Pertanika J. Sci. & Technol. 25 (S): 151 - 162 (2017)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Nonlinear Regression in Tax Evasion with Uncertainty: a Variational Approach

Mohamad Mobasher-Kashani[1]*, Masri Ayob[1], Azuraliza Abu Bakar[1], Razieh Tanabandeh[2], Kourosh Taheri[3] and Mohammad Hassan Tayarani Najaran[4]

[1]Datamining and Optimization Research Group, Center for Artificial Intelligence Technolgy, Universiti Kebangsaan Malaysia, 46300 Bangi, Selangor, Malaysia
[2]School of Economics, Faculty of Economics and Management, Universiti Kebangsaan Malaysia, 46300 Bangi, Selangor, Malaysia
[3]Faculty of Computer Science, Khayyam University of Mashhad, Mashhad, Iran
[4]School of Computer Science, University of Glasgow, Glasgow, United Kingdom

ABSTRACT

One of the major problems in today's economy is the phenomenon of tax evasion. The linear regression method is a solution to find a formula to investigate the effect of each variable in the final tax evasion rate. Since the tax evasion data in this study has a great degree of uncertainty and the relationship between variables is nonlinear, Bayesian method is used to address the uncertainty along with 6 nonlinear basis functions to tackle the nonlinearity problem. Furthermore, variational method is applied on Bayesian linear regression in tax evasion data to approximate the model evidence in Bayesian method. The dataset is collected from tax evasion in Malaysia in period from 1963 to 2013 with 8 input variables. Results from variational method are compared with Maximum Likelihood Estimation technique on Bayeisan linear regression and variational method provides more accurate prediction. This study suggests that, in order to reduce the tax evasion, Malaysian government should decrease direct tax and taxpayer income and increase indirect tax and government regulation variables by 5% in the small amount of changes (10%-30%) and reduce direct tax and income on taxpayer and increment indirect tax and government regulation variables by 90% in the large amount of changes (70%-90%) with respect to the current situation to reduce the final tax evasion rate.

Keywords: Bayesian inference, Linear regression, Nonlinear problem, Tax evasion, Uncertainty, Variational approximation

ARTICLE INFO

Article history:
Received: 15 August 2016
Accepted: 18 May 2017

E-mail addresses:
mohamadnet@gmail.com (Mohamad Mobasher-Kashani),
masri@ukm.edu.my (Masri Ayob),
azuraliza@ukm.edu.my (Azuraliza Abu Bakar),
razieh_tabandeh@yahoo.com (Razieh Tanabandeh),
taheri@ferdowsi.um.ac.ir (Kourosh Taheri),
Mohammad.tayarani@glasgow.ac.uk
(Mohammad Hassan Tayarani Najaran)
*Corresponding Author

ISSN: 0128-7680

INTRODUCTION

Tax evasion is one of the most challenging issue facing governments. Failure to tackle all aspects of tax evasion in their respective countries can give rise to numerous social problems. Several researches have been done on the tax evasion problem in Malaysia. Tax evasion in Malaysia is studied from different aspects (Choong & Lai, 2008; Jaffar, Bakar, & Tahir, 2011; Kasipillai, Aripin, & Amran, 2003; Kasipillai, Baldry, & Prasada Rao, 2000; Miskam, Noor, Omar, & Aziz, 2013; Tabandeh & Tamadonnejad, 2015). Although Tabandeh and Tamadonnejad (2015) suggest a regression model for tax evasion data in Malaysia, there is also the need to consider uncertainty and nonlinearity in tax evasion data in Malaysia.

Bayesian inference methods have various features where the major one is the uniform way they handle uncertainties in data through the modelling process (Jaakkola & Jordan, 2000). The procedure also provides a monolithic combination of prior knowledge and data observation to infer the posterior knowledge (Bernardo & Smith, 2009; Gelman, Carlin, Stern, & Rubin, 2014). Byesian approaches in practice are intractable even for simple applications; Bayesian inference methods take two major approaches to handle this problem. They either sample from the exact solution, e.g. Markov Chain Monte Carlo (or MCMC) approaches, or approximate, e.g. variational approximation. Although MCMC approaches in general and the Gibbs sampler tech-nique in particular have gained a widespread popularity as tools for modelling complex systems, there are major disadvantages from practical end-user perspective.

- MCMC sampling methods can be computationally expensive, even for rather small-scale statistical problems (Barber, 2012). This is aggravated for large-scale datasets where supplementary hardware might be needed.

- Real-time data assimilation adds a further complication to MCMC methods. An intuitive way to deals with this problem is to restart and run the method when new sample or batch of samples arrives. An alternative solution is to adopt a sequential method, but this presents a further challenge to the inference procedure.

- Determining the length of "pre-convergence" (Hjort, Holmes, Müller, & Walker, 2010) and deciding when it is safe to stop the sampling (Winn & Bishop, 2005) pose further challenges to MCMC which cause great difficulties in implementing the method. Moreover, it is reasonable here to question whether samples are drawn correspond to the distribution of the Markov chain (Cowles & Carlin, 1996).

Attias (2000) suggests the variational Bayes (VB) approach which facilitates analytical calculations of the posterior distributions over a model. The proposed method uses the mean field approximation theory by adopting a factorized approximation to the true posterior distribution, although in contrast to the Laplace approximation these factorized posteriors are not limited to a Gaussian form. Jan Drugowitsch (2013) has applied automatic relevance determination (ARD) on linear and logistic regression for a randomly generated dataset. In

this study, we follow the presentation of (J Drugowitsch, 2008; Jan Drugowitsch, 2013) for variational regression formula to solve the problem of tax evasion regression in Malaysia.

In the domain of economics, measuring uncertainty is a challenging issue; hence Bayesian methods have been applied on a variety of economical applications (Jackson, 1991; Palfrey & Srivastava, 1987; Punt & Hilborn, 1997; Sun & Shenoy, 2007; Tan & da Costa Werlang, 1988). However, uncertainty in the tax evasion prediction has not been addressed. This study deals with the uncertainty that exists in tax evasion dataset in Malaysia by using variational linear regression (VLR) approach with ARD technique. Furthermore, by examining the performance of 6 different basis functions, nonlinearity of Malaysia tax evasion dataset is studied. This paper is organized as follows. In section II variational linear regression technique is discussed and the basis functions and case study used in this research are introduced. Section III is dedicated to comparing VLR with ARD on 6 basis functions and between VLR and MLE method on Malaysia dataset. The final section concludes the paper.

MATERIALS AND METHODS

Linear regression model is widely used in many practical applications and is employed to deduce the trend. The main idea behind the linear regression is to find a linear combination of input variables that fits the most to the output variable. In this sense, the following equation de-fines simple linear regression model,

$$y(X) = \sum_{j=1}^{D} w_j x_j + \varepsilon \quad , \tag{1}$$

where w is the weight of input variable, which is always linear; x is the input variable, which can be a nonlinear function (i.e. basis function) of the input variable; and D is the dimension of input variable x; finally, ε is the residual error between the true response and the predicted values and should be minimized by adopted model. The parameters of linear regression mould can be estimated by VB in an efficient way which the final method is called variational linear regression or VLR.

A system can be formalized by defining on certain parameters which is called system modelling. Suppose these parameters are defined by θ then the prime issue is to obtain the distribution that governs these parameters. A central task in the Bayesian applications is to calculate the posterior distribution in a way that we are able to infer extra information from it. In this regard, uncertainty can be addressed using the shape of posterior distribution; the narrower the posterior distribution is around the mean, the more the method is certain about the final result. However, the posterior distribution proved to be intractable and the most computationally expensive part of the posterior distribution is the model evidence $P(\mathrm{D})$. The variational Bayes or VB method approximates posterior distribution by factorizing parameters, $P(\theta \mid \mathrm{D}) \approx \prod_k \mathrm{Q}(\theta_k)$ and estimating distribution for each parameter. The variational approximation method attempts to evaluate the posterior distribution $P(w,\tau,\alpha \mid \mathrm{D})$ by factorizing into $Q(\mathrm{w},\tau)\mathrm{Q}(\alpha)$. Hence, the lower bound of model evidence is estimated by:

$$\mathrm{L}(Q) = \iiint Q(w,\tau,\alpha) ln \frac{P(Y|\Phi_X, w,\tau)P(w,\tau|\alpha)P(\alpha)}{Q(w,\tau,\alpha)} dw d\tau d\alpha \leq \ln P(\mathrm{D}), \tag{2}$$

The first part of our modelling is to indicate the prior form which consists of Gaussian distribution over weights and two Gamma distributions for calculating the precision.

$$P(\mathrm{w},\tau,\alpha) = \mathrm{N}\left(w|0,(\tau\alpha)^{-1}I\right)Gam(\tau|a_0,b_0)Gam(\alpha|c_0,d_0), \tag{3}$$

Since α is hyper-prior in our model, no analytic solution exists to the posterior distribution. The optimal form of posterior is approximated as,

$$Q^*(w,\tau,\alpha) = \mathrm{N}\left(w|w_N,\tau^{-1}V_N\right)Gam(\tau\,|\,a_N,b_N)Gam(\alpha\,|\,c_N,d_N), \tag{4}$$

The variational bound is formed is given by

$$\begin{aligned}&\mathrm{L}(Q) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\sum_n\left(\frac{a_N}{b_N}\left(y_n - w_N^T\varphi_{x_n}\right)^2 + \varphi_{x_n}^T V_N\varphi_{x_n}\right)\\&+\frac{1}{2}\ln|V_N| + \frac{D}{2} - \ln\Gamma(a_0) + a_0\ln b_0 - b_0\frac{a_N}{b_N} + \ln\Gamma(a_N)\\&-a_N\ln b_N + a_N - \ln\Gamma(c_0) + c_0\ln d_0 + \ln\Gamma(c_N) - c_N\ln d_N,\end{aligned} \tag{5}$$

In order to change the method to ARD, the first step is to modify priors and then the rest of the procedure will follow the changes. As seen in equation (6), ARD method alters the Gamma distribution over α,

$$P(w,\tau,\alpha) = \mathrm{N}\left(w|0,(\tau A)^{-1}\right)Gam(\tau|a_0,b_0)\prod_i Gam(\alpha_i|c_0,d_0), \tag{6}$$

where the optimal variational factorization for the approximation of posterior distribution consists of:

$$Q^*(w,\tau,\alpha) = \mathrm{N}\left(w|w_N,\tau^{-1}V_N\right)Gam(\tau|a_N,b_N)\prod_i Gam(\alpha_i\,|\,c_N,d_{Ni}), \tag{7}$$

Consequently, based on Jan Drugowitsch (2013), the equation (5) for ARD changes to:

$$\begin{aligned}&\mathrm{L}(Q) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\sum_n\left(\frac{a_N}{b_N}\left(y_n - w_N^T\varphi_{x_n}\right)^2 + \varphi_{x_n}^T V_N\varphi_{x_n}\right)\\&+\frac{1}{2}\ln|V_N| + \frac{D}{2} - \ln\Gamma(a_0) + a_0\ln b_0 - b_0\frac{a_N}{b_N} + \ln\Gamma(a_N)\\&-a_N\ln b_N + a_N + \sum_i\left(-\ln\Gamma(c_0) + c_0\ln d_0 + \ln\Gamma(c_N) - c_N\ln d_{Ni}\right),\end{aligned} \tag{8}$$

The predictive distribution is evaluated in the form of a Student distribution using variational distribution $Q(w,\tau)$ as an approximation to the posterior $P(w,\tau|\mathrm{D})$ which involves uncertainty of the method and results in the following equation,

$$\begin{aligned}&P(y|x,\mathrm{D}) = \iint P(y|\varphi_x,w,\tau)P(w,\tau|\mathrm{D})\,dw\,d\tau \approx\\&\iint P(y|\varphi_x,w,\tau)Q(w,\tau)\,dw\,d\tau = St(y\,|\,w_N^T x,\left(1+\varphi_x^T V_N\varphi_x\right)^{-1}\frac{a_N}{b_N},2a_N),\end{aligned} \tag{9}$$

In the simplest case of linear regression models, it is assumed that the model is a linear function for both weights and input variables. However, the model can be made more complex using basis functions in order to be efficient on nonlinear data. Since the weights remain linear, the

model is considered to be linear. In this study, 6 basis functions are tested on tax evasion data in Malaysia and results are compared in next section. Equations (10-15) present the basis functions that are applied in this study.

Eiffel Tower Basis Function:

$$\varphi(x) = e^{-|x-t|}, \tag{10}$$

Radial Basis Function (RBF):

$$\varphi(x) = e^{-\frac{1}{u}(x-t)^2}, \tag{11}$$

Fourier Basis Function

$$\varphi(x) = \cos(t \times x),\ \sin(t \times x)\ , \tag{12}$$

V Basis Function:

$$\varphi(x) = |x-t| + t\ , \tag{13}$$

Step Basis Function:

$$\varphi(x) = -1 + 2 \times \theta(x-t)\theta(t-x)\ , \tag{14}$$

Polynomial Basis Function:

$$\varphi(x) = x^t, \tag{15}$$

Basis functions define the model by appropriately identifying the number of functions within input variable domain. In this study, interval values between basis functions are assigned based on minimum and maximum values of input data. However, the number (or shape in the RBF case) of basis functions needs to be calculated using model selection procedure.

This study aims to compute nonlinear regression model for tax evasion with regard to uncertainty of data in Malaysia. Here, tax evasion data from Malaysia by Tabandeh and Tamadonnejad (2015) has been employed; this dataset consists of 8 predictor variables and tax evasion as target variable in the GDP scale. Tax evasion is function of predictor variables as in the following:

$$\text{TE} = f(\text{TB1, TB2, GR, I, OP, IR, U, II}) \tag{17}$$

where TB1 is Direct tax and TB2 is Indirect tax in GDP, GR represents Government regulations, I is Income of taxpayers, OP is Trade openness, IR represents Inflation rate, U is Unemployment Rate, and II is Income inequality.

RESULTS AND DISCUSSION

The final results from VLR are compared with MLE algorithm which is a well-known sampling technique. Figure (1-6) compare of VLR with ARD method to MLE algorithm. The horizontal axis represents the year and vertical axis demonstrate the true respond value. The uncertainty value is shown by grey margin in these plots. Clearly, VLR with RBF and V basis functions provides the most certain results because their uncertainty margins are less. Nevertheless, Eiffel Tower basis function provides the most uniform uncertainty for the final results. Obviously, MLE method suffers from over-fitting in the Step and Fourier basis functions (Figures 4,6), however, the final results of MLE and VLR with RBF and V basis functions are more reasonable (Figures 1,2).

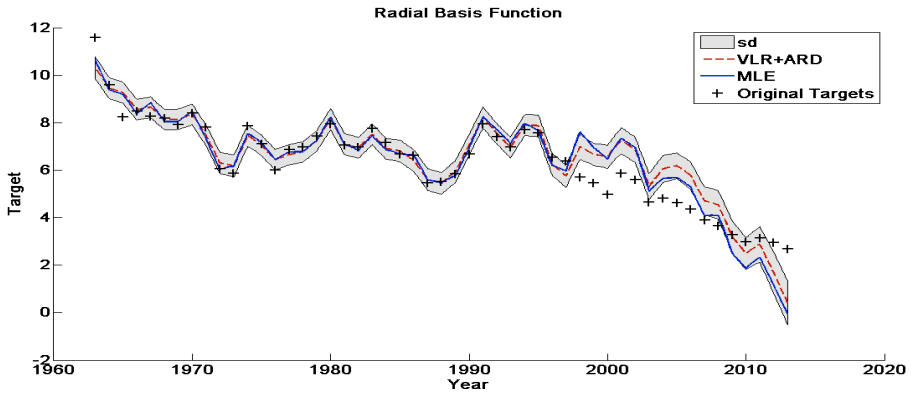

Figure 1. Comparison between VLR with ARD and MLE on RBF

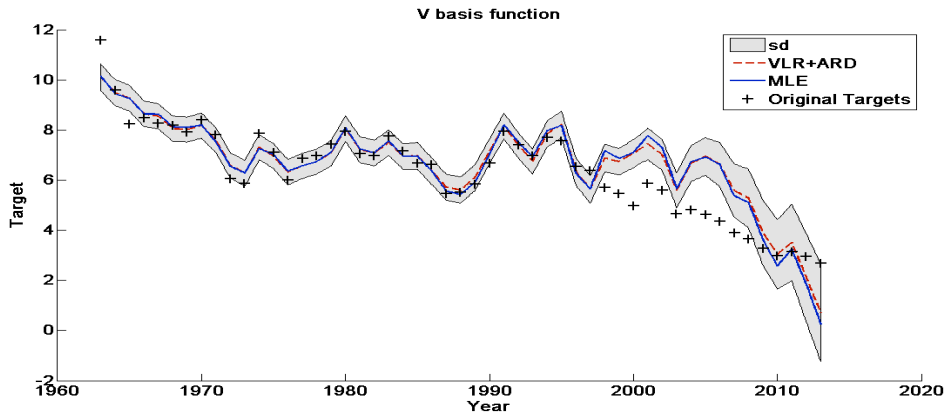

Figure 2. Comparison between VLR with ARD and MLE on V basis function

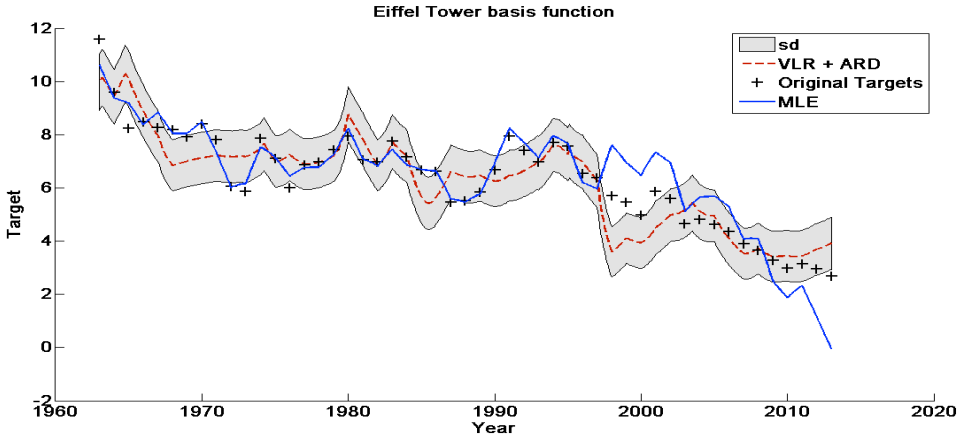

Figure 3. Comparison between VLR with ARD and MLE on Eiffel Tower basis function

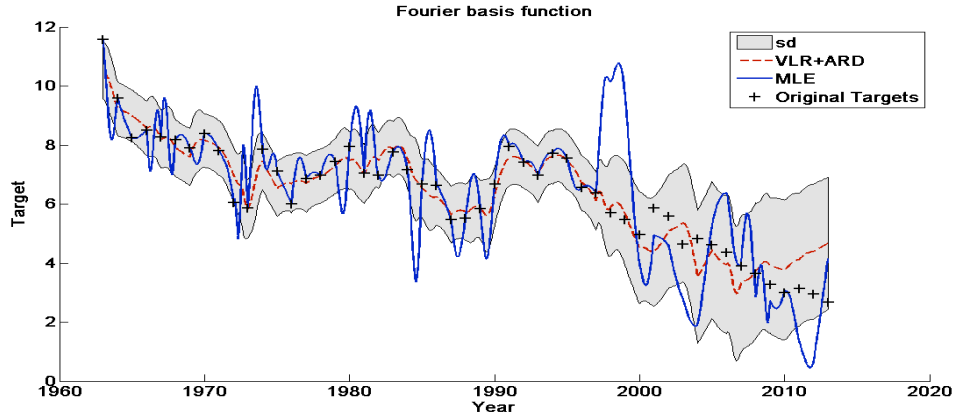

Figure 4. Comparison between VLR with ARD and MLE on Fourier basis function

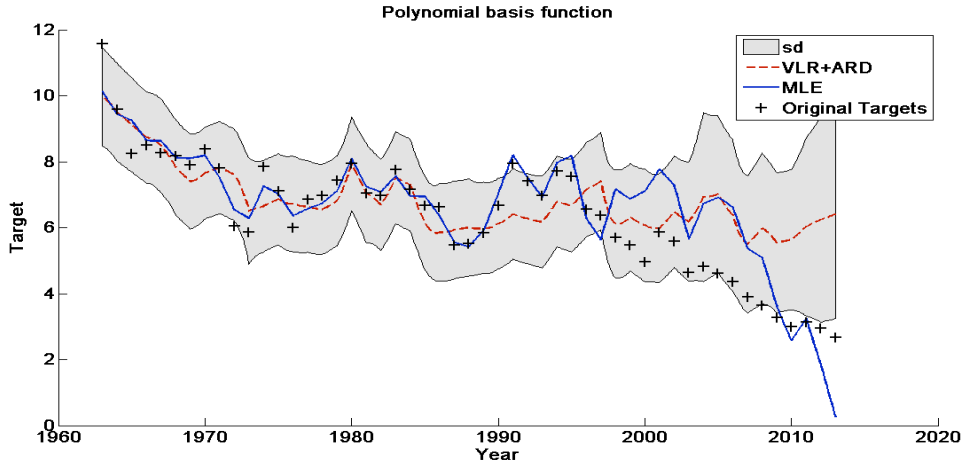

Figure 5. Comparison between VLR with ARD and MLE on Polynomial basis function

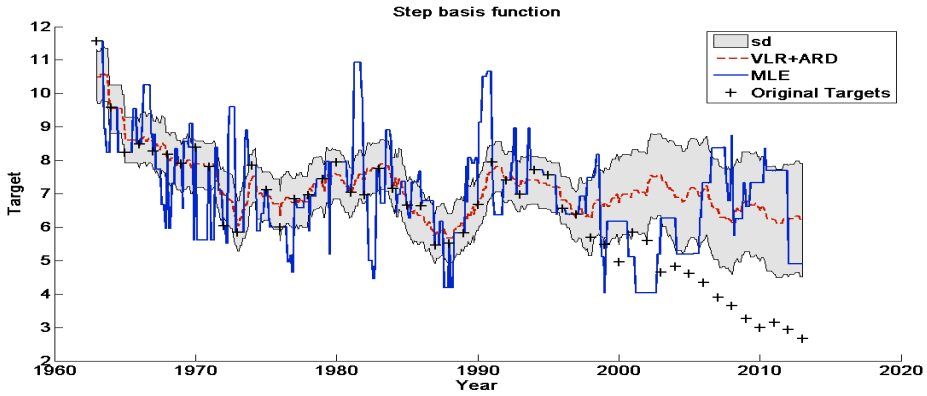

Figure 6. Comparison between VLR with ARD and MLE on Step basis function

The sample output for Eiffel Tower basis function on VLR with ARD technique is exemplified in Table 1. The main purpose of Table 1 is to estimate tax evasion with regard to changes on each predictor variable individually. Each column represents the percentage of changes in tax evasion in the way that only one particular variable changes and the rest of variables remain unchanged. For instance, with 10% increase in TB1 there is 0.8139% decrease in tax evasion. Likewise, 30% decline in the government regulation (GR) will lead to 1.4335% reduction in tax evasion. Furthermore, any change in tax evasion rate below 0.1% is negligible due to the residual error in the VLR. Hence, based on Table 1, the first four variables (i.e. TB1, TB2, GR, and I) affect tax evasion significantly and the contribution of last four variables (i.e. OP, IR, U, II) to tax evasion in Malaysia is minimal.

Table 1
The amount of small changes in tax evasion for any single change in valuables

		TB1	TB2	GR	I	OP	IR	U	II
5%↑	MSE (%)	-3.576	7.929	-2.627	-2.273	0.002	0.0001	-0.012	-0.019
	sd (%)	4.410	2.758	0.039	4.644	0.059	0.009	0.017	0.024
5%↓	MSE (%)	-1.433	-4.483	-3.775	-1.283	-0.010	-0.031	0.008	0.011
	sd (%)	1.486	-0.379	-0.948	2.147	0.002	-0.041	-0.017	-0.003
10%↑	MSE (%)	-0.814	1.274	-0.556	-0.364	0.001	0.003	-0.002	-0.004
	sd (%)	0.197	0.277	-0.124	0.436	0.010	-0.001	-0.0004	0.002
10%↓	MSE (%)	0.92	-1.136	-0.291	0.280	-0.002	-0.005	0.002	0.004
	sd (%)	0.490	-0.186	-0.209	0.026	-0.006	-0.003	-0.003	-0.003
20%↑	MSE (%)	-1.543	2.700	-1.112	-0.836	0.002	0.004	-0.005	-0.009
	sd (%)	0.893	0.672	-0.154	1.223	0.025	-0.001	0.002	0.007
20%↓	MSE (%)	0.735	-2.144	-0.768	0.152	-0.005	-0.011	0.004	0.007
	sd (%)	0.538	-0.301	-0.419	0.33	-0.009	-0.012	-0.007	-0.003
30%↑	MSE (%)	-2.293	4.294	-1.713	-1.319	0.003	0.003	-0.007	-0.012
	sd (%)	1.846	1.211	-0.157	2.263	0.042	0.0003	0.005	0.012
30%↓	MSE (%)	-0.173	-3.034	-1.433	-0.429	-0.007	-0.018	0.007	0.009
	sd (%)	0.230	-0.36	-0.587	0.507	-0.007	-0.026	-0.007	-0.004

Suppose the Malaysian government is liable to change each predictor variable to a maximum 30% every year, this study recommends the government raise TB1 and I by 5%, reduce TB2 and GR by 5% and keep the other parameters unchanged. Nevertheless, for changes above 30% these results are not reliable and reduction and increase in tax evasion should be calculated according to VLR formula. For instance, in a more dramatic alteration in predictor variables in Table 2, different outcomes appear; the results suggest that regarding the government's decision to change variables from 70% to 90% it is preferable to shrink TB2 and GR by 90% and increase TB1 and I by 90%. Interestingly, almost the same pattern appears in Table 1, although the level of changes is different.

Table 1
The amount of great changes in tax evasion for any single change in valuables

		TB1	TB2	GR	I	OP	IR	U	II
70%↑	MSE (%)	-4.803	8.957	-3.388	-3.129	-0.0005	-0.002	-0.014	-0.024
	sd (%)	6.966	2.550	0.322	7.192	0.077	0.028	0.025	0.036
70%↓	MSE (%)	-1.957	-5.522	-6.667	-1.631	-0.012	-0.041	0.009	0.006
	sd (%)	4.158	-0.323	-1.362	4.877	0.018	-0.045	-0.022	-0.010
80%↑	MSE (%)	-5.403	9.158	-3.511	-3.494	-0.002	-0.006	-0.015	-0.026
	sd (%)	8.110	2.362	0.574	8.488	0.087	0.036	0.029	0.040
80%↓	MSE (%)	-1.931	-5.896	-7.557	-1.611	-0.013	-0.046	0.01	0.007
	sd (%)	5.794	-0.269	-1.421	6.487	0.027	-0.046	-0.023	-0.010
90%↑	MSE (%)	-5.948	9.303	-3.532	-3.852	-0.003	-0.011	-0.016	-0.027
	sd (%)	9.240	2.196	0.868	9.685	0.098	0.039	0.033	0.043
90%↓	MSE (%)	-1.697	-6.171	-8.139	-1.455	-0.013	-0.0497	0.010	0.002
	sd (%)	7.599	-0.200	-1.419	8.238	0.038	-0.045	-0.023	-0.008

CONCLUSION

In this study, the nonlinear feature of tax evasion data is identified by testing 6 nonlinear basis functions on VLR in both with and without ARD cases. In the case of Malaysia tax evasion dataset, Eiffel Tower basis function on VLR with ARD and Fourier basis function in both with and without ARD define the pattern of data better than the other mentioned basis functions. Moreover, sample changes in each predictor variable is calculated to maximum 30% for minor changes in variables and between 70% to 90% for major changes and, based on results, TB1, TB2, GR, and I affect tax evasion more significantly than OP, IR, U, and II in the same situation. From computer science viewpoint, one of the important conclusions of this study is that, by choosing an appropriate basis function and prior parameters for VLR, complexity of a nonlinear data in tax evasion data can be detected by VLR and each basis function treats uncertainty in a different way. For instance, RBF tends to underestimate uncertainty, while polynomial basis function overestimates it. Furthermore, based on figures (1-6), stability in results in VLR+ARD is higher than MLE; in the cases of step and Fourier functions MLE provides poor results and results fluctuate rapidly dur-ing times. For future work the authors of

this study will consider non-parametric methods as a robust technique to tackle the uncertainty and nonlinearity of Malaysian tax evasion data.

REFERENCES

Attias, H. (2000). A variational Bayesian framework for graphical models. *Advances in neural information processing systems, 12*(1-2), 209-215.

Barber, D. (2012). *Bayesian reasoning and machine learning.* Cambridge University Press.

Bernardo, J. M., & Smith, A. F. (2009). *Bayesian theory* (Vol. 405). John Wiley & Sons.

Choong, K., & Lai, M. (2008). Tax practitioners' perception on tax audit and tax evasion: Survey evidence in Malaysia. *8th International Business Research Conference*, Dubai.

Cowles, M. K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association, 91*(434), 883-904.

Drugowitsch, J. (2008). Bayesian linear regression: Technical report, University of Rochester, Rochester, NY.

Drugowitsch, J. (2013). Variational Bayesian inference for linear and logistic regression. *arXiv preprint arXiv:1310.5438.*

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Taylor & Francis.

Hjort, N. L., Holmes, C., Müller, P., & Walker, S. G. (2010). *Bayesian nonparametrics* (Vol. 28). Cambridge University Press.

Jaakkola, T. S., & Jordan, M. I. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing, 10*(1), 25-37.

Jackson, M. O. (1991). Bayesian implementation. *Econometrica: Journal of the Econometric Society,* 461-477.

Jaffar, R., Bakar, M. J. A., & Tahir, I. M. (2011). Ethics on tax evasion: Do accounting and business students' opinions differ? *International Business and Management, 2*(1), 122-128.

Kasipillai, J., Aripin, N., & Amran, N. A. (2003). The influence of education on tax avoidance and tax evasion. *eJournal of Tax Research, 1*(2), 134-146.

Kasipillai, J., Baldry, J., & Prasada Rao, D. (2000). Estimating the size and determinants of hidden income and tax evasion in Malaysia. *Asian Review of Accounting, 8*(2), 25-42.

Miskam, M., Noor, R. M., Omar, N., & Aziz, R. A. (2013). Determinants of tax evasion on imported vehicles. *Procedia Economics and Finance, 7*, 205-212.

Palfrey, T. R., & Srivastava, S. (1987). On Bayesian implementable allocations. *The Review of Economic Studies*, 193-208.

Punt, A. E., & Hilborn, R. (1997). Fisheries stock assessment and decision analysis: The Bayesian approach. *Reviews in Fish Biology and Fisheries, 7*(1), 35-63.

Sun, L., & Shenoy, P. P. (2007). Using Bayesian networks for bankruptcy prediction: Some methodological issues. *European Journal of Operational Research, 180*(2), 738-753.

Tabandeh, R., & Tamadonnejad, A. (2015). The application of artificial neural network method to investigate the effect of unemployment on tax evasion. *Journal of Research in Business, Economics and Management, 4*(3), 393-402.

Tan, T. C.-C., & da Costa Werlang, S. R. (1988). The Bayesian foundations of solution concepts of games. *Journal of Economic Theory, 45*(2), 370-391.

Winn, J. M., & Bishop, C. M. (2005). Variational message passing. *Journal of Machine Learning Research,* (6), 661-694.