - 2.7 the t-ratio $\hat{t}$: (i) is asymptotically standard normal under* (1′)*; and (ii) converges to $+\infty$ in probability under* (2′).

Smith (1992) also proposed a test for (1′) against (2′). Let $\hat{g}_N$ and $\hat{h}_N$ denote the sample averages of $g(z_i, \hat{\beta})w_i$ and $h(z_i, \hat{\alpha})w_i$ respectively. Smith's test statistic (Smith (1992, pp. 975, equation (2.6)) is $\hat{g}_N'\hat{W}\hat{h}_N$, where $\hat{W}$ is an estimator of some matrix. Note that $\hat{g}_N \to 0$ in probability under (1′) and $\hat{h}_N \to 0$ in probability under (2′). We have $\hat{g}_N'\hat{W}\hat{h}_N \to 0$ in probability under both (1′) and (2′). This is in contrast with the proposed test which has divergent limits. Thus, our test is more likely to reject (1′) when (2′) is true than Smith's test.

The proposed test is in fact a test for $E\{uw\} = 0$. Thus, it can be viewed as a conditional moment test of Newey (1985) and Tauchen (1985). It differs from their tests in that in our test: (i) $\hat{\alpha}$ and $\hat{\beta}$ can be any sensible estimators; (ii) these estimators do not affect the asymptotic distribution of the proposed test statistic; and (iii) the proposed test statistic is just

[6]This follows because, if (1′) and (2′) hold simultaneously, γ^* (and hence θ^*) is zero.
[7]Derivation of this test is straightforward and shall not be done here to conserve space.

the t-ratio of the 2SLS estimator provided that White heteroskedasticity consistent covariance matrix is computed.

For dependent but stationary time series data, the 2SLS estimator $\sqrt{N}\hat{\delta}$ under (1′) is still asymptotically normal but its standard errors will be incorrectly estimated. Consequently, the proposed test statistic $\hat{t}$ is asymptotically normal but not standard normal. But, if the Newey- West autocorrelation and heteroskedasticity consistent covariance matrix is used to compute the standard error of the 2SLS estimator, Theorem 1 still holds.[8]

Theorem 2.1 can be easily extended to the case of multiple alternatives. To see this, suppose that the true model when (1′) is rejected is one of the following:

$$E\{h_j(z,\alpha_{j0})w\} = 0 \quad \text{for } j = 1, 2, \cdots, J.$$

Then the counterpart of (5) is

$$g(z,\hat{\beta}) = g_\beta(z,\hat{\beta})'\lambda_0 + \sum_{j=1}^{J} \delta_{j0}(h_j(z,\hat{\alpha}_j) - g(z,\hat{\beta})) + \hat{\eta}, \tag{7}$$

where $\hat{\alpha}_j$ are some $\sqrt{N}-$consistent estimators for α_{j0}. (1′) is true if and only if all $\delta_{j0} = 0$. Apply 2SLS to (7). The 2SLS estimators $\hat{\delta}_j \to 0$ for all j under (1′) and at least one $\hat{\delta}_j$ converges to some nonzero constant under (7). Let $\hat{\delta} = (\hat{\delta}_1, \cdots, \hat{\delta}_J)'$ and let $\hat{V}$ denote the covariance matrix of $\hat{\delta}$. Then $\hat{\delta}'\hat{V}^{-1}\hat{\delta}$ has a central Chi-Square distribution with degrees of freedom J under (1′) and a noncentral Chi-Square distribution if (1) does not hold.

Theorem 2.1 also can be extended to the case of a simultaneous equations system. Consider testing the null: $E\{g_m(z_i,\beta_{m0})w\} = 0$ for $m = 1, 2, \cdots, M$ against the alternative: $E\{h_m(z,\alpha_{m0})w\} = 0$ for $m = 1, 2, \cdots, M$. Let $\hat{\beta}_m$ and $\hat{\alpha}_m$ denote some $\sqrt{N}-$consistent estimators of β_{m0} and α_{m0} respectively. Let $g_{m\beta}(.)$ denote the derivatives with respect to β. Then, (5) becomes

$$g_{m\beta}(z,\hat{\beta}_m) = g_m(z,\hat{\beta}_m)'\lambda_0 + \delta_{m0}(h_m(z,\hat{\alpha}_m) - g_m(z,\hat{\beta}_m)) + \hat{\eta}_m. \tag{8}$$

Apply 3SLS to (8) and save the covariance $\hat{V}$ of the 3SLS estimators $\hat{\delta} = (\hat{\delta}_1, \cdots, \hat{\delta}_M)'$. Then $\hat{\delta}'\hat{V}^{-1}\hat{\delta}$ has a central Chi-Square distribution with degree of freedom M under the null and noncentral Chi-Square distribution under the alternative.

[8]The Newey-West covariance matrix requires extra programming effort.

3. NONPARAMETRIC PROCEDURE I

The proposed test, though consistent for $(1')$ against $(2')$, is not consistent for (1) against (2). This is because $(1')$ and $(2')$ do not imply (1) and (2). To assure consistency of the test for (1) against (2), we must allow the dimension of w to increase with sample size so that $(1')$ and $(2')$ are equivalent to (1) and (2) in the limit. In other words, we must assure that $\Gamma^* w$ and $w'\gamma^*$ approximate $E\{g_\beta(z,\beta^*)|x\}$ and $E\{h(z,\alpha^*)-g(z,\beta^*)|x\}$ respectively arbitrarily well. To accomplish this, we need to be specific about w. Let $\{a_s(x), s \geq 1\}$ denote a prespecified family of base functions.[9] This can be any family of functions satisfying the conditions below. There are no other restrictions on these base functions other than that they must be dense in a functional space. For some integer k, define $w = (a_1(x), \cdots, a_k(x))'$.[10] We now study the limiting behavior of the t- ratio of Section 2 when k tends to infinity as N tends to infinity.

As k increases with N, the number of parameters in γ^* and Γ^* also increases with N. This implies that $\hat{\gamma}$ and $\hat{\Gamma}$ cannot be $\sqrt{N}$ consistent. However, we will show that $\hat{\Gamma} w$ and $w'\hat{\gamma}$ converge to $E\{g_\beta(z,\beta^*)|x\}$ and $E\{h(z,\alpha^*)-g(z,\beta^*)|x\}$ respectively at rate $o_p(N^{-1/4})$. Moreover, we will show that $\hat{\theta}$ is still $\sqrt{N}$ consistent (Lemma B.4). This implies that, with $f(x) = E\{h(z,\alpha^*) - g(z,\beta^*) - g(z,\beta^*) - g_\beta(z,\beta^*)'\theta^*|x\}$,

$$N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma} - (\Gamma w_i)'\hat{\theta})^2 = E\{f(x)^2\} + o_p(1).$$

We also will show that, under (1), $\sqrt{N}\hat{\delta}$ is asymptotically normal with mean zero and variance: $\sigma_\delta^2 = (E\{f(x)^2\})^{-2}E\{f(x)^2u^2\}$, which can be estimated consistently by the estimator $\hat{\sigma}_\delta^2$ defined in (6). Thus, the t-ratio $\hat{t}$ is still asymptotically standard normal under (1). Under (2), $\hat{t}$ converges to $+\infty$.

To assure these results, we must impose all assumptions of Section 2 with Assumption 2.5 now replaced by:

Assumption 3.1. *$E\{h(z,\alpha^*)|x\}$ is not a linear combination of $E\{g_\beta(z,\beta^*)|x\}$ and $E\{g(z,\beta^*)|x\}$.*

In addition, we also impose conditions on the base functions and on the number of base functions used in the approximations. Let $R_{1k}(x) = E\{h(z,\alpha^*)-g(z,\beta^*)|x\} - w'\gamma^*$ and $R_{2k}(x) = E\{g_\beta(z,\beta^*)|x\} - \Gamma^* w$, where γ^* and Γ^* are now the coefficients in the series approximations. Define $\nu_{1i} = h(z_i,\alpha^*) - g(z_i,\beta^*) - E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\}$ and $\nu_{2i} = g_\beta(z_i,\beta^*) -$

[9] Examples of the family of functions include the polynominal, trigonometric and FFF families.

[10] Clearly, w depends on k. We omit the argument k for convenience.

$E\{g_\beta(z,\beta^*)|x_i\}$. Let $|.|$ denote the absolute value operator. For any matrix A_i, suppose that $|A_i|$, $\max_{j\leq N}\{|A_i|\}$, and A_i^s for any integer s are element by element operations.

ASSUMPTION 3.2. $\max_{i\leq N} E\{\nu_{1i}^2|x_i\} = O_p(1)$ *and* $\max_{i\leq N} E\{\nu_{2i}^2|x_i\} = O_p(1)$.

ASSUMPTION 3.3. *(i)* $k \to +\infty$ *as* $N \to +\infty$*; (ii)* $k/\sqrt{N} \to 0$.

ASSUMPTION 3.4. $\sqrt{N}\sup_x\{|R_{1k}(x)|\} \to 0$ *and* $\sqrt{N}\sup_x\{|R_{2k}(x)|\} \to 0$ *as* $N \to +\infty$.

Assumption 3.1 requires that $h(z,\alpha^*)$ must not be a local approximation to $h(z,\beta^*)$. It also requires that (1) and (2) must be intrinsically nonnested in the sense that the probability of the event: $E\{g(z,\beta^*)|x\} = 0$ and $E\{h(z,\alpha^*)|x\} = 0$ is less than one. Assumption 3.2 permits heteroskedasticity but requires that the variances are bounded. Assumption 3.3 imposes an upper bound on the number of instruments while Assumption 3.4 imposes a lower bound. These conditions require that k be large enough to kill the bias in the series approximations but not large enough to increase the variance of the 2SLS estimator. Under these conditions, we show:

THEOREM 3.1. *Under Assumptions 2.1 - 2.4, 2.6 - 2.7, and 3.1 - 3.4, the t-ratio* $\hat{t}$*:(i) is asymptotically standard normal under (1); and (ii) converges to* $+\infty$ *in probability under (2).*

Theorem 3.1 extends the results of Theorem 2.1 to the case where the number of instruments (k) increases with sample sizes. It shows that the one-sided t-test proposed earlier is consistent for (1) against (2). This result generalizes those of Davidson and Mackinnon (1981, 1985) and Wooldridge (1994) to the more general conditional moment setting.

Although the second test is theoretically better than the first one, in practice both tests use the ratio of the 2SLS estimator and hence are identical. Both tests improve upon the existing tests. To see this, consider the case: $g(z,\beta_0) = y - f(x,\beta_0)$ and $h(z,\alpha_0) = y - q(x,\alpha_0)$. Equation (5) now becomes:

$$y - f(x,\hat{\beta}) = f_\beta(x,\hat{\beta})'\lambda_0 + \delta_0(f(x,\hat{\beta}) - q(x,\hat{\alpha})) + \hat{\eta}.$$

Since $f_\beta(.)$ and $f(.) - q(.)$ are uncorrelated with η, they can be used as instruments. With $w = (f_\beta(x,\hat{\beta})', f(x,\hat{\beta}) - q(x,\hat{\alpha}))'$, 2SLS is identical to

OLS and $\hat{t}$ is the t-ratio of the OLS estimator of δ_0. In this case, our test is identical to the P-test suggested by Davidson and MacKinnon (1981).

As a second example, consider the case: $g(z,\beta_0) = y - f(x,\beta_0)$ and $h(z,\alpha_0) = \varphi(y) - q(x,\alpha_0)$, where $\varphi(.)$ is known. Then (5) becomes.

$$y - f(x,\hat{\beta}) = f_\beta(x,\hat{\beta})'\lambda_0 - \delta_0(\varphi(y) - y + f(x,\hat{\beta}) - q(x,\hat{\alpha})) + \hat{\eta}. \tag{9}$$

Since $f_\beta(x,\hat{\beta})$ is uncorrelated with η, it can be used as instrument. The second regressor in (9) is correlated with η, so 2SLS is called for. The PE-test proposed by MacKinnon, White and Davidson (1985) is based on

$$y - f(x,\hat{\beta}) = f_\beta(x,\hat{\beta})'\lambda_0 + \delta_0(\varphi(f(x,\hat{\beta})) - q(x,\hat{\alpha})) + \hat{\eta}.$$

Their test statistic is the t-ratio of the OLS estimator of δ_0. As Wooldridge (1994) pointed out, the PE-test is generally inconsistent because, by Jensen's inequality, $E\{\varphi(y)|x\} \neq \varphi(E\{y|x\})$ under (2). Wooldridge modified the PE-test by assuming that $E\{y|x\} = \tau_0\varphi^{-1}(E\{\varphi(y)|x\})$ under (2). His testing procedure is based on:

$$y - f(x,\hat{\beta}) = f_\beta(x,\hat{\beta})'\lambda_0 + \delta_0(f(x,\hat{\beta}) - \hat{\tau}\varphi^{-1}(q(x,\hat{\alpha}))) + \hat{\eta},^{11}$$

where $\hat{\tau}$ is the OLS estimator of regressing y on $\varphi^{-1}(q(x,\hat{\alpha}))$. The Wooldridge test is consistent under his condition. However, we argue that not all distributions satisfy his condition. In particular, his condition may not be satisfied if y is heteroskedastic. Besides, his test does not apply to the case of endogenous regressors.

As a third example, consider testing the fixed-effects panel regression model:

$$y_{it} = x_{1it}'\beta_0 + \mu_i + u_{it}, \quad \text{for } i = 1,2,\cdots,N; t = 1,2, \tag{10}$$

against the nonnested fixed effects panel regression model:

$$\varphi(y_{ti}) = x_{2it}'\alpha_0 + \nu_i + \varepsilon_{it}, \text{ for } i = 1,2,\cdots,N; t = 1,2, \tag{11}$$

where the variables μ_i and ν_i denote individual effects. Let x_{it} denote the union of x_{1it} and x_{2it}. Suppose that the disturbances satisfy $E\{u_{it}|x_{it}\} = 0$ and $E\{\varepsilon_{it}|x_{it}\} = 0$. Because of the presence of the individual effects, the proposed test procedures cannot be applied to (10)-(11) directly. Instead, we first eliminate the individual effects through differencing:

$$\Delta y_{i2} = \Delta x_{1i2}'\beta_0 + \Delta u_{i2}, \tag{12}$$
$$\Delta\varphi(y_{i2}) = \Delta x_{2i2}'\alpha_0 + \Delta\varepsilon_{i2}, \tag{13}$$

[11] For details on the Wooldrige test, see Wooldridge (1994).

where Δ is the first difference operator. We then apply our procedure to (12) - (13). The counterpart of (5) is:

$$\Delta(y_{i2} - x'_{1i2}\hat{\beta}) = \Delta x'_{1it}\lambda_0 + \delta_0 \Delta(\varphi(y_{i2}) - x'_{2i2}\hat{\alpha} - y_{i2} + x'_{1i2}\hat{\beta}) + \Delta\eta_{i2}. \quad (14)$$

Note that $E\{u_{it}|x_{it}\} = 0$ and $E\{\varepsilon_{it}|x_{it}\} = 0$ do not necessarily imply $E\{\Delta u_{it}|\Delta x_{it}\} = 0$ and $E\{\Delta\varepsilon_{it}|\Delta x_{it}\} = 0$. This is because x_{it} may include lagged dependent variables. Thus, Davidson and Mackinnon test and Wooldridge test do not apply. But if we assume $E\{\Delta u_{i2}|x_{i1}\} = 0$ and $E\{\Delta\varepsilon_{i2}|x_{i1}\} = 0$, we have $E\{\Delta\eta_{i2}|x_{i1}\} = 0$. So (14) can be estimated by 2SLS with x_{i1} as instruments. If x_{i1} cannot be used as instruments, external instruments are then called for.

4. NONPARAMETRIC PROCEDURE II

The proposed test in Section 3, though consistent for (1) against a particular alternative model (2), is not consistent against all alternative specifications. We now present a test that is consistent for (1) against all alternatives. The test is closely related to those of De Jong and Bierens (1994) and Hong and White (1996) and is based on $E\{(u/\sigma(x))E\{(u/\sigma(x))|x\}\}$, where $u = g(z, \beta^*)$ and $\sigma(x)^2 = E\{u^2|x\}$. Note that $E\{(u/\sigma(x))E\{(u/\sigma(x))|x\}\} = 0$ if (1) holds and $E\{(u/\sigma(x))E\{(u/\sigma(x))|x\}\} > 0$ if $P\{E\{u|x\} \neq 0\} > 0$. A consistent test can be constructed using the sample average:

$$\sum_{i=1}^{N}(u_i/\sigma(x_i))E\{u_i/\sigma(x_i))|x_i\}.$$

The residuals u_i are unknown and shall be estimated by $g(z_i, \hat{\beta})$ The variance also is unknown and shall be estimated by the fitted values of regressing $g(z_i, \hat{\beta})^2$ on w_{1i}, where $w_{1i} = (a_1(x_i), \cdots, a_{k_1}(x_i))'$. Let $\hat{\sigma}(x_i)^2$ denote the fitted values. $\hat{\sigma}(x_i)^2$ is not guaranteed to be positive. Hence, some sort of trimming is necessary. Define

$s_b(\mu) = 1\sqrt{\mu}$ for $\mu > b$; $s_b(\mu) = 3.5^*\mu^2/b^{2.5} - 2.5^*\mu^3/b^{3.5}$ for $0 \leq \mu \leq b$; and $s_b(\mu) = 0$ for $\mu < 0$.

Then $s_b(\mu)$ is an approximation of $1/\sqrt{\mu}$. Clearly, $s_b(\mu)$ controls small values of and trims negative values of μ. It can be verified that $s_b(\mu)$ has first continuous derivative with respect to μ. Moreover, $|s_b(\mu)| = O(1/\sqrt{b})$ and $|s'_b(\mu)/d\mu| = O(1/b^{1.5})$, where $s'_b(\mu)$ denotes the first derivative with respect to μ. We now estimate $u_i/\sigma(x_i)$ by $g(z_i, \hat{\beta})s_b(\hat{\sigma}(x_i)^2)$.

Next, we estimate $E\{u_i/\sigma(x_i)|x_i\}$ by the fitted values of regressing $g(z_i, \hat{\beta})s_b(\hat{\sigma}(x_i)^2)$ on w_i. Let $\hat{E}\{g(z, \hat{\beta})s_b(\hat{\sigma}(x)^2|x_i\}$ denote the fitted values. Then, our test statistic is given by:

$$\hat{r} = \sum_{i=1}^{N} g(z_i, \hat{\beta}) s_b(\hat{\sigma}(x_i)^2) \hat{E}\{g(z, \hat{\beta}) s_b(\hat{\sigma}(x)^2)|x_i\}.$$

We shall show that, under some conditions, $(\hat{r} - k)/\sqrt{2k}$ is asymptotically standard normal under (1). When (1) does not hold, we show $(\hat{r} - k)/\sqrt{2k} \to +\infty$ in probability.

Conditions that assure these asymptotic results are those imposed in Section 2 and Section 3 as well as those on the series approximation of $\sigma(x^2) = E\{g(z, \beta^*)^2|x\}$. Let $w_1'\gamma_s^*$ denote the series approximation of $E\{g(z, \beta^*)^2|x\}$ and let $R_{sk_1}(x) = E\{g(z, \beta^*)^2|x\} - w_1'\gamma_s^*$ denote the approximation error. Let $\nu_{3i} = g(z_i, \beta^*)^2 - E\{g(z, \beta^*)^2|x_i\}$. Let $w'\gamma_4^*$ denote the series approximation of $E\{g(z, \beta^*)/\sigma(x)|x\}$ and let $R_{4k}(x) = E\{g(z, \beta^*)/\sigma(x)|x\} - w'\gamma_4^*$ denote the approximation error. Let $\nu_{4i} = g(z_i, \beta^*)/\sigma(x_i) - E\{g(z, \beta^*)/\sigma(x)|x_i\}$. The following conditions shall be imposed:

ASSUMPTION 4.1. *(i)* $k \to +\infty$ *and* $N/\sqrt{k} \to +\infty$ *as* $N \to +\infty$*; (ii)* $\sum_{j=1}^{N} w_j w_j'$ *is nonsingular; (iii)* $\sup_i\{w_i'(\sum_{j=1}^{N} w_j w_j')^{-1} w_i\} \to 0$ *as* $N \to +\infty$*; and (vi)* $E\{u^4/\sigma(x)^4\}$ *is finite.*

ASSUMPTION 4.2. *(i)* $b \to 0, kb^6\infty + \infty$*, and* $k_1 \to +\infty$ *as* $N \to +\infty$*; (ii)* $k_1/\sqrt{kb^6} \to 0$*; (iii)* $\sup_x\{R_{3k_1}(x)^2\}^* N/\sqrt{kb^6} \to 0$ *as* $N \to +\infty$*; and (vi)* $\sup_x\{R_{4k}(x)^2\} \to 0$ *as* $N \to +\infty$.

ASSUMPTION 4.3. *(i)* $E\left\{\sup_{\beta\in B}\{g(z,\beta)^2|x\}\right\}$ *is bounded;*
(ii) $E\left\{\sup_{\beta\in B}\{g_\beta(z,\beta)g_\beta(z,\beta)'\}\right\}$ *is finite;*
(iii) $\max_i\left\{E\{\nu_{si}^2|x_i\}\right\} = O_p(1)$*; and*
(vi) $\max_i\left\{E\{\nu_{4i}^2|x_i\}\right\} = O_p(1)$.

Assumption 4.1 is the same as Assumptions A.5(a,b), A.7 and A.8 of Hong and White (1996). This assumption is sufficient for the central limit result:

$$\left(\left(\sum_{i=1}^{N} u_i w_i'/\sigma(x_i)\right)\left(\sum_{j=1}^{N} w_j w_j^{-1}\left(\sum_{i=1}^{N} u_i w_i/\sigma(x_i)\right) - k\right)\right)/\sqrt{2k}$$

is asymptotically standard normal (Hong and White (1996), Proposition 2.3). Assumption 4.2 imposes conditions on the series estimator of the variance function. Part (i) requires an infinite series be used to approximate the variance function. Part (ii) requires the dimension of w_1 grow at a much slower rate than the dimension of w while part (iii) requires the approximation error shrink to zero rapidly. Assumption 4.3 is a stochastic dominance condition. It is needed to show that the estimated variance function has no effects on the asymptotic distribution of $\hat{r}$. Under all these conditions, we have:

THEOREM 4.1. *Under Assumptions 2.1 - 2.4, 2.6 - 2.7, 3.1 - 3.4, and 4.1 - 4.2, the statistic $(\hat{r}-k)/\sqrt{2k}$: (i) is asymptotically standard normal under (1); and (ii) converges to $+\infty$ in probability if (1) does not hold.*

The test statistic $\hat{r}$, though cannot be obtained directly from regression output, is computable from LS regressions output via auxiliary LS regressions. For instance, the following steps can be followed to obtain $\hat{r}$:
Step 1. obtain $\hat{\beta}$ with any sensible estimation methods;
Step 2. regress $g(z_i,\hat{\beta})^2$ on w_{1i}, retrieve the fitted values $\hat{\sigma}(x_i)^2$;
Step 3. regress $g(z_i,\hat{\beta})s_b(\hat{\sigma}(x_i)^2)$ on w_i, retrieve the fitted values $\hat{E}\{g(z,\hat{\beta})s_b(\hat{\sigma}(x)^2)|x_i\}$. The sum of squared fitted values is $\hat{r}$.

5. CONCLUSION

This paper presents three testing procedures for testing the specifications of nonlinear regression models. The first two procedures are for testing nonnested hypotheses while the third one is a consistent test for a parametric model. All procedures are easy to compute. A common limitation of these procedures is that they all require the null specification be parametric and the alternative specification be either parametric or completely nonparametric. Recent developments in the literature have been in semiparametric and nonparametric models. Thus, an extension of the proposed tests to the semiparametric or nonparametric null against the semiparametric or nonparametric alternative would be desirable.

APPENDIX A

Some useful results

In this appendix, we derive some results that will be used to prove Theorems 2 and 3. These results are similar to those of Ai and McFadden (1997, Lemma A.1). Recall that $w=(a_1(x),\cdots,a_k(x))'$. For a known function $\varphi(x)$, let $\varphi(x)=w'\gamma+R_k(x)$, where $w'\gamma$ is the series approximation of $\varphi(x)$ and $R_k(x)$ is the approximation error. Let ν_i denote sample realizations of ν. Suppose that $E\{\nu_i|x_i\}=0$. Let $\phi(z)$ denote a known function. Let $\hat{\varphi}(x_i), \hat{E}\{\phi(z)|x_i\}$ and $\hat{E}\{\nu|x_i\}$ denote the fitted values of regressing $\phi(x_i), \phi(z_i)$, and ν_i respectively on w_i.

LEMMA A.1. *Suppose that* (z_i, x_i, ν_i) *for* $i=1,2,\cdots,N$ *are i.i.d.. Then:*

(i) $\sum_{i=1}^{N}(\hat{\varphi}(x_i)-\varphi(x_i))^2 = O(N * \sup_{x\in\chi}\{R_k(x)^2\})$;

(ii) if $\max_{i\le N} E\{\nu_i^2|x_i\}=O_p(1), \sum_{i=1}^{N}\hat{E}\{\nu|x_i\}^2 = O_p(k)$;

(iii) With $\sup_{x\in\chi}\{R_k(x)^2\}\to 0$ *as* $N\to+\infty$,

$$N^{-1/2}\sum_{i=1}^{N}\varphi(x_i)\hat{E}\{\nu|x_i\} = N^{-1/2}\sum_{i=1}^{N}\varphi(x_i)\nu_i + o_p(1);$$

(iv) suppose that $k/\sqrt{N}\to 0$ *and* $\sqrt{N} * \sup_{x\in\chi}\{(R_k(x))^2\}\to 0$ *as* $N\to+\infty$. *We obtain:*

$$N^{-1/2}\sum_{i=1}^{N}\varphi(z_i)\hat{E}\{\nu|x_i\} = N^{-1/2}\sum_{i=1}^{N}E\{\varphi(z)|x_i\}\nu_i + o_p(1);$$

Proof. Define $W=(w_1,w_2,\cdots,w_N)'$ and $P=W(W'W)^{-1}W'$. Let P_{ij} denote the (i,j)-th element of P. Then $\sum_{i=1}^{N}P_{ii}=k$. Part (i) follows from:

$$\sum_{i=1}^{N}(\hat{\varphi}(x_i)-\varphi(x_i))^2 = \sum_{i=1}^{N}R_k(x_i)^2 \le N * \sup_{x\in\chi}\{R_k(x)^2\}.$$

Part (ii) follows from:

$$\begin{aligned} E\left\{\sum_{i=1}^{N}\hat{E}\{\nu|x_i\}^2\right\} &= E\left\{\sum_{i=1}^{N}\sum_{j=1}^{N}\nu_i\nu_j P_{ij}\right\} = E\left\{\sum_{i=1}^{N}E\{\nu_i^2|x_i\}P_{ii}\right\} \\ &\le E\left\{\max_i E\{\nu_i^2|x_i\}\right\}k. \end{aligned}$$

To prove part (iii), write

$$
\begin{aligned}
N^{-1/2}\sum_{i=1}^{N}\varphi(x_i)\hat{E}\{\nu|x_i\} &= N^{-1/2}\sum_{i=1}^{N}(\hat{\varphi}(x_i)\nu_i \\
&= N^{-1/2}\sum_{i=1}^{N}(\hat{\varphi}(x_i)-\varphi(x_i))\nu_i + N^{-1/2}\sum_{i=1}^{N}\varphi(x_i)\nu_i.
\end{aligned}
$$

Note that

$$
E\left\{[N^{-1/2}\sum_{i=1}^{N}(\hat{\varphi}(x_i)-\varphi(x_i))\nu_i]^2\right\} = E\left\{N^{-1}\sum_{i=1}^{N}(\hat{\varphi}(x_i)-\varphi(x_i))^2\right\} * \max_i E\{\nu_i^2|x_i\},
$$

which converges to zero by part (i) if $\sup_{x\in\chi}\{R_k(x)^2\}\to 0$ as $N\to+\infty$.

To prove part (iv), write

$$
\begin{aligned}
N^{-1/2}\sum_{i=1}^{N}\phi(z_i)\hat{E}\{\nu|x_i\} &= N^{-1/2}\sum_{i=1}^{N}(\hat{\phi}(z_i)-E\{\phi(z)|x_i\})\hat{E}\{\nu|x_i\} \\
&+ N^{-1/2}\sum_{i=1}^{N}E\{\phi(z)|x_i\}\hat{E}\{\nu|x_i\}.
\end{aligned}
$$

Note that $\phi(z) = E\{\phi(z)|x\}+\nu$ and $E\{\nu|x\}=0$. With $\varphi(x)=E\{\varphi(z)|x\}$, applying part (i) and (ii) gives

$$
N^{-1/2}\sum_{i=1}^{N}(\hat{\phi}(z_i)-E\{\phi(z)|x_i\})^2 = O_p(\max\{k, N^*\sup_{x\in\chi}\{R_k(x)^2\}\}/\sqrt{N}),
$$

which converges to zero if $k/\sqrt{N}\to 0$ and $\sqrt{N^*}\sup_{x\in\chi}\{R_k(x)^2\}\to 0$.

Write

$$
\begin{aligned}
N^{-1/2}\sum_{i=1}^{N}E\{\phi(z)|x_i\}\hat{E}\{\nu|x_i\} &= N^{1/2}\sum_{i=1}^{N}(\hat{E}\{\phi(z)|x_i\}-E\{\phi(z)|x_i\})\nu_i \\
+\ N^{-1/2}\sum_{i=1}^{N}E\{\phi(z)|x_i\}\nu_i
\end{aligned}
$$

Applying part (i) to $\varphi(x) = E\{\varphi(z)|x\}$ proves $N^{-1/2}\sum_{i=1}^{N}(\hat{E}\{\phi(z)|x_i\} - E\{\phi(z)|x_i\})\nu_i = o_p(1)$. ∎

Parts (i) and (ii) can be used to determine a convergence rate of any series estimator, while parts (iii) and (iv) can be used to determine the asymptotic distribution of a series estimator. The difference between (iii) and (iv) is that in (iii) ? and x are conditional mean independent but in (iv) z and u may be correlated.

APPENDIX B

Omitted proofs

To prove Theorem 2.1, we first prove some useful lemmas.

LEMMA B.1. *Suppose that Assumptions 2.1 - 2.4 and 2.7 are satisfied. Then we can show:*
(i) $\sqrt{N}(\hat{\Gamma}-\Gamma^*) = O_p(1)$*;*
(ii) $\sqrt{N}(\hat{\gamma}-\gamma^*) = O_p(1)$*;*
(iii) $\sqrt{N}(\hat{\theta}-\theta^*) = O_p(1)$.

Proof. Recall that $\hat{\Gamma} = N^{-1}\sum_{i=1}^{N} g_\beta(z_i,\hat{\beta})w_1'\left(N^{-1}\sum_{i=1}^{N} w_i w_i'\right)^{-1}$. By Assumption 2.2 and the weak law of large numbers, we obtain $N^{-1}\sum_{i=1}^{N} w_i w_i' = E\{ww'\} + o_p(1)$. By Assumption 2.7 and a mean value expansion, we obtain:

$$N^{-1}\sum_{i=1}^{N} g_\beta(z_i,\hat{\beta})w_i' = N^{-1}\sum_{i=1}^{N} g_\beta(z_i,\beta^*)w_i' + N^{-1}\sum_{i=1}^{N} g_{\beta\beta}(z_i,\overline{\beta})(\hat{\beta}-\beta^*)w_i',$$

where $g_{\beta\beta}(.)$ denotes the second derivatives with respect to β and where $\overline{\beta}$ lies between $\hat{\beta}$ and β^*. By the Lindeberg-Levy central limit theorem, the first summation on the right-hand side converges to $E\{g_\beta(z,\beta^*)w'\}$ at rate $\sqrt{N}$. By Assumption 2.1, we have: $\sqrt{N}(\hat{\beta}-\beta^*) = O_p(1)$ and $\sqrt{N}(\overline{\beta}-\beta^*) = O_p(1)$. Let β_j denote the j-th component of β. By Assumption 2.3, we obtain:

$$N^{-1}\sum_{i=1}^{N} g_{\beta\beta_j}(z_i\beta)w_i' = E\{g_{\beta\beta_j}(z_i,\beta)w'\} + o_p(1) \text{ uniformly over } \beta \text{ for every} j.$$

This implies that $N^{-1}\sum_{i=1}^{N} g_{\beta\beta_j}(z_i,\overline{\beta})w_i' = E\{g_{\beta\beta_j}(z_i,\beta)w'\}+o_p(1)$. Hence, the second summation converges to zero at rate $\sqrt{N}$. This proves part (i). Parts (ii) and (iii) can be proved using similar arguments. ▮

LEMMA B.2. *Under Assumptions 2.1 - 2.5 and 2.7, we obtain:*

(i) $N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})^2 = E\{(w'\gamma^*-w'\Gamma^{*\prime}\theta^*)^2\}+o_p(1)$.

Moreover, if model $(1')$ *holds, we have:*

(ii) $N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})g(z_i,\hat{\beta}) = N^{-1/2}\sum_{i=1}^{N}(w_i'\gamma^*-(\Gamma^*w_i)'\theta^*)u_i+ o_p(1)$.

Proof. Part (i) follows immediately from $\hat{\gamma},\hat{\Gamma}$, and $\hat{\theta}$ being $\sqrt{N}$ consistent. To show (ii), Taylor expansion around $\hat{\beta}$ to the second order and noting that $u_i = g(z_i,\beta^*)$, we obtain:

$$
\begin{aligned}
&N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})u_i = N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})g(z_i,\hat{\beta})\\
&+\ N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})g_\beta(z_i,\hat{\beta})'(\hat{\beta}-\beta^*) \qquad \text{(B.1)}\\
&+\ (\hat{\beta}-\beta^*)'N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})g_{\beta\beta}(z_i,\hat{\xi})(\hat{\beta}-\beta^*)/2,
\end{aligned}
$$

where $\hat{\xi}$ is between $\hat{\beta}$ and β^*. Since $\hat{\Gamma}$ is the OLS estimator of $g_\beta(z_i,\hat{\beta})$ on w_i and $\hat{\theta}$ is the OLS estimator of $w_i'\hat{\gamma}$ on $\hat{\Gamma}w_i$, $g_\beta(z_i,\hat{\beta})-\hat{\Gamma}w_i$ is orthogonal to w_i while $w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta}$ is orthogonal to $\hat{\Gamma}w_i$. Thus the second summation on the right-hand side of (B.1) is zero. The third summation converges to zero in probability because $\hat{\beta}$ converges to β^* at rate $\sqrt{N}$. Hence, (B.1) implies

$$
N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})g(z_i,\hat{\beta}) = N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})u_i+o_p(1). \qquad \text{(B.2)}
$$

Further, since $\hat{\gamma}, \hat{\Gamma}$ and $\hat{\theta}$ are $\sqrt{N}$ consistent and $E\{wu\} = 0$ by model $(1')$, we obtain:

$$N^{-1/2} \sum_{i=1}^{N} (w_i'\hat{\gamma} - (\hat{\Gamma} w_i)'\hat{\theta}) u_i = N^{-1/2} \sum_{i=1}^{N} (w_i'\gamma^* - (\hat{\Gamma} w_i)'\theta^*) u_i + o_p(1). \quad \text{(B.3)}$$

Combining (B.1) - (B.3) proves part (ii). ∎

LEMMA B.3. *Under Assumptions 2.1 - 2.7 and model* $(1')$*, we obtain:*

$$N^{-1} \sum_{i=1}^{N} (w_i'\hat{\gamma} - (\hat{\Gamma} w_i)'\hat{\theta})^2 \hat{u}_i^2 = E\{(w'\gamma^* - (\hat{\Gamma} w)'\theta^*)^2 u_i^2\} + o_p(1).$$

Proof. Recall that

$$\hat{\delta} = \left\{ \sum_{i=1}^{N} (w_i'\hat{\gamma} - (\hat{\Gamma} w_i)'\hat{\theta})^2 \right\}^{-1} \sum_{i=1}^{N} (w_i'\hat{\gamma} - (\hat{\Gamma} w_i)'\hat{\theta}) g(z_i, \hat{\beta}).$$

By Lemma B.1 and equation (B.2), $\hat{\delta}$ converges to the true value at rate $\sqrt{N}$. Using similar arguments, we also can show that the 2SLS estimator for λ_0 converges to the true value at rate $\sqrt{N}$. This implies that the estimated residuals $\hat{u}_i = g(z, \hat{\beta}) - g_\beta(z, \hat{\beta})'\hat{\lambda} - \hat{\delta}(h(z, \hat{\alpha}) - g(z, \hat{\beta}))$ converge to u_i at rate $\sqrt{N}$. These convergence results imply that we can safely replace $\hat{\Gamma}, \hat{\gamma}, \hat{\theta}, \hat{\alpha}, \hat{\beta}, \hat{\delta}$, and $\hat{\lambda}$ by their respective limits. Hence,

$$N^{-1} \sum_{i=1}^{N} (w_i'\hat{\gamma} - (\hat{\Gamma} w_i)'\hat{\theta})^2 \hat{u}_i^2 = N^{-1} \sum_{i=1}^{N} (w_i'\gamma^* - (\Gamma^* w_i)'\theta^*)^2 u_i^2 + o_p(1).$$

Invoking the weak law of large numbers proves the lemma. ∎

Proof of Theorem 2.1

Proof of part (i) Note that $E\{uw\} = 0$ under $(1')$. Applying Lemma B.1 and Lemma B.2 proves that $\sqrt{N}\hat{\delta}$ is asymptotically normal with mean zero and variance σ_δ^2. By Lemma B.2 and Lemma B.3, the estimator $\hat{\sigma}_\delta^2$ defined in (6) is consistent for σ_δ^2. Hence $\hat{t}$ is asymptotically standard normal.

Proof of part (ii) Unless stated otherwise, the derivations below are obtained under $E\{w\varepsilon\} = 0$, where $\varepsilon = h(z, \alpha^*)$. Under $(2')$, Lemma B.1 and part (i) of Lemma B.2 still hold. Let $\hat{\gamma}_1$ denote the OLS estimator obtained by regressing $g(z_i, \hat{\beta})$ on w_i and let $\hat{\theta}_1$ denote the OLS estimator obtained by regressing $w_i'\hat{\gamma}$ on $\hat{\Gamma} w_i$. Then, because $h(z_i, \hat{\alpha})$ converges to ε_i

in probability and because $E\{w_i\varepsilon_i\} = 0$ by model (2′), we obtain $\hat{\gamma} - \hat{\gamma}_1 = o_p(1)$ and $\hat{\theta} - \hat{\theta}_1 = o_p(1)$. This implies

$$N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma} - (\hat{\Gamma}w_i)'\hat{\theta})g(z_i,\hat{\beta}) = N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma}_1 - (\hat{\Gamma}w_i)'\hat{\theta}_1)g(z_i,\hat{\beta}) + o_p(1). \tag{B.4}$$

Because $g(z_i,\hat{\beta}) - w_i'\hat{\gamma}_1$ is orthogonal to w_i and $w_i'\hat{\gamma}_1 - (\hat{\Gamma}w_i)'\hat{\theta}_1$ is orthogonal to $\hat{\Gamma}w_i$, we obtain:

$$N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma}_1 - (\hat{\Gamma}w_i)'\hat{\theta}_1)g(z_i,\hat{\beta}) = N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma}_1 - (\hat{\Gamma}w_i)'\hat{\theta}_1)^2 > 0. \tag{B.5}$$

By Assumption 2.5, $\hat{\gamma}_1 - \hat{\Gamma}'\hat{\theta}_1$ does not converge to zero in probability. Part (ii) now follows immediately from part (i) of Lemma B.2 and equations (B.4) - (B.5). ∎

To prove Theorem 3.1, we need to show several lemmas similar to Lemmas B.1 - B.3.

Lemma B.4. *Under Assumptions 2.1 - 2.4, 2.7, and 3.1 - 3.4, we can show that* $\sqrt{N}(\hat{\theta} - \theta^*) = O_p(1)$.

Proof. Let $\tilde{\gamma}$ and $\tilde{\Gamma}$ denote the OLS estimators obtained by regressing $h(z_i,\alpha^*) - g(z_i,\beta^*)$ and $g_\beta(z_i,\beta^*)$ on w_i respectively. Note that the conditions of Lemma A.1 are satisfied by Assumptions 3.2 - 3.4 for $\phi(z) = h(z,\alpha^*) - g(z,\beta^*)$ and $\phi(z) = g_\beta(z,\beta^*)$ respectively. Applying Lemma A.1 gives

$$N^{-1/2}\sum_{i=1}^{N}(w_i'\tilde{\gamma} - E\{h(z,\alpha^*) - g(z,\beta^*)|x\})^2 = o_p(1); \tag{B.6}$$

$$N^{-1/2}\sum_{i=1}^{N}(\tilde{\Gamma}w_i - E\{g_\beta(z,\beta^*)|x\})(\tilde{\Gamma}w_i - E\{g_\beta(z,\beta^*)|x\})' = o_p(1). \tag{B.7}$$

Note that the difference between $(\hat{\gamma},\hat{\Gamma})$ and $(\tilde{\gamma},\tilde{\Gamma})$ is that the former use $\hat{\alpha}$ and $\hat{\beta}$ while the latter use α^* and β^*. Since $\hat{\alpha}$ and $\hat{\beta}$ are $\sqrt{N}$ consistent, when $\tilde{\gamma}$ and $\tilde{\Gamma}$ in (B.6) and (B.7) are replaced by $\hat{\gamma}$ and $\hat{\Gamma}$ respectively, the $o_p(1)$ on the right-hand side must be replaced by $O_p(1)$:

$$N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma} - E\{h(z,\alpha^*) - g(z,\beta^*)|x_i\})^2 = O_p(1); \tag{B.8}$$

$$N^{-1/2}\sum_{i=1}^{N}(\hat{\Gamma}w_i - E\{g_\beta(z,\beta^*)|x_i\})(\hat{\Gamma}w_i - E\{g_\beta(z,\beta^*)|x_i\})' = O_p(1). \tag{B.9}$$

Write

$$N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma})\hat{\Gamma}w_i - N^{-1/2}\sum_{i=1}^{N}E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\}*E\{g_\beta(z,\beta^*)|x_i\}$$
$$= N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\})\hat{\Gamma}-E\{g_\beta(z,\beta^*)|x_i\})$$
$$+ N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\})*E\{g_\beta(z,\beta^*)|x_i\}$$
$$+ N^{-1/2}\sum_{i=1}^{N}(E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\})(\hat{\Gamma}w_i-E\{g_\beta(z,\beta^*)|x_i\}).$$

The first summation on the right-hand side is $O_p(1)$ by (B.8) and (B.9). The second summation writes:

$$N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\})^*E\{g_\beta(z,\beta^*)|x_i\}$$
$$= N^{-1/2}\sum_{i=1}^{N}(w_i'(\hat{\gamma}-\tilde{\gamma}))*E\{g_\beta(z,\beta^*)|x_i\}$$
$$+ N^{-1/2}\sum_{i=1}^{N}(w_i'\tilde{\gamma}-E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\})^*E\{g_\beta(z,\beta^*)|x_i\}.$$

With $\hat{\nu}_i = h(z_i,\hat{\alpha})-g(z_i,\hat{\beta})-h(z_i,\alpha^*)+g(z_i,\beta^*)$, write

$$N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-\tilde{\gamma})*E\{g_\beta(z,\beta^*)|x_i\} = N^{-1/2}\sum_{i=1}^{N}\hat{\nu}_i\hat{E}\{g_\beta(z,\beta^*)|x_i\}$$

which is $O_p(1)$ because $\hat{\nu}_i \to 0$ at rate $\sqrt{N}$ and $\hat{E}\{g_\beta(z,\beta^*)|x_i\} \to E\{g_\beta(z,\beta^*)|x_i\}$ in probability.

$$N^{-1/2}\sum_{i=1}^{N}(w_i'\tilde{\gamma}-E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\})^*E\{g_\beta(z,\beta^*)|x_i\} = O_p(1)$$

by part (iii) of Lemma A.1 and $\sqrt{N}*\sup_x|R_{1k}(x)| = O(1)$. This proves that the second summation is $O_p(1)$. Similarly, the third summation is $O_p(1)$. This completes the proof of the lemma. ▮

LEMMA B.5. *Under Assumptions 2.1 - 2.4, 2.6 - 2.7, and 3.1 - 3.4, we can show that:*

(i) $N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})^2 = E\{[E\{h(z,\alpha^*)-g(z,\beta^*)|x\}-E\{g_\beta(z,\beta^*)|x\}'\theta^*]^2\}+O_p(1)$.

Moreover, if model (1) holds, we have:

(ii) $N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})g(z_i,\hat{\beta}) = N^{-1/2}\sum_{i=1}^{N}[E\{h(z,\alpha^*)-g(z,\beta^*)|x\}'\theta^*]u_i+o_p(1)$.

Proof. Part (i) follows immediately from Lemma B.4 and equations (B.8) - (B.9). To show part (ii), we write:

$$
\begin{aligned}
& N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-(\hat{\Gamma}w_i)'\hat{\theta})g(z_i,\hat{\beta}) \\
= \;& N^{-1/2}\sum_{i=1}^{N}[E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\}-(E\{g_\beta(z,\beta^*)|x_i\})'\theta^*]u_i \\
+ \;& N^{-1/2}\sum_{i=1}^{N}[E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\}-(E\{g_\beta(z,\beta^*)|x_i\})'(\hat{\theta}-\theta^*)]u_i \\
+ \;& N^{-1/2}\sum_{i=1}^{N}[w_i'\hat{\gamma}-E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\}-(\hat{\Gamma}w_i-E\{g_\beta(z,\beta^*)|x_i\})'\hat{\theta}]u_i \\
+ \;& N^{-1/2}\sum_{i=1}^{N}[w_i'\hat{\gamma}-E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\} \\
- \;& (\hat{\Gamma}w_i-E\{g_\beta(z,\beta^*)|x_i\})'\hat{\theta}](g(z_i,\hat{\beta})-u_i) \\
+ \;& N^{-1/2}\sum_{i=1}^{N}[E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\}-(E\{g_\beta(z,\beta^*)|x_i\})'\hat{\theta}](g(z_i,\hat{\beta})-u_i) \\
= \;& A1+A2+A3+A4+A5.
\end{aligned}
$$

$A2 = o_p(1)$ because $\sqrt{N}(\hat{\theta}-\theta^*) = O_p(1)$ and $E\{u_i|x_i\} = 0$ by model (1). $A4 = o_p(1)$ because $\hat{\beta}$ is $\sqrt{N}$ consistent and (B.8) and (B.9) hold. A mean-value expansion gives:

$$A5 = N^{-1/2}\sum_{i=1}^{N}[E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\}-(E\{g_\beta(z,\beta^*)|x_i\}'\hat{\theta}]g_\beta(z_i,\xi)'(\hat{\beta}-\beta^*)$$

where ξ is between $\hat{\beta}$ and β^*. Since $(E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\}-E\{g_\beta(z,\beta^*)|x_i\}'\theta)$ is the LS projection residual,

$$N^{-1}\sum_{i=1}^{N}(E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\}-E\{g_\beta(z,\beta^*)|x_i\}'\theta^*)g_\beta(z_i,\beta^*)=o_p(1).$$

Thus, $A5=o_p(1)$ since $\hat{\theta}$ and ξ are $\sqrt{N}$ consistent.

It remains to show $A3=o_p(1)$. Let $\hat{E}\{u|x_i\}$ denote the OLS fitted values of regressing u_i on w_i. Applying part (ii) of Lemma A.1 to $\nu=u$ and noting that $k/\sqrt{N}\to 0$, we obtain $N^{-1/2}\sum_{i=1}^{N}\hat{E}\{u|x_i\}^2=o_p(1)$. Since $u_i-\hat{E}\{u|x_i\}$ is orthogonal to w_i, we obtain:

$$\begin{aligned}
&N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma})u_i=N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma})^*\hat{E}\{u|x_i\}\\
=\ &N^{-1/2}\sum_{i=1}^{N}[w_i'\hat{\gamma}-E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\}]^*\hat{E}\{u|x_i\}\\
+\ &N^{-1/2}\sum_{i=1}^{N}[E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\}]^*\hat{E}\{u|x_i\}.
\end{aligned}$$

The first summation on the right-hand side is $o_p(1)$ by equation (B.8) and $N^{-1/2}\sum_{i=1}^{N}\hat{E}\{u|x_i\}^2=o_p(1)$. Applying part (iii) of Lemma A.1 to $\varphi(x)=E\{h(z,\alpha^*)-g(z,\beta^*)|x\}$ and $\nu=u$, the second summation on the right-hand side writes:

$$\begin{aligned}
&N^{-1/2}\sum_{i=1}^{N}(E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\})\hat{u}_i\\
=\ &N^{-1/2}\sum_{i=1}^{N}(E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\})u_i+o_p(1).
\end{aligned}$$

This proves that

$$N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}-E\{h(z,\alpha^*)-g(z,\beta^*)|x_i\})u_i=o_p(1).$$

Similarly, we can show that $N^{-1/2}\sum_{i=1}^{N}(\hat{\Gamma}w_i - E\{g_\beta(z,\beta^*)|x_i\})u_i = o_p(1)$. Combining these results prove $A3 = o(1)$ and hence part (ii).

LEMMA B.6. *Under Assumptions 2.1 - 2.4, 2.6 - 2.7, and 3.1 - 3.4 and model (1), we can show that:*

$$
\begin{aligned}
N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma} - (\hat{\Gamma}w_i)'\hat{\theta})^2\hat{u}_i^2 &= E\{(E\{h(z,\alpha^*) - g(z,\beta^*)|x\} \\
&- E\{g_\beta(z,\beta^*)|x\}'\theta^*)^2 u_i^2\} + o_p(1).
\end{aligned}
$$

Proof. It follows from equation (B.8) and (B.9) and $\hat{u}_i$ converges to u_i at rate $\sqrt{N}$. ▌

Proof of Theorem 3.1 By Assumption 2.2 and the Lindeberg-Levy central limit theorem, applying Lemma B.4 shows that $\sqrt{N}\hat{\delta}$ is asymptotically normal with mean zero and variance

$$E\{(E\{h(z,\alpha^*) - g(z,\beta^*)|x\} - (E\{g_\beta(z,\beta^*)|x\})'\theta^*)^2 u^2\}.$$

By Lemma B.5, $\hat{\sigma}_\delta^2$ defined in equation (6) is consistent for σ_δ^2. Hence, part (i) holds.

To prove part (ii), let $\hat{\gamma}_1$ denote the OLS estimator obtained from regressing $g(z_i,\hat{\beta})$ on w_i, $\hat{\gamma}_2$ denote the OLS estimator obtained from regressing $h(z_i,\hat{\alpha})$ on w_i, and $\hat{\theta}_1$ denote the OLS estimator obtained from regressing $w_i'\hat{\gamma}_1$ on $\hat{\Gamma}w_i$. Then, $\hat{\gamma} = \hat{\gamma}_2 - \hat{\gamma}_1$. Applying Lemma A.1 and noting that $\hat{\alpha}$ and $\hat{\beta}$ are $\sqrt{N}$ consistent, we can show

$$
\begin{aligned}
N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}_1 - E\{g(z,\beta^*)|x_i\})^2 &= O_p(1); \\
N^{-1/2}\sum_{i=1}^{N}(w_i'\hat{\gamma}_2 - E\{h(z,\alpha^*)|x_i\})^2 &= O_p(1).
\end{aligned}
$$

Moreover, we can show $\hat{\theta}_1$ is $\sqrt{N}$ consistent. Because $E\{h(z,\alpha^*)|x\} = 0$ under (2), we have $\sqrt{N}(\hat{\theta}_1\hat{\theta}) = O_p(1)$ and

$$N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma} - (\hat{\Gamma}w_i)'\hat{\theta})g(z_i,\hat{\beta}) = N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma}_1 - (\hat{\Gamma}w_i)'\hat{\theta}_1)g(z_i,\hat{\beta}) + o_p(1).$$

Because $g(z_i,\hat{\beta})-w_i'\hat{\gamma}_1$ is orthogonal to w_i and $w_i'\hat{\gamma}_1-(\hat{\Gamma}w_i)'\hat{\theta}_1$ is orthogonal to $\hat{\Gamma}_1 w_i$, we have

$$N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma}_1-(\hat{\Gamma}w_i)'\hat{\theta}_1)g(z_i,\hat{\beta}) = N^{-1}\sum_{i=1}^{N}(w_i'\hat{\gamma}_1-(\hat{\Gamma}w_i)'\hat{\theta}_1)^2 > 0.$$

Part (ii) now follows immediately. ▌

Proof of Theorem 4.1

Proof of part (i) Let $\hat{E}\{g(z,\beta^*)s_b(\sigma(x)^2|x\}$ denote the fitted values from regressing $g(z_i,\beta^*)s_b(\sigma(x_i)^2)$ on w_i. Then, we can write

$$\begin{aligned}
\hat{r} &= \sum_{i=1}^{N}[\hat{E}\{g(z,\hat{\beta})s_b(\hat{\sigma}(x)^2)|x_i\} \\
&= \sum_{i=1}^{N}[\hat{E}\{g(z,\beta^*)s_b(\sigma(x)^2)|x_i\}]^2 \\
&+ 2\sum_{i=1}^{N}[\hat{E}\{g(z,\hat{\beta})s_b(\hat{\sigma}(x)^2)|x_i\}-\hat{E}\{g(z,\beta^*)s_b(\sigma(x)^2)|x_i\}]\hat{E}\{g(z,\beta^*)s_b(\sigma(x)^2)|x_i\} \\
&+ \sum_{i=1}^{N}[\hat{E}\{g(z,\hat{\beta})s_b(\hat{\sigma}(x)^2)|x_i\}-\hat{E}\{g(z,\beta^*)s_b(\sigma(x)^2)|x_i\}]^2 \\
&= B1+B2+B3.
\end{aligned}$$

First, we show that $B3/\sqrt{2k} = o_p(1)$. Recall that P_{ij} is the (i,j)-th element of $P = W(W'W)W'$. Write

$$B3 = \sum_{i=1}^{N}\sum_{j=1}^{N}[g(z_i,\hat{\beta})s_b(\hat{\sigma}(x_i)^2)-g(z_i,\beta^*)s_b(\sigma(x_i)^2)]^*P_{ij}^*[g(z_j,\hat{\beta})s_b(\sigma(x_j)^2)].$$

Taylor expansion around the true values β^* and $\sigma(x_i)$ yields:

$$\begin{aligned}
B3 &= \sum_{i=1}^{N}\sum_{j=1}^{N}[(\hat{\beta}-\beta)'g_\beta(z_i,\overline{\beta})s_b(\overline{\sigma}(x_i)^2)+g(z_i,\overline{\beta})s_b'(\overline{\sigma}(x_i)^2-\sigma(x_i)^2] \\
&* \ P_{ij} * [(\hat{\beta}-\beta)'g_\beta(z_j,\overline{\beta})s_b(\overline{\sigma}(x_j)^2)+g(z_j,\overline{\beta})s_b'(\overline{\sigma}(x_j)^2-\sigma(x_j)^2)],
\end{aligned}$$

where "-" denotes the mean value between the estimators and the true values. To show $B3/\sqrt{2k} = o_p(1)$, it suffices to show:

$$\begin{aligned} B31 &= \sum_{i=1}^{N}\sum_{j=1}^{N}(\hat\beta-\beta)' g_\beta(z_i,\overline{\beta}) s_b(\overline{\sigma}(x_i)^2) * P_{ij} \\ &\times (\hat\beta-\beta)' g_\beta(z_j,\overline{\beta}) s_b(\overline{\sigma}(x_j)^2)/\sqrt{2k} = o_p(1); \\ B32 &= \sum_{i=1}^{N}\sum_{j=1}^{N} g(z_i,\overline{\beta}) s_b'(\overline{\sigma}(x_i)^2)(\hat\sigma(x_i)^2 - \sigma(x_i)^2) \\ &\times P_{ij} * g(z_j,\overline{\beta}) s_b'(\overline{\sigma}(x_j)^2)(\hat\sigma(x_j)^2 - \sigma(x_j)^2)/\sqrt{2k} \\ &= o_p(1). \end{aligned}$$

Note that

$$\begin{aligned} B31 &\le \sum_{i=1}^{N}[(\hat\beta-\beta)' g_\beta(z_i,\overline{\beta}) s_b(\overline{\sigma}(x_i)^2)]^2/\sqrt{2k} \\ &\le \sqrt{N}(\hat\beta-\beta)' * N^{-1}\sum_{i=1}^{N} g_\beta(z_i,\overline{\beta})'\sqrt{N}(\hat\beta-\beta)/\sqrt{2kb^6} = o_p(1) \end{aligned}$$

because $\sqrt{N}(\hat\beta-\beta) = O_p(1)$ by Assumption 2.1, $N^{-1}\sum_{i=1}^{N} g_\beta(z_i,\overline{\beta}) g_\beta(z_i,\overline{\beta})' = O_p(1)$ by Assumption 4.3, and $kb^6 \to +\infty$ by Assumption 4.2.

Define $v(z) = \sup_{\beta\in B}\{g(z,\beta)^2\}$. From the construction of $s_b(\mu)$, we obtain:

$$B32 \le \sum_{i=1}^{N} \nu(z_i)(\hat\sigma(x_i)^2 - \sigma(x_i)^2)^2/\sqrt{2kb^6}.$$

Let P_{1ij} denote the (i,j)-th element of $P_1 = W_1(W_1'W_1)^{-1}W_1$, where $W_1 = (w_{11},\cdots,w_{1N})'$. Let $\nu_{3i} = [g(z,\beta^*)^2 - E\{g(z,\beta^*)^2|x\}]$. Taylor expansion around β^* gives:

$$\begin{aligned} \sum_{i=1}^{N} v(z_i)(\hat\sigma(x_i)^2 - \sigma(x_i)^2)^2 &= \sum_{i=1}^{N} v(z_i)(\sum_{j=1}^{N} P_{1ij}\nu_{3j} + (\hat\beta-\beta^*)'\sum_{j=1}^{N} g_\beta(z_j,\overline{\beta})P_{1ij} \\ &+ \sum_{j=1}^{N} P_{1ij}R_{3k_1}(x_j) - R_{3k_1}(x_i))^2 \end{aligned}$$

Note that

$$E\left\{\sum_{i=1}^{N}v(z_i)\left(\sum_{j=1}^{N}P_{1ij}\nu_{3j}\right)^2\right\}=E\left\{\sum_{i=1}^{N}\sum_{j=1}^{N}v(z_i)P_{1ij}^2\nu_{3j}^2\right\}$$
$$=\;E\left\{\sum_{i=1}^{N}\sum_{j=1}^{N}E\{v(z)|x_i\}P_{1ij}^2E\{\nu_{3j}^2|x_j\}\right\}=O(k_1)$$

because $E\{v(z)|x_i\}$ and $E\{\nu_{3j}^2|x_j\}$ are bounded by Assumption 4.3. Let $\hat{E}\{g_\beta(z,\overline{\beta})|x_i\}$ denote the fitted values from regressing $g_\beta(z_i,\overline{\beta})$ on w_{1i}. Then,

$$\sum_{i=1}^{N}v(z_i)((\hat{\beta}-\beta^*)'\sum_{j=1}^{N}g_\beta(z_j,\overline{\beta})P_{1ij})^2$$
$$=\;\sqrt{N}(\hat{\beta}-\beta^*)'[N^{-1}\sum_{i=1}^{N}v(z_i)*\hat{E}\{g_\beta(z,\overline{\beta})|x_i\}*\hat{E}\{g_\beta(z,\overline{\beta})|x_i)']\sqrt{N}(\hat{\beta}-\beta^*)$$
$$=\;O_p(1)$$

because $\sqrt{N}(\hat{\beta}-\beta^*)=O_p(1)$ by Assumption 2.1 and $N^{-1}\sum_{i=1}^{N}v(z_i)*\hat{E}\{g_\beta(z,\overline{\beta})|x_i\}=O_p(1)$ by applying Lemma A.1. Also, note that

$$\sum_{i=1}^{N}v(z_i)\left(\sum_{j=1}^{N}P_{1ij}R_{3k_1}(x_j)-R_{3k_1}(x_i)\right)^2\leq\sum_{i=1}^{N}v(z_i)R_{3k_1}(x_i)^2$$
$$\leq\;N*\sup_x\{R_{3k_1}(x)^2\}*O_p(1).$$

Combing these results proves $B32=o_p(1)$ because $N*\sup_x\{R_{3k_1}(x)^2\}/\sqrt{2kb^6}\to 0$ and $k_1/\sqrt{2kb^6}\to 0$ by Assumption 4.2. This completes the proof of $B3/\sqrt{2k}=o_p(1)$.

Applying Proposition 2.3 of Hong and White (1996), $(B1-k)/\sqrt{2k}$ is asymptotically standard normal. Part (i) now follows from $|B2|/\sqrt{2k}\leq\sqrt{2}\sqrt{B1/k}\sqrt{B3}$ and $B3=o_p(1)$.

Proof of part (ii) Note that model (1) is not used in the most part of the proof of part (i). Thus, we have:

$$\begin{aligned}\hat{r}/\sqrt{k} &= \sum_{i=1}^{N}[\hat{E}\{g(z,\beta^*)s_b(\sigma(x)^2)|x_i\}]^2/\sqrt{k} + o_p(1)\\ &= \sum_{i=1}^{N}[\hat{E}\{g(z,\beta^*)/\sigma(x)|x_i\}]^2/\sqrt{k} + o_p(1),\end{aligned}$$

because $\sigma(x)$ is bounded away from zero by Assumption 2.6 and $b \to 0$. Applying parts (i) and (ii) of Lemma A.1 to $\varphi(x) = E\{g(z,\beta^*)/\sigma(x)|x_i\}$ and $\nu_i = \nu_{4i} = g(z_i,\beta^*)/\sigma(x_i) - E\{g(z,\beta^*)/\sigma(x)|x_i\}$, and $R_{1k}(x) = R_{4k}(x) = E\{g(z,\beta^*)/\sigma(x)|x\} - w'\gamma_4$, we obtain:

$$\sum_{i=1}^{N}[\hat{E}\{g(z,\beta^*)\sigma(x)|x_i\}]^2 = O_p(\max\{k, N * \sup_x R_{4k}(x)^2\}).$$

Hence, $N^{-1}\sum_{i=1}^{N}[\hat{E}\{g(z,\beta^*)/\sigma(x)|x_i\}]^2 = 0$ since $k/N \to 0$ and $\sup_x R_{4x}(x)^2 \to 0$. Part (ii) now follows from $N/\sqrt{k} \to +\infty$ ▮

REFERENCES

Ai, C. and D. McFadden, 1997, Estimation of some partially specified nonlinear models. *Journal of Econometrics.*

Atkinson, A. C., 1970, A method for discriminating between models. *Journal of the Royal Statistical Society* **Series B, 32**, 211-243.

Box, G. E. P. and D. R. Cox, 1964, An analysis of transformations. *Journal of the Royal Statistical Society* **Series B, 26**, 211-252.

Dadkhah, K. M. and S. Valbuena, 1985, Non-nested test of new classical vs Keynesian models: Evidence from european economies. *Applied Economics* **17**, 1083-1098.

Davidson, R. and J. G. MacKinnon, 1981, Several tests of model specification in the presence of alternative hypotheses. *Econometrica* **49**, 781-793.

Davidson, R. and J. G. MacKinnon, 1985, Testing linear and loglinear regressions against Box-Cox alternatives. *Canadian Journal of Economics* **25**, 499-517.

Fan, Y. and Q. Li, 1996, Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* **64**, 865-890.

Godfrey, L. G., M. McAleer, and C. R. McKenzie, 1988, Variable addition and Lagrange multiplier tests for linear and loglinear regression models. *Review of Economics and Statistics* **70**, 492-503.

Godfrey, L. G. and M. R. Wickens, 1981, Testing linear and loglinear regressions for functional form. *Review of Economic Studies* **48**, 487-496.

Gourieroux, C. and A. Monfort, 1994, Testing non-nested hypotheses. *Chapter 44 in the Handbook of Econometrics IV*, edited by R. Engle and D. McFadden.

Hansen, L. P., 1982, Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029-1054.

Hansen, L. P. and K. J. Singleton, 1982, Generalized instrumental variable estimation of nonlinear rational expectations models. *Econometrica* **50**, 1269-1286.

Hong Y. and H. White, 1996, Consistent specification testing via nonparametric series regressions. *Econometrica* **63**, 1133-1160.

MacKinnon, J. G., H. White, and R. Davidson, 1983, Tests for model specification in the presence of nonnested alternatives: Some further results. *Journal of Econometrics* **21**, 53-70.

Newey, W. K., 1985, Maximum likelihood specification testing and conditional moment tests. *Econometrica* **53**, 1047-1070.

Quandt, R. E., 1974, A comparison of methods for testing non-nested hypotheses. *Review of Economics and Statistics* **56**, 92-99.

Smith, R., 1992, Non-nested tests for competing models estimated by generalized method of moments. *Econometrica* **4**, 973-980.

Tauchen, G., 1985, Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* **30**, 415-443.

Utts, J., 1982, The rainbow test for lack of fit in regression. *Communications in Statistics - Theory and Method* **11**, 2801-2815.

White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica* **48**, 817-838.

Wooldridge, J. M., 1991, On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. *Journal of Econometrics* **48**, 5-46.

—, 1992, Some alternatives to the box-cox regression model. *International Economic Review* **33**, 935-955.

—, 1994, A simple specification test for the predictive ability of transformation models. *Review of Economics and Statistics* **76**, 59-65.