This is the paper presented in 22 Nov 2005 at the National Sustainable Development Seminar, Shah Alam organized by UiTM

The Dynamic Interaction between Road Density and Land Uses in West Malaysian Cities

Muhammad Zaly Shah Bin Muhammad Hussein, Ph.D. [1]*
Mohd Noor Bin Awang [2]
Ho Chin Siong, Ph.D. [1]
Othman Bin Che Puan, Ph.D. [3]
Hishamuddin Bin Mohd Ali, Ph.D. [4]
Mohd Nor Bin Said [4]
Norizam Bin Katmon [5]

[1] Dept. of Urban and Regional Planning, Faculty of Built Environment
[3] Faculty of Civil Engineering
[4] Faculty of Engineering and Geoinformation
[5] Faculty of Computer Science and Information Resources

Universiti Teknologi Malaysia
81310 Skudai, Johor, Malaysia

[2] MN Consult, Taman Universiti, 81300 Skudai, Johor

* This is the paper presented in 22 Nov 2005 at the National Sustainable Development Seminar, Shah Alam organized by UiTM

The Dynamic Interaction between Road Density and Land Uses in West Malaysian Cities

Muhammad Zaly Shah Bin Muhammad Hussein, Mohd Noor Bin Awang [2]
Ho Chin Siong, Othman Bin Che Puan, Hishamuddin Bin Mohd Ali, Mohd Nor Bin Said
Norizam Bin Katmon [5]

National Sustainable Development Seminar, Shah Alam organized by UiTM

22 Nov 2005

Abstract

Urban planning theories state that changes in land uses affect the provision of road infrastructure and vice versa. Invariably, this interaction between land uses and road infrastructure, represented by road density, is a dynamic one. Many studies abroad have attempted to quantify this dynamic interaction. Using a two-steps approach, an empirical model that describes the interaction between land uses and road density has been developed for selected West Malaysian cities. The two steps require that, first, a regression model is developed that functionally linked road density and land uses, and second, simulating land use developments to predict changes in road density. In this paper, the evidence for the interaction is shown for two West Malaysian Cities – Johor Bahru and Kuantan. The calibration of the interaction models for these two cities shows promising result. The main advantage of such interaction models is that the model provides transport planners a planning (and, budgeting) tool to plan for future provision of road infrastructure.

Keywords: Urban growth, land use planning, transport planning, quantitative model

1 Introduction

Road density level inevitably influences many aspects of daily life of a community including the location, density and types of residence, social-cultural, economic and leisure activities. It also influences the range and quality of the provision of goods and services to the community, which eventually determine the life style and the quality of life of the community.

Many literatures pointed out that the transportation-land use relationship is reciprocal and dynamic in nature. As our towns and cities have grown, both in terms of their population size and functions, the relative advantage of locations and accessibility within them has changed through time because of the development of new roads in response to the current pressure for development. Inevitably, these developments provide different level of accessibility to new locations. This in turn will ease the movements or reduce the travel costs between locations, which open up new opportunities and contribute to the economic vitality of specific economic and social activities as well as residential locations.

Recent development in highway development in Malaysia especially toll highways have produced some significant effects in term of accessibility to certain locations resulting in land use changes in certain localities along these highways. The completion of the 848 km North-South Expressway (NSE) in 1994 and other urban toll highways especially in the Klang Valley area have enhanced accessibility to the previously remote areas and created great potential for new developments. Fast development that took place in Rawang (Rawang industrial park), Bukit Beruntung, Serendah (national car project and auto industrial park) in Selangor, Bukit Merah in Perak and Nilai in Negeri Sembilan, to mention a few, were a direct impact of improved accessibility resulting from the NSE development. The same has occurred to Puchong, in which Puchong has been transformed from a small rural town to a modern township accommodating a population of more than fifty thousand within a short span of time induced by the development of Damansara-Puchong Highway.

This study then attempts to empirically quantify the dynamic interaction between land use and road infrastructure developments. The intended results would be a mathematical model that explains such interaction, if it exists, that would be beneficial to planners and decision makers alike.

2 Theoretical Foundation

Existing transportation economic theories consider transportation as a derived demand of land uses. What this really means is that transportation in itself cannot exist except for providing accessibility to land uses or as a medium to move goods and services between land uses. Rarely do people travel for the sake of travelling. More often than not, people travel to obtain some perceived benefit at the destination, which is offered by the different land uses. Hence, the modelling of the interaction between land use and road transportation starts with a premise that future road transportation requirements are based on what land use changes that will be taking place. This relationship between land use and transportation activities is described graphically in Figure 1.

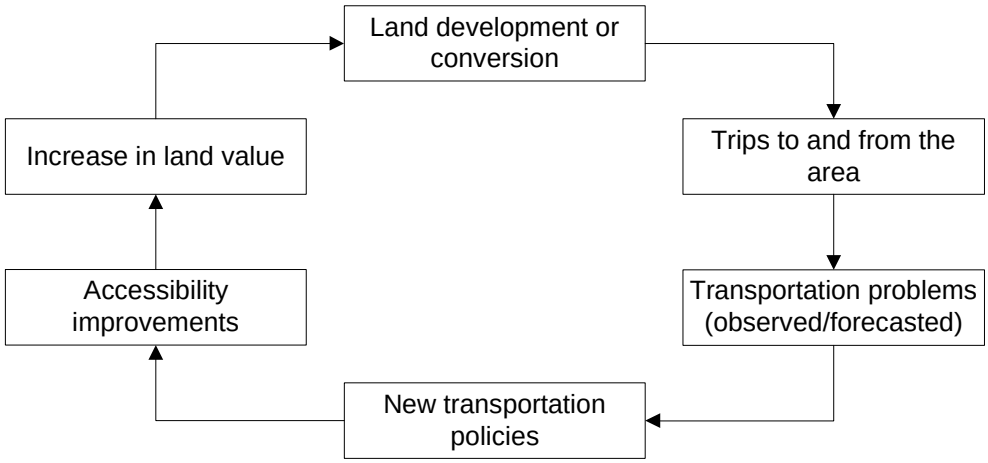

Figure 1: Relationship between land use and transportation activities

In this study, the relationship between land use and transportation is described using multivariate linear regression. Given that the transportation variable is represented by road density, the general form of the regression equation then is given in Eq. (1) as:

$$\mathbf{Y}_{i,t} = f(\mathbf{X}_{i,t}) \tag{1}$$

where:

$\mathbf{Y}_{i,t}$ = a vector of road density for town *i*, *i* = 1, 2 at time *t*

$\mathbf{X}_{i,t}$ = a matrix of land use variables for town *i*, *i* = 1, 2 at time *t*

The subscript *i* in Eq. (1) denotes the two towns involved in this study, namely Johor Bahru and Kuantan. With Eq. (1), it is clear from the relationship that transportation, represented by road density, is dependent upon land use.

The land use matrix $\mathbf{X}_{i,t}$, however, is represented by several components described below:

$$\mathbf{X}_{i,t} \in \left\{\mathbf{E}_{i,t}, \mathbf{H}_{i,t}, \mathbf{P}_{i,t}\right\} \tag{2}$$

where:

$\mathbf{E}_{i,t}$ = a matrix of employment variables for town *i*, *i* = 1, 2 at time *t*

$\mathbf{H}_{i,t}$ = a matrix of households variables for town *i*, *i* = 1, 2 at time *t*

$\mathbf{P}_{i,t}$ = a matrix of property variables for town *i*, *i* = 1, 2 at time *t*

Simply replacing Eq. (2) into Eq. (1) will give us:

$$\mathbf{Y}_{i,t} = f(\mathbf{E}_{i,t}, \mathbf{H}_{i,t}, \mathbf{P}_{i,t}) \tag{3}$$

which now states that the road density vector $\mathbf{Y}_{i,t}$ is functionally related to the independent variable matrices of employment $\mathbf{E}_{i,t}$, households $\mathbf{H}_{i,t}$ and property $\mathbf{P}_{i,t}$.

Individually, each of the land use components – employment, households and property – can be further defined into more specific subcomponents. For employment matrix $\mathbf{E}_{i,t}$, it can be defined more specific as:

$$\mathbf{E}_{i,t} \in \left\{\mathbf{E}_{1i,t}\right\} \tag{4}$$

where $\mathbf{E}_{1i,t}$ is a vector of total workers for town *i*, *i* = 1, 2 at time *t*. For the household matrix $\mathbf{H}_{i,t}$, the specific subcomponents are:

$$\mathbf{H}_{i,t} \in \left\{\mathbf{H}_{1i,t}, \mathbf{H}_{2i,t}, \mathbf{H}_{3i,t}, \mathbf{H}_{4i,t}\right\} \tag{5}$$

where:

$\mathbf{H}_{1i,t}$ = a vector of households' average income for town *i*, *i* = 1, 2 at time *t*

$\mathbf{H}_{2i,t}$ = a vector of total number of cars for town *i*, *i* = 1, 2 at time *t*

$\mathbf{H}_{3i,t}$ = a vector of total number of households for town *i*, *i* = 1, 2 at time *t*

$\mathbf{H}_{4i,t}$ = a vector of households' average income for town *i*, *i* = 1, 2 at time *t*

Finally, for property matrix $\mathbf{P}_{i,t}$, the subcomponents are:

$$\mathbf{P}_{i,t} \in \left\{\mathbf{P}_{1i,t}, \mathbf{P}_{2i,t}, \mathbf{P}_{3i,t}, \mathbf{P}_{4i,t}, \mathbf{P}_{5i,t}, \mathbf{P}_{6i,t}, \mathbf{P}_{7i,t}, \mathbf{P}_{8i,t}, \mathbf{P}_{9i,t}, \mathbf{P}_{10i,t}\right\} \tag{6}$$

where, given town *i*, *i* = 1, 2:

$\mathbf{P}_{1i,t}$ = a vector of nonresidential land value (in RM) at time *t*

$\mathbf{P}_{2i,t}$ = a vector of government office floor space area (in sq. ft.) at time *t*

$\mathbf{P}_{3i,t}$ = a vector of commercial building floor space area (in sq. ft.) at time *t*

$\mathbf{P}_{4i,t}$ = a vector of commercial improvement value (in RM) at time *t*

$\mathbf{P}_{5i,t}$ = a vector of industrial building floor space area (in sq. ft.) at time *t*

$\mathbf{P}_{6i,t}$ = a vector of residential land value (in RM) at time *t*

$\mathbf{P}_{7i,t}$ = a vector of government building improvement value (in RM) at time *t*

$\mathbf{P}_{8i,t}$ = a vector of residential building improvement value (in RM) at time *t*

$\mathbf{P}_{9i,t}$ = a vector of industrial building improvement value (in RM) at time *t*

$\mathbf{P}_{10i,t}$ = a vector of number of residential unit at time *t*

Therefore, with Eqs. (4), (5) and (6), the functional relationship between road density and land use components given in Eq. (3) now becomes:

$$\mathbf{Y}_{i,t} = f(\mathbf{E}_{1i,t}, \mathbf{H}_{1i,t}, \mathbf{H}_{2i,t}, \mathbf{H}_{3i,t}, \mathbf{H}_{4i,t}, \mathbf{P}_{1i,t}, \mathbf{P}_{2i,t}, \ldots, \mathbf{P}_{10i,t}) \tag{7}$$

As each town *i* is divided into grids of 1 sq. km, a town *i* with 150 grids will then define the functional relationship in Eq. (7) to be:

$$\begin{bmatrix} Y_{i_1,t} \\ Y_{i_2,t} \\ \vdots \\ Y_{i_{150},t} \end{bmatrix} = f \begin{bmatrix} E_{1i_1,t} & H_{1i_1,t} & H_{2i_1,t} & \cdots & H_{4i_1,t} & P_{1i_1,t} & P_{2i_1,t} & \cdots & P_{10i_1,t} \\ E_{1i_2,t} & H_{1i_2,t} & H_{2i_2,t} & \cdots & H_{4i_2,t} & P_{1i_2,t} & P_{2i_2,t} & \cdots & P_{10i_2,t} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ E_{1i_{150},t} & H_{1i_{150},t} & H_{2i_{150},t} & \cdots & H_{4i_{150},t} & P_{1i_{150},t} & P_{2i_{150},t} & \cdots & P_{10i_{150},t} \end{bmatrix} \tag{8}$$

where, for example, the variables $Y_{i_2,t}$ and $E_{1i_2,t}$ are the road density and the number of total workers in grid 2 of town *i* at time *t*, respectively. The modelling process therefore is concerned to define the function $f(\bullet)$ in Eq. (8) for each town *i* as it describes the relationship between road density and land use components in that town. For our multiple linear regression model, the best fitted regression line for this multivariate problem is identified using the ordinary least squares (OLS) method. Once the functional relationship $f(\bullet)$ is defined, the predicted future value of $\mathbf{Y}_{i,t}$, denoted as $\hat{\mathbf{Y}}_{i,t+1}$, is obtained from:

$$\hat{\mathbf{Y}}_{i,t+1} = f(\hat{\mathbf{E}}_{1i,t+1}, \hat{\mathbf{H}}_{1i,t+1}, \hat{\mathbf{H}}_{2i,t+1}, \hat{\mathbf{H}}_{3i,t+1}, \hat{\mathbf{H}}_{4i,t+1}, \hat{\mathbf{P}}_{1i,t+1}, \hat{\mathbf{P}}_{2i,t+1}, \ldots, \hat{\mathbf{P}}_{10i,t+1}) \tag{9}$$

Hence, from Eq. (9), the predicted future value of road density at time *t* + 1, i.e. $\hat{\mathbf{Y}}_{i,t+1}$, requires that the predicted values of the independent variables must first be obtained. In this situation, the procedure to get the predicted values of the independent variables must be established before we can proceed to predict the future values of road density. At this juncture, an urban simulation application called UrbanSim will be utilised to provide the future values of the independent variables.

3 Modelling Process

The development of the interaction model between land uses and road density is a multi-step process. Schematically, the entire process of developing such an empirical model is illustrated in Figure 2.

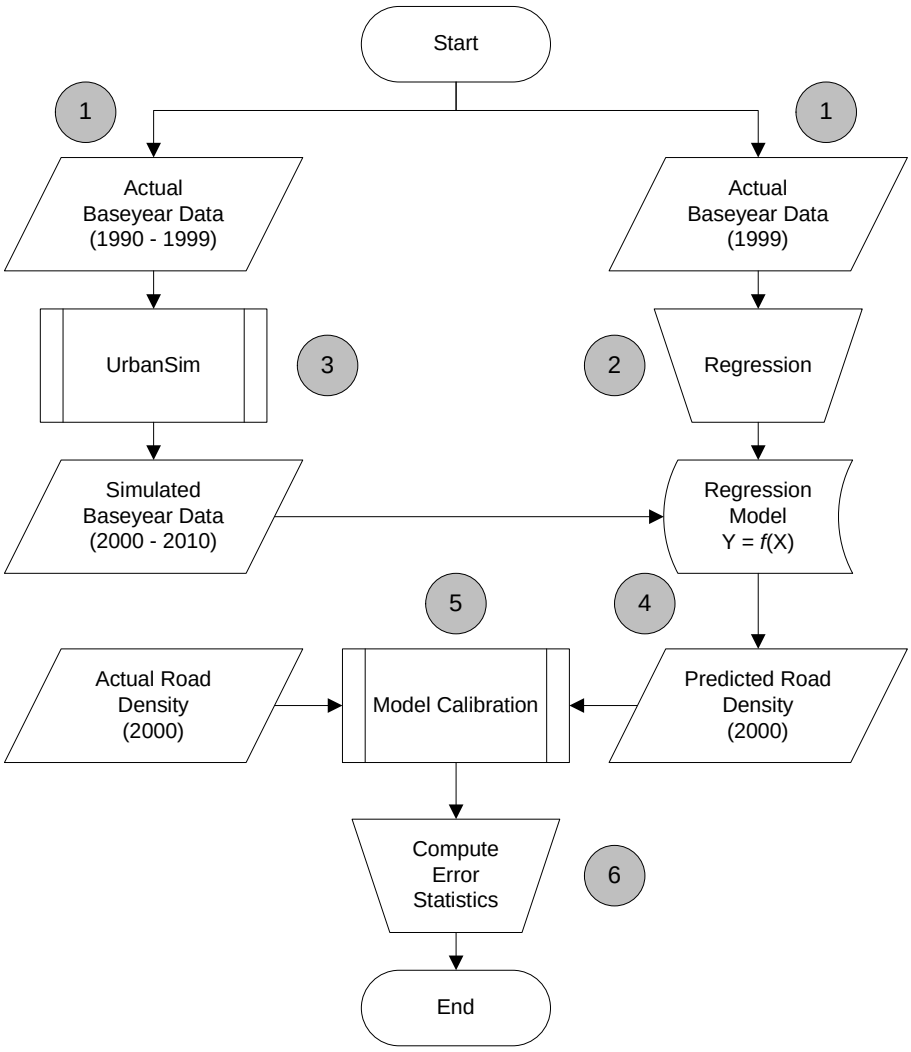

Figure 2: Modelling Process

The modelling process shown in Figure 2 is applied to all the study sites. The detailed explanation of each the different steps are provided below:

Step 1: Data Collection and Preparation

The entire process starts of with the collection of the data required for modelling the interaction between road transport and land use. In this project, the data, called the baseyear data, are collected for the period between 1990 and 2000. The baseyear data for the year 2000 is also used later as calibration data.

Step 2: Development of Regression Model

Using the year 1999 baseyear data, a multivariate regression model is constructed. This regression model will define the interaction between the dependent variable, road density, and the independent variables, which are the land use variables.

However, in order for Eq. (3) to be mathematically correct, any collinearity between the variables $\mathbf{E}_{i,t}$, $\mathbf{H}_{i,t}$ and $\mathbf{P}_{i,t}$ must be removed. Collinearity in this case is defined as the simultaneous interaction between the independent variables. The removal of collinear variables is necessary to ensure that the final set of independent variables is truly independent of each other, and the only interaction that exists is that between the independent variables and the dependent variable.

The removal of collinear independent variables is done through a statistic called the Variance Inflation Factor (VIF) where a large VIF value, i.e. VIF > 4, indicates the existence of collinearity in a specific variable. Thus, any independent variable that produces a VIF value greater than 4 is removed from the final regression equation. The computations of the VIF for each of the independent variables are done using SPSS statistical software.

As a final step in getting the regression equation, the independent variables that are not significant at $\alpha = 0.05$ level are removed. The rule adopted is that if the variable's significant value p is greater than $\alpha = 0.05$, then the variable is not significant and is removed from the final regression equation. Otherwise, if $p < \alpha$, the variable is considered significant and is retained in the final regression equation. Thus, the final regression equation should only

contain variables that are significant as well as variables that are not collinearly related with other independent variables.

Step 3: Prediction of Baseyear Data (2000 – 2010) using UrbanSim

Concurrent to the development of the regression model is the prediction of baseyear data for the period between year 2000 and 2010. This is done using UrbanSim.

Step 4: Prediction of Road Density (2000)

With the year 2000 simulated baseyear data obtained from UrbanSim in Step 3, the dependent variable, i.e. road density, for year 2000 is predicted using the regression model developed in Step 2.

Step 5: Model Calibration

In this step, the predicted road density for the year 2000 obtained in Step 4 is compared to the actual road density for the same year. This will give the prediction error – which will form the basis for model calibration. The aim of the calibration process is to observe the prediction error, defined as:

$$e_t = Y_t - \hat{Y}_t \tag{10}$$

Where e_t is the error term, Y_t is the actual data at time *t* while $\hat{Y}_t$ is the forecasted value for time *t* obtained by applying our regression model. Here, for calibration purposes, *t* is taken to be the year 2000.

Once the error term e_t has been obtained, several common accuracy measures can be applied to gauge the validity and reliability of the regression function $f(\bullet)$. In this study, the accuracy measures used are:

1. Mean Errors, $ME = \frac{1}{n}\sum_{n} e_{t,n}$ (11)

2. Mean Squared Error, $MSE = \frac{1}{n}\sum_{n} e_{t,n}^{2}$ (12)

3. Root Mean Squared Error, $RMSE = \sqrt{\frac{1}{n}\sum_{n} e_{t,n}^{2}} = \sqrt{MSE}$ (13)

It is with the above accuracy measures that the team ascertain the degrees of confidence in the regression models developed for the study sites.

Step 6: Computation of Error Statistics

The measures of accuracy described in the previous section identify the usefulness and reliability of the regression models in predicting future road density. However, these accuracy measures, e.g. MSE and RMSE, are not capable of testing the degree of closeness of the predicted values to the actual ones. In other words, given e_t as the prediction error, as defined in Eq. (10), there is a need to see if the error is close to zero, i.e.:

$$e_t \rightarrow 0 \tag{14}$$

To achieve this, a test that evaluates Eq. (14) must be adopted. This is accomplished through testing the following hypothesis:

$$\mathrm{H}_0 : \bar{e}_t = 0 \tag{15}$$

where:

$$\bar{e}_t = \mu_{t,a} - \mu_{t,f} \tag{16}$$

which states that the average prediction error $\overline{e}_t$ is the difference between the average actual values of road density at time *t*, denoted as $\mu_{t,a}$, and the average forecasted values of road density at time *t*, denoted as $\mu_{t,f}$. In this study, the time *t* is set at the year 2000 – the baseyear data.

In fact, the null hypothesis in Eq. (15) is a special case of the following, more general, null hypothesis:

$$\mathrm{H}_0 : \overline{e}_t = d_0 \tag{17}$$

where $d_0 = 0$. Therefore, such a hypothesis in Eq. (17) is not limited to testing $d_0 = 0$ but also capable of testing other differential d_0 values. In any hypothesis test, a null hypothesis is accompanied by an alternative hypothesis. For this study, the alternative hypothesis, denoted as H_1 is defined as:

$$\mathrm{H}_1 : \overline{e}_t \neq d_0 \tag{18}$$

or, when Eq. (16) is replaced into Eq. (18), the alternative hypothesis above can also be written as:

$$\mathrm{H}_1 : \mu_{t,a} - \mu_{t,f} \neq d_0 \tag{19}$$

It is noted from Eq. (19) that this test of equality is a two-tailed test. Also, if d_0 is set at 5%, i.e. $d_0 = 5$, the alternative hypothesis of Eq. (18) is testing if the difference between the actual road density and the forecasted road density is not equal to 5%.

The test of equality will be tested at a significance level of 0.05, i.e. α = 0.05, corresponding to a confidence level of 95%. The setting of α = 0.05 also means that the study only allows a maximum of 5% probability of committing Type I error, which is the error of rejecting the null hypothesis of Eq. (17) when in fact it is true.

The test statistic employed for this test of equality is the two samples *t*-test with unequal and unknown population variances. This *t* test statistics has the following form:

$$t = \frac{\left(\bar{x}_{t,a} - \bar{x}_{t,f}\right) - d_0}{\sqrt{\left(\frac{s_{t,a}^2}{n_{t,a}}\right) + \left(\frac{s_{t,f}^2}{n_{t,f}}\right)}} \tag{20}$$

with the degrees of freedom ν given as:

$$\nu = \frac{\left(\frac{s_{t,a}^2}{n_{t,a}}\right)^2 + \left(\frac{s_{t,f}^2}{n_{t,f}}\right)^2}{\left(\frac{s_{t,a}^2}{n_{t,a}-1}\right)^2 + \left(\frac{s_{t,f}^2}{n_{t,f}-1}\right)^2} \tag{21}$$

where:

t = computed student's *t* value

$\bar{x}_{t,a}$ = mean actual road density at time *t*, *t* = 2000

$\bar{x}_{t,f}$ = mean forecasted road density at time *t*, *t* = 2000

d_0 = postulated difference between mean actual and forecasted road densities at time *t*, *t* = 2000

$n_{t,a}$ = sample size for actual road density at time *t*, *t* = 2000

$n_{t,f}$ = sample size for forecasted road density at time *t*, *t* = 2000

$s_{t,a}^2$ = sample variance for actual road density at time *t*, *t* = 2000

$s_{t,f}^2$ = sample variance for forecasted road density at time *t*, *t* = 2000

The value computed using Eq. (20) determines the significance value, denoted as *p*, which is used to guide the decision of either accepting or rejecting the null hypothesis H_0. The rule used is that if $p > \alpha$, then the null hypothesis of Eq. (17) is accepted. However, if $p < \alpha$, then the null hypothesis is rejected subjected to $\alpha = 0.05$ or 5% probability of committing Type I error.

If the decision is to accept the null hypothesis, then the test concludes that there is evidence to show that the mean forecasted road density is statistically equal to the mean of actual road density. This is similar to saying that d_0 is indeed statistically equals to zero.

On the other hand, if the null hypothesis is rejected (i.e. accepting the alternative hypothesis H_1), then the test has concluded, subject to $\alpha = 0.05$, that the evidence shows that the mean forecasted road density is statistically not equal to the mean actual road density. In the event that the null hypothesis is rejected, the test of equality proceeds to determining the value d_0 that will finally bring about the acceptance of the null hypothesis.

4 Findings

In this section, the regression model for each the study sites are described together with their statistical measures. The equation is obtained by regressing the dependent variables, which is road density, against a set of 18 independent variables listed earlier. These regression models described are for the two study sites – the town of Johor Bahru and Kuantan.

4.1 Johor Bahru

The process of identifying the regression equation starts of with checking for multicollinearity among the independent variables. As stated in Step 2 of Section 3, the rule adopted is that if the Variance Inflation Factor (VIF) is greater than four, i.e. VIF > 4, the variable is considered to be collinearly related with other independent variables. For Johor Bahru, the result of this step is shown in Table 1.

From Table 1, it can be seen that the following variables is suspected to be collinearly related with other independent variables since their VIF values are greater than 4, i.e. VIF > 4:

- Residential Improvement Value
- Residential Land Value

- Residential Units
- Total Persons
- Total Workers
- Total Cars

Therefore, based on the VIF statistics, the above independent variables will be eliminated from the list of independent variables.

Table 1: Collinearity statistics for Johor Bahru

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	2.330	.977		2.386	.018		
	RESIDENTIAL_IMPROVEMENT_VALUE	7.100E-05	.000	.242	1.730	.085	.104	9.620
	RESIDENTIAL_LAND_VALUE	-4.91E-02	.064	-.110	-.773	.440	.100	9.981
	COMMERCIAL_SQFT	-5.57E-06	.000	-.027	-.579	.563	.947	1.056
	INDUSTRIAL_SQFT	4.445E-06	.000	.066	1.370	.172	.865	1.156
	GOVERNMENTAL_SQFT	3.305E-05	.000	.095	1.192	.235	.319	3.137
	RESIDENTIAL_UNITS	1.237E-02	.024	.763	.526	.599	.001	1036.411
	COMMERCIAL_IMPROVEMENT_VALUE	2.231E-05	.000	.089	1.276	.203	.419	2.384
	INDUSTRIAL_IMPROVEMENT_VALUE	3.358E-06	.000	.075	1.616	.108	.948	1.055
	GOVERNMENTAL_IMPROVEMENT_VALUE	-3.51E-08	.000	-.036	-.450	.653	.322	3.102
	NONRESIDENTIAL_LAND_VALUE	1.637E-05	.000	.084	1.157	.249	.388	2.574
	FRACTION_RESIDENTIAL_LAND	18.832	5.897	.284	3.194	.002	.257	3.888
	sum(persons)	1.541E-03	.006	.502	.261	.795	.001	1828.930
	sum(workers)	-6.90E-03	.015	-.699	-.462	.645	.001	1128.390
	avg(income)	1.507E-03	.000	.184	3.451	.001	.711	1.407
	sum(cars)	-1.64E-02	.013	-.517	-1.271	.205	.012	81.509

a. Dependent Variable: PERCENT_ROADS

In the following step, the remaining independent variables are tested for their significance. The hypothesis for this test is:

$$H_0 : \beta_i = 0$$

which specifically test if the coefficient of the independent variable *i*, β_i, is equal to zero. The alternative to this hypothesis is:

$$\mathrm{H}_1 : \beta_i \neq 0$$

which shows that the coefficient β_i is significantly different from zero. This test is conducted at $\alpha = 0.05$ which denotes a confidence level of 95%. The rule used in this significance test is that if the significance values, i.e. the *p*-value, is less than the significance level α then the null hypothesis of $\mathrm{H}_0 : \beta_i = 0$ is rejected. Table 2 shows the significance value (i.e. the *p*-value) for each of the remaining independent variables.

Table 2: Significance level for the independent variables

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.852	1.035		3.722	.000
	COMMERCIAL_SQFT	-1.45E-06	.000	-.007	-.138	.890
	INDUSTRIAL_SQFT	4.029E-06	.000	.060	1.152	.251
	GOVERNMENTAL_SQFT	4.886E-05	.000	.141	1.625	.106
	COMMERCIAL_IMPROVEMENT_VALUE	5.123E-05	.000	.204	2.889	.004
	INDUSTRIAL_IMPROVEMENT_VALUE	4.345E-06	.000	.097	1.932	.055
	GOVERNMENTAL_IMPROVEMENT_VALUE	-5.56E-08	.000	-.057	-.656	.513
	NONRESIDENTIAL_LAND_VALUE	2.407E-05	.000	.123	1.792	.074
	FRACTION_RESIDENTIAL_LAND	26.193	3.889	.395	6.735	.000
	avg(income)	1.293E-03	.000	.158	2.799	.006

a. Dependent Variable: PERCENT_ROADS

Based on Table 2, it was found that the following variables are not significant as their corresponding coefficients have *p*-values greater than $\alpha = 0.05$:

- Commercial Square Feet
- Industrial Square Feet
- Government Square Feet

- Industrial Improvement Value
- Governmental Improvement Value
- Non-residential Land Value

Once the above non-significant independent variables are removed, the model are left with only the following independent variables as shown in Table 3:

- Commercial Improvement Value
- Fraction Residential Land
- Average Household Income

Table 3: List of significant independent variables for Johor Bahru

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.072	1.045		3.896	.000
	COMMERCIAL_IMPROVEMENT_VALUE	7.392E-05	.000	.294	5.712	.000
	FRACTION_RESIDENTIAL_LAND	27.026	3.870	.407	6.984	.000
	avg(income)	1.387E-03	.000	.170	2.974	.003

a. Dependent Variable: PERCENT_ROADS

Based on the information in Table 3, the final regression model for Johor Bahru, having only the significant independent variables, is given as:

$$\textbf{Road Density} = 4.072 + 7.329E\text{-}5 * \textbf{P7} + 27.026 * \textbf{P4} + 1.387E\text{-}3 * \textbf{H3} \quad (22)$$

where:

P7 = Commercial Improvement Value

P4 = Fraction Residential Land

H3 = Average household income

4.1.1 Model Calibration

Given that the final regression model for Johor Bahru has been identified, the process of calibrating the regression model against the baseyear data of 2000 can begin. This is important as it provide information on how good the regression model is at predicting road density in Johor Bahru.

Based on baseyear data at year 2000, using Eqs. (11) - (13), the calibration yields the following accuracy measures for the regression model of Johor Bahru:

Mean Error, ME	= -3.5469
Mean Squared Error, MSE	= 104.5491
Root Mean Squared Error, RMSE	= 10.2249

The Mean Error (ME) shows a low error level at only -3.5%. The negative sign in the ME indicates that the regression model of Eq. (22), on the average, tends to over-estimate the road density level in Johor Bahru. Also, the ME is lower in magnitude than the RMSE since the different signs (i.e. positive and negative) can potentially cancel each other out.

However, ignoring the sign by using a sign-free measures like MSE and RMSE, it is found that the regression model for Johor Bahru, in fact, under-predicts the actual road density for Johor Bahru by approximately 10%.

4.1.2 Testing the Hypothesis of Equality

In this final step, the model is tested to see if the difference between the actual and the predicted road density, on the average, is statistically similar or is it significantly different. The procedure for this test is described in detail in Step 6 of Section 3.

The complete result of the test of hypothesis is described below, using the usual 5-step approach of hypothesis testing:

1. $\mathrm{H}_0 : \mu_{t,a} - \mu_{t,f} = 0$

   $\mathrm{H}_1 : \mu_{t,a} - \mu_{t,f} \neq 0$
2. $\alpha = 0.05$
3. Using the test statistics defined in Eqs. (20) - (21). With *n* = 237 grids, the degrees of freedom is equal to 472.
4. The computed *t*-statistic is -4.3472. The *p*-value is 1.72E-5.
5. Since the *p*-value is less than $\alpha = 0.05$, the null hypothesis of equality is therefore rejected. The conclusion is that there is a significant difference between the actual and the predicted road density for Johor Bahru. Since the computed t-statistic is a negative value, it is suspected that the mean predicted road density is greater than the mean actual road density.

Since the null hypothesis is rejected when d_0 = 0 [refer Eq. (17)] signifying equality between the actual and the predicted road density, it is therefore necessary to determine the degree of d_0 that would bring about the acceptance of the null hypothesis $\mathrm{H}_0 : \mu_{t,a} - \mu_{t,f} = d_0$.

In the next step, the value d_0 is decreased to 2.5, i.e. d_0 = -2.5, resulting in the following test of hypothesis:

1. $\mathrm{H}_0 : \mu_{t,a} - \mu_{t,f} = -2.5$

   $\mathrm{H}_1 : \mu_{t,a} - \mu_{t,f} \neq -2.5$
2. $\alpha = 0.05$
3. Using the test statistics defined in Eqs. (20) - (21). With *n* = 237 grids, the degrees of freedom is equal to 472.
4. The computed t-statistic is 1.4873. The *p*-value is 0.1377.
5. Since the *p*-value is greater than $\alpha = 0.05$, the null hypothesis of equality is therefore accepted. The conclusion is that the difference between the actual and the predicted road density for Johor Bahru is at 2.5%.

Based on the two tests of hypothesis above, it is concluded that the mean predicted road density by the final regression model of see Eq. (22) is not equal to the mean actual road density. However, the difference between them is no more than 2.5%.

4.2 Kuantan

The same process explained earlier for the town of Johor Bahru is repeated again for the town of Kuantan. For simplicity sake, only the final regression model for Kuantan is presented below:

$$\textbf{Road Density} = 1.721E\text{-}5 * \textbf{P5} + 8.555E\text{-}7 * \textbf{P8} + 3.788E\text{-}3 * \textbf{H3} \quad (23)$$

where:

P5 = Commercial Square Feet

P8 = Industrial Square Feet

H3 = Average household income

4.2.1 Model Calibration

Based on baseyear data at year 2000, the calibration yields the following accuracy measures for the regression model of Kuantan:

Mean Error, ME	= -0.4171
Mean Squared Error, MSE	= 25.5916
Root Mean Squared Error, RMSE	= 5.0588

From the above statistics, it can be seen that the Mean Error yields an exceptionally low value. This can be due the possibility that the negative and the positive prediction error *e* cancels each other out. However, looking at the resulting value, what can be said is that, on average, the regression model of Eq. (23) for Kuantan produces predicted road density that tends to over-estimate the actual road density.

To circumvent the issue of negative and positive values cancelling each other out, a sign-free accuracy measures like the MSE and the RMSE are used instead. The values for these two statistics (MSE and RMSE) shows that, ignoring the sign, the regression model produces predicted road densities for Kuantan that tend to underestimate the actual road density, which

is quite the reverse than from what has been measured by the Mean Error statistic. However, the prediction error is not more than 5% for Kuantan.

4.2.2 Hypothesis of Equality

The calibration for Kuantan with reference to the baseyear 2000 data undoubtedly shows a promising result when the capability of the resulting regression equation is concerned. However, there is still a need to test the hypothesis that, even though the accuracy measures shows that there is some difference between the actual and the predicted road density, these difference are not significant. This aim is achieved through the following test of hypothesis when $d_0 = 0$ and tested at $\alpha = 0.05$ (corresponding to a 95% confidence level):

1. $\mathrm{H}_0 : \mu_{t,a} - \mu_{t,f} = 0$

   $\mathrm{H}_1 : \mu_{t,a} - \mu_{t,f} \neq 0$
2. $\alpha = 0.05$
3. Using the test statistics defined in Eqs. (20) - (21) with *n* = 536 grids.
4. The computed *t*-statistic is -1.1040. The *p*-value is 0.2699.
5. Since the *p*-value is greater than $\alpha = 0.05$, the null hypothesis of equality is failed to be rejected. The conclusion is that there is no significant difference between the actual and the predicted road density for Kuantan.

Based on the result of the above test of hypothesis, it can be concluded that the final regression model of Eq. (23) is robust at predicting the road density for Kuantan. This conclusion is supported by the measures of accuracy and the test of hypothesis.

5 Discussion and Conclusion

Based on the two regression models for each of the study site, there is a great promise of developing an empirical model that explains the dynamic interaction between road density and land uses. The results shows that the interaction model identified for each of the two sites – Johor Bahru and Kuantan – is parsimonious in that only three out of the possible 18 land

uses variables are found to significantly describe the interaction between the dependent variable (i.e. the road density) and the independent variables. However, the two models – each for Johor Bahru and Kuantan – are not similar which indicates that different level of urban hierarchy might play some roles in the definition of the interaction model. It is hopeful that further study can be conducted using more towns to ascertain this hypothesis.