Introductory Econometrics ECOM20001

Assignment 1

Kwang Bum (Kevin) KIM 692192

Tutor Name: Chin Yong QUEK

Tutorial: 12pm ~ 1pm Spot 3013

Question 1

1.a. i)

Using the OLS estimates,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \rightarrow \hat{y}_i = b_0 + b_1 x_i \ldots \textit{(1)}$$

and

$$c_1 y_i = \gamma_0 + \gamma_1 (c_2 x_i) + v_i \rightarrow c_1 \hat{y}_i = g_0 + g_1 (c_2 x_i) \ldots \textit{(2)}$$

Now, $c_1 \times (1) = c_1 \hat{y}_i = c_1 b_0 + b_1 c_1 x_i$ ...*(3)*

Hence, *(2)* is equal to *(3)*. Now equating coefficients,

$$c_1 b_0 + b_1 c_1 x_i = g_0 + g_1 (c_2 x_i)$$

$\therefore \boldsymbol{g_0 = c_1 b_0}$ and $b_1 c_1 x_i = g_1 (c_2 x_i) \rightarrow \boldsymbol{g_1} = \frac{b_1 c_1 x_i}{c_2 x_i} = \boldsymbol{\frac{c_1}{c_2} b_1}$ as required.

1.a. ii)

Using the OLS estimates,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \rightarrow \hat{y}_i = b_0 + b_1 x_i \ldots \textit{(4)}$$

and

$$(c_1 + y_i) = \theta_0 + \theta_1 (c_2 + x_i) + \omega_i \rightarrow (c_1 + \hat{y}_i) = q_0 + q_1 (c_2 + x_i) = q_0 + q_1 c_2 + q_1 x_i \ldots \textit{(5)}$$

Now, $adding\ c_1 in\ both\ sides\ of\ (4) will\ give: (c_1 + \hat{y}_i) = c_1 + b_0 + b_1 x_i$ ...*(6)*

Hence, *(5)*is equal to *(6)*. Now equating coefficients,

$$q_0 + q_1 c_2 + q_1 x_i = c_1 + b_0 + b_1 x_i$$

$$\therefore\ q_1 x_i = b_1 x_i \rightarrow \boldsymbol{q_1 = b_1}\ and\ q_0 + q_1 c_2 = c_1 + b_0 \rightarrow q_0 = b_0 + c_1 - q_1 c_2 \rightarrow$$

$\boldsymbol{q_0 = b_0 + c_1 - c_2 b_1}$(using $q_1 = b_1$) as required.

1.a. iii)

Using variance properties

$\text{var}(\sigma_\varepsilon) = \sigma_\varepsilon^2$

$\text{var}(v_i) = \sigma_v^2 = \text{var}(c_1 \varepsilon_i) = c_1^2 var(\varepsilon_i) = c_1^2 \sigma_\varepsilon^2$

$\text{var}(\omega_i) = \sigma_\omega^2 = var(c_1 + \varepsilon_i) = \sigma_\varepsilon^2$

$$\therefore \boldsymbol{\sigma_v^2 = c_1^2 \sigma_\varepsilon^2 = c_1^2 \sigma_\omega^2}$$

1.b)

Using the OLS estimates,

$$\log(y_i) = \beta_0 + \beta_1 x_i + \varepsilon_i \rightarrow \log(\widehat{y_i}) = b_0 + b_1 x_i \ldots \textit{(1b)} \text{for } y_i > 0$$

and

$$\log(c_1 y_i) = \gamma_0 + \gamma_1 x_i + v_i \rightarrow \log(c_1 \widehat{y_i}) = g_0 + g_1 x_i \ldots \textit{(2b)} \text{ for } c_1 > 0$$

add $\log(c_1)$ in both sides of (1b) will give: $\log(c_1 \widehat{y_i}) = \log(c_1) + b_0 + b_1 x_i$ ...*(3b)*

Hence, *(2b)*is equal to *(3b)*.Now equating coefficients,

$$g_0 + g_1 x_i = \log(c_1) + b_0 + b_1 x_i$$

$\therefore\ \boldsymbol{g_0 = \log(c_1) + b_0}andg_1 x_i = b_1 x_i \rightarrow \boldsymbol{g_1 = b_1}$as required.

1.c)

Using the OLS estimates,

$$y_i = \beta_0 + \beta_1 \log(x_i) + \varepsilon_i \rightarrow \hat{y}_i = b_0 + b_1 \log(x_i) \ldots \textit{(1c)} \text{for} x_i > 0$$

and

$$y_i = \gamma_0 + \gamma_1 \log(c_2 x_i) + v_i \rightarrow \hat{y}_i = g_0 + g_1 \log(c_2 x_i) \text{ ...(2c) for } c_2 > 0$$

expanding *(2c)*using logarithm property gives: $\hat{y}_i = g_0 + g_1 \log(c_2) + g_1 log(x_i)$ ...*(3c)*
Hence, *(1c)*is equal to *(3c)*. Now equating coefficients,

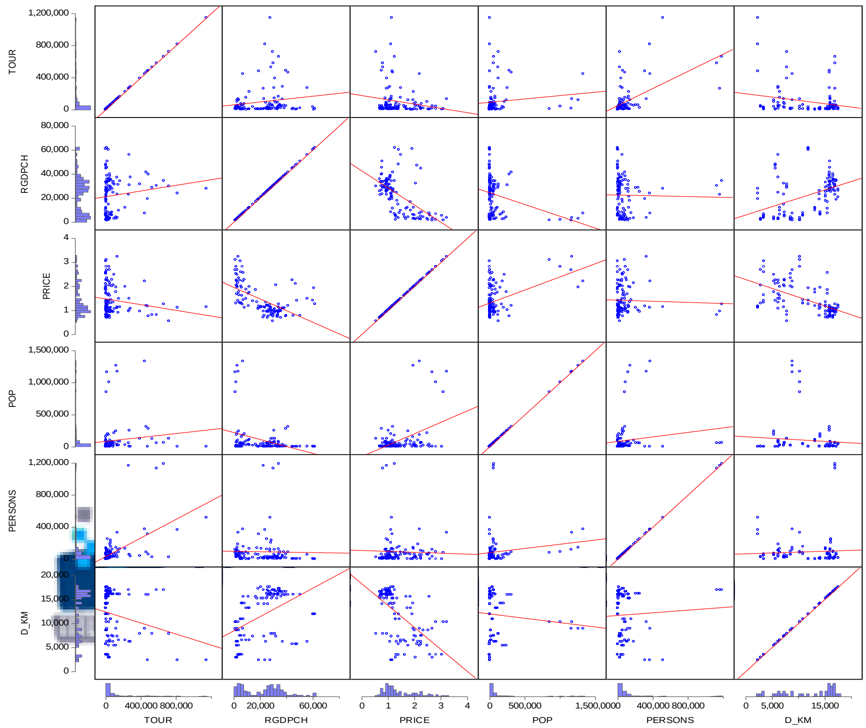
$$b_0 + b_1 \log(x_i) = g_0 + g_1 \log(c_2) + g_1 log(x_i)$$

$\therefore$ $\boldsymbol{b_0 = g_0 + g_1 \log(c_2)}$ and $b_1 \log(x_i) = g_1 \log(x_i) \rightarrow \boldsymbol{b_1 = g_1}$

Question 2

2.a

Scatter plot matrix



Screen-shot from EViews

1. Income vs. Tourist level

There seems to be no obvious relationship between income and the number of tourists, as the numbers of observations are located at the bottom of the graph, showing slight positive to no slope. This implies that the number of tourists is not truly dependent on tourist countries' income level. However, some of the observations show that the number of tourists are relatively high when the income level is around the \$US 23000~28000 (GDP per Capita). This is evident, as some of the observations are located at high tourist level.

Now, the independent variable'scontribution towards the independent variable (tour) can be explained using simple linear regression and finding the goodness of fit, which is coefficient of determination (R^2).

After running simple linear regression by setting the equation: $\widehat{tour} = b_1 + b_2 rgdpch$

R^2 = 0.021204 →this implies that approximately 2.12% variation in the number of tourists is explained by the income variable

2. Price vs. Tourist level

Based on scatter plot, the observations show convex trend, as the number of tourists are relatively high at low price level and fall with diminishing slope as price level increases. Also, the observations are concentrated at low price level with low number of tourists. Low price level could explain that the touristcountries' economy is already developed and the basket price is relatively more expensive than in Australia. This notion can be vice versa. Hence, price trend shows negative slope, as expected. This is because, the tourists from the developed nations will have increased purchasing power in Australia compared to tourists from undeveloped/developing countries that will exhibit decreased purchasing power in Australia.

The R^2 for the estimated model: $\widehat{tour} = b_1 + b_2 price$→ R^2= 0.031741
This implies that approximately 3.17% variation in the number of tourists is explained by the price variable.

3. Population vs. Tourist level

Using scatter plot, large numbers of observations are located at low population(in tourist's home country) with low number of tourists level with some observations are located at the very high level of population level.This concludes that the relationship between the population levels in the tourists' country with the number of tourists cannot be determined clearly.Hence, the regression line may not be accurate. But, as most observations were located at low population level, one might conclude that the tourists, regardless of the size,are mostly from the low populated country.
The R^2 for the estimated model: $\widehat{tour} = b_1 + b_2 pop$→ R^2= 0.012814
This implies that approximately 1.28% variation in the number of tourists is explained by the population variable.

4. Persons vs. Tourist level

Analysing scatter plot, the observations show high concentration at low level of number of persons in Australia born in tourist's country with low tourist level. Other observations are sparse but most of the observations are located at low 'persons' level. This implies that low or even large numbers of tourists are observed at low level of 'persons' – low number of persons in Australia born in tourist's country.
The R^2 for the estimated model: $\widehat{tour} = b_1 + b_2 persons$→ R^2= 0.310421
This implies that approximately 31.04% variation in the number of tourists is explained by the persons variable.

5. Distance(km) vs. Tourist level

The observations are relatively more concentrated at low number of tourists, but some of the observations show vertical variation sporadically. For example, some of the observations containing large number of tourists are located at variousdistances to Canberra-tourist's capital city, mainly, around 2300km, 7800km and 16000km.But, the vertical variation's spread tends to become smaller, as the distance increases. Nevertheless, it is still difficult to find the relationship between the distance and corresponding number of tourists from tourists' countries.
The R^2 for the estimated model: $\widehat{tour} = b_1 + b_2 d\_km$→ R^2= 0.047547
This implies that approximately 4.75% variation in the number of tourists is explained by the population variable.

Based on the goodness of fit, the regressor, *persons*, can explain the model most confidently. It is supportedby having highest R^2value relative to other variables. But, this statementcannot be concluded with confidence yet. As can be seen from the scatter matrix, some of the red linear regression line represents the observations inaccurately. The previous models have used simple linear model, whereas more accurate conclusion can be reached by embracing non-linear regressionwith multiple factors (variables).

By using*simpleone variable* model for this question, the possible problems are:
1. the coefficients of the parameters can be incorrect, as there could be omitted variable biasedness.
2. there could be aninteractive independent variables, which cannot be captured using a simple model.
2. better fit can be made by estimating using non-linear model (e.g. quadratic).

EXPECTATIONS on coefficient sign

Finally, the linear regression from EViews (red linear line) slope shows that Income, population and persons coefficients should be positive, whilst price and distance (km) coefficients should be negative (economic reasoning continued at **2.b i)**)

2.b.i)

Using the estimated model $\widehat{tour} = b_0 + b_1 rgdpch + b_2 price + b_3 pop + b_4 d_{km} + b_5 persons$

1. by restricting year = 1991,

Year	B_0 (Constant)	B_1 (Income)	B_2 (Price)	B_3 (Population)	B_4 (Distance)	B_5 (persons)	R^2
1991	300995.5	1.873425	-95566.38	0.137829	-14.46339	0.248351	0.451901
p-value	0.0012	0.1974	0.0065	0.1050	0.0008	0.0061	
Significance	Significant	-	Significant	-	Significant	Significant	

at $\alpha = 0.05$ $for\ 2\ tailed\ test,$
b_2 , b_4 *and* b_5 have p-values (two-tailed) of 0.0065, 0.0008 and 0.0061 respectively. Hence, they are significantly different from zero at 0.05 level and this implies that the price, distance and persons coefficients cannot be zero at 5% level.
Furthermore, the R^2 for this model is 0.451901, which states that: approximately 45.19% variation in the number of tourists is explained by the regressors (income, price, population, distance_km and persons).

2. by restricting year = 2000

Year	B_0 (Constant)	B_1 (Income)	B_2 (Price)	B_3 (Population)	B_4 (Distance)	B_5 (persons)	R^2
2000	637985	1.322315	-237458.6	0.223461	-26.98134	0.535742	0.619785
p-value	0.0001	0.5147	0.0009	0.0302	0.0000	0.0001	
Significance	Significant	-	Significant	-	-	-	

at $\alpha = 0.05$ $for\ 2\ tailed\ test,$
b_2 , b_3, b_4 *and* b_5 have p-values of 0.0009, 0.0302, 0.0000 and 0.0001 respectively. Hence, they are significantly different from zero at 0.05 level and this implies that the price, population, distance and persons coefficients cannot be zero at 5% level.
Furthermore, the R^2 for this model is 0.619785, which states that: approximately 61.98% variation in the number of tourists is explained by the regressors (income, price, population, distance_km and persons).

3. by restricting year = 2010

Year	B_0 (Constant)	B_1 (Income)	B_2 (Price)	B_3 (Population)	B_4 (Distance)	B_5 (persons)	R^2
2010	676311.8	2.563547	-223308.5	0.238545	-31.77598	0.694172	0.752482
p-value	0.0000	0.1192	0.0002	0.0085	0.0000	0.0000	
Significance	Significant	-	Significant	Significant	Significant	Significant	

at $\alpha = 0.05\ for\ 2\ tailed\ test$,
b_2 , b_3, b_4 and $b_5$have p-values of 00002, 0.0085, 0.0000 and 0.0000 respectively. Hence, itis significantly different from zero at 0.05 level and this implies that the price, population, distance and persons coefficients cannot be zero at 5% level.
Furthermore, the R^2 for this model is 0.752482, which states that: approximately 75.25% variation in the number of tourists is explained by the regressors (income, price, population, distance_km and persons).

Comparing 1991 and 2000, the coefficients for income, price, population and distance(km) have decreased, whilst the intercept term, population and persons have increased. Overall, the total number of tourists visited Australia has increased. This can be explained by near two-fold increase in 'mean dependent var' from EViews Output. The main contribution based on the model may be substantial increase in the intercept term by approximately 340,000 tourists regardless of the factors. Sydney Olympics held in year 2000 can explain this change. Olympics could explain the increase in the constant term, which is not accounted in any of the parameters from the model.

The sign for parameter estimate is that:

1. positive income coefficient

This is consistent as expected. The tourists' country'sGDP per capita will likely to have positive effect on the number of tourists from the country. This is because higher average income will likely to lead tourists to make overseas travel. Thus, citizens from a country with relatively higher GDP per capita will more likely to become tourists in Australia.

2. negative price coefficient

This is consistent as expected.Price variable in this model is defined as PPP(AU)/Exchange Rate (AU).Based on this, price > 1 implies that the tourists' countries have price of basket of goods cheaper than in Australia. This is likely to happen in countries that have less developed economies. Hence, citizens that live in a country that has relatively cheap good price will need to pay more to purchase the same bundle of goods in Australia. Therefore, the negative coefficient isalso logical in a way that the citizens from other countries will less likely to come to Australia as tourists. This notion is applicable in vice versa, where the citizens from countries that have relatively more expensive basket good will more likely to come to Australia.

3. positive population coefficient

This is consistent as expected. If the population in tourists' countries are relatively large (e.g. China, India), the total number of tourists that will travel to Australia will naturally increase given other conditions.

4. negative distance coefficient

This is consistent as expected. If the tourists' country is further from Australia, due to time devotion, increase in transportation costs (airplane costs) and the likely option to travel countries that are in close proximity, the chances of tourists coming to Australia will decrease and hence, the total number of tourists will decrease.

5. positive person coefficient

This is consistent as expected. If the number of persons in Australia born in tourist's country increases, they are more likely to travel back and forth. This category will mainly include international students, non-Australia-born permanent residents/Australian citizens. As these people will learn or own jobs in Australia, yet relatives can be in their homeland, they will more likely to become categorised as tourists to Australia.

2b ii)

Elasticity using mean of the data

→ $elasticity = \frac{dy}{dx} \times \frac{\bar{X}}{\bar{Y}}$

Using EViews, the elasticity at means for each coefficient are as follow:

Scaled Coefficients
Date: 05/02/15 Time: 09:16
Sample: 1 114 IF YEAR=2000
Included observations: 38

Variable	Coefficient	Standardized Coefficient	Elasticity at Means
C	637985.7	NA	5.516270
RGDPCH	1.322315	0.102432	0.261756
PRICE	-237458.6	-0.671670	-2.612460
POP	0.223461	0.287941	0.198055
D_KM	-26.98134	-0.679998	-2.754873
PERSONS	0.535742	0.508314	0.391252

Scaled Coefficients
Date: 05/02/15 Time: 09:14
Included observations: 38

Variable	Coefficient	Standardized Coefficient	Elasticity at Means
C	300995.5	NA	5.189812
RGDPCH	1.873425	0.209748	0.605182
PRICE	-95566.38	-0.538911	-2.420211
POP	0.137829	0.257869	0.218284
D_KM	-14.46339	-0.592508	-2.944865
PERSONS	0.248351	0.390456	0.351798

For year 1991,
Elasticity of distance (km) = -2.944865 (elastic)
Elasticity of price = -2.420211 (elastic)
Elasticity of income = 0.605182 (inelastic)

For year 2000,
Elasticity of distance(km) = -2.754873 (elastic)
Elasticity of price = -2.612460 (elastic)
Elasticity of income = 0.261756 (inelastic)

Scaled Coefficients
Date: 05/02/15 Time: 09:16
Sample: 1 114 IF YEAR=2010
Included observations: 38

	Standardized	Elasticity

Variable	Coefficient	Coefficient	at Means
C	676311.8	NA	5.042350
RGDPCH	2.563547	0.183857	0.491644
PRICE	-223308.5	-0.614252	-2.519144
POP	0.238545	0.297623	0.200530
D_KM	-31.77598	-0.706039	-2.797618
PERSONS	0.694172	0.649649	0.582237

For year 2010,
Elasticity of distance(km) = -2.797618 (elastic)
Elasticity of price = -2.519144 (elastic)
Elasticity of income = 0.491644 (inelastic)

Over 19 years, the elasticity of distance, price and income have numerically fluctuated. But, the signs or the elasticity in terms of whether the factor has elastic influence or inelastic influence has been maintained. The reasoning can be derived from whether the factor has direct or indirect impact to the number of tourists. For example, income elasticity is positive but inelastic. This can be due to the indirect impact that 1% point increase in income of a typical citizenwill not necessarily spend the extra 1%point purely on travel (more likely to have direct relationship with disposable income or propensity to save for a trip, α).
Moreover, the elasticity ofdistance and price has a direct and strong impact on deciding the number of tourists coming to Australia.If the distance increases by 1% point, the opportunity costs (e.g. transportation costs, time and increase in option to travel closer country) will increase to a large extent (>1% point)
Lastly, the elasticity ofprice also has large impact on the independent variable. Increase in 1% point of price variable will discourageoverseas citizens to come to Australia (by >1% point), asprice for the basket of goods and services in Australia became more expensive than before.

2b iii)
Use of distance parameter in terms of miles or kilometres will not change any outcome values. This can be shown in two ways: consistency in the total number of tourists (dependent values) and the consistency in the elasticity.
1.
Let's say 1 mile ≈ 1.61km and use simple linear model for simplification purposes

$$\hat{y}_i = b_0 + b_1 d\_miles_i \text{ ...(1)}$$

From this estimate, d_miles is the regressor and b_1 is corresponding coefficient. Now, say the regressor data is in kilometres instead of miles. The new estimate equation, using d_km as regressor:

$$\hat{y}_i = b_0 + b_1' d\_km_i$$

Mathematically, d_miles = 1.61*d_km. The equation becomes:

$$\hat{y}_i = b_0 + b_1' 1.61(d\_miles_i) \text{ ...(2)}$$

Comparing (1) with (2), the parameter has been multiplied by 1.61. In order to maintain the idea *(1)* = (2), b_1' has to be:

$$b_1' = \frac{b_1}{1.61}$$

Now, the equation becomes:

$$\hat{y}_i = b_0 + \frac{b_1}{1.61} 1.61(d\_miles_i)$$

As 1.61 cancels out, the outcome (independent variable) value must be unchanged.

2.

using the linear relation and the definition: $\beta_{mile} = 1.6\beta_{km}$,

$$tour = \beta_1 + \beta_2 rgdpch + \beta_3 price + \beta_4 pop + \beta_5 d_{km} + \beta_6 persons + e_i$$

By writing d_miles in terms of d_km,

$$=tour = \beta_1 + \beta_2 rgdpch + \beta_3 price + \beta_4 pop + 1.6\beta_5 \frac{d_{km}}{1.6} + \beta_6 persons + e_i$$

Hence, the elasticity of the variable in miles at the mean: $\frac{dy}{dx} \times \frac{\bar{x}}{\bar{y}}$

$\rightarrow 1.6\beta_5 \left(\frac{\bar{x}(in\ miles)}{\bar{y}}\right) = 1.6\beta_5 \left(\frac{\bar{x}(in\ km)}{1.6\bar{y}}\right) = \beta_5 \left(\frac{\bar{x}}{\bar{y}}\right)$

$$\therefore elasticity\ is\ consistent; does\ not\ change$$

2.c

As mentioned before, there are limitations or possible problems using linear regression, where in fact, some of the variables could generate non-linear influence. This could lead in acquiringinaccurate or misleadingconclusion. In order to test the non-linearity of each variable, it is important to include the 'square (^2)' all variables and examine the corresponding p-values for significance.

$$\text{New model: } \widehat{tour} = b_0 + b_1 rgdpch + b_1' rgdpch^2 + b_2 price + b_2' price^2 + b_3 pop + b_3' pop^2 + b_4 d_{km} + b_4' {d_{km}}^2 + b_5 persons + b_5' persons^2$$

Based on the output from EViews (see below)

In order to test the significanceof the non-linear (quadratic) model, hypothesis testing can be done.

$$H_0: \beta_j = 0 \ \& H_1: \beta_j \neq 0$$

For 10% significance test, the p-value is greater than 0.1 for price[2]and d_km[2].This implies that these variables fail to reject the null that the coefficient for the squared variable is in fact zero. Hence, price and distance have no non-linear (quadratic) influence on the number of tourists. Note that this does not necessarily mean that the price and distance regressors will be linear functions, as other non-linear form may show better fit (e.g. cubic. Asymptotic etc.)

Dependent Variable: TOUR
Method: Least Squares
Date: 05/02/15 Time: 08:54
Sample: 1 114 IF YEAR=2010
Included observations: 38

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	411535.7	211412.5	1.946601	0.0621
RGDPCH	16.19813	4.982709	3.250867	0.0031
RGDPCH^2	-0.000198	8.05E-05	-2.456694	0.0207
PRICE	-64546.54	184759.0	-0.349355	0.7295
PRICE^2	-18206.09	47146.39	-0.386161	0.7024
POP	0.975740	0.299825	3.254364	0.0031
POP^2	-6.12E-07	2.30E-07	-2.661354	0.0129
D_KM	-71.66655	24.74532	-2.896166	0.0074
D_KM^2	0.001923	0.001201	1.600738	0.1211
PERSONS	1.271067	0.306392	4.148495	0.0003
PERSONS^2	-5.98E-07	2.67E-07	-2.243332	0.0333

R-squared	0.863358	Mean dependent var	134126.3
Adjusted R-squared	0.812750	S.D. dependent var	225529.8
S.E. of regression	97592.21	Akaike info criterion	26.05218
Sum squared resid	2.57E+11	Schwarz criterion	26.52622

Log likelihood	-483.9914	Hannan-Quinn criter.	26.22084
F-statistic	17.05966	Prob(F-statistic)	0.000000

Meanwhile, other squared regressors, income, population and persons failed to accept the null hypothesis and hence,is in favour of the alternative hypothesis. This shows that the regressors have non-linear (quadratic) influence.

Possible economics rationale behind the linearity/non-linearity can be as follows:

First, it is important to note that the coefficient for the quadratic model is negative: the parabola will show inverse 'U' shape that contains maximum-turning point.

1.income^2

According to statistics, it is likely for income variable to have non-linear relationship to the number of tourists. Assuming that the inverse 'U' shaped quadratic is an adequate regression,

1. at a low level of income (GDP per capita), citizens in a typical country will less likely to travel to Australia. This is a simple assumption that people who spend most of their income on purchasing basic needs cannot or will be difficult to afford traveling. This population within this category will be biggest for countries with undeveloped economies and large income-inequality.

2. at a medium level of income (taking account standard deviation), the likelihood and affordability to travel to Australia increases, as citizens can afford to travel to an extent. Also, the total population within thisgroup can bethe biggest,assuming country with average economy.

3. at a high level of income, the population within this group can most certainly afford to travel to Australia. But, the total numbers of people who earn significantly large income, so-called 'rich people', are too small compared to the population of a typical country. Hence, the total tourists within this income range are relatively small compared to medium level income range.

2. population^2

Non-linearity of population has direct link to income variable, as income is a function of population (per capita). As mentioned above, low population category has low number of tourists, as the total population within the group is relatively too small (e.g. highest earning people). The total number of tourists will marginally increase for medium population, who are likely to represent the medium level of income.

Finally, the tourist level will decrease for relatively large population, who represent medium-low level of income.

3. persons^2

If inverse 'U' shape best represents persons variable than linear model, the possible reasons could be:
1. if the number of persons in Australia, born in tourist's country are far too small, the potential visitors (regarded as tourists) to visit Australia-residing people will be diminished.
2. if the number of persons in Australia, born in tourist's country are far too large, this implies that the whole relative and potential visitors would be already living in Australia (e.g. migration) as well. Hence, the number of tourists would inevitably decrease as well.
Thus, the combinations of migration and temporary non-Australian citizens would result in maximum number of tourists.

2.d

1. Comparison of elasticity with linear model from q2b and log-log model

Formatting the original equation (yr=2010) to a log-log model:

$$\log(\widehat{tour}) = \log(b_0) + b_1\log(rgdpch) + \ b_2 log(price) + b_3\log(pop) + b_4\log(d_{km}) \ \ b_5\log(persons)$$

$$for\ b_0 > 0$$

It is important to emphasise that the coefficient for each variable is equal to the elasticity and hence, elasticity for log-log model is consistent throughout. This can be proved mathematically with elasticity = $\frac{dy}{dx} \times \frac{x}{y}. \frac{dy}{dx}$ for log-log model in general is $\beta_i \frac{y}{x}$→ elasticity =$\beta_i \frac{y}{x} \times \frac{x}{y} = \beta_i$ (coefficient)
This is not the case for linear model from *q2b*, where the elasticity of the model changes at different points. Thus, in order to compare linear model to log-log model, elasticity at the means($\bar{X}, \bar{Y}$) for linear model is used as a representation.

Variable (year 2010)	Elasticity at means (linear model)	Elasticity = Coefficient (log-log model)
RGDPCH	0.491644 (I, +ve)	0.885061 (I, +ve)
PRICE	-2.519144 (E, -ve)	-0.814077 (I, -ve)
POP	0.200530 (I, +ve)	0.455247 (I, +ve)
D_KM	-2.797618 (E, -ve)	-1.936983 (E, -ve)
PERSONS	0.582237 (I, +ve)	0.399901 (I, +ve)
		Legend: E: elastic, **I=** inelastic, **+ve =** positive slope and **–ve =** negative slope

As can be seen from previous table, elasticity for the linear model and log-log model show some discrepancies. Apart from price regressor, the sign and the elasticity (in terms of E & I) have not changed. But it should be noted that the elasticity are still different by relatively large magnitude.
Notably,

- Income and population variables' elasticity has increased in log-log model
- Price, distance and persons variables' elasticity have decreased in log-log model

Overall, the use of linear model and log-log model can represent different elasticity and hence, different interpretations. Finally, the elasticity differences can be emphasised using arc and point elasticity. For linear model, elasticity change at different values and hence, point elasticity at means are used to represent elasticity only at the mean, hence point elasticity. Meanwhile, arc elasticity is used for the log-log model and yet, the elasticity is maintained throughout the model.

2. Comparison with the log-log model with the quadratic model from q2c

at 10% significance level, using the same hypothesis testing of linearity,

$$H_0: \beta_j = 0 \; \& H_1: \beta_j \neq 0$$

all parameters' p-value are less than 0.1 → the coefficients are statistically significantly different from zero and hence, all log parameters' coefficients cannot be zero. This implies that the log-log model can still hold, which also supports the argument that <u>log-log model linearises the regression function</u>. This can be compared tothe non-linear model from q2c, where price and distance variables showed non-linear influence.

Dependent Variable: LOG(TOUR)
Method: Least Squares
Date: 05/02/15 Time: 10:08
Sample: 1 114 IF YEAR=2010
Included observations: 38

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.59474	2.768261	4.188454	0.0002
LOG(RGDPCH)	0.885061	0.174552	5.070475	0.0000
LOG(PRICE)	-0.814077	0.419970	-1.938419	0.0614
LOG(POP)	0.455247	0.079531	5.724125	0.0000
LOG(D_KM)	-1.936983	0.284281	-6.813620	0.0000
LOG(PERSONS)	0.399901	0.082042	4.874358	0.0000

R-squared	0.819962	Mean dependent var	10.87633
Adjusted R-squared	0.791831	S.D. dependent var	1.348484
S.E. of regression	0.615253	Akaike info criterion	2.010372
Sum squared resid	12.11315	Schwarz criterion	2.268939
Log likelihood	-32.19708	Hannan-Quinn criter.	2.102368
F-statistic	29.14808	Prob(F-statistic)	0.000000

Estimation output for log-log model