THE PHILIPPINE REVIEW OF ECONOMICS
VOL. XLVI NO. 2 DECEMBER 2009 PP. 47-61

Testing for omitted variables in partially linear regression models

Lawrence Dacuycuy*
De La Salle University

This study conducts Monte Carlo simulation exercises to determine the finite sample properties of the Fan-Li test [1996] when the interest focuses on uncovering the impact of omitted variables on the semiparametric partially linear model (PLM). It essentially subjects the said test to various empirical situations such as the case of variable omissions in the linear or nonlinear components of the model or both. The study finds that the Fan-Li test achieves considerable power in rejecting the invalid null hypothesis. However, actual sizes still deviate from their respective nominal sizes.

JEL classification: C15, C14
Keywords: partially linear model, semiparametric models, nonparametric consistent tests

1. Introduction

Consistent test procedures have been developed to tackle issues concerning the validity of assumed functional forms and adapt to specific classical econometric problems like omitted and irrelevant variable problems. For parametric models, the applicability of the Zheng [1996] and Elisson and Elisson [2000] frameworks to test for functional-form adequacy and the relative ease with which they can be applied to omitted-variable testing have been noted. For instance, Dacuycuy [2005, 2006] analyzed the implications of omitted and irrelevant variable problems using the Zheng [1996] and Elisson and Elisson [2000] tests. There is also a growing interest in determining variable significance for nonparametric regression models wherein dimensionality problems render

*The author acknowledges his intellectual debt to Dr. Rolando Danao during his stint as a graduate student at the UP School of Economics. He likewise thanks the DLSU-University Research Coordination Office (URCO) for the project grant, Ms. Aizelle Andrade for research assistance, and an anonymous referee for invaluable comments and suggestions.

computational effort imprecise, thereby warranting the examination of an admissible set of regressors (see Lavergne and Vuong [1996]; Fan and Li [1996]). However, limited research has been done on the finite sample properties of the Fan-Li test when applied to testing for omitted variables in partially linear models.

In an important study, Fan and Li developed a test to ascertain the validity of the partially linear model as a representation of the conditional mean against the nonparametric alternative. Li [1999] studied the finite sample properties of the said test in a time-series setting. In other empirical settings, the test has been recently applied to verify the adequacy of semiparametric representations of the conditional mean as shown in the studies of Horowitz and Lee [2002] on specifications for baseball salaries and Bellemare, Melenberg, and van Soest [2002] on modeling satisfaction with income. However, no study has been undertaken investigating the properties of the said test in instances wherein the null may be flawed by failing to include a relevant regressor.[1]

In this empirical paper, we examine the finite sample properties of the Fan-Li test when applied to testing for omitted variables in partially linear models. To our knowledge, no direct application of the said test has been made. More specifically, we want to determine whether omissions in the linear or nonlinear components would result in significant differences in terms of the power and size of the test. The distinction here is necessary since each component, whether linear or nonlinear, has a material impact on the model's ability to define and empirically support various relationships. The simulation study also allows us to determine the impact of bandwidth choices for both the null and alternative hypotheses. Thus, a clear understanding of the extent of the test's performance would benefit the researcher in applications wherein specifications are beset with problems involving variable omission.

This paper is organized as follows: section 2 discusses the problem of omitted variables in the context of the semiparametric partially linear model. Section 3 discusses the simulation design for determining the empirical size and power of the Fan-Li test situated to handle variable omissions. Section 4 discusses the results, and the final section concludes.

[1]In other studies, the interest lies in determining the statistical adequacy of the partially linear model as the valid representation of the conditional mean relative to a parametric model belonging to a known family of models.

2. The econometric theory of omitted variable problems in partially linear models

2.1. Defining the omitted variables problem

In defining omitted variable problems in regression models, Ramsey [1969] used the concept of true model under ideal conditions, which differentiates it from the estimating model that suffers from omitted variable bias. Following Ramsey [1969], let the usual linear parametric model be written in matrix form as

$$y = Z\beta + W\delta + \varepsilon \tag{1}$$

where y is a $n\times 1$ vector dependent variable; Z is the data matrix with a column vector of 1s in the first column; δ is the coefficient vector of any d dimensional matrix W; β is the unknown coefficient vector and ε is a vector of stochastic disturbances or errors.

In relation to the model in (1), the estimating equation is given by

$$y = Z\beta + \upsilon. \tag{2}$$

According to Ramsey and Gilbert [1972], if equation (1) is true, then it follows that specification (2) suffers from omitted variables wherein the error term $\upsilon = W\delta + \varepsilon$. This econometric problem spawns inconsistency problems in estimation.

Similarly, Fan and Li [1996] have characterized omitted variables by referring to conditional expectations of the model. Again consider both equations (1) and (2). After taking their conditional expectations and noting that $E\left[\upsilon|X\right] = E\left[\upsilon|Z,W\right]$ are both zero, omitted variable bias exists when $E\left[y|X\right] \neq E\left[y|Z,W\right]$. As written in Fan and Li, there is no omitted variable problem when the opposite holds, that is, $E\left[y|Z\right] = E\left[y|Z,W\right]$. The latter definition may prove to be useful in characterizing omitted variable problems in partially linear models.

2.2. The partially linear model

Over the years, the partially linear model proposed by Robinson [1988] has gained importance in applied econometric research. The indispensability of the partially linear model partly stems from the incorporation of both linear and nonlinear components, with the nonlinear component being data determined via kernel and other methods as compared with ad hoc procedures involving

the increase of polynomial orders or other variable transformations. In this section, we examine the possible sources of omitted variable bias. We begin by defining the partially linear model.

The partially linear model is specified as

$$y = Z\beta + g(W) + \varepsilon \tag{3}$$

where exclusion restrictions are imposed on the data matrices Z and W. The model's intercept is not identified and is subsumed in the nonlinear component, $g(W)$. The econometric object of interest concerns the estimation of the unknown coefficient vector β given a nonparametric function. Robinson [1988] has shown that it is possible to arrive at a $\sqrt{N}$-consistent estimation of the coefficient vector by using kernel-based nonparametric methods.[2] To carry out the estimation, condition equation (3) on W. This results in

$$E[y|W] = E[Z|W]\beta + g(W) \tag{4}$$

where $E[\varepsilon|W] = 0$. Then following Yatchew [1998] and Robinson [1988], subtract equation (4) from (3). This results in

$$y - E[y|W] = (Z - E[Z|W])\beta + \upsilon. \tag{5}$$

Thus, to estimate β one needs to apply ordinary least squares. Based on Robinson, one of the criteria ensuring estimation feasibility is the nonsingularity of the matrix $(Z - E[Z|W])(Z - E[Z|W])'$. The estimator of the nonparametric component proceeds as if the coefficient vector is known. To compute for the said nonparametric component, generate the partial residual vector $\hat{\varepsilon} = (y - E[y|W]) - (Z - E[Z|W])\beta$ and apply a kernel-smoothing technique, the nature of which depends on the dimension of the nonlinear component.

2.3. The Fan-Li test

2.3.1. Preliminaries

In studies that consider the specification for partially linear model, the alternative is represented by the nonparametric model. In this case, the alternative calls for the consolidation of the effects of covariates rather than opting to have linear and nonlinear components, which should be orthogonal. Thus, the hypotheses, as shown in Fan and Li [1996], are specified as follows:

---

[2]There exist other estimation alternatives as well. One estimation technique known as differencing dispenses with the necessity of estimating nonparametric components of the partially linear model. For reference, please see Yatchew [2003].

$$H_0^{PLM}: E\left[y_i|x_i\right] = g(w_i) + \theta_0 z_i \text{ for some } \theta_0 \in \Theta \text{ and } g(w_i): \square^{q1} \to \square \quad (6)$$
$$H_1^{PLM}: E\left[y_i|x_i\right] \neq g(w_i) + \theta_0 z_i \text{ for all } \theta_0 \in \Theta \text{ and } g(w_i): \square^{q1} \to \square$$

where $g(W_i)$ is the nonlinear component and $\theta_0 z$ represents the linear component and $X = [Z\ W]$.

Consider the true model in equation (3). Define the residual vector as $e = (Y - \hat{Y}) - (Z - \hat{Z})\beta$ where $\hat{Y}$ and $\hat{Z}$ are the estimates for the nonparametric regression of Y and Z on W, respectively.[3] Also define the respective nonparametric density functions of W and Z as

$$f(W_j) = \frac{1}{nh^{q1}} \sum_{i=1} \sum_{j \neq 1} K\left(\frac{w_i - w_j}{h_q}\right) \quad (7)$$

$$f(Z_j) = \frac{1}{nh^{d}} \sum_{i=1} \sum_{j \neq 1} K\left(\frac{x_i - x_j}{h}\right) \quad (8)$$

where K is the symmetric and nonnegative kernel function, which is usually represented by the Gaussian kernel function, $q1$ and d are the respective dimensions of the nonparametric component in the model (W) and both W and Z, and h is the bandwidth parameter that governs the smoothness of the kernel estimates. The density function of X is defined similarly. As mentioned in Fan and Li [1996], the presence of the density functions ensures that the random denominator problem evident in the formula for the test statistic can be addressed. Given the residual vector and using the conditional moment, $E\left[E\left[\varepsilon|Z,W\right]f(X)\right]$ is written in U-statistic form.

$$1 = \frac{1}{n(n-1)h^{d}} \sum_{i=1} \sum_{j \neq 1} K\left(\frac{x_i - x_j}{h}\right) \hat{\varepsilon}_i f(w_i) \hat{\varepsilon}_j f(w_j). \quad (9)$$

As given in Fan and Li [1996], the estimator for the variance of the test statistic is given as

$$V = \frac{1}{n(n-1)h^{d}} \sum_{i=1} \sum_{j \neq 1} K\left(\frac{x_i - x_j}{h}\right) \hat{\varepsilon}_i^2 f(w_i)^2\, \hat{\varepsilon}_j^2 f(w_j)^2 \int K(u)\,du. \quad (10)$$

[3]This follows how Fan and Li defined the error in their study.

The normalized test statistic $\tau = nh^{d/2} I/\sqrt{2V}$ has a standard normal distribution.

The application of the test to omitted variables is justified as the Fan-Li test follows the structure of tests for significance considered in Yatchew [2003]. Yatchew noted the versatility of residual-based conditional moment test in testing for variable significance. In the null hypothesis, the dependent variable is assumed to vary with only a set of the regressors. In the alternative, the dependent variable is assumed to vary jointly with the entire set of regressors.

2.3.2. Hypotheses under omitted variables

The presence of the linear and nonlinear components in a single model paves the way for amply characterizing the impact of variable omission in partially linear models. We consider three scenarios that require closer econometric scrutiny. The first scenario concerns the correct specification of the nonlinear component, but some variables are left out of the linear component. In this case, the set of hypotheses is written as follows:

$$H_0^a : E\left[y_i | x_i\right] = E\left[y_i | z_i^*, w_i\right] = g\left(w_i\right) + \theta_0 z_i^* \qquad (11)$$

$$\text{for some } \theta_0 \in \Theta \text{ and } g\left(w_i\right) : \square^{q1} \to \square$$

$$H_1^a : E\left[y_i | x_i\right] \neq E\left[y_i | z_i^*, w_i\right] = g\left(w_i\right) + \theta_0' z_i^*$$

$$\text{for all } \theta_0 \in \Theta \text{ and } g\left(w_i\right) : \square^{q1} \to \square$$

where z_i^* implies that the linear component is defined on a regressor set characterized by a variable omission problem. Similarly, we denote w_i^* as the regressor set that suffers from omitted variables. The second scenario shows that the problem of variable omission lies in the nonlinear component, but no omission is observed for the linear part. This results in the following set of hypotheses:

$$H_0^b : E\left[y_i | x_i\right] = E\left[y_i | w_i^*, z_i\right] = g\left(w_i^*\right) + \theta_0 z_i \qquad (12)$$

$$\text{for some } \theta_0 \in \Theta \text{ and } g\left(w_i^*\right) : \square^{q1} \to \square$$

$$H_1^b : E\left[y_i | x_i\right] \neq E\left[y_i | w_i^*, z_i\right] = g\left(w_i\right) + \theta_0' z_i$$

$$\text{for all } \theta_0 \in \Theta \text{ and } g\left(w_i\right) : \square^{q1} \to \square$$

Finally, the third scenario shows that the problem of omission occurs in both components. Again, the hypotheses are

$$H_0^c : E\left[y_i \middle| x_i\right] = E\left[y_i \middle| w_i^*, z_i^*\right] = g\left(w_i^*\right) + \theta_0 z_i^* \tag{13}$$

$$\text{for some } \theta_0 \in \Theta \text{ and } g\left(w_i^*\right) : \square^{q1} \to \square$$

$$H_1^c : E\left[y_i \middle| x_i\right] \neq E\left[y_i \middle| w_i^*, z_i\right] = g\left(w_i\right) + \theta_0' z_i$$

$$\text{for all } \theta_0 \in \Theta \text{ and } g\left(w_i\right) : \square^{q1} \to \square$$

3. Investigating test performance: a proposed experiment

In this section, we investigate the statistical properties of the Fan-Li test by examining instances wherein omitted variables occur in the linear, nonlinear, or both linear and nonlinear components.

To analyze the power and size of the Fan-Li test for omitted variables, we employ several data-generating processes (DGPs) for testing the consistency of the partially linear model.[4]

The DGPs on which the alternative and null hypotheses are based are specified as follows:

$$DGP_0^a : Y_i = 1 + W_{1i} + Z_{1i} + \varepsilon_i$$

$$DGP_1^a : Y_i = 1 + W_{1i} + Z_{1i} + Z_{2i} + \varepsilon_i$$

$$DGP_2^a : Y_i = 1 + W_{1i} + Z_{1i} + Z_{2i} + Z_{1i} \times Z_{2i} + \varepsilon_i$$

$$DGP_3^a : Y_i = 1 + W_{1i} + Z_{1i} + Z_{2i} + W_{1i} \times Z_{2i} + \varepsilon_i$$

where $W = [W_1\ W_2]$ and $Z = [Z_1\ Z_2]$. ε is i.i.d $N(0, 1)$. To estimate, we use normal kernels. Similar to Li [1999], we fixed the bandwidth of the linear component but proceed to investigating how sensitive the bandwidth for the nonlinear component to changes in the constant c. Suppose that $E\left[y_i \middle| w_i, z_{1i}, z_{2i}\right] = E\left[y_i \middle| w_i, z_{1i}\right]$. This implies that z_{2i} has no material or significant explanatory power. Similar to Li [1999], DGP_0^a can be rewritten as $Y_i = 1 + g\left(W_{1i}\right) + Z_{1i} + \varepsilon_i$, wherein W_i is the nonlinear component. DGP_1^a states that variable z_{2i} has material linear effects on y_i. In this case, W_i is still the nonlinear component. Thus, the omitted variable problem lies in the linear component of the partially linear model. In relation to the above DGPs, alternative DGP processes may exploit the interaction between the linear components, the omitted linear component, and the included nonlinear components, etc.

---

[4]Some of the data-generating processes are similar in form to Li [1999] except that they are situated within the cross-sectional environment.

In the next scenario, the omitted variable plays a role in the nonlinear component of the model. In this case, we have the following data-generating processes:

$$DGP_0^b : Y_i = 1 + Z_{1i} + W_{1i} + W_{1i}^2 + \varepsilon_i$$
$$DGP_1^b : Y_i = 1 + Z_{1i} + W_{1i} + W_{1i}^2 + W_{2i} + \varepsilon_i$$
$$DGP_2^b : Y_i = 1 + Z_{1i} + W_{1i} + W_{1i}^2 + W_{2i} + W_{2i}^2 + \varepsilon_i$$
$$DGP_3^b : Y_i = 1 + Z_{1i} + W_{1i} + W_{1i}^2 + W_{2i} + W_{2i} \times Z_{1i} + \varepsilon_i$$

Based on the above, the alternative hypotheses assert that the omitted variable is important. For instance, DGP_1^b underscores the importance of W_2, and in DGP_2^b its interaction with the linear component is highlighted.

To determine the performance when both linear and nonlinear components suffer from omitted variable problems, we have the following data-generating processes:

$$DGP_0^c : Y_i = 1 + Z_{2i} + W_{2i} + \varepsilon_i$$
$$DGP_1^c : Y_i = 1 + Z_{1i} + Z_{2i} + W_{1i} + W_{2i} + \varepsilon_i$$
$$DGP_2^c : Y_i = 1 + Z_{1i} + Z_{2i} + W_{1i} + W_{2i} + W_{1i} \times W_{2i} + \varepsilon_i.$$

To carry out the procedure, we propose to combine GAUSS programs written by Horowitz and Lee [2002] and Miles and Mora [2003]. Horowitz and Lee's program computes for the estimates of the partially linear model and evaluates the test statistic. Miles and Mora's program provides the backbone of the simulation design. Consistent with the test's requirements, we employ the Gaussian kernel function. Following Horowitz and Lee, the bandwidth for the null model is specified as $h_0 = c_1 n^{-1/5}$ while for the alternative nonparametric model it is specified as $h_a = c_2 n^{-1/(d+4)}$. Following the simulation strategy of Li [1999], we simply allow c_2 to vary and peg the value of c_1 at 1. The problem with this type of test lies in the selection of the bandwidth parameter because of the sensitivity of the estimates to the specification of the parameter value. The simulation design also employs the leave-one-out strategy in estimating the kernel functions. The number of replications is 2,000, and various sample sizes (50, 100, 200, and 500) are considered. The Gauss program may be requested from the author.

4. Simulation results

Results are presented in Tables 1- 4. As shown in Table 1, the test is empirically undersized. This means that the actual size is less than the nominal size. Similar to other empirical exercises, the size of the test appears to decline as the bandwidth increases via c_2. However, surprisingly, when only one component is omitted at a time and as the sample size increases, the empirical size approximates the theoretical size. This is evident even when the omitted variable belongs to the nonlinear part. But the deviation appears greater when both omissions occur in the linear and nonlinear components of the model.

As expected, power increases with sample size and bandwidth. However, an interesting case occurs when omission occurs in both model components. Relative to misspecified models, the power is unexpectedly low at smaller bandwidths. This highlights the sensitivity of the test to bandwidth choices, especially in cases where omissions occur in both components. Results also show that the power converges to 1 at a much slower rate relative to the other types of omission.

Results indicate further that the test's power increases if the omitted variable is highly nonlinear as in the case of DGP_2^b. The performance of the test when omissions occur in both components, however, pales in comparison to omissions in either linear or nonlinear component.

5. Concluding remarks

The paper delves into the empirical examination of the finite sample properties of the Fan-Li test when applied to the partially linear model. Simulation results confirm the power of the Fan-Li test when the semiparametric null suffers from omitted variable problem. The test performs rather reasonably well, given that more than two continuous variables were used in validating the alternative hypothesis.

The test performs reasonably well when only one component suffers from variable omission. However, the test performs less satisfactorily in small samples and low bandwidths when both components suffer from variable omissions.

Table 1. Empirical size of test

	50			100			200			500		
c_2	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
DGP_0^a												
0.5	0.002	0.017	0.049	0.006	0.029	0.054	0.007	0.029	0.059	0.009	0.029	0.066
1	0.001	0.004	0.012	0.001	0.006	0.016	0.003	0.011	0.025	0.006	0.021	0.038
2	0	0.001	0.001	0	0	0	0	0.001	0.001	0	0.002	0.005
3	0	0	0	0	0	0	0	0	0	0	0	0
DGP_0^b												
0.5	0.005	0.023	0.047	0.009	0.023	0.046	0.01	0.027	0.057	0.013	0.037	0.07
1	0.001	0.004	0.012	0.003	0.008	0.019	0.007	0.017	0.028	0.007	0.026	0.046
2	0	0	0	0	0.001	0.002	0	0.001	0.003	0.001	0.004	0.01
3	0	0	0	0	0	0	0	0	0	0	0	0.001
DGP_0^c												
0.5	0.002	0.017	0.058	0.004	0.027	0.074	0.007	0.042	0.075	0.01	0.036	0.076
1	0.009	0.026	0.049	0.004	0.021	0.044	0.009	0.032	0.061	0.012	0.032	0.059
2	0.001	0.005	0.009	0	0.005	0.012	0.002	0.008	0.014	0.004	0.012	0.023
3	0	0	0	0	0	0	0.001	0.002	0.002	0.001	0.003	0.005

Table 2. Proportion of rejections of model alternatives: omitted in linear component

	50			100			200			500		
c	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$DGP_1^a : Y_i = 1 + W_{1i} + Z_{1i} + Z_{2i} + \varepsilon_i$												
0.5	0.107	0.322	0.473	0.492	0.709	0.817	0.941	0.99	0.996	1	1	1
1	0.604	0.781	0.847	0.973	0.99	0.997	1	1	1	1	1	1
2	0.823	0.916	0.952	1	1	1	1	1	1	1	1	1
3	0.68	0.857	0.914	1	1	1	1	1	1	1	1	1
$DGP_2^a : Y_i = 1 + W_{1i} + Z_{1i} + Z_{2i} + Z_{1i} \times Z_{2i} + \varepsilon_i$												
0.5	0.119	0.387	0.558	0.571	0.824	0.906	0.984	0.995	1	1	1	1
1	0.705	0.869	0.918	0.991	0.998	0.999	1	1	1	1	1	1
2	0.821	0.919	0.949	1	1	1	1	1	1	1	1	1
3	0.497	0.729	0.823	0.997	1	1	1	1	1	1	1	1
$DGP_3^a : Y_i = 1 + W_{1i} + Z_{1i} + Z_{2i} + W_{1i} \times Z_{2i} + \varepsilon_i$												
0.5	0.119	0.377	0.521	0.557	0.79	0.876	0.971	0.995	0.998	1	1	1
1	0.65	0.809	0.873	0.982	0.996	0.998	1	1	1	1	1	1
2	0.782	0.886	0.935	1	1	1	1	1	1	1	1	1
3	0.547	0.748	0.838	0.998	1	1	1	1	1	1	1	1

Table 3. Proportion of rejections of model alternatives: omitted in nonlinear component

	50			100			200			500		
c_2	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$DGP_1^b: Y_i = 1 + Z_{1i} + W_{1i} + W_{1i}^2 + W_{2i} + \varepsilon_i$												
0.5	0.113	0.345	0.488	0.499	0.732	0.835	0.949	0.985	0.994	1	1	1
1	0.631	0.796	0.855	0.98	0.993	0.996	1	1	1	1	1	1
2	0.819	0.909	0.945	1	1	1	1	1	1	1	1	1
3	0.657	0.846	0.908	1	1	1	1	1	1	1	1	1
$DGP_2^b: Y_i = 1 + Z_{1i} + W_{1i} + W_{1i}^2 + W_{2i} + W_{2i}^2 + \varepsilon_i$												
0.5	0.32	0.677	0.82	0.892	0.981	0.994	1	1	1	1	1	1
1	0.92	0.98	0.99	1	1	1	1	1	1	1	1	1
2	0.92	0.975	0.991	1	1	1	1	1	1	1	1	1
3	0.487	0.719	0.829	1	1	1	1	1	1	1	1	1
$DGP_3^b: Y_i = 1 + Z_{1i} + W_{1i} + W_{1i}^2 + W_{2i} + W_{2i} \times Z_{1i} + \varepsilon_i$												
0.5	0.114	0.399	0.577	0.596	0.84	0.914	0.983	0.998	1	1	1	1
1	0.714	0.866	0.92	0.995	0.998	0.999	1	1	1	1	1	1
2	0.799	0.913	0.951	1	1	1	1	1	1	1	1	1
3	0.46	0.698	0.81	1	1	1	1	1	1	1	1	1

Table 4. Proportion of rejections of model alternatives: omitted in both linear and nonlinear component

	50			100			200			500		
c_2	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$DGP_1^c : Y_i = 1 + Z_{1i} + Z_{2i} + W_{1i} + W_{2i} + \varepsilon_i$												
0.5	0.029	0.209	0.398	0.297	0.656	0.801	0.881	0.974	0.993	1	1	1
1	0.792	0.924	0.958	0.995	0.998	1	1	1	1	1	1	1
2	0.972	0.991	0.996	1	1	1	1	1	1	1	1	1
3	0.93	0.974	0.988	1	1	1	1	1	1	1	1	1
$DGP_2^c : Y_i = 1 + Z_{1i} + Z_{2i} + W_{1i} + W_{2i} + W_{1i} \times W_{2i} + \varepsilon_i$.												
0.5	0.028	0.214	0.419	0.319	0.681	0.823	0.92	0.984	0.994	1	1	1
1	0.808	0.925	0.964	0.997	1	1	1	1	1	1	1	1
2	0.963	0.987	0.994	1	1	1	1	1	1	1	1	1
3	0.871	0.955	0.976	1	1	1	1	1	1	1	1	1

References

Bellemare, C., B. Melenberg, and A. van Soest [2002] "Semiparametric models for satisfaction with income", CEMMAP Working Paper No. 12/02.

Dacuycuy, L. [2005] "A note on the comparative performance of the Zheng and Elisson and Elisson tests for omitted variables in regression models", *Economics Bulletin* **3**: 1-6.

Dacuycuy, L. [2006] "On the finite sampling properties of the Zheng test for omitted and irrelevant variable problems", *Applied Economics Letters* **13**: 681-684.

Ellison, G. and S. F. Ellison [2000] "A simple framework for nonparametric specification testing", *Journal of Econometrics* **96**: 1-23.

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

Greene, W. [2003] *Econometric analysis*. New Jersey: Prentice-Hall.

Horowitz, J. and S. Lee [2002] "Semiparametric methods in applied econometrics: do the models fit the data", *Statistical Modeling* **2**: 3-22.

Lavergne, P. and Q. Vuong [1996] "Nonparametric selection of regressors: the nonnested case", *Econometrica* **64**: 207-219.

Li, Q. [1999] "Consistent model specification tests for time series econometric models", *Journal of Econometrics* **92**: 101-147.

Li, Q. and S. Wang [1998] "A simple consistent bootstrap test for a parametric regression function", *Journal of Econometrics* **87**: 145-165.

Miles, D. and J. Mora [2003] "On the performance of nonparametric specification tests in regression models", *Computational Statistics and Data Analysis* **42**: 477-490.

Pagan, A. and A. Ullah [1999] *Nonparametric econometrics*. Cambridge: Cambridge University Press.

Ramsey, J.B. [1969) "Tests for specification errors in classical linear least squares regression analysis", *Journal of Royal Statistics Society Series B.* **31**(2): 350-371.

Ramsey, J. and R. Gilbert [1972] "A Monte Carlo study of some small sample properties of tests for specification error", *Journal of the American Statistical Association* **67**: 180-186.

Robinson, P. [1988] "Root N consistent semiparametric regression", *Econometrica* **56**: 931-954.

Yatchew, A. [1998] "Nonparametric regression techniques in economics", *Journal of Economic Literature* **36**(2): 669-721.

Yatchew, A. [2003] Semiparametric regression for the applied econometrician. Cambridge: Cambridge University Press.

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