Timothy, D., and S. Boyd
2003 Heritage Tourism. Harlow: Prentice Hall.
Tunbridge, J., and G. Ashworth
1996 Dissonant Heritage: The Management of the Past as a Resource in Conflict. Chichester: Wiley.
Walsh, K.
1992 The Representation of The Past: Museums and Heritage in the Post-Modern World. London: Routledge.

Yaniv Poria: Department of Hotel and Tourism Management, Ben-Gurion University of the Negev, Beer-Sheva, Israel. Email: <yporia@som.bgu.ac.il>

Submitted 11 November 2008. Resubmitted 27 January 2009. Final version 1 March 2009. Accepted 26 March 2009. Refereed anonymously. ***Coordinating Editor: Juergen Gnoth***

doi:10.1016/j.annals.2009.03.004

Annals of Tourism Research, Vol. 36, No. 3, pp. 525–532, 2009
0160-7383/$ - see front matter
Printed in Great Britain

On the Use of Model Averaging in Tourism Research

Alan T.K. Wan
Xinyu Zhang
City University of Hong Kong, Hong Kong

Econometric modeling now plays a vital role in tourism research. For example, forecasting is an essential element in the process of tourism planning. Quantitative tools are also commonly used to evaluate the economic impacts of tourism. The last several decades have seen a large increase in the number of published studies on tourism research using econometric and other quantitative approaches. Among the various econometric techniques, the most widely used is probably linear regression by least squares, which attempts to determine how variables of interest, such as tourism demand, relate to other socioeconomic factors such as economic growth, exchange rates and so on. While the statistical properties of the least squares technique are well-documented, the technique does not provide researchers with a means of specifying the model; there is almost always a list of potential explanatory variables to consider. Presumably it is safe to ignore university entrance scores in modeling tourist arrivals, but whether the ultraviolet index has any significant impact cannot be known with certainty before the study is undertaken. In practice, we often face data with an imperfectly specified model and learn of our imperfection from the data themselves. While data-based non-parametric modeling and optimization methods such as the genetic algorithm have gained some popularity among tourism researchers (e.g., Potter and Coshall 1988; Hurley, Moutinho, and Witt

1998; Chen and Wang 2007; Valdés, Torres and Domínguez 2007), it is fair to say that, due to ease of implementation and the analysis being easy to interpret, linear least squares regression remains the most widely used modeling method in tourism research, as it does in other social science disciplines.

When confronted by the problem of variable selection in a linear regression, one commonly proceeds by pretesting. For example, to determine the significance of the ultraviolet index in a model that seeks to explain tourist arrivals, a t test is conducted and the ultraviolet index is either retained or dropped accordingly. The final specification of the model depends on the outcome of pretesting, as do the estimates of all other coefficients in the model. The popular ''general-to-specific'' econometric modeling approach, which involves the formulation of a general model and the application of a testing down process, eliminating variables that are not significant, leading to a simpler specific model, also involves extensive use of pretest strategies. Specification search by pretesting is widely practiced in tourism research. Vanegas and Croes (2000), for example, searched for the preferred specification of a model that seeks to explain U.S. tourist arrivals in Aruba based on t tests of significance of the regressors, Song, Witt and Gang (2003) examined the demand for Thai tourism using the general-to-specific model selection approach, and Song and Witt (2003) applied the same approach to estimate a model that explains inbound tourism to Korea.

An issue often raised in criticism of the general-to-specific approach is the lack of understanding of the effects of the pretest strategies that come into play. Indeed, econometricians have known for a long time that pretest estimators, all being discontinuous functions of the data, possess rather poor sampling properties and distort the usual properties of the least squares estimator in the sense that the end results are not what they appear to be. See, for example, Danilov and Magnus (2004a). In applied studies, however, investigators typically report estimates and associated precision statistics (e.g., standard errors) as if the estimation had not been preceded by pretesting. Obviously, the reported precision statistics are incorrect. Reporting estimate precision should clearly account for the pretesting that has been integrated into the procedure.

Recent papers by Danilov and Magnus (2004a,b) attempted to do that, and they showed that ignoring pretesting in reporting the precision of the least squares estimator can lead to very substantial errors. Another widely practiced method closely related to common pretesting procedures is model selection by schemes such as the AIC and BIC (the Akaike and the Bayesian information criteria); these methods are routinely applied, sometimes along with stepwise regressions. Again, investigators using these procedures typically proceed as if the final model had been decided in advance, without acknowledging the additional uncertainty introduced by model selection. There is also a growing collection of literature that discusses the effects of model selection on inference. See, for instance, Leeb and Pötscher (2005).

An alternative approach to pretesting and model selection is model averaging, where one averages across least squares estimates obtained from different models, rather than using only one model arrived at by pretesting or a model selection criterion. Model averaging has long been a popular technique among Bayesian statisticians. See Hoeting, Madigan, Raftery and Volinsky (1999) for a non-technical discussion. In the econometrics literature, several methods for implementing model averaging in the context of linear regression have recently emerged. Magnus (2002) and Danilov (2005) advocated a weighted average least squares (WALS) estimator with weights based on a Laplace prior density. The WALS estimator assumes a set of ''key'' explanatory variables which the investigator wants in the model on theoretical or other grounds irrespective of statistical significance, as well as another set of ''auxiliary'' explanatory variables of which the investigator is less certain. The role of the auxiliary explanatory variables is primarily to improve the estimation

of the coefficients of the key explanatory variables. Unlike the pretest estimator, the resultant WALS estimator is a continuous function of the data and is admissible. More recently, Hansen (2007, 2008) proposed a model average estimator with weights selected by minimizing the Mallows criterion. One advantage of model averaging over model selection is that it pays due attention to the problem of model uncertainty. Thus, model averaging reconciles the disparities caused by pretesting and model selection, and will undoubtedly be used more extensively in the future. The latest edition of the New Palgrave Dictionary of Economics, for example, includes a chapter on model averaging (see, Doppelhofer, 2008).

Model averaging tools have been applied widely in biostatistics, and have recently found applications in economics, finance and sociology. See, for example, Raftery (1995), Levin and Williams (2003), Sala-i-Martin, Doppelhofer and Miller (2004), Bird and Gerlach (2006), Doppelhofer and Weeks (2009) and Magnus, Powell and Prüfer (2008a). There is considerable appeal for using model averaging in other social science disciplines, and this technical note is meant to broaden its appeal by using the application in the context of tourism research. Another purpose of this note is to alert tourism researchers to the dangers of pretesting and model selection in terms of underreporting the variability of estimates by way of a practical example. Our examination of model averaging centers on the WALS estimator. More computationally intensive methods can be employed but the WALS estimator has the distinct advantage of being simple and is readily applicable without an undue computational burden; the calculation of the WALS estimates does not require computationally intensive methods such as the Markov Chain Monte Carlo techniques which are common with Bayesian model averaging. Moreover, appropriate formulae for computing the standard errors of the WALS estimates are available, while the practical application of some other model average estimators has been limited to some degree by the fact that they produce only point estimates.

Our practical example is taken from a recent paper by Reeder and Brown (2005), who used linear regression to determine the degree to which recreation and the development of tourism affected a range of socioeconomic indicators (e.g., earnings per job, income per capita, rent levels, death rate, etc.) in 311 rural U.S. counties during the 90s and 2000. Corresponding to each of the socioeconomic indicators is a multiple regression model with the socioeconomic indicator as the dependent variable. The key explanatory variable of the regressions is dependency on recreation and tourism as measured by a so-called Z-score. The Z-score, as developed by Johnson and Beale (2002), covers tourism-related employment and income shares of the local economy, as well as the share of total county homes dependent on recreational use. The higher the Z-score the more dependent a county is on recreation and tourism. Reeder and Brown (2005) were primarily interested in the coefficient estimate of the Z-score, though they also included 21 other explanatory variables such as dummy variables on county types and regional subdivisions and a range of demographic variables in each regression. We will reconsider the regression models formulated by Reeder and Brown (2005), not to re-evaluate the conclusions reached by the authors, but to analyze the effect of pretesting, highlighting the merits of model averaging as an alternative to model selection.

We have performed analyses based on data of both 90s and 2000, but in the interest of brevity we only present the results for 2000 here. The general conclusions are the same with the data of 90s and the results are available upon request. Table 1 lists the 16 dependent and 22 explanatory variables used in Reeder and Brown's (2005) study. Our analysis treats the intercept term and the Z-score as the key explanatory variables and all others as auxiliary explanatory variables. In illustrating the effects of pretesting we adopt a stepwise selection procedure that begins like backward elimination. However, after two or more variables have been removed from the model, a forward selection procedure is employed to allow vari-

Table 1. List of Variables

Dependent variables	Explanatory variables (excluding intercept)
1. Employment-population ratio of ages 16-24	1. Z-score
2. Employment-population ratio of ages 25-64	2-9. Eight dummy variables identifying the Census regional subdivisions
4. Employment-population ratio of ages 65 and over	10-19. Ten dummy variables reflecting Johnson and Beale's (2002) recreation county types
8. Earnings per job	
9. Income per capita	
10. Median rent	20. Dummy variable indicating whether the county is influenced by a nearby metropolitan area
11. Population growth rate	21. Percentage of county population residing in rural areas
12. Travel time to work	22. County population density
13. Percentage of population without HS diploma	
14. Percentage of population with bachelor's degree	
15. Physicians per 100,000 population	
16. Age-adjusted death rate per 100,000 population	

ables that have been eliminated to be reconsidered for inclusion. The procedure continues until no additions or deletions of variables are indicated. In our application we choose the significance level for adding a variable to be 0.05 and for removing a variable to be 0.10. Note that the significance level set for entering a variable should always be smaller than that for removing a variable otherwise cycling is possible where a variable is continually entered and removed. Stepwise selection is a popular model selection procedure and automated routines for this procedure are available in most statistical software packages like SAS or Matlab. The estimation results for the coefficient of the Z-score are shown in Table 2. Column 2 in Table 2 gives the estimates of the coefficients of the Z-score based on the model selected by stepwise selection for each of the 16 regressions. The third column gives the 95% confidence bounds of the coefficients when pretesting is not taken into account. These are the confidence bounds usually reported in applied work when the researcher assumes (erroneously) that the model has been chosen in advance. The numbers in Column 4, on the other hand, are the ''correct'' 95% confidence bounds when one pays due attention to the effect of stepwise selection on the variability of the estimates. The formulae for computing the correct confidence bounds are available from Danilov and Magnus (2004b). We see that in all cases the commonly reported confidence bounds that ignore pretesting underreport the true confidence bounds. In some cases the difference between the reported and the correct confidence bounds can be very large. In the worst case, the true confidence bounds are 10.87 times as wide as the bounds that ignore pretesting; on average they are 2.95 times as wide. The correct confidence bounds are typically much wider than is apparent and the reported bounds are far too optimistic.

The WALS coefficient estimates and the 95% confidence bounds appear in Columns 5 and 6 respectively. The formulae for computing the WALS confidence

Table 2. Estimates of Coefficient of Z-score in Each Regression and Confidence Bounds

Model[1]	Pre-test estimate	Reported Pre-test confidence bounds	Correct Pre-test confidence bounds	WALS estimate	WALS confidence bounds
1	1.19	(0.50, 1.87)	(−0.68, 2.96)	1.19	(0.48, 1.90)
2	0.89	(0.20, 1.59)	(−0.76, 2.60)	1.00	(0.29, 1.71)
3	1.01	(0.56, 1.46)	(0.08, 2.01)	1.13	(0.67, 1.59)
4	1.63	(−340.92, 344.18)	(−456.81, 440.90)	37.8	(−303.22, 378.82)
5	880.47	(514.87, 1246.07)	(−93.26, 1786.26)	898.9	(520.14, 1277.66)
6	1073.54	(558.91, 1588.18)	(−222.67, 2311.72)	1077.42	(550.79, 1604.05)
7	1538.95	(965.59, 2112.31)	(431.10, 2511.70)	1563.95	(992.77, 2135.14)
8	32.83	(23.84, 41.82)	(16.47, 48.72)	34.28	(25.29, 43.27)
9	4.48	(2.89, 6.07)	(−0.64, 9.82)	4.84	(3.11, 6.58)
10	−0.27	(−0.67, 0.13)	(−1.58, 1.07)	−0.26	(−0.69, 0.17)
11	−0.84	(−1.34, −0.33)	(−2.20, 0.50)	−0.81	(−1.33, −0.30)
12	−1.42	(−1.93, −0.90)	(−2.62, −0.13)	−1.41	(−1.94, −0.88)
13	2.22	(1.59, 2.85)	(1.14, 3.36)	2.36	(1.73, 2.99)
14	1.59	(−7.49, 10.68)	(−21.42, 22.82)	1.32	(−7.88, 10.52)
15	−25.88	(−37.60, −14.16)	(−62.03, 13.62)	−25.45	(−37.93, −12.97)
16	0.67	(0.50, 0.83)	(−1.05, 2.51)	0.65	(0.47, 0.82)

[1] Models 1-16 are based on, respectively, dependent variables 1-16 listed in Table 1.

bounds are available in Magnus et al (2008a). In all cases the WALS and pretest coefficient estimates have the same sign, and the difference in magnitude between the two estimates is usually not large. However, without exception the WALS estimates produce confidence bounds with a decreased width from the (true) pretest confidence bounds. On average the WALS confidence bounds are 42.84% as wide as the correct pretest confidence bounds; thus notable reductions in estimator variability are achievable with the WALS approach. This is expected because model averaging usually leads to estimates that are of superior precision than those achieved by selecting a single model, as demonstrated in the theoretical literature. While these results are, of course, specific to the data example considered here, the evidence does provide an indication of the performance gains that are possible over a range of models involving tourism data. The WALS estimator is very easy to implement—the steps involved in implementing the WALS estimator and the Matlab codes written to produce the estimates are available online at http://fbstaff.cityu.edu.hk/msawan/research1.htm. Admittedly, one disadvantage of WALS is that it is not strictly suitable outside the standard linear regression context; also, the optimality properties established for the WALS estimator do not apply to time series models such as ARIMA or transfer function models which are also common tools for tourism research. These issues are currently being addressed; for example, work in process by Magnus, Wan and Zhang (2008b) is developing a variant of the WALS estimator for models with autocorrelated or heteroscedastic errors.

Finally the following illustration sheds some light on the advantage afforded by model averaging in forecasting with common time series models. This is highly relevant in light of the large (and growing) tourism forecasting literature. Our illustration uses data on the number of long-stay visitor arrivals in Barbados between 1956 and 1992 given in Dharmaratne (1995). The author estimated two Box-Jenkins ARIMA models, namely, ARIMA(2,1,1) and ARIMA(2,1,1)$(1,1,1)_5$, with data from 1956 to 1987, and based on forecasted values generated by these two models for the remaining years he concluded that the ARIMA(2,1,1) specification is bet-

Table 3. Forecast Comparisons

Period	Actual Value	Forecasts (absolute percentage errors in brackets)		
		ARIMA(2,1,1)	ARIMA(2,1,1)(1,1,1)$_5$	Averaged model
1988	451443	453721 (0.50%)	409067 (9.39%)	451980 (0.12%)
1989	461259	463343 (0.45%)	367310 (20.37%)	459600 (0.36%)
1990	432092	462613 (7.06%)	348448 (19.36%)	458163 (6.03%)
1991	394222	464243 (17.76%)	298296 (24.33%)	457775 (16.12%)
1992	385979	473952 (22.79%)	302413 (21.65%)	467266 (21.06%)

ter. Our purpose is to demonstrate that substantial gains in forecast accuracy can be achieved by compromising across the two models. WALS is not applicable here because the models involved do not fall within the framework of linear regression. Our subsequent analysis is based on another well-known model weighting scheme, namely, the Smooth AIC weight introduced in Buckland, Burnham and Augustin (1997). This weight is proportional to the value of exp(-AIC$_S$ $*$0.5), where AIC$_S$ is the AIC score for candidate model S. Dharmaratne (1995) reported the coefficient estimates and AIC scores for both models; for the ARIMA(2,1,1) model, the AIC score is 430.26 while for ARIMA(2,1,1)(1,1,1)$_5$, it is 436.67. While these two AIC values are very comparable, they still point to a preference for ARIMA(2,1,1). We now apply model averaging using the Smooth AIC scheme and generate forecasts for 1988 to 1992. Table 3 presents the forecasts and their absolute percentage errors. The forecast performance of the ARIMA(2,1,1)(1,1,1)$_5$ model is rather poor—except for 1992 its predictions are always worse, and usually by a large margin, than those based on the ARIMA(2,1,1) model. The forecast performance of the averaged model is quite remarkable—in all cases under consideration the forecasts obtained from the averaged model are closer to the true values than those obtained from the better of the two single models. The averaged model yields a mean absolute percentage forecast error of 8.74%, while the corresponding figures for the ARIMA(2,1,1) and ARIMA(2,1,1)(1,1,1)$_5$ models are 9.71% and 19.02% respectively.

The techniques that have been illustrated are just two of the available model averaging techniques and one can adopt other more complex combining techniques in practice. Nonetheless we view the results presented here as being very promising. Certainly, further exploration of the model averaging approach in tourism research seems to be justified. A

Acknowledgements—The authors thank Richard Reeder and Dennis Brown of the U.S. Department of Agriculture for supplying the data used in this study, and the referees, editor and associate editor for comments and suggestions.

REFERENCES

Bird, R., and R. Gerlach
2006 A Bayesian Model Averaging Approach to Enhance Value Investment. International Journal of Business and Economics 5:93–110.

Buckland, S. T., K. P. Burnham, and N. H. Augustin
1997 Model Selection: An Integral Part of Inference. Biometrics 53:603–618.

Chen, K. Y., and C. H. Wang
2007 Support Vector Regression with Genetic Algorithms in Forecasting Tourist Demand. Tourism Management 28:215–226.

Danilov, D.
2005 Estimation of the Mean of a Univariate Normal Distribution When the Variance is Not Known. Econometrics Journal 8:277–291.
Danilov, D., and J. R. Magnus
2004a On the Harm that Ignoring Pretesting Can Cause. Journal of Econometrics 122:27–46.
2004b Forecast Accuracy after Pretesting with an Application to the Stock Market. Journal of Forecasting 23:251–274.
Dharmaratne, G. S.
1995 Forecasting Tourist Arrivals in Barbados. Annals of Tourism Research 22:804–818.
Doppelhofer, G.
2008 Model Averaging. *In* The New Palgrave Dictionary of Economics, L. Blume and S. Durlauf, eds., (2nd ed.). London: Palgrave Macmillan.
Doppelhofer, G., and M. Weeks
2009 Jointness of Growth Determinants. Journal of Applied Econometrics 24:209–244.
Hansen, B. E.
2007 Least Squares Model Averaging. Econometrica 75:1175–1189.
2008 Least Squares Forecast Averaging. Journal of Econometrics 146:342–350.
Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky
1999 Bayesian Model Averaging: A Tutorial. Statistical Science 14:382–417.
Hurley, S., L. Moutinho, and S. F. Witt
1998 Genetic Algorithms for Tourism Marketing. Annals of Tourism Research 25:498–514.
Johnson, K. M., and C. L. Beale
2002 Nonmetro Recreation Counties: Their Identification and Rapid Growth. Rural America 17:12–19.
Leeb, H., and B. M. Pötscher
2005 Model Selection and Inference: Facts and Fiction. Econometric Theory 7:163–185.
Levin, A. T., and J. C. Williams
2003 Robust Monetary Policy with Competing Reference Models. Journal of Monetary Economics 50:947–975.
Magnus, J. R.
2002 Estimation of the Mean of a Univariate Normal Distribution with Known Variance. Econometrics Journal 5:225–236.
Magnus, J.R., O. Powell, and P. Prüfer
2008a A Comparison of Two Averaging Techniques with an Application to Growth Empirics, mimeo, Department of Econometrics and Operations Research, Tilburg University, the Netherlands.
Magnus, J.R., A.T.K. Wan, and X. Zhang
2008b WALS Estimation with Non-Spherical Disturbances and an Application to the Hong Kong Housing Market, mimeo, Department of Econometrics and Operations Research, Tilburg University, the Netherlands.
Potter, R. B., and J. Coshall
1988 Sociopsychological Methods for Tourism Research. Annals of Tourism Research 15:63–75.
Raftery, A. E.
1995 Bayesian Model Selection in Social Research. Sociological Methodology 25:111–163.
Reeder, R.J., and D.M. Brown
2005 Recreation, Tourism and Rural Well-Being. Economic Report Number 7, Economic Research Service, U.S. Department of Agriculture. (Paper available online at http://www.ers.usda.gov/Publications/ERR7/)
Sala-i-Martin, X., G. Doppelhofer, and R. M. Miller
2004 Determinants of Economic Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. American Economic Review 94:813–835.
Song, H., and S. F. Witt
2003 Tourism Forecasting: The General to Specific Approach. Journal of Travel Research 42:65–74.

Song, H., S. F. Witt, and L. Gang
2003 Modeling and Forecasting the Demand for Thai Tourism. Tourism Economics 9:363–387.
Valdés, L., E. Torres, and J. S. Domínguez
2007 A Model to Study the Economic Impact of Collective Accommodation in a Region. Tourism Management 28:152–161.
Vanegas, M., and R. R. Croes
2000 Evaluation of Demand: US Tourists to Aruba. Annals of Tourism Research 27:946–963.

Alan T.K. Wan: Department of Management Sciences, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong. Email: <msawan@cityu.edu.hk>

Submitted 18 July 2008. Resubmitted 26 November 2008. Resubmitted 11 February 2009. Final version 3 March 2009. Accepted 26 March 2009. Refereed anonymously. ***Coordinating Editor: Juergen Gnoth***

doi:10.1016/j.annals.2009.03.004