JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **19**: 533–535 (2004)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.777

PREDICTOR RELEVANCE AND EXTRAMARITAL AFFAIRS

QI LI[a] AND JEFF RACINE[b]*

[a] *Department of Economics, Texas A&M University, USA*
[b] *Department of Economics and Center for Policy Research, Syracuse University, USA*

SUMMARY

We revisit Fair's (1978) 'theory of extramarital affairs' using robust nonparametric methods developed for the analysis of categorical data. We find evidence suggesting that the number of years married is *not* a relevant predictor of the propensity to engage in extramarital affairs having controlled for other factors. This finding runs counter to the prevailing wisdom gleaned from misspecified parametric models. Copyright © 2004 John Wiley & Sons, Ltd.

In a widely cited paper, Fair (1978) proposes a 'theory of extramarital affairs' and considers the allocation of an individual's time among work and two types of leisure activities: time spent with one's spouse and time spent with one's 'paramour'. This unique dataset is taken from two magazine surveys, and Fair (1978) considers a parametric tobit estimator for modelling the number of extramarital affairs per year. This dataset and the econometric methodology continue to generate controversy, as witnessed by Wells (2003) and references therein.

One of the rather fascinating aspects of this study is the potential influence of an individual's personal characteristics on their propensity to engage in extracurricular activities. In particular, Fair (1978), as well as the more in-depth analysis undertaken by Wells (2003), found that infidelity increased significantly with the number of years married.

The data consist of 601 observations on nine covariates, Sex (0/1), Age (nine groups), Years married (eight groups), Children (0/1), How religious (1–5), Level of education (seven groups), Occupation (seven groups), Marriage rating (1–5), and number of times subject engaged in extramarital sexual intercourse during the past year. All variables are categorical in nature. The tobit specification employed by Fair (1978) has a positive and significant parameter associated with number of years married, as does a Poisson count model, simple linear model, ordered probit and ordered logit model. However, each of these models share a common feature—a linear index specification. In addition, the discrete and unordered nature of some of the covariates calls into question the appropriateness of such specifications, as does the lack of interaction terms.

* Correspondence to: Jeff Racine, Department of Economics, Eggero 426, Syracuse University, Syracuse, NY 13244, USA.
E-mail: jracine@maxwell.syr.edu

Received 17 November 2003
Revised 6 January 2004

Nonparametric methods, being robust to misspecification, may add value in this setting. The traditional nonparametric frequency approach for handling discrete covariates requires the number of observations to be much larger than the number of discrete cells; for the data at hand, this is clearly not the case. In light of this, we use the recently proposed kernel estimator of Hall *et al*. (2004), Li and Racine (2003) and Racine and Li (2004) who suggest smoothing the discrete covariates. One of the features of Hall *et al*.'s (2003) approach is its ability to automatically remove irrelevant predictors. That is, when an explanatory covariate is in fact an irrelevant predictor, the data-driven cross-validation method they propose is shown to select a large value for the covariate's smoothing parameter; consequently, this irrelevant predictor will be automatically removed. We treat Sex, Children and Occupation as unordered categorical covariates and the remaining covariates as ordered.

In Table I, the 'Upper bound' column denotes the maximal admissible value for λ. We observe from Table I that the cross-validated smoothing parameter ($\hat{\lambda}$) associated with the number of years married coincides with its upper bound, which implies that number of years married is *not* a relevant predictor of the number of affairs having controlled for other covariates. In fact, the only covariates that appear to be relevant from a predictive standpoint are age, how religious a respondent claims to be and, not surprisingly, how they rate their marriage. All others are deemed to be irrelevant and are automatically removed from the nonparametric conditional probability estimator.

To better understand why nonparametric and parametric methods yield differing results, we consider the nonparametric estimation results in more detail. Figure 1 plots the expected number of affairs conditional upon age from the nonparametric model when all remaining covariates are held constant at their median values. We observe that the relationship between age and number of affairs is nonlinear, being flat and/or upward sloping for younger ages, and exhibiting a downward trend for people over 40. Thus, the nonparametric results suggest that age has a nonlinear effect on the number of affairs, while the number of years of marriage is independent of the number of affairs conditional on age and other relevant covariates. Years of marriage and age are highly correlated covariates, and simple parametric models assume that each covariate has only either a globally positive or negative effect. When the true relationship is nonlinear, say in age, age has a positive effect over one range and a negative effect over another. It appears that the parametric models erroneously attribute a negative effect to the age covariate, and in an attempt to model nonlinearity in age with a linear index function (the upward sloping portion for younger ages), the parametric model picks an irrelevant covariate (which is highly correlated with young age) and assigns a significant positive sign to it.

The above analysis is based on the assumption that the nonparametric approach is more faithful to the data than the parametric specifications. In order to assess whether this is indeed the case, we consider the in-sample predictions for the kernel method and best-performing parametric model

Table I. Cross-validation selected smoothing parameters

Covariate	$\hat{\lambda}$	Upper bound	Covariate	$\hat{\lambda}$	Upper bound
Sex	0.500	0.500	Years married	1.000	1.000
Children	0.500	0.500	Religious	0.290	1.000
Occupation	0.857	0.857	Education	1.000	1.000
Age	0.886	1.000	Marriage rating	0.329	1.000

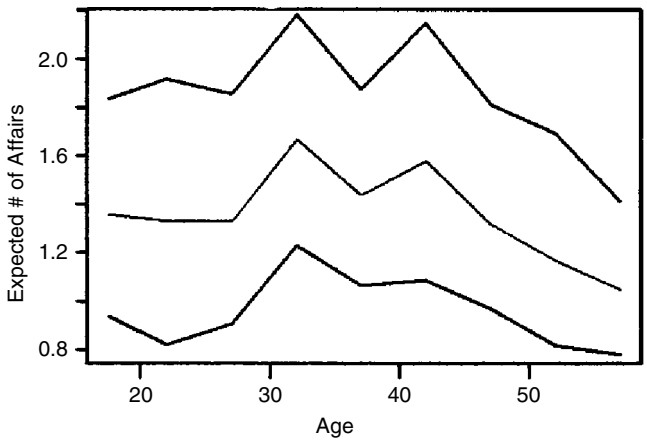

Figure 1. The expected number of affairs conditional on age, remaining covariates held constant at their median values, along with resampled 10th and 90th quantiles

(ordered logit specification). The nonparametric method and the ordered logit give 79.2% and 75.0% correct predictions for the number of affairs, respectively. Also, for the 150 people recording a positive number of affairs, the nonparametric method correctly predicts the number of affairs for 25 of them (16.7%), where the ordered logit model correctly predicts those for only three (2.0%). The prediction results indicate that the parametric models are misspecified.

The extramarital affairs data studied by Fair (1978) is a rich dataset that continues to generate controversy. Existing work has focused mainly on specification tests based on *parametric* linear single index models. The possible nonlinearity of relevant covariates is largely ignored. We focus instead on the impact of parametric misspecification on assessing predictor relevance, and conclude that one covariate long thought to be a 'significant' predictor of affairs, the number of years married, does not in fact appear to be associated with an increase in one's propensity to engage in these types of extracurricular activities.

REFERENCES

Fair RC. 1978. A theory of extramarital affairs. *Journal of Political Economy* **86**(1): 45–61.

Hall P, Racine J, Li Q. 2004. Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association (tentatively accepted).*

Li Q, Racine J. 2003. Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis* **86**: 266–292.

Racine J, Li Q. 2004. Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics* **119**(1): 99–130.

Wells C. 2003. Retesting Fair's (1978) model on infidelity. *Journal of Applied Econometrics* **18**: 237–239.

