International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in

Non-Nested Encompassing Tests for Functional Form Selection in Linear Regression: Box-Cox Transformation, Voung's Likelihood Ratio, and a Nonparametric Encompassing Criterion

G. Mokesh Rayalu[1] , K. Murali[2], Bandi Ramanjineyulu[3*]
Assistant Professor, Statistics & OR Division, School of Advanced Sciences, VIT University, Vellore
Assistant Professor, Sri Padmavathi College of Computer Science and Technology, Tiruchanoor, Tirupati
Senior Process Associate ,TCS , Bangalore , India
Corresponding Author : ramanji.bandi@gmail.com

Abstract

The problem of selecting among competing functional form specifications for linear regression models—including linear-in-levels, log-linear, Box-Cox transformed, polynomial, and semiparametric alternatives—constitutes a fundamental challenge in applied econometric research that cannot be adequately addressed by information criteria or classical F-tests alone, because functional form alternatives are typically non-nested: neither specification is a special case of the other. The present study develops and evaluates a unified framework for non-nested functional form testing that integrates Cox-type and J-type non-nested hypothesis tests, the Davidson-MacKinnon PE test, Vuong's (1989) likelihood ratio test for non-nested models, Box-Cox transformation tests, and a newly proposed Nonparametric Encompassing Criterion (NP-Enc) based on kernel regression residuals. The NP-Enc statistic tests whether a parametric functional form encompasses a nonparametric fit—in the sense that the parametric residuals contain no systematic structure that a kernel smoother can detect—providing a direct test of functional form adequacy against the nonparametric alternative. The unified framework is implemented as a sequential selection protocol for applied researchers facing functional form uncertainty. Monte Carlo simulation experiments across sample sizes of 30, 60, and 120 confirm that the proposed NP-Enc achieves the lowest size distortion and highest power against functional form misspecification among all procedures evaluated. An empirical application to a health expenditure demand function—with five competing functional form specifications—illustrates the practical advantage of the unified protocol, with the Box-Cox model ($\hat{\lambda} = 0.431$) emerging as the empirically preferred specification, correctly identified by AIC, the PE test, and the NP-Enc, while the standard linear model is rejected by RESET, Box-Cox LR test, and all non-nested tests. The findings contribute to the methodological literature on specification testing and provide applied researchers with a structured, evidence-based protocol for functional form selection.

Keywords: non-nested hypothesis tests, functional form, Box-Cox transformation, encompassing test, J-test, PE test, Vuong likelihood ratio, nonparametric specification test, RESET, functional form misspecification

1. Introduction

The specification of the correct functional form for a regression model—the mathematical relationship assumed to hold between the dependent variable and its explanatory variables—is among the most consequential and least routinely addressed decisions in applied econometric practice. While variable selection has attracted an enormous literature and is supported by a rich toolkit of information criteria, stepwise procedures, and cross-validation methods, the equally important question of whether to model the relationship as linear, log-linear, semi-logarithmic, polynomial, or through some nonlinear transformation has received comparatively sparse methodological attention in the applied statistics literature, despite the well-documented sensitivity of coefficient estimates, elasticity calculations, and policy inferences to functional form assumptions.

The fundamental difficulty in functional form selection is that the competing specifications are typically non-nested: the log-linear model $Y = \exp(X\beta)\varepsilon$ cannot be obtained from the linear model $Y = X\beta + \varepsilon$ by restricting the parameter space, nor vice versa. This means that standard F-tests and likelihood ratio tests—which are valid for testing nested hypotheses under the Neyman-Pearson framework—are inapplicable to the functional form comparison problem. Cox (1961, 1962) provided the foundational theoretical treatment of testing between non-nested hypotheses, demonstrating that the classical likelihood ratio statistic requires a bias correction to have known asymptotic distribution when the null and alternative models are non-nested, and deriving an asymptotically normal test statistic based on this correction.

Subsequent developments by Davidson and MacKinnon (1981), Fisher and McAleer (1981), and Vuong (1989) substantially extended the operational toolkit for non-nested hypothesis testing. However, the specific problem of testing between parametric functional form specifications—particularly when these specifications differ in their transformation of the dependent variable as well as the regressors—

raises additional complications because the likelihood functions of different dependent variable transformations are not directly comparable without Jacobian correction. The Box-Cox (1964) transformation framework addresses this by nesting many common functional forms within a single parametric family indexed by the transformation parameter λ, enabling likelihood-based tests for particular functional forms as restrictions on λ. However, the Box-Cox approach imposes the constraint that the correct functional form is within the Box-Cox family, which excludes semiparametric and nonparametric specifications.

The present study proposes a unified framework for functional form selection that combines parametric non-nested tests (J-test, PE test, Cox test), the Vuong likelihood ratio test for non-nested model comparison, Box-Cox transformation tests, and a newly proposed Nonparametric Encompassing Criterion (NP-Enc) into a sequential protocol that provides evidence-weighted guidance on functional form selection across the full range of practically relevant alternatives.

2. Review of Literature

2.1 Non-Nested Hypothesis Testing: Foundations

Cox (1961, 1962) established the theoretical foundations of non-nested hypothesis testing within the classical likelihood framework, demonstrating that when two hypotheses H_1 and H_2 are non-nested—neither being a special case of the other—the classical likelihood ratio test statistic does not have a known limiting distribution under H_1. Cox proposed a modified test statistic based on the expected value of the log-likelihood ratio under H_1, which is asymptotically normally distributed under H_1 and provides a valid test of H_1 against H_2. Cox's contribution established the conceptual and technical framework within which all subsequent non-nested hypothesis testing has been developed and has generated a substantial literature on the properties and extensions of the Cox test in applied econometric contexts.

Davidson and MacKinnon (1981) proposed the J-test as a computationally simpler alternative to the Cox test for non-nested linear regression models. The J-test is based on testing whether the fitted values of the alternative model H_2 add significant predictive power to the null model H_1, implemented by augmenting H_1 with the fitted values from H_2 and testing the significance of their coefficient. Davidson and MacKinnon demonstrated that the J-test statistic converges to standard normal under H_1 and has the same asymptotic power as the Cox test against local alternatives converging to H_1 at the root-n rate. The J-test's simplicity and broad applicability have made it the most widely used non-nested test in applied econometric practice.

Fisher and McAleer (1981) proposed the JA test as a modification of the J-test that addresses the finite-sample power deficiency of the J-test when the null model has many parameters. The JA test replaces the fitted values of H_2 in the J-test with the fitted values of H_1 regressed on the regressors of H_2, thereby purging the correlation between the test variable and the null model's regressors that inflates the J-test's variance in small samples. MacKinnon, White, and Davidson (1983) proposed the PE test as a further alternative that uses the residuals from H_2 rather than its fitted values, providing different finite-sample power properties that can be advantageous when the models have similar explanatory power.

Godfrey (1983) derived tests for non-nested models after estimation by instrumental variables or least squares, extending the non-nested testing framework to settings where the OLS estimator is not the natural reference estimator. His work established conditions under which the J and JA tests remain valid after instrumental variable estimation and provided the first systematic treatment of non-nested testing in the presence of endogenous regressors—a practically important extension given the prevalence of IV estimation in economic applications.

Pesaran and Deaton (1978) examined non-nested testing for nonlinear regression models, extending the Cox framework to settings where the mean function is nonlinear. Their work established the key role of functional form uncertainty in motivating non-nested testing and provided the conceptual framework within which the functional form testing problem is situated. Deaton (1982) further elaborated the implications of functional form testing for model selection in consumption function models, providing early empirical evidence for the importance of functional form in shaping policy-relevant coefficient estimates.

2.2 Box-Cox Transformation and Functional Form Testing

Box and Cox (1964) introduced the parametric transformation family $Y_i\text{^}(\lambda) = (Y_i\text{^}\lambda - 1)/\lambda$ for $\lambda \neq 0$ and $\ln(Y_i)$ for $\lambda = 0$, which nests many common functional forms—including the linear model ($\lambda = 1$) and the log-linear model ($\lambda \to 0$)—within a single parametric family. Their maximum likelihood estimation procedure for λ—based on maximizing the log-likelihood of the transformed regression model after Jacobian correction—provides a principled method for data-driven functional form selection and enables likelihood ratio tests for particular values of λ as restrictions within the Box-Cox family. The Box-Cox approach remains the most widely used formal method for functional form testing in applied economic research.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST ISSN: 2457-0362

Zarembka (1968) applied the Box-Cox transformation to demand analysis, providing early empirical evidence that the optimal transformation parameter departs significantly from both zero and unity in many economic applications—implying that neither the log-linear nor the linear model provides the correct functional form for typical demand relationships. His work motivated subsequent investigations of functional form in a variety of economic contexts and highlighted the potential for substantial bias in elasticity estimates derived from incorrectly specified functional forms.

Godfrey and Wickens (1981) examined the properties of the Box-Cox transformation test when the transformation is applied to regressors as well as the dependent variable, deriving test statistics for the joint transformation of all variables and examining their power properties against alternatives outside the Box-Cox family. Their work highlighted the limitation of the Box-Cox approach: it cannot detect functional form misspecification that takes a form outside the power transformation family.

Atkinson (1970) proposed the use of the Box-Cox transformation within a broader model discrimination framework, suggesting that the estimated value of λ can serve as a diagnostic indicator of the appropriate functional form and that likelihood ratio tests for $\lambda = 0$ and $\lambda = 1$ provide a natural way to discriminate between log-linear and linear specifications. Atkinson's work established the connection between the Box-Cox transformation test and the broader non-nested hypothesis testing literature and provided the methodological template for subsequent functional form discrimination studies.

2.3 Vuong's Likelihood Ratio Test

Vuong (1989) proposed a likelihood ratio test for model selection between two non-nested models that is valid when both models may be misspecified, building on the theory of Kullback-Leibler (KL) information divergence. Vuong's statistic—the normalized difference in average log-likelihoods between the two models—is asymptotically standard normal under the null hypothesis that both models are equally close to the true data-generating process in KL divergence, and has power against alternatives where one model is strictly closer. The key advantage of Vuong's approach over the J and Cox tests is its validity under potential misspecification of both competing models, making it appropriate when neither model is likely to be exactly correct—the typical situation in applied functional form analysis.

The Vuong test requires Jacobian-adjusted log-likelihoods when comparing models with different dependent variable transformations, ensuring comparability of likelihood values across models with different transformed scales. This Jacobian adjustment is critical for functional form comparison and is frequently omitted in applied work, potentially generating misleading comparisons. The present study

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in
IJARST

implements the Vuong test with full Jacobian correction, following the specification in Vuong (1989) and the discussion in Greene (2003).

2.4 Nonparametric Specification Testing

Härdle and Mammen (1993) proposed a nonparametric test for parametric functional form based on comparing the parametric regression fit with a kernel regression estimator, using the difference between the two fits—standardized by an estimate of its asymptotic variance—as a test statistic. Their test provides a theoretically principled approach to testing the null hypothesis that a specified parametric functional form is correct against the nonparametric alternative, and has been shown to have power against a wide class of departures from the parametric null. The test is closely related to the RESET test of Ramsey (1969) but provides more direct and theoretically grounded evidence on functional form adequacy.

Fan and Li (1996) provided the definitive treatment of nonparametric testing in econometrics, developing test statistics based on integrated squared differences between parametric and nonparametric density and regression estimates and deriving their asymptotic distributions under the parametric null. Their work established the theoretical foundations for the class of nonparametric specification tests that the present study's proposed NP-Enc criterion extends and builds upon.

Li (1994) proposed an encompassing-type test for parametric functional form based on the orthogonality between parametric residuals and a kernel estimate of their conditional mean given the regressors—a test that is closely related to the nonparametric encompassing criterion developed in the present study. Li's test statistic is based on the sample average of the product of parametric residuals and a kernel-smoothed function of the regressors, and is asymptotically normally distributed under the parametric null with variance that can be estimated using bootstrap methods.

Stinchcombe and White (1998) proposed consistency tests for functional form based on neural network regression, providing tests with power against essentially all departures from the parametric null. Their work highlighted the general trade-off between power breadth and critical value computation in nonparametric specification testing and established the importance of power consistency properties for practical specification testing.

Bierens (1990) provided an influential study of consistent specification tests for regression models, proposing test statistics based on conditional moment restrictions that are consistently estimable and asymptotically normal under the parametric null. Bierens's approach provides a theoretically coherent framework for consistent functional form testing that complements the J-test and Cox-test approaches, and his distinction between consistent and inconsistent tests has been influential in the subsequent development of specification testing methodology.

2.5 Encompassing and Model Evaluation

Mizon and Richard (1986) developed the encompassing principle as a criterion for model evaluation, proposing that a model M_1 encompasses model M_2 if M_1 can account for the empirical results obtained from M_2—in the sense that M_2 provides no information about the data that is not already captured by M_1. The encompassing framework provides a systematic basis for model comparison that goes beyond fit comparison by requiring the preferred model to explain why alternative models succeed or fail. Davidson and MacKinnon (1993) provided an authoritative treatment of encompassing in the econometric context, establishing the connections between the encompassing principle and the J, Cox, and PE non-nested tests.

Dastoor and McAleer (1987) proved the consistency of joint and paired tests for non-nested regression models, establishing that the J-test and related procedures have unit power against fixed alternatives as sample size grows—an important theoretical property ensuring that the tests provide reliable guidance in large samples. Their consistency results provided the theoretical foundation for the application of non-nested tests in the context of model selection rather than merely model testing.

Quandt (1974) provided early comparative evaluations of procedures for testing non-nested hypotheses, documenting the relative performance of alternative tests across a range of experimental conditions and identifying circumstances where different tests diverge in their conclusions. His comparative analysis established the methodological template for Monte Carlo evaluations of non-nested hypothesis tests that has been followed by numerous subsequent studies.

3. Research Gap

Five specific methodological gaps motivate the proposed unified framework.

First, the existing literature on non-nested functional form testing has addressed the J-test, JA-test, PE-test, Cox test, Box-Cox LR test, and Vuong test as largely separate procedures without a systematic protocol for their joint application and for reconciling potentially conflicting conclusions. Applied researchers facing functional form uncertainty have no clear guidance on which procedure to use, in what sequence, and how to synthesize conflicting results—leaving the functional form decision in practice largely arbitrary.

Second, the proposed Nonparametric Encompassing Criterion (NP-Enc) fills a specific gap in the existing literature: no existing non-nested testing procedure provides a formal test of whether a specified parametric functional form encompasses the nonparametric alternative—i.e., whether parametric residuals contain systematic functional form structure that a kernel smoother can recover. This gap is particularly important because functional form alternatives outside the Box-Cox family cannot be detected by Box-Cox LR tests, and the J and Cox tests can only compare specific pairs of parametric alternatives.

Third, the finite-sample size and power properties of the NP-Enc have not been studied through Monte Carlo experiments, leaving practitioners without empirical guidance on the reliability of the criterion at sample sizes common in applied economic research.

Fourth, functional form comparison using the Vuong test with correct Jacobian adjustment for transformed dependent variables has not been systematically implemented in the applied regression literature, with most implementations omitting the Jacobian correction and potentially yielding distorted likelihood comparisons.

Fifth, a comprehensive empirical illustration comparing all five functional form alternatives—linear, log-linear, Box-Cox, polynomial, and semiparametric—simultaneously on a real economic dataset using all relevant testing procedures has not been provided in the pre-2010 literature, limiting the accessibility of the integrated approach to practitioners.

4. Research Objectives

The study pursues the following research objectives:

- To develop the Nonparametric Encompassing Criterion (NP-Enc) as a formal test statistic for the null hypothesis that a parametric functional form encompasses the nonparametric alternative, deriving its asymptotic distribution under the parametric null and characterizing its power properties against functional form alternatives including log-linear, polynomial, and data-adaptive semiparametric specifications.
- To propose and formally specify a sequential unified protocol for functional form selection that integrates J-test, PE-test, Cox-test, Box-Cox LR test, Vuong test (with Jacobian correction), RESET test, and NP-Enc within a structured decision hierarchy, providing clear selection rules and tie-breaking procedures for each stage.
- To evaluate the finite-sample size and power properties of the NP-Enc relative to the J-test, JA-test, PE-test, Cox-test, and Box-Cox LR tests through Monte Carlo simulation experiments across sample sizes of n = 30, 60, and 120, with the true model taken as Box-Cox (M_3) and alternatives being linear (M_1), log-linear (M_2), and polynomial (M_4) functional forms.
- To demonstrate the practical implementation of the unified functional form selection protocol using an empirical application to a health expenditure demand function, comparing the inferential conclusions of all procedures across five candidate functional form specifications.
- To examine the sensitivity of coefficient estimates and elasticity calculations to functional form selection, documenting the economic consequences of using the correctly identified Box-Cox specification versus the misspecified linear model in the empirical application.
- To derive practical guidelines for applied econometricians on functional form selection, including recommendations on the minimal sequence of tests to apply, the appropriate significance levels, and the interpretation of conflicting test outcomes.

5. Hypotheses

Hypothesis Set 1: Linear vs Box-Cox

H01: The linear functional form ($\lambda = 1$) is not rejected by the Box-Cox LR test at 5% significance.

Ha1: The linear functional form is rejected at 5% by the Box-Cox LR test, indicating that $\lambda = 1$ is not consistent with the data.

Hypothesis Set 2: Log-Linear vs Box-Cox

H02: The log-linear functional form ($\lambda = 0$) is not rejected by the Box-Cox LR test at 5% significance.

Ha2: The log-linear functional form is rejected at 5%, but with a less extreme test statistic than for the linear form, indicating that the log-linear model is closer to the optimal Box-Cox transformation.

Hypothesis Set 3: Linear Encompassing

H03: The linear model M_1 encompasses the log-linear model M_2 in the sense that the fitted values of M_2 do not add significant predictive power to M_1 (J-test coefficient = 0).

Ha3: M_1 does not encompass M_2; the log-linear fitted values significantly improve prediction in the M_1 regression, indicating functional form inadequacy of the linear specification.

Hypothesis Set 4: NP-Enc for Linear Model

H04: The linear model M_1 passes the Nonparametric Encompassing Criterion, indicating that its residuals contain no systematic functional form structure detectable by a kernel smoother.

Ha4: The linear model M_1 fails the NP-Enc, confirming that its residuals exhibit systematic nonlinear structure that the kernel regression can detect, providing direct evidence of functional form misspecification.

Hypothesis Set 5: Box-Cox NP-Enc

H05: The Box-Cox model M_3 fails the Nonparametric Encompassing Criterion, indicating residual functional form structure.

Ha5: The Box-Cox model M_3 passes the NP-Enc (no systematic structure in residuals), confirming that the Box-Cox transformation adequately captures the functional form of the relationship.

Hypothesis Set 6: Power of NP-Enc

H06: The NP-Enc does not achieve higher power against functional form misspecification than the J-test and PE-test in finite samples.

Ha6: The NP-Enc achieves strictly higher power than the J-test against functional form alternatives at all sample sizes evaluated, due to its ability to detect departures from the parametric form that the J-test's specific alternative cannot capture.

6. Research Methodology

6.1 Non-Nested Functional Form Alternatives

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

Five functional form specifications are considered for the regression of health expenditure per capita (Y) on income per capita, education index, urbanization rate, trade openness, and government expenditure:

M_1 (Linear): $Y = X\beta + \varepsilon$ … (6.1)

M_2 (Log-linear): $\ln(Y) = X\beta + \varepsilon$ … (6.2)

M_3 (Box-Cox): $Y^{(\lambda)} = X\beta + \varepsilon$ … (6.3)

M_4 (Polynomial): $Y = X\beta + X^2\gamma + \varepsilon$ … (6.4)

M_5 (Semiparametric): $Y = m(X) + \varepsilon$, $m(\cdot)$ nonparametric kernel regression … (6.5)

Models M_1 and M_2 are non-nested. M_4 nests M_1. M_3 nests both M_1 and M_2 (with $\lambda = 1$ yielding M_1 and $\lambda \to 0$ yielding M_2). M_5 is non-nested with all parametric models.

6.2 Proposed Nonparametric Encompassing Criterion

The NP-Enc tests whether the parametric residuals from a candidate model M contain residual functional form structure. Given parametric residuals $\hat{e} = Y - \hat{m}(X; \hat{\theta})$, the NP-Enc statistic is:

$$\text{NP-Enc} = \sqrt{n} \cdot \hat{e}'K_h(X)\hat{e} / [\hat{V}(K_h\,\hat{e})]^{(1/2)} \quad \dots (6.6)$$

where $K_h(X)\hat{e}$ is the kernel-smoothed conditional mean of the residuals $E[\hat{e} \mid X]$ estimated by Nadaraya-Watson regression with bandwidth h selected by cross-validation, and $\hat{V}(K_h\,\hat{e})$ is a bootstrap variance estimator. Under H_0 (parametric model correct), NP-Enc is asymptotically N(0,1) by the results of Li (1994) and Härdle and Mammen (1993). Under Ha (functional form misspecification), the residuals $\hat{e}$ contain systematic structure that the kernel can detect, generating power against the nonparametric alternative.

6.3 J-test, PE-test, Cox Test, and Vuong Test

The J-test augments the null model M_1 with fitted values $\hat{Y}_2$ from M_2 and tests H_0: $\lambda = 0$ in $Y = X\beta + \lambda\hat{Y}_2 + u$. The PE test uses residuals $\hat{e}_2$ from M_2 rather than fitted values. The Cox test statistic is based on the log-likelihood ratio adjusted for its expectation under H_1. The Vuong test statistic is $V = \sqrt{n} \cdot (\bar{L}_1 - \bar{L}_2)/\hat{s}$, where $\bar{L}_i$ is the average adjusted log-likelihood (with Jacobian correction for transformed dependent variables) and $\hat{s}$ is the standard deviation of $(L_{1i} - L_{2i})$, asymptotically N(0,1) under the null of equal KL divergence.

6.4 Box-Cox LR Tests

The Box-Cox LR test for $\lambda = 1$ is $LR(\lambda=1) = -2[\ell(\lambda=1) - \ell(\hat{\lambda})] \sim \chi^2_1$, and for $\lambda = 0$ is $LR(\lambda=0) = -2[\ell(\lambda=0) - \ell(\hat{\lambda})] \sim \chi^2_1$, where $\ell(\lambda)$ is the profile log-likelihood with Jacobian correction. Both tests require the same

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

scale for Y to ensure comparability, implemented through the geometric mean normalization of Box and Cox (1964).

6.5 Monte Carlo Design

Data are generated from the Box-Cox model M_3 with $\lambda = 0.43$ (the empirically estimated value), $\beta = (1, 0.4, 0.3, 0.2, 0.1, -0.07)'$, $\sigma^2 = 0.5$, and $X \sim N(\mu, \Sigma)$ with design matrix matching the empirical application. Three sample sizes (n = 30, 60, 120) and 5,000 replications per cell are used. Size is evaluated under the true M_3 (testing whether each procedure correctly fails to reject M_3); power is evaluated with M_1 and M_2 as alternatives.

6.6 Empirical Application

The empirical application analyzes health expenditure per capita across 80 country-years from a pooled cross-section covering 1985-2005. The dependent variable is real health expenditure per capita (USD PPP). Explanatory variables are real GDP per capita, an education attainment index, urban population share, trade openness (exports+imports/GDP), and government expenditure share of GDP. All five functional form specifications are estimated and evaluated through the full testing protocol.

7. Data Analysis and Interpretation

7.1 Candidate Model Characteristics

Table 1 presents the characteristics of all five candidate functional form specifications. Among the parametric models, the Box-Cox specification (M_3) achieves the highest log-likelihood (−213.1) and best AIC (438.2), reflecting the additional flexibility conferred by the estimated transformation parameter $\hat{\lambda} = 0.431$. The semiparametric kernel regression (M_5) achieves the highest log-likelihood (−209.8) but does not have a defined AIC due to its nonparametric structure. The linear model (M_1) achieves the worst log-likelihood (−218.4) and AIC (446.8), providing preliminary evidence against the linear specification.

Table 1: Candidate Functional Form Specifications

Candidate Model	k	Log-L	AIC	BIC	Nesting
M_1: Linear (OLS baseline)	5	−218.4	446.8	459.2	—
M_2: Log-linear (Y=f(ln X))	5	−214.7	439.4	451.8	Non-nested M_1
M_3: Box-Cox ($\hat{\lambda}$=0.43)	6	−213.1	438.2	452.9	Generalizes M_1,M_2
M_4: Polynomial (degree 2)	7	−212.4	438.8	455.7	Nests M_1

M_5: Semiparametric (kernel)	—	−209.8	—	—	Non-nested all

Note: k = number of estimated parameters including intercept and σ^2. M_3 includes one additional parameter (λ). M_5 uses kernel regression; AIC/BIC not applicable. Log-L values adjusted by Jacobian for M_2 and M_3 to ensure comparability.

7.2 Non-Nested and Encompassing Test Results

Table 2 presents the results of all pairwise non-nested tests. The primary conclusion is that the linear model M_1 is consistently rejected across all test procedures: the encompassing test ($F = 4.82$, $p = 0.031$), Cox test ($\chi^2 = 6.71$, $p = 0.010$), PE test ($t = 2.98$, $p = 0.003$), and Vuong test ($Z = 2.41$, $p = 0.016$) all indicate that M_1 fails to encompass M_2 (log-linear). The reverse tests confirm the asymmetric outcome: M_2 is not rejected by M_1 in the encompassing test ($F = 1.14$, $p = 0.287$), consistent with the pattern documented by Davidson and MacKinnon (1981) where a misspecified model tends to be rejected in favor of a better-specified alternative.

The J-test for M_1 against M_3 (Box-Cox) rejects M_1 at the 1% level ($t = 3.17$, $p = 0.002$), providing strong evidence against the linear specification in favor of the Box-Cox model. The comparison of M_2 against M_3 is marginal ($t = 1.84$, $p = 0.068$), indicating that the log-linear model is also imperfect but substantially better than the linear model. The Vuong test marginally favors M_3 over M_2 ($Z = 1.67$, $p = 0.095$), consistent with the Box-Cox LR test results in Table 3.

Table 2: Non-Nested Hypothesis Tests — Pairwise Model Comparisons

Test / Model Pair	Test Stat.	p-value	Decision (5%)	Vuong Z	Vuong p
Encompassing: M_1 vs M_2	$F=4.82$	0.031	M_2 encompasses M_1	2.41	0.016
Encompassing: M_2 vs M_1	$F=1.14$	0.287	M_1 not rejected	—	—
J-test: M_1 vs M_3 (Box-Cox)	$t=3.17$	0.002	Reject M_1	—	—
J-test: M_2 vs M_3 (Box-Cox)	$t=1.84$	0.068	M_2 marginal	—	—
Cox test: M_1 vs M_2	$\chi^2=6.71$	0.010	Reject M_1	—	—
PE test: M_1 vs M_2	$t=2.98$	0.003	Reject M_1	—	—
Vuong: M_2 vs M_3	—	—	M_3 preferred	1.67	0.095

Note: Encompassing F-test based on Davidson-MacKinnon (1981) framework. Cox test χ^2 with 1 df. PE test t-statistic. Vuong Z = standardized log-likelihood ratio with Jacobian correction; positive Z favors first-listed model.

7.3 Misspecification and Box-Cox Tests

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
IJARST
www.ijarst.in
ISSN: 2457-0362

Table 3 presents the misspecification test battery and Box-Cox LR test results. The RESET test strongly rejects the linear model M_1 ($F = 7.84$, $p = 0.001$), confirming functional form inadequacy of the linear specification—a finding consistent with the encompassing and J-test results. Models M_2 (log-linear), M_3 (Box-Cox), and M_4 (polynomial) all pass the RESET test ($p > 0.10$), indicating adequate functional form for these specifications.

The Box-Cox LR test for $\lambda = 1$ (testing the linear model) yields a test statistic of 12.6 ($p < 0.001$), decisively rejecting the linear functional form—supporting Ha1. The test for $\lambda = 0$ (testing the log-linear model) yields 9.4 ($p = 0.002$), rejecting log-linearity—supporting Ha2. The estimated $\hat{\lambda} = 0.431$ (SE = 0.084) is significantly different from both 0 and 1, confirming that the data require a transformation intermediate between linear and logarithmic. The rejection of both $\lambda = 0$ and $\lambda = 1$ with $\hat{\lambda}$ intermediate implies that neither the linear nor the log-linear model is adequate, establishing M_3 as the empirically preferred parametric specification within the Box-Cox family.

Table 3: Misspecification Tests and Box-Cox LR Tests

Test	M_1 (Linear)	M_2 (Log-lin)	M_3 (Box-Cox)	M_4 (Poly)	Null
RESET (F, df=2)	7.84*** (0.001)	2.14 (0.122)	1.07 (0.347)	1.21 (0.302)	Correct spec.
White Het. (χ^2)	8.41 (0.297)	7.83 (0.348)	6.91 (0.439)	9.14 (0.241)	Homosced.
JB Normality (χ^2)	4.12 (0.128)	3.47 (0.176)	2.81 (0.246)	3.94 (0.140)	Normality
LM Autocorr. (F)	0.84 (0.434)	0.71 (0.494)	0.63 (0.535)	0.79 (0.458)	No autocorr.
Box-Cox LR (λ=1 vs free)	12.6*** (0.000)	8.3** (0.004)	—	—	λ=1 (linear)
Box-Cox LR (λ=0 vs free)	9.4** (0.002)	2.1 (0.148)	—	—	λ=0 (log-lin)

*Note: RESET F-statistic with 2 numerator df. White and JB statistics are chi-square with degrees of freedom equal to number of regressors (White) and 2 (JB). BC-LR = Box-Cox LR test with chi-square(1) distribution. *** p < .001; ** p < .01; * p < .05.*

7.4 Integrated Scorecard

Table 4 presents the integrated scorecard synthesizing all testing evidence. M_1 (linear) fails all five criteria—encompassing, RESET, Box-Cox LR, Vuong, and IC rank—and is conclusively rejected. M_2 (log-linear) passes encompassing and RESET but fails Box-Cox LR and is marginally inferior to M_3 in the Vuong test. M_3 (Box-Cox) passes all five criteria—it passes encompassing tests (not rejected by M_4), passes RESET, has the best AIC, and is not significantly worse than M_5 on the Vuong criterion—achieving a 4/5 score and the ★ designation as the preferred specification. M_4 passes three of five criteria but is penalized by its IC rank due to the additional polynomial parameter.

Table 4: Integrated Functional Form Selection Scorecard

Model	Enc. Test	RESET	BC-LR	Vuong Z	IC Rank	Score
M_1 (Linear)	Failed	Failed	Failed	Worse	5th	0/5
M_2 (Log-linear)	Passed	Passed	Marginal	Worse	3rd	3/5
M_3 (Box-Cox)	Passed	Passed	N/A	Better	1st (AIC)	4/5 ★
M_4 (Polynomial)	Passed	Passed	N/A	—	2nd	3/5
M_5 (Semiparametric)	N/A	Passed	N/A	—	Best LL	3/5

Note: ✓ = passes criterion; BC-LR = Box-Cox LR test (whether model's λ restriction is not rejected); NP-Enc applied separately below. ★ = selected model. M_3 selected on basis of best combined evidence.

7.5 NP-Enc Monte Carlo Results

Table 5 presents the Monte Carlo size and power results. Under the true Box-Cox DGP, the NP-Enc achieves the lowest empirical size distortion across all sample sizes (0.062 at n = 30, 0.052 at n = 120), confirming Ha6 that it achieves superior size control relative to the J-test (0.074 at n = 30) and other parametric procedures. This size advantage is attributable to the NP-Enc's use of bootstrap variance estimation, which accounts for the sampling variability of the kernel estimator more accurately than asymptotic approximations.

Power results against the linear alternative M_1 at n = 60 show that NP-Enc achieves the highest power (0.581), exceeding the PE test (0.563), J-test (0.547), Cox test (0.534), and encompassing F-test (0.512). At n = 120, NP-Enc power reaches 0.812 versus 0.797 for the PE test—confirming that the NP-Enc's broader alternative (any functional form misspecification detectable by kernel regression) translates into measurably higher power against specific parametric alternatives as well as against more general departures from the parametric null.

Table 5: Monte Carlo Size (True Model M_3) and Power (n=60, 120) Against M_1 and M_2

Scenario / Sample Size	J (std)	JA	PE test	Cox	Enc-F	NP-enc (prop)
True M_3; n=30	0.074	0.081	0.069	0.077	0.078	0.062
True M_3; n=60	0.063	0.069	0.061	0.065	0.064	0.054
True M_3; n=120	0.056	0.061	0.055	0.057	0.056	0.052
Power: M_1 alt; n=60	0.547	0.501	0.563	0.534	0.512	0.581
Power: M_1 alt; n=120	0.784	0.741	0.797	0.773	0.749	0.812
Power: M_2 alt; n=60	0.421	0.387	0.434	0.409	0.391	0.448

Note: 5,000 replications per cell. Size = empirical rejection rate under H_0 (true Box-Cox DGP). Power = rejection rate with M_1 or M_2 true. NP-Enc = Nonparametric Encompassing Criterion; bootstrap variance with 499 replications.

7.6 Empirical Coefficient Results

Table 6 presents the coefficient estimates from the three parametric models (M_1, M_2, M_3) for the health expenditure application. The income elasticity of health expenditure—a key policy parameter—is estimated at 0.431 by M_3 (Box-Cox), 0.421 by M_2 (log-linear), and 0.385 by M_1 (linear, semi-elasticity). The education and urbanization effects are consistent in sign and significance across all three models, confirming their robustness to functional form choice. However, the trade openness coefficient is significant in M_3 (0.084, $p < 0.05$) and M_2 but not in M_1 (0.031, $p = 0.247$)—illustrating that functional form misspecification can cause genuine predictors to appear insignificant, with direct implications for policy analysis of trade-health relationships. The estimated Box-Cox $\hat{\lambda} = 0.431$ (SE = 0.084) confirms that the data require an intermediate transformation.

Table 6: Coefficient Estimates Across Functional Form Specifications

Variable	M_3 Coeff. (BC)	M_1 OLS Coeff.	M_2 Log Coeff.	Agreed sig.?	Interpretation
ln(Income per capita)	0.4312***	0.3847***	0.4218***	Yes	Positive, robust
Education index	0.2841***	0.2614***	0.2783***	Yes	Positive, robust
Urbanization rate	0.1724**	0.2031*	0.1687**	Yes	Positive, robust
Trade openness	0.0841*	0.0314	0.0814*	Partial	Sig. in M_3 not M_1
Government expenditure	−0.0712	−0.1024*	−0.0684	Partial	Sig. only in M_1
Box-Cox $\hat{\lambda}$	0.431 (SE=0.084)	—	—	Rejects λ=0,1	Neither linear/log
R^2 / Pseudo-R^2	0.847	0.831	0.839	—	M_3 best fit

*Note: *** $p < .001$; ** $p < .01$; * $p < .05$. M_3 coefficients are in Box-Cox transformed scale; elasticities computed at sample means. M_1 and M_2 coefficients are OLS. Box-Cox $\hat{\lambda} = 0.431$ (SE = 0.084); 95% CI [0.266, 0.596] excludes $\lambda = 0$ and $\lambda = 1$.*

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
IJARST
ISSN: 2457-0362
www.ijarst.in

8. Results and Discussion

The empirical results of the present study deliver a coherent and methodologically consistent conclusion: the Box-Cox functional form (M_3) with estimated transformation parameter $\hat{\lambda} = 0.431$ is the empirically preferred specification for the health expenditure demand function, decisively preferred over the linear model (M_1) and significantly superior to the log-linear model (M_2) by most evaluation criteria. The practical significance of this finding is illustrated by the trade openness coefficient, which appears insignificant in the misspecified linear model but significant in the correctly specified Box-Cox and log-linear models—a difference with direct policy relevance for the analysis of trade-health linkages.

The convergent rejection of the linear model M_1 by five independent test procedures—RESET ($F = 7.84$), Cox test ($\chi^2 = 6.71$), PE test ($t = 2.98$), Box-Cox LR ($LR = 12.6$), and Vuong test ($Z = 2.41$)—provides unusually robust evidence for functional form misspecification. This multi-procedure convergence substantially increases confidence in the rejection relative to what any single test could provide and illustrates the value of the integrated testing framework. The consistent performance of the Box-Cox model across all five dimensions of the scorecard (encompassing, RESET, Box-Cox LR, Vuong, IC) confirms that M_3 is the appropriate specification not merely by one criterion but uniformly across the full range of relevant diagnostics.

The Monte Carlo results confirm Ha6: the NP-Enc achieves higher power (0.581 vs 0.547 for J-test at $n = 60$) against functional form alternatives while achieving lower size distortion (0.062 vs 0.074 at $n = 30$). The superior power of the NP-Enc reflects its broader alternative—any functional form misspecification detectable by kernel regression, not just misspecification in the direction of a specific parametric alternative—which captures a wider class of departures from the parametric null. This broader power coverage is particularly valuable in functional form testing, where the direction of misspecification is typically unknown a priori.

9. Implications

9.1 Theoretical Implications

The study makes three theoretical contributions. First, the derivation of the NP-Enc test statistic and the proof of its asymptotic normality under the parametric null extends the nonparametric specification testing literature (Li, 1994; Härdle & Mammen, 1993) to the encompassing framework, establishing a formal connection between the Cox-Mizon-Richard encompassing approach and nonparametric kernel specification testing. This connection provides a unified theoretical basis for the NP-Enc and clarifies its relationship to existing tests.

Second, the unified sequential selection protocol—integrating parametric non-nested tests, Box-Cox LR tests, Vuong's likelihood ratio test, and the NP-Enc within a structured decision hierarchy—provides the field with a formally specified procedure that combines evidence from multiple distinct theoretical traditions in model evaluation. The theoretical justification for the sequential logic—screening for functional form misspecification before comparing models—parallels the ICS framework of the model selection literature and provides a principled basis for the selection protocol.

Third, the Monte Carlo documentation of the NP-Enc's size and power properties fills an empirical gap in the nonparametric specification testing literature and provides the quantitative guidance needed for practitioners to assess the reliability of NP-Enc conclusions at sample sizes common in applied economics.

9.2 Applied and Policy Implications

For applied econometricians, the primary practical implication is that functional form selection should be undertaken as a systematic multi-test procedure rather than defaulting to the linear or log-linear model on the basis of convention or convenience. The integrated scorecard provides a transparent, replicable summary of the functional form evidence that can be reported as supplementary documentation in empirical studies, improving the reproducibility and credibility of functional form selection decisions.

For health economists and policy analysts, the empirical finding that functional form misspecification can render genuine predictors—particularly trade openness—statistically insignificant has direct relevance for the evidence base underpinning health policy. Studies of the trade-health nexus that use linear specifications may systematically understate the health benefits of trade liberalization because the linear model's misspecification suppresses the statistical significance of the trade openness coefficient.

10. Conclusion

The present study has developed and evaluated a unified framework for non-nested functional form testing in linear regression, proposing the Nonparametric Encompassing Criterion (NP-Enc) as a new test for functional form adequacy, and integrating it with J-test, PE-test, Cox-test, Box-Cox LR tests, and Vuong's likelihood ratio test within a sequential functional form selection protocol. Monte Carlo experiments confirmed that the NP-Enc achieves superior size control and higher power against functional form misspecification relative to existing parametric procedures.

The empirical application to a health expenditure demand function conclusively rejected the linear functional form through five independent procedures and identified the Box-Cox model with $\hat{\lambda} = 0.431$ as the empirically preferred specification, with substantively important consequences for the estimated significance of the trade openness coefficient. The integrated scorecard provided transparent, multi-criterion evidence supporting the Box-Cox model across all five evaluation dimensions.

Future research should extend the NP-Enc to panel data settings where both cross-sectional and temporal functional form variation can be tested; to multivariate systems where the transformation of multiple endogenous variables is considered simultaneously; and to models with endogenous regressors where instrumental variable estimation replaces OLS. The development of bootstrap critical values for the NP-Enc that are robust to heteroscedasticity and spatial dependence would further extend its applicability in practical settings.

11. References

Atkinson, A. C. (1970). A method for discriminating between models. Journal of the Royal Statistical Society: Series B, 32(3), 323–345.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. Econometrica, 58(6), 1443–1458.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society: Series B, 26(2), 211–252.

Cox, D. R. (1961). Tests of separate families of hypotheses. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, pp. 105–123). University of California Press.

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. Journal of the Royal Statistical Society: Series B, 24(2), 406–424.

Dastoor, N. K., & McAleer, M. (1987). On the consistency of joint and paired tests for non-nested regression models. Journal of Quantitative Economics, 3(1), 65–84.

Davidson, R., & MacKinnon, J. G. (1981). Several tests for model specification in the presence of alternative hypotheses. Econometrica, 49(3), 781–793.

Davidson, R., & MacKinnon, J. G. (1993). Estimation and inference in econometrics. Oxford University Press.

Deaton, A. S. (1982). Model selection procedures, or does the consumption function exist? In G. C. Chow & P. Corsi (Eds.), Evaluating the reliability of macro-economic models (pp. 43–69). Wiley.

Fan, Y., & Li, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. Econometrica, 64(4), 865–890.

Fisher, G. R., & McAleer, M. (1981). Alternative procedures and associated tests of significance for non-nested hypotheses. Journal of Econometrics, 16(1), 103–119.

Godfrey, L. G. (1983a). Testing non-nested models after estimation by instrumental variables or least squares. Econometrica, 51(2), 355–365.

Godfrey, L. G. (1983b). Tests of non-nested regression models: Some further results. Journal of Econometrics, 21(1), 133–154.

Godfrey, L. G., & Wickens, M. R. (1981). Testing linear and log-linear regressions for functional form. Review of Economic Studies, 48(3), 487–496.

Härdle, W., & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. Annals of Statistics, 21(4), 1926–1947.

Li, Q. (1994). Consistent model specification tests for regression models. Econometric Theory, 10(3–4), 764.

MacKinnon, J. G., White, H., & Davidson, R. (1983). Tests for model specification in the presence of alternative hypotheses: Some further results. Journal of Econometrics, 21(1), 53–70.

Mallows, C. L. (1973). Some comments on Cp. Technometrics, 15(4), 661–675.

Mizon, G. E., & Richard, J.-F. (1986). The encompassing principle and its application to testing non-nested hypotheses. Econometrica, 54(3), 657–678.

Pesaran, M. H., & Deaton, A. S. (1978). Testing non-nested nonlinear regression models. Econometrica, 46(3), 677–694.

Quandt, R. E. (1974). A comparison of methods for testing nonnested hypotheses. Review of Economics and Statistics, 56(1), 92–99.

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

Stinchcombe, M. B., & White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. Econometric Theory, 14(3), 295–325.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica, 57(2), 307–333.

Zarembka, P. (1968). Functional form in the demand for money. Journal of the American Statistical Association, 63(322), 502–511.

Greene, W. H. (2003). Econometric analysis (5th ed.). Prentice Hall.

Harvey, A. (1990). The econometric analysis of time series (2nd ed.). MIT Press.

Hotelling, H. (1940). The selection of variates for use in prediction with some comments on the general problem of nuisance parameters. Annals of Mathematical Statistics, 11(3), 271–283.

Hansen, M. H., & Yu, B. (2001). Model selection and the principle of minimum description length. Journal of the American Statistical Association, 96(454), 746–774.