Determining the Significant Factors Affecting the Physical and Mental Components of Academicians Using Robust Linear Regression Models

S. Norin Rahayu, M. R. Norazan, A. H. Az'lina, I. Adriana, M. R. Nornadiah and A. Nur Niswah Naslina

***Abstract*—** This paper proposed linear regression models for determinig the contributing factors of health related quality of life (HRQoL) of academicians in a public university in Malaysia. The two models were used, namely the classical and robust MM linear regression models. Robust MM model is proposed in this study as data is with outlier problem. The results indicate that the robust MM model is the best model for determining the contributing factors affecting the physical and mental components of academicians with higher adjusted R square value as compared to the classical linear model.

***Keywords*—** Academician, HRQoL, linear model, robust, MM-method, SF-36.

I. Introduction

ONE of the main objectives of National Mission stated under The 9th Malaysia Plan is to ensure a comprehensive human capital development in Malaysia [1]. In fact, the government's main mission is to produce a first class human capital. This desire can only be achieved with the help of lecturers as they are assets of the universities and government and the achievement of a university greatly depend on the work performance of their staff mainly the lecturers.

Nowadays, the tasks of lecturers are not only educating people through classes and tutorials, giving supervision and consultation, involving in students' activities, but also conducting research, providing professional services [2-3] and getting involved in community works or services [2]. With the increasing number of students' enrolment each year and with the projection number of enrolment of approximately 1.4 million throughout Malaysia by the year 2020 [4], the tense to the lecturers in various aspects are also expected to increase. This scenario may lessen the quality of life as well as the health status of a lecturer.

Currently the academic excellence of lecturers is being measured a number of ways which include the number of published articles in the high impact journals. In terms of research works, the researcher is judged by the quality of research that have an impact in marketing the 'products' or that can be used and implemented in public policy [5]. According to [6], for the past 2 decades it is reported that the occupational stress have increased among the university staff as the government decided to cut the university research funding. The academicians were forced to find the funding through other source such as from the industrial or international linkages.

The diversity of lecturers' job would indirectly affect the health and quality of life of a lecturer. Lecturers who are facing problems with their health might affect their performance and their teaching ability. Thus, lecturers must have good quality of life since it is related with productivity and job satisfaction. In the work performance, productivity is the main outcome measure.

The Health Related Quality of Life (HRQoL) is used to measure the quality of life of a patient with specific disease and for a different level of severity [7]. However this instrument has become one of the most widely used nowadays for a healthy people. The HRQOL measures such as Short Form – 36 (SF-36) consists of 36 questions that yield eight different dimensions; physical functioning (PH), general health (GH), bodily pain (BP), vitality (VT), social functioning (SF), mental health (MH), role physical (RP) and role emotion (RE). These eight domains can be summarized into two categories which are physical component summary (PCS) and mental component summary (MCS). It is reported that the PCS component score were contributed from RP, PF and BP. These three domains are known as physical component. For the MCS component, the MH, RE and SF contributed significantly for the scoring of MCS. These three domains were clustered as mental component. However VT, GH and SF correlated positively with both PCS and MCS [8]. Both of the PCS and MSC were derived using the standardized scale (using norm

Norin Rahayu, S. is with the Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA (corresponding author to provide phone: 603-55435454; fax: 603-55435501; e-mail: norin@tmsk.uitm.edu.my).

Norazan, M.R., Az'lina, A.H., Adriana, I., Nornadiah, M.R. and Nur Niswah Naslina, A. are with the Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA (e-mail: norazan@tmsk.uitm.edu.my, azlina@tmsk.uitm.edu.my, adriana@tmsk.uitm.edu.my, nornadiah@tmsk.uitm.edu.my, nurniswah@tmsk.uitm.edu.my)

based (NBS) with mean = 50 and standard deviation = 10) [9].

The scale of measurement used in HRQoL instrument is an ordinal. The value of the total score from the summated value from the similar questions generated the HRQoL scores. Usually this score is treated as it is normally distributed and the parametric test is being applied [10].

II. Modeling the Quality of Life Data

A. Data Description

The instruments used in the study consist of three sections, namely

i. the demographic information such as gender, age, teaching workload, years of servicing, gender, income, department and servicing status;
ii. the job stress from the Occupational Stress Inventory Revised Edition (OSI-R) [11]. Te OSI is a set of questionnaire-based, and does not require on the job analysis. Each element of the OSI is scored on a scale from 0 to 2, with zero being not present and 2 as strongly present. The General OSI Questionnaire can be used among workers of any occupational profile.
iii. the SF-36 questionnaire [8]. The instrument has been tested for validity and reliability. It has been translated into more than 60 languages including Malay Language, Malaysian Chinese and Tamil. There are two versions of SF-36 which are standard (four weeks) and acute (one week) [8].

A cross sectional study was conducted from September to November 2011, in the middle of semester, on lecturers in a certain public university in Malaysia. The sample size of the respondents was determined using IBM Complex Samples 19 using a multistage sampling technique. A total of 227 respondents completed the questionnaire; however only 193 respondents were selected while another 34 is excluded since the lecturers were on study leaves.

The main objective of this paper is to investigate the contributing factors on the HRQoL of these 193 lecturers and to compare the result between the classical and robust MM multiple linear regression models. We consider the following model to explain the variation in the lecturers' PCS as weel as for MCS:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i \qquad (1)$$

where

$i = 1, 2, \ldots, 193$

$y =$ PCS; MCS

$x_1 =$ number of years in service

$x_2 =$ workload (in hours)/semester

$x_3 =$ working hours/day

$x_4 =$ working hours/week

$x_5 =$ gender (female = 0, male = 1)

$x_6 =$ hold administrative post (yes=1, no=0)

$x_7 =$ teaching postgraduate (yes=1, no=0)

$x_8 =$ teaching more than one programs (yes=1, no=0)

$x_9 =$ maximum working hours without rest

All the factors included in the regression are the main factors that contributed to the job stress among lecturers. The following assumptions are made for equation (1):

i. The relationship between the y and x is linear.
ii. All x are uncorrelated.
iii. The error term is normally distributed for tests of significance of coefficients and other statistics of the regression equation.

We first performed a diagnostic to check whether the given data meet the normality assumption required for model (1). Diagnostic plots in Figure 1(a) and (b) indicates that the PCS scores are not normally distributed while for MCS scores, it is slightly skewed. We suspect the PCS scores may contain outliers. The presence of outliers in data set not only causes skewness that violates the required assumption of least square estimation, but more importantly it affects the precision of parameter estimation in the linear regression model. Thus, an alternative approach which is called MM method, for estimating parameters in model (1) is proposed in this paper.

The method is expected to be robust against outliers. Compare to the LS method, the MM method is expected to be still good for data that do not conform to normality assumptions.

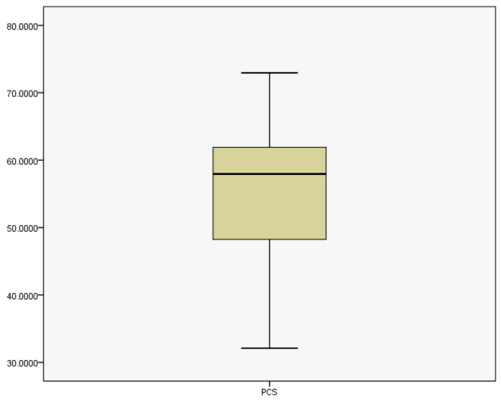

Figure 1(a): Boxplot for PCS

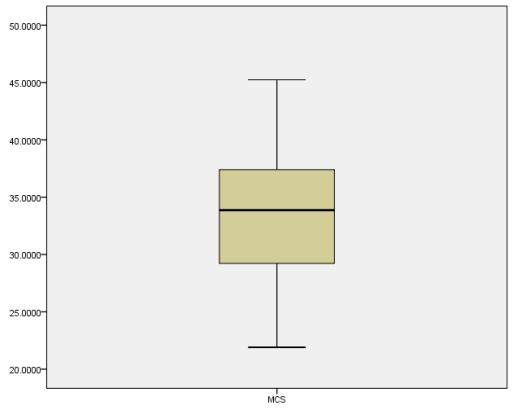

Figure 1(b): Boxplot for MCS

B. LS and Robust MM Methods

The least squares (LS) method is the simplest and most commonly applied parameter estimation method for linear regression model. The basic idea in the method is to find the estimate of β by minimizing the sum of squares of residual,

$$S(\beta) = S(\beta_0, \beta_1, \ldots, \beta_r) = \sum_{i=1}^{n} \varepsilon_i^2 \quad (2)$$

where

$$\sum_{i=1}^{n} \varepsilon_i^2 = (y_i - \hat{y}_i)^2 \quad (3)$$

and

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_k x_{ik} \quad (4)$$

Meanwhile, the MM method is proposed by Yohai [12]. This method produces estimates with high-efficiency and high breakdown values. The MM-estimates can be calculated as follows:

Stage 1: Calculate an S-estimate using an influence function ρ where

$$\rho(x) = 3\left(\frac{x}{c}\right)^2 - 3\left(\frac{x}{c}\right)^4 + \left(\frac{x}{c}\right)^6 \text{ for } |x \le c|$$

or otherwise, $\rho(x) = 1$. c is constant value and is selected as 1.548.

Stage 2: Calculate the MM parameters that give the minimum value of $\sum_{i=1}^{n} \rho\left(\frac{r_i(\hat{\beta}_{MM})}{\hat{\sigma}_0}\right)$ where $\rho(x)$ is the one that being used in Stage 1 with tuning constant 4.687 and $\hat{\sigma}_0$ is actually the standard deviation of the residuals obtained from Stage 1.

Stage 3: Compute the MM-estimate of scale, which is obtained from the solution of $\frac{\sum_{i}^{n} \rho\left(\frac{y_i - x'_i \beta}{s}\right)}{(n-p)} = 0.5$.

Except for the case of high leverage points or outliers in the predictor variables, generally the MM-method is resistant to outliers. In this paper, the robust MM-estimator which is available in the standard Robust Package of S-PLUS is used.

III. Result and Discussion

A. Demographic Profile

Table 1 presents the demographic profile of the data. The mean age of the respondents was 42.42 years age (with standard deviation, SD = 8.16). The minimum age is 28 years and maximum of 58 years. On average, we notice that the lecturers have worked for 14.87 years in the university (*SD* = 9.53). The workload per semester for the lecturers were 17.58 hours (*SD* = 5.48) which comprises of teaching, consultation and supervision for at most six days per week. The duration or working per day for the lecturers last for at most ten hours and they can work on average three hours without taking a break.

B. Fitted Models

Table 2 and Table 3 present the results of the regression analysis from the LS and robust MM methods. The adjusted R^2 value indicates how much variance in PCS and MCS is explained by the identified factors. Generally we observe that the robust MM regression method is more efficient in determining the contributing and significant factors of both the PCS and MCS variables. These can be seen from their adjusted R^2 values which are much higher compared to the one from the OLS model. In fact, we can conclude that so far, the value of R^2 obtained in this model from this study is the highest if compared with other studies [13-15].

Table 1: Demographic characteristics.

Variables	Frequency (%)
Age	
≤ 40 years	62 (32.1)
> 40 years	131 (67.9)
Gender	
Male	66 (34.2)
Female	127 (65.8)
Position hold	
Yes	94 (48.7)
No	99 (51.3)
Programme level teach	
Undergraduate	97 (50.5)
Postgraduate/other	22 (11.4)
> 1programme level	74 (38.3)

The long working hours in a day have a positive effect, however the result contradict with the study by Wu et al. [14]. Years of service in a university, the holding a post (role overload) and teaching more than one programs inversely associated with PCS. The years of servicing may be associated with age of the lecturers. The tendency to cope with the changes in the policies and management [16] and the ability to cope with competition among each other led may lessen the role physical and general health of the lecturers. As stated before, the tasks of the lecturers are huge. They need to prepared for the lesson, construct test and examination question and at the same time fulfilling the administrative work [13]. At the same time, the lecturers need to equip themselves with a new knowledge to ensure the students being exposed to the real world situation.

Table 2 also indicates that "workload (in hours)/semester" and "hold administrative post" variables are not significant in the least square model, but found to be significant in the robust MM model. The smaller value of residual standard errors in the robust MM model shows that the method is resistant to outliers. The goodness of the robust MM model is confirmed by the good fit of residual QQ-plot in Figure 2(a) and density plot in Figure 2(b).

Now let us focus on the figures in Table 3. Again, the long working hours have a positive effect on the MCS in the linear model, but this is very much contradicted to the MM model. The long working hours in a day give a negative effect on the MCS, and this finding is consistent with other studies [14, 17]. Factors like working hours/week, holding a post (role overload) and maximum working hours without rest are found to have inverse effect on the MCS in both models. Long working hours in university may affect mentality and the vitality of a person as they also have other responsibilities such as family matters [17].

Table 2: LS and Robust MM- Estimates for PCS

	LS				**Robust MM**			
Estimates	**Coefficient**	**Std. Error**	**t-value**	**p-value**	**Coefficient**	**Std. Error**	**t-value**	**p-value**
Intercept	64.5578	6.7625	9.5465	0.0000	103.968	2.0882	49.7886	0.0000
x1	-0.3223	0.0823	-3.9145	0.0001	-1.483	0.0317	-46.7472	0.0000
x2	-0.0645	0.1178	-0.5476	0.5850	-1.3681	0.044	-31.128	0.0000
x3	3.2014	0.4000	8.0042	0.0000	6.158	0.1323	46.5475	0.0000
x4	-5.6355	1.0648	-5.2928	0.0000	-9.8302	0.2868	-34.2701	0.0000
x5	6.9444	0.8484	8.1855	0.0000	9.0486	0.2357	38.3866	0.0000
x6	-1.7583	0.81136	-2.1671	0.0320	-15.957	0.3461	-46.102	0.0000
x7	3.3532	1.3038	2.5719	0.0011	12.3898	0.424	29.2202	0.0000
x8	-3.5820	0.7432	-4.8194	0.0000	-11.6258	0.2807	-41.4127	0.0000
x9	0.6251	0.336	1.8606	0.0644	0.4363	0.082	5.3209	0.0000
Residual Std. Error	8.666 (df = 183)				5.476 (df = 183)			
adjusted R^2	0.4491				0.7444			

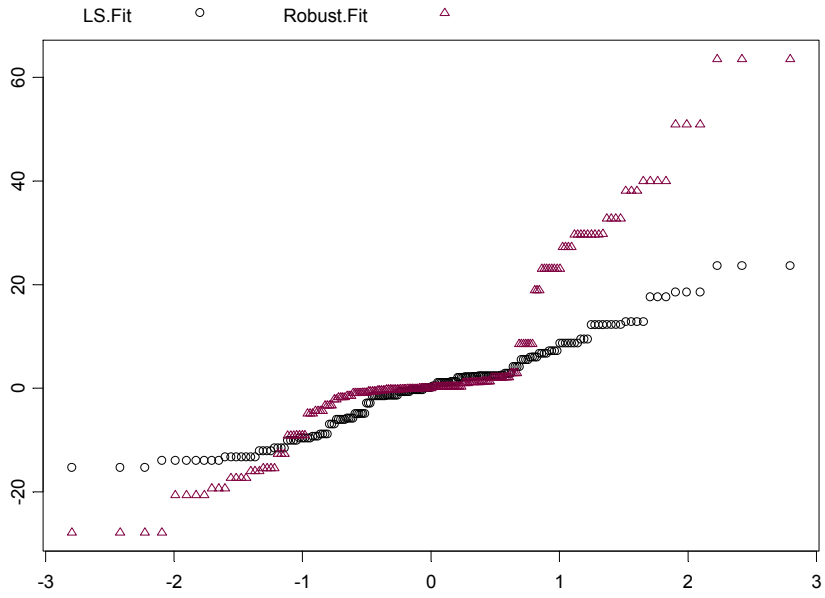

Figure 2(a): Comparison of QQ plot for LS residuals and Robust MM residuals for PCS.

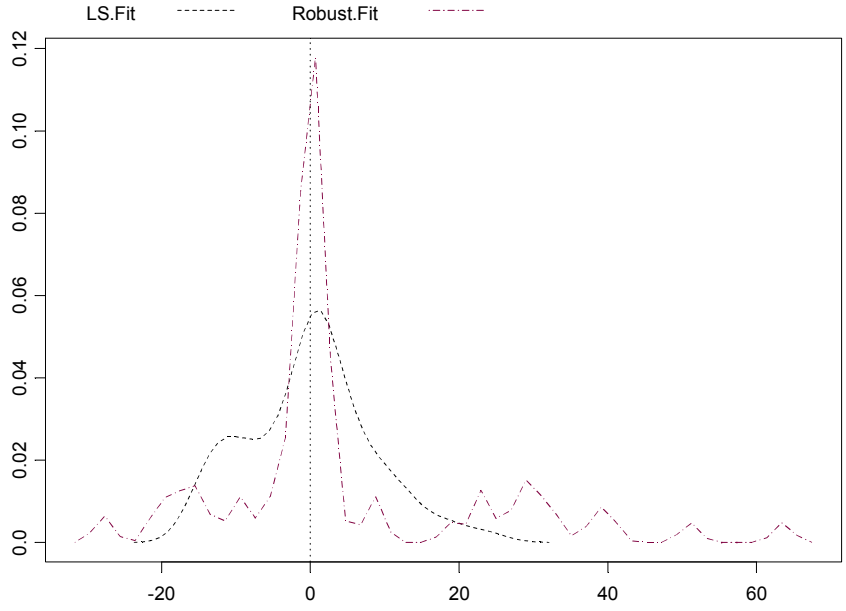

Figure 2(b): Comparison of density plots for LS residuals and Robust MM residuals for PCS

Table 3 also concludes that for both models, teaching more than one programs does not affect the MCS scores. The result contradict with the other studies, where excessive workload had lead to the occupational stress which affect the capabilities of a person [17-18]. Again, the goodness of the robust MM model is confirmed by the good fit of residual QQ-plot in Figure 3(a) and density plot in Figure 3(b).

Overall, result shows the male lecturers' scores higher in both PCS and MCS component summary compare to female. A few study revealed a similar result in comparing the health

status between gender [13, 19-20] This may be because female has a multitask role that need her attention, not only in the workplace but also at home as a mother and wife, which lessen her physical and mental condition. Moreover the characteristics of women which differ especially in emotional condition and vitality are more vulnerable position compared to male.

IV. CONCLUSIONS

In this paper we propose to use robust MM linear regression models to determine the significant contributing factors to the lecturers' PCS and MCS scores for quality of life data. The ordinary least square linear model is not recommended in this case as our data contain outliers and thus the model is not good enough as compared to the MM model [21]. In real situation, data usually are not normally distributed. Therefore other alternative is needed to ensure the end result will give more reliable information.

The MM model has shown that all factors contribute significantly to the PCS scores of the lecturers. All factors, except teaching more than one programs, are significantly contribute to the MCS scores based on the MM linear model.

Based on these findings, proper and immediate actions must be taken by this particular university to treat this problem. Giving social support, sharing the latest knowledge and information and providing the counseling session to the lecturers are among the good activities that should be introduced and implemented immediately to cater this problem.

Table 3: LS and Robust MM- Estimates for MCS

	LS				Robust MM			
Estimates	**Coefficient**	**Std. Error**	**t-value**	**p-value**	**Coefficient**	**Std. Error**	**t-value**	**p-value**
Intercept	52.3925	2.8997	18.0685	0.0000	43.1402	0.8295	52.0082	0.0000
x1	-0.0348	0.0353	-0.9865	0.3252	0.0414	0.0092	4.5169	0.0000
x2	0.2023	0.0505	4.0040	0.0000	0.2999	0.0139	21.5817	0.0000
x3	-0.5178	0.1715	-3.0194	0.0029	0.6073	0.0599	10.1245	0.0000
x4	-2.2739	0.4566	-4.9806	0.0000	-2.3822	0.1127	-21.1417	0.0000
x5	3.9898	0.3638	10.9678	0.0000	3.7607	0.0860	43.7132	0.0000
x6	-0.6715	0.3479	-1.9301	0.0552	-0.8058	0.0529	-9.7218	0.0000
x7	2.9879	0.5590	5.3447	0.0000	3.6818	0.1628	22.6112	0.0000
x8	0.4448	0.3187	1.3957	0.1645	0.0039	0.0801	0.0488	0.9611
x9	-0.1771	0.1441	-1.2295	0.2205	-0.9438	0.0369	-25.5811	0.0000
Residual Std. Error	3.716 (df = 183)				1.577 (df = 183)			
adjusted R^2	0.5572				0.9261			

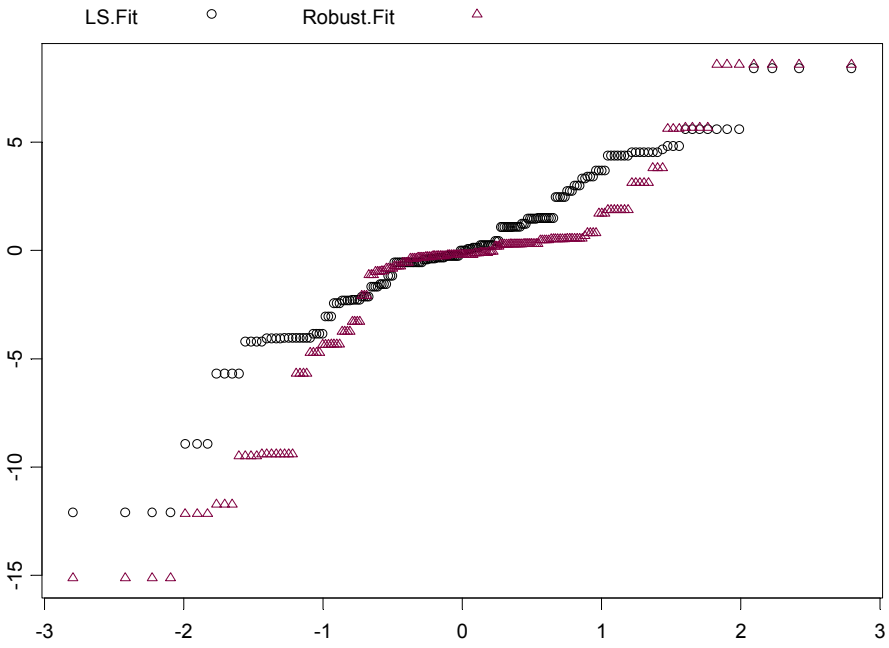

Figure 3(a): Comparison of QQ plot for LS residuals and Robust MM residuals for MCS.

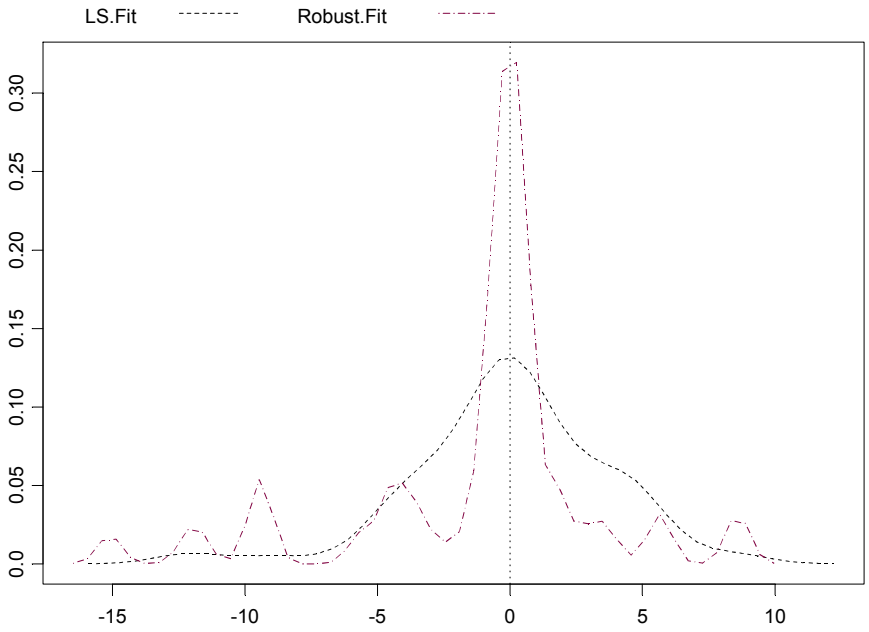

Figure 3(b): Comparison of density plots for LS residuals and Robust MM residuals for MCS.

ACKNOWLEDGMENT

The authors wish to thank the Ministry of Higher Education Malaysia and the Research Management Institute of Universiti Teknologi MARA Malaysia for providing the grant (600-RMI/ST/FRGS 5/3/Fst (239/2010) for this research.

REFERENCES

[1] E. P. Unit, "The 9th Malaysia plan: 2006-2010," *Putra Jaya: Economic Planning Unit, Prime Ministers Department,* p. 258, 2006.

[2] N. Z. Abidin, *Kerjaya sebagai pensyarah*: PTS Professional, 2008.

[3] K. Jusoff and H. S. A. A. Samah, "Reconciling Challenges and Opportunities in Academic Scientific Writing," *Academic Leadership Journal,* vol. 8, pp. 85-87, 2010.

[4] Ministry of Higher Education Malaysia, "The National Higher Education Strategic Plan: Laying the Foundation Beyond 2020," vol. 1, 2007.

[5] R. Omar, "Academic Point: Academic Excellent and Nation Building," *Malaysian Edutrend,* vol. 1, pp. 1-2, 2010.

[6] A. H. Winefield*, et al.*, *Job stress in university staff: An Australian research study*: Australian Academic Pr, 2008.

[7] I. Moleavin and C. E. Dobrescu, "The usage of a software to process the SF-36 data," in *The 4th EUROPEAN COMPUTING CONFERENCE (ECC '10)*, Bucharest, Romania, 2010, p. 22.

[8] J. E. Ware*, et al.*, *How to score version 2 of the SF-36 health survey (standard & acute forms)*: QualityMetric Inc., 2000.

[9] J. E. Ware and M. Kosinski, "Interpreting SF&-36 summary health measures: A response," *Quality of Life Research,* vol. 10, pp. 405-413, 2001.

[10] S. J. Walters and M. J. Campbell, "The use of bootstrap methods for analysing health-related quality of life outcomes (particularly the SF-36)," *Health and quality of life outcomes,* vol. 2, p. 70, 2004.

[11] K. Belkic, *The occupational stress index: an approach derived from cognitive ergonomics and brain research for clinical practice*: Cambridge International Science Pub., 2003.

[12] V. J. Yohai, "High breakdown-point and high efficiency robust estimates for regression," *The Annals of Statistics,* pp. 642-656, 1987.

[13] X. Yang*, et al.*, "Relationship between quality of life and occupational stress among teachers," *Public health,* vol. 123, pp. 750-755, 2009.

[14] S. Wu*, et al.*, "Health-related Quality of Life and Its Main Related Factors among Nurses in China," *Industrial health,* p. 1012100035, 2010.

[15] D. Jurakić*, et al.*, "Physical activity in different domains and health-related quality of life: a population-based study," *Quality of Life Research,* vol. 19, pp. 1303-1309, 2010.

[16] J. E. Ilmarinen, "Aging workers," *Occupational and environmental medicine,* vol. 58, p. 546, 2001.

[17] M. Teichmann and J. Ilvest Jr, "Sources of occupational stress in technical university academics," in *International Conference on Education and Education Technology*, Corfu Island, Greece, 2010, pp. 448-453.

[18] M. Teichmann and P. Dondon, "Sources of stress in Bordeaux University academics," in *Recent Researches in Educational Technologies*, Corfu Island, Greece, 2011.

[19] D. Cherepanov*, et al.*, "Gender differences in health-related quality-of-life are partly explained by sociodemographic and socioeconomic variation between adult men and women in the US: evidence from four US nationally representative data sets," *Quality of Life Research,* vol. 19, pp. 1115-1124, 2010.

[20] H. C. Hsu, "Gender Differences in Health-related Quality of Life among the Elderly in Taiwan," *Asian Journal of Health and Information Sciences,* vol. 1, pp. 366-376, 2007.

[21] M. Habshah*, et al.*, "A Robust Optimum Response Surface Methodology based On MM-estimators," in *Recent Reseaches in Applied Informatics and Remote Sensing*, Penang, Malaysia, 2011.