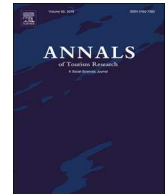




ELSEVIER

Contents lists available at ScienceDirect

Annals of Tourism Research

journal homepage: www.elsevier.com/locate/annals

Research Note

On proper specification in tourism research

A. George Assaf^{a,*}, Mike G. Tsionas^{b,1}^a Isenberg School of Management, University of Massachusetts-Amherst, 90 Campus Center Way, 209A Flint Lab, Amherst, MA 01003, United States^b Lancaster University Management School, United Kingdom

ARTICLE INFO

Associate editor: Sara Dolnicar

Introduction

Despite the excessive use of regression modelling in the field, it is surprising that not much attention has been given to some critical issues that affect the process of hypothesis testing. In this note, we argue that extensive diagnostic testing of the model is required before proceeding with hypothesis testing or other usages of the model. Problems like autocorrelation and heteroscedasticity should be interpreted as *misspecification* tests rather than as problems that merely affect standard errors (which is the case only in *correctly specified* models). Moreover, we suggest out-of-sample forecasting as a tool for model validation.

To illustrate, we use a simple regression model with an interaction term. Often, we test hypotheses about moderating variables using the following:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u. \quad (1)$$

For instance, when $\beta_3 \neq 0$ we conclude that the moderating impact is significant.

Prior to making such conclusions, it is important to go through several and necessary important steps that are often not reported in tourism papers and may significantly affect the results.

1. First, the issue of specification: there is, in fact, no guarantee that the functional form in (1) is acceptable (or “true”). While such model may produce “reasonable” results, this may be misleading because of misspecification problems. *Misspecification may be detected using diagnostic tests (RESET, heteroscedasticity, and autocorrelation)*. Such diagnostics tests are, fortunately, sensitive in various forms of misspecification. Reporting heteroscedasticity tests before testing hypotheses (such as $\beta_3 = 0$ in (1)²) is important. For instance, models such as (1) may fail several diagnostics *because* of misspecification problems.
2. Second, the issue of reverse causality, which occurs when x_1 , x_2 and y are jointly determined, leading to endogeneity problems that may affect adversely the issue of estimating models like (1).
3. Third, and related to point 1 above, we rarely see nonparametric models like $y = f(x_1, x_2)$ being estimated.

We emphasize and discuss the above in more detail and illustrate using a tourism application. We examine the performance of specification tests for panel data which, to our knowledge is novel. The baseline model is a regression where “sales” depends on advertising, size and the interaction of advertising and size. Although the reported results are sensible, we look deeper into the

* Corresponding author.

E-mail addresses: assaf@isenberg.umass.edu (A.G. Assaf), m.tsionas@lancaster.ac.uk (M.G. Tsionas).

¹ Both authors have contributed equally to the paper.

² In other instances, we may be interested in testing economic restrictions like $\beta_1 + \beta_2 = 1$, inequality restrictions etc.

<https://doi.org/10.1016/j.annals.2018.11.007>

Received 6 March 2018; Received in revised form 16 November 2018; Accepted 20 November 2018

0160-7383/ © 2018 Elsevier Ltd. All rights reserved.

problem and uncover problems with the specification. The model is misspecified in terms of the functional form based on the RESET test. In turn, this motivates us to investigate two other models; a model with fixed effects and a nonparametric model. The fixed-effects model also fails the RESET test. Although the nonparametric model passes the RESET test it fails a heteroscedasticity test: If heteroscedasticity is not merely an issue that requires light-hearted application of robust standard errors (such as HAC) but is, instead, indicative of specification error (omitted variables for example) then we have a warning that the nonparametric model is misspecified as well. This motivates us to investigate a dynamic models and reverse causality issues. A dynamic panel data model passes both RESET and heteroscedasticity tests and, therefore, can be considered as a valid working platform for further analysis.

How to proceed?

Multiple regression models are prone to functional form misspecification. The specified model does not always account properly for the relationship between the outcome and observed variables. For instance, there is no reason to assume that a relationship is linear. Omitting a quadratic term or an interaction term when they *should* be included, may lead to functional form misspecification or omitted variable bias.

Specification should be the first and most important critical step.³ Tests such a Ramsey's RESET are designed to detect non-linearities in the model (arising from possibly excluding powers of the *existing* regressors) but not omitted variables *per se*. *Misspecification will often result in heteroscedasticity and / or autocorrelation. Reverse causality should be also tested when appropriate.* One can also use non-parametric regression to avoid imposing restrictive *a priori* specifications on the functional form. However, there is no guarantee that the non-parametric regression will pass all these tests if there are omitted variables. We discuss below all these issues in more detail using a real example.

Data and results

We use a well-known application on the relationship between advertising and sales. We use data on 21 US hotels and restaurants covering a balanced sample of publicly traded companies from 2003 to 2012 (21*10 = 210 observations). We propose the following models for comparison:

$$\ln sales = \alpha_0 + \alpha_1 Adv + \alpha_2 Adv^2 + \alpha_3 Size + \alpha_4 Adv * Size + \alpha_5 Leverage + \varepsilon \quad (\text{Model 1})$$

$$\ln sales = \alpha_0 + \alpha_1 Adv + \alpha_2 Adv^2 + \alpha_3 Adv^3 + \alpha_4 Size + \alpha_5 Size^2 + \alpha_6 Adv * Size + \alpha_7 Leverage + \varepsilon \quad (\text{Model 2})$$

where *sales* is total sales,⁴ *Adv* is advertising expenses, *Size* is hotel size, *Leverage* is financial leverage and is used as a control variable. In line with the previous literature that suggests an S-shaped relationship between advertising and sales, we include both the square and cubic term of advertising in Model 2.

As we have panel data, we tried estimating the models using fixed effects in a static and dynamic framework:

$$y_{it} = \alpha_i + x'_{it}\beta + u_{it}, \quad (2)$$

$$y_{it} = \alpha_i + \rho y_{i,t-1} + x'_{it}\beta + u_{it}.$$

For comparison we also estimated the following model in a nonparametric fashion: $\ln sales = f(Adv, Size, Leverage)$, but as can we see from [Table 1](#), the nonparametric model failed the heteroscedasticity test (H_0 : Homoscedasticity). The fixed effect estimation also fails one of the two tests in either model. The dynamic specification (using the estimator of [Arellano & Bond, 1991](#)), using two lags for $\ln sales$), on the other hand, passes both tests for Model 2.⁵

The results confirm that sales in Model 2 should be modeled using a dynamic specification, supporting other related studies in the area ([Assaf, Josiassen, Ratchford, & Barros, 2012](#)). Along with the specification tests, we also tested for reverse causality between sales and advertising using the [Dumitrescu and Hurlin \(2012\)](#) test but we did not find any reverse causality problem (p-value = 0.9133).⁶ For additional evidence, we also tested for reverse causality using the following DPD model:

$$\begin{aligned} y_{it} &= \alpha_i + \beta_{11}y_{i,t-1} + \beta_{12}x_{i,t-1} + u_{it}, \\ x_{it} &= \lambda_i + \beta_{21}y_{i,t-1} + \beta_{22}x_{i,t-1} + v_{it}, \end{aligned} \quad (3)$$

where α_i and λ_i represent fixed effects, and u_{it} , v_{it} are only contemporaneously correlated. We have reverse causality when $\beta_{21} \neq 0$. Of course, more lags can be included and weakly exogenous regressors may be present in this specification. This model is a panel vector

³ In general, the specification of regression models should be driven by theoretical considerations, but we can never pretend that we have "too much theory".

⁴ We used the log of sales due to the highly skewed nature of this variable. We also tried "sales" as a dependent variable but the model failed all specification tests.

⁵ For panel data, we constructed the RESET and heteroscedasticity tests manually. The regressors of these tests are the variables in Model 2, their squares and interactions. The dependent variables are the residuals and the logarithm of absolute residuals, respectively (we provide in Appendix B more validation about the performance of these tests). We also tested for autocorrelation in STATA and we did not find an autocorrelation problem. The static model, on the other hand, failed the autocorrelation test.

⁶ Previous studies have also found evidence of one-way causality from advertising to sales (Kim, 2007).

Table 1
Heteroscedasticity and RESET tests.

	Heteroscedasticity Test	RESET Test
Non-Parametric Regression	p-value = 0.00	p-value = 0.35
Panel Regression-Fixed Effect	p-value = 0.65 (Model 1)	p-value = 0.00 (Model 1)
	p-value = 0.06 (Model 2)	p-value = 0.04 (Model 2)
Panel Regression-Dynamic	p-value = 0.00 (Model 1)	p-value = 0.80 (Model 1)
	p-value = 0.29 (Model 2)	p-value = 0.48 (Model 2)

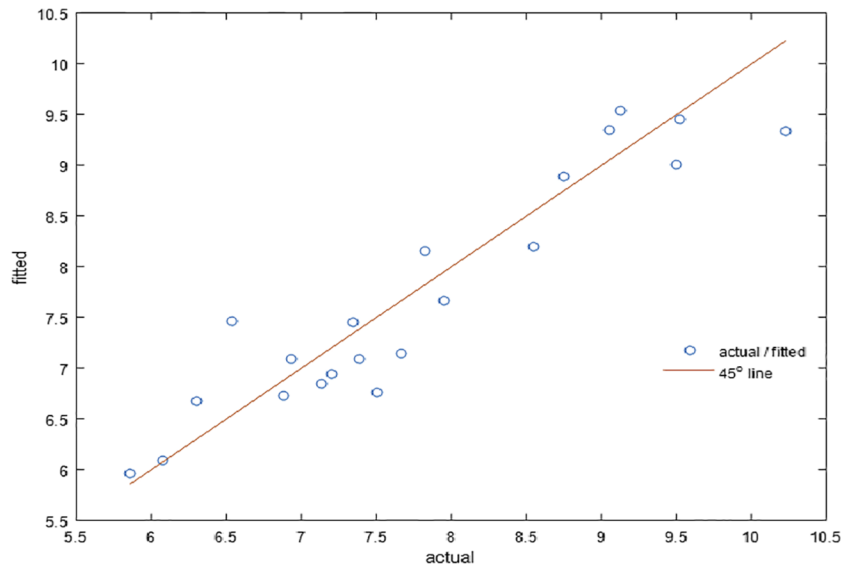


Fig. 1. Comparison of actual and fitted values in out-of-sample validation-Dynamic Model.

autoregression (PVAR) and, in fact, can serve as a prototype for several empirical applications to control for both dynamics as well as reverse causality.⁷ We tried the model in (3) using all variables in Model 2, but again, we failed to find evidence in favor of reverse causality between sales and advertising.

To further test the performance of our dynamic specification, we examined whether it is *acceptable* (leaving the notions of “correct” and “true” behind) in out-of-sample forecasting. We believe that any unacceptable model is very likely to fail this exercise. We estimated the model for the first seven years of our sample, and we forecast out-of-sample for the remaining three years. The results are shown in Fig. 1, where it is clear that the deviations between actual and predicted values are, apparently, small and randomly scattered around the 45° line.⁸

Parameter estimates for the dynamic specification are reported in Table 2. In Fig. 2, we also report the elasticity of advertising across our whole sample. The average elasticity is around 0.36, indicating that a 10% increase in advertising leads to around 3.6% increase in sales. Such impact seems to be reasonably substantial, although we recommend deriving the standard errors of these elasticity for a more concrete evidence of the impact of advertising on sales. Obviously, one method to derive those standard errors is the delta method.

Concluding remarks

In this note, we highlighted the importance of testing for specification before proceeding with further regression analysis. Often in tourism research we report the results from a single model without first confirming that the functional form and related diagnostics⁹ of such model are acceptable.

The message from our Monte Carlo simulations is that standard diagnostic tests and simple modifications when we have panel data, can reveal problematic aspects of the data. For example, autocorrelation and heteroscedasticity diagnostics are sensitive to

⁷ Of course, the issues of missing variables and correct functional form are still at large with this model, so extensive diagnostic testing is required to validate the model.

⁸ As a benchmark example, we provide in Appendix A an example of a model that fails the out-of-sample forecasting. This model is based only on a simple linear relationship between the log sales and the log of the following covariates: advertising, size and leverage.

⁹ By “diagnostics” we refer always to heteroscedasticity, autocorrelation, and RESET tests.

Table 2
Parameter estimates-dynamic model.

	Coefficient	Z	P > z
Ln Sales (t-1)	0.465	2.98	0.003
Ln Sales (t-2)	-0.089	-1.33	0.184
Adv	0.0087	3.10	0.002
Adv2	-4.07×10^{-6}	-0.54	0.590
Adv3	2.39×10^{-9}	0.44	0.663
Size	-1.0759	-1.31	0.191
Size2	0.3016	2.25	0.025
Adv*Size	-0.0014	-1.96	0.050
Leverage	0.1138	1.57	0.116
_cons	4.6049	4.31	0.000

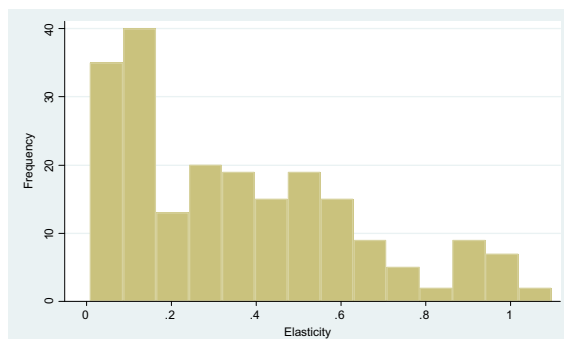


Fig. 2. Elasticity of advertising-dynamic model.

various forms of misspecification and the same is true for the RESET test -although it needs a simple modification in the case of panel data.

Although there is some confusion over specification tests versus residual diagnostics, we should point out that, for the most part, specification tests *are* residual diagnostics and, fortunately, such diagnostics are sensitive to specification problems. If a model does not pass the diagnostics we presented, then it cannot be used. If it does, then it can be subjected to further scrutiny. Our recommendation is to perform diagnostic testing, and subject the model to out-of-sample tests such as the one in Fig. 1.

Appendix A. Out-of-sample validation

This as an example of a simple linear model $y = f(x_1, x_2, x_3)$ that we estimate using the first 160 observations of our data, and we forecast the remaining 50 observations the results are shown in the following Fig. A.1. Evidently, the model overestimates the dependent variable relative to the actual value, so it cannot be acceptable in any practical sense of the term as the deviations between actual and predicted values are, apparently, systematic.

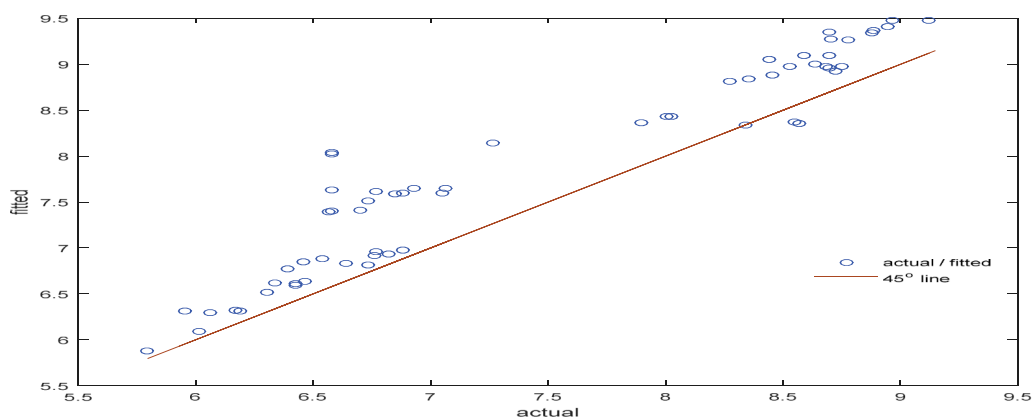


Fig. A.1. Actual and fitted values in out-of-sample validation for a poor fitting model.

Appendix B. Simulation experiments

Simulation experiment and specification tests

In this Appendix we use artificial data to examine how diagnostic tests perform under various forms of misspecification. We use Monte Carlo to see how various diagnostics perform in this context. We tried the performance of diagnostic tests for both cross-sectional and panel contexts but we confine our attention here to panel data due to space limitation¹⁰. The main message is that most diagnostics perform well. For example, if an omitted variable is heteroscedastic, the heteroscedasticity tests are able to detect different variance of the error terms. RESET tests are able to detect problems when the omitted variable is a nonlinear function of the included variables. Autocorrelation tests detect a specification problem when the omitted variable is serially correlated.

What is less known is how such tests perform when we have panel data where, additionally, we have to account for heterogeneity by using, for example, fixed effects. Diagnostic testing of panel data models is not well developed in econometrics and, certainly, such diagnostics are rarely if ever used in applied research. Our panel data model is:

$$y_{it} = \alpha_i + \beta_1 x_{it,1} + \beta_2 x_{it,2} + \beta_3 x_{it,3} + u_{it}, u_{it} \sim iidN(0, 0, 1^2), i = 1, \dots, n, t = 1, \dots, T,$$

where the true values are $\beta_1 = \beta_2 = \beta_3 = 1$. We simply set $\alpha_i = i$. Moreover, $n = 20$, $T = 5$, and

$$x_{it,1} \sim iidN(0, 1), x_{it,2} \sim iidN(0, 1)$$

We consider an omitted variable which has stochastic heteroscedasticity of the following forms:

$$\text{DGP 1: } x_{it,3} \sim N(0, \sigma_{it}^2), \ln \sigma_{it}^2 = x_{it,1} + x_{it,2} + e_{it}, e_{it} \sim N(0, 0, 1^2)$$

$$\text{DGP 2: } x_{it,3} \sim N(0, \sigma_{it}^2), \ln \sigma_{it}^2 = x_{it,1}x_{it,2} + e_{it}, e_{it} \sim N(0, 0, 1^2)$$

$$\text{DGP 3: } x_{it,3} \sim N(0, \sigma_{it}^2), \ln \sigma_{it}^2 = x_{it,1}^2 + x_{it,2}^2 + e_{it}, e_{it} \sim N(0, 0, 1^2)$$

$$\text{DGP 4: } x_{it,3} \sim N(0, \sigma_{it}^2), \ln \sigma_{it}^2 = |x_{it,1}| + |x_{it,2}| + e_{it}, e_{it} \sim N(0, 0, 1^2)$$

$$\text{DGP 5: } x_{it,3} \sim N(0, \sigma_{it}^2), \ln \sigma_{it}^2 = \sin(x_{it,1}) + \cos(x_{it,2}) + e_{it}, e_{it} \sim N(0, 0, 1^2)$$

The method of estimation is the standard fixed-effects estimator (also known as Least Squares with Dummy Variables, LSDV) and we apply the heteroscedasticity test to the squared residuals. The heteroscedasticity test, generally, is able to detect the misspecification but the RESET test cannot as we see from [Table A.1](#) below. For the heteroscedasticity test most p-values greater than 5% occur 13% of the time under DGP2. In DGP5, the test fails to detect a problem as 55.5% of the time p-values exceed 5%. On the other hand, the RESET test has p-values in excess of 5% almost all the time and only in DGP3 the exceedance occurs 27.3% of the time.

Next, we consider omitted variables related to x_1 and x_2 as follows:

Table A.1
Heteroscedasticity and RESET tests.

	Heteroscedasticity test p-values of F-test greater than 0.05	RESET test p-values of F-test greater than 0.05
DGP 1	2.30%	95.5%
DGP 2	13.0	95.5
DGP 3	0.00	27.30
DGP 4	10.90	97.80
DGP 5	55.50	100

$$\text{DGP 1: } x_{it,3} = x_{it,1}^2 + x_{it,2}^2 + e_{it}, e_{it} \sim iidN(0, 0, 1^2)$$

$$\text{DGP 2: } x_{it,3} = |x_{it,1}| + |x_{it,2}| + e_{it}, e_{it} \sim iidN(0, 0, 1^2)$$

$$\text{DGP 3: } x_{it,3} = x_{it,1}x_{it,2} + e_{it}, e_{it} \sim iidN(0, 0, 1^2)$$

$$\text{DGP 4: } x_{it,3} = \sin(x_{it,1}) + \cos(x_{it,2}) + e_{it}, e_{it} \sim iidN(0, 0, 1^2)$$

$$\text{DGP 5: } x_{it,3} = \exp(x_{it,1}) + \exp(-x_{it,2}^2) + e_{it}, e_{it} \sim iidN(0, 0, 1^2)$$

$$\text{DGP 6: } x_{it,3} = (x_{it,1} + x_{it,2})^2 + e_{it}, e_{it} \sim iidN(0, 0, 1^2)$$

The RESET test will fail to detect any problems, although the omitted variables are nonlinear functions of the regressors (details are omitted). As this is a significant problem, we propose to regress residuals on levels, squares and interactions of x_1 and x_2 . This is

¹⁰ The results from the cross-sectional simulations can be obtained from the authors upon request.

Table A.2

Heteroscedasticity, RESET and modified RESET tests.

	Heteroscedasticity test p-values of F-test greater than 0.05	RESET test p-values of F-test greater than 0.05	Modified RESET p-values of F-test greater than 0.05
DGP 1	0.000%	96.40%	0.200%
DGP 2	15.40	99.00	0.200
DGP 3	0.000	97.10	0.300
DGP 4	0.300	98.10	0.497
DGP 5	0.000	67.30	0.200
DGP 6	0.000	88.90	0.000

equivalent to considering the fitted values but excluding the fixed effects. From [Table A.2](#), the modified RESET test works now much better (and the heteroscedasticity test is also quite useful, as well). The p-values of the modified RESET test are almost never above 5% even under DGP2 where the p-values of the heteroscedasticity test can be larger than 5%, 15.4% of the time, on average. This shows that the modified RESET test we use in this paper will perform well in panel data, revealing specification problems when excluded regressors are nonlinear functions of the included explanatory variables.

References

- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2), 277–297.
- Assaf, A. G., Josiassen, A., Ratchford, B. T., & Barros, C. P. (2012). Internationalization and performance of retail firms: A Bayesian dynamic model. *Journal of Retailing*, 88(2), 191–205.
- Dumitrescu, E.-I., & Hurlin, C. (2012). Testing for Granger non-causality in heterogeneous panels. *Economic Modelling*, 29(4), 1450–1460.