# Panel data analysis in Tourism Research

Seetaram, Neelu and Petit, Sylvain

Bournemouth University, IDP (EA 1384), University of Valenciennes

2012

# PANEL DATA ANALYSIS

Neelu Seetaram and Sylvain Petit

## Introduction

Panel data sets are also known as longitudinal data or cross sectional time-series data. They have spatial (N) and temporal (T) dimensions. They constitute of a number of observations overtime on a number of cross sectional units such as individuals, firms, and countries allowing researchers to analyse the dynamics of change in short time series data. According to Baltes and Nesselroade (1979), longitudinal data and techniques involve "a variety of methods connected by the idea that the entity under investigation is observed repeatedly as it exists and evolves over time". These methods have been applied in different research disciplines. Frees (2004) posits that they have developed because important databases have become available to empirical researchers.

The term panel data was introduced by Lazarsfeld and Fiske (1938, in Frees 2004) in their study of the effect of the relationship between radio advertising and product sales, where they proposed to interview a 'panel' of consumer overtime. Toon (2000 in Frees 2004), acknowledges Engel's 1857 budget surveys as one earliest application of longitudinal data. In this survey Engel collected data on the expenditure pattern from the same set of subjects over a period time. The aim was to study expenditure on food as a function of income. Panel data modelling and estimation techniques developed in the second half of the twentieth century. Early application of this technique are those Kuh (1959), Johnson (1960), Mundlak (1961) and Hoch (1962) who used the fixed effect models (explained later in the text) and Balestra and Nerlove (1966) and Wallace and Hussain (1969) who used the random effect models (explained later in the text).

This chapter focuses on the application of panel data techniques in the tourism literature. The chapter is organised as follows: The next section explains panel data modelling technique emphasising on the difference between fixed and random effects. Dynamism in modelling is introduced and the implication for model estimation is discussed. The chapter then devotes a section to unit root and cointegration tests which is followed by an illustration of the application of panel data analysis in tourism research namely, in tourism demand modeling and in explorations of the relationship between tourism and economic growth. The conclusion spells out the limitations of this technique and directions for future research.

**Panel Data Modelling Techniques and Benefits**

Panel data analysis offers several advantages. The most obvious is that inferences are performed using a larger sample and the lack of degrees of freedom is fairly unlikely to occur. According to Baltagi (2005), more complex relationships can be modelled, for example temporal changes in cross-section can be analysed. One of the most important advantages however, is that panel data modelling allows for the control of heterogeneity in the sample.

A standard approach to model the relationship between **Y** (dependent variable) and **X,** a set of explanatory variables, is given below where $\boldsymbol{\varepsilon_{it}}$ is the stochastic error term which takes into account the variation in the expected value of **Y** which cannot be explained by the **X**'s.

$$\mathbf{y_{it} = x'_{it}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_{it}} \tag{1}$$

For example in a tourism demand model, **Y** can stand for the number of arrivals to a particular destination while the **X**'s include factors affecting demand, such as income in the home country, relative prices, marketing expenditure, transportation cost and so on. The **X**'s or explanatory variables can be included in the model so long as they are observable and measureable. There are however, factors such as culture and other unique characteristics of the individuals or groups under study which are not observable or measurable but which influence the outcome of expected **Y**. These factors are referred to as the heterogeneity and are not directly part of Equation 1. The effect is incorporated in the one of the $\boldsymbol{\beta}$'s, should they be correlated with the respective **X** or otherwise included in $\boldsymbol{\varepsilon_{it}}$. As a result, in the first case, the estimated $\boldsymbol{\beta}$ will not reveal the true effect of the variation in **X** on expected **Y**. By modeling the relationship between **X** and **Y**, using the panel data technique, the researcher is able to separate the effect of the heterogeneity from that of $\boldsymbol{\beta}$.

Suppose K number of subjects are observed over time. The response subject of $\mathbf{Y_{1t}}$ will tend to be similar to responses in previous years but different to the rest $\mathbf{Y_{2t}}, \mathbf{Y_{3t}}\ldots\ldots\mathbf{Y_{k-1,t}}$. That is, there is uniqueness in the behaviour of this subject and this is assumed to be constant over time. The uniqueness of $\mathbf{Y_1}$ can be attributed to factors such as culture, and past experience (for a person) business practices, (for a firm), history, or system of government (for a country). For example, the cultural and historical links that Australia and UK share may be expected to promote international tourism flows between the two countries. In this respect, tourism flows to Australia from UK are expected to be influenced by factors which are unique to travellers from this source and not pertinent for travellers from other home countries. In a panel data model, the uniqueness of each market in the sample is captured by $\boldsymbol{\mu_i}$ which is the unobserved heterogeneity in the sample as given in Equation (2) below.

$$\mathbf{y_{it} = \mu_i + x'_{it}\beta_i + \varepsilon_{it}} \tag{2}$$

The additional benefit of including $\boldsymbol{\mu_i}$ in the model is that it offers a potential solution to the problem of omitted variables and measurement errors in data. Lack of data is the most common problem faced by researchers. In regression analysis, it often results in the omission of relevant variables from the model. This may gives rise to potential model specification errors for example, due to difficulties in obtaining data on marketing expenditure, this variable is often omitted from tourism demand model, although *a priori*, it is expected to be relevant in explaining demand.

Consider Equation 3, where $\mathbf{Y_{it}}$ is determined by three variables.

$$\mathbf{y_{it} = \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + \varepsilon_{it}} \tag{3}$$

omitting $\mathbf{x_3}$ from the model will reduce Equation 3 to Equation 4, where the effect of $\mathbf{X_3}$ is soaked by the $\boldsymbol{\varepsilon_{it}}$. The actual model estimated is

$$\mathbf{y_{it} = \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it}} \tag{4}$$

where $\mathbf{U_{it} = x_{3t} + \varepsilon_{it}}$

In this case the residuals include the effect of $\mathbf{x_{3t}}$ and will display patterns leading to the conclusion that the model may be suffering from serial correlation (Green, 1999). The implications for the $\hat{\boldsymbol{\beta}}$'s will depend on whether the $\mathbf{X_{it}}$'s are correlated with $\mathbf{X_{3t}}$. If so, the estimated coefficients will be biased. If on the other hand, the covariance between $\mathbf{X_{3t}}$ and the $\mathbf{X_{it}}$'s is equal to zero, $\hat{\boldsymbol{\beta}}$'s will be unbiased and consistent. According to Wooldridge (2002), panel data modelling technique offers an effective solution to this problem. The inclusion of $\mu_i$ in Equation 4 will solve the problem as it will absorb the effect of $X_{3t}$. This solution is also applicable when measurement errors are present in the data.

Taking the example of tourism model, Seetaram (2010) explains the complexities that arise when faced with the computation of airfare elasticities. Airfare data are often plagued with measurement errors which arise mainly because of the wide array of airfares and travel class categories which are prevalent in the market. This makes the task of the researcher complex as often no choice is left but to use an average airfare to represent the transportation cost to the destination. Average airfare is not always a good representation of actual airfare. Suppose that airfare, $x_{3t}$ is measured with errors such that the actual variable which is included in the model is $x^*_{3t} = x_{3t} + v_t$ .The model estimated is given by:

$$y_t = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_{3t}^* + \eta_{it} \tag{5}$$

where $\eta_{it} = \varepsilon_{it} + v_t$.

In Equation 5, the stochastic error term is given by $\eta_t$. The cov $(\eta_i, x_{3t}^*) = - \beta \delta_v^2$ where $\delta_v^2$ is the variance of the measurement error (Green, 1999). This violates the crucial assumption of non-correlation between any explanatory variable and the residuals of the equation (Green, 1999). As a consequence, all the Ordinary Least Squares (OLS) estimators, $\hat{\beta}$'s will be biased and inconsistent and $\hat{\beta}_3$ will be attenuated. If instead the relationship between $Y$ and the explanatory variables are modelled using the format of Equation 2, the measurement errors, $v_t$, will be absorbed by the unobserved heterogeneity $\mu_i$ leaving $\varepsilon_{it}$ free from its effect.

*Fixed Effect and Random Effects*

There are two ways of modeling $\mu_i$, namely the fixed and random effect. The choice between these two depends on whether $\mu_i$ is correlated to any of the other explanatory variables of the model (Wooldridge, 2002). Equation 2 is formulated using the fixed effect (FE) technique. This method assumes that the heterogeneity in the model is $\mu_i$ time invariant and specific to the individual group. In Equation 2, the slopes are fixed but the intercepts vary for each cross section. This is equivalent to adding a dummy specific for each cross section which is why it is also referred to as the Least Square Dummy Variable (LSDV) method. The slopes are treated as constant across group and across time. It is however, possible to allow the slope to vary across groups, across time or both (Hsaio, 2003). The rationale behind this modelling approach is that since $\mu_i$ accounts for time invariant characteristics of the group, it removes the pernicious effect of omitted variables (Allison, 2005). Fixed effect is often chosen as a precaution against omitted variable bias. The drawback is that $\mu_i$ cannot be used to assess the effect of characteristics which changes overtime.

The FE technique explores the relationship between explanatory and dependant variables within one individual group. For each group, the variations of the all variables from their mean values are considered and the estimated coefficients are also known as the within estimates. This can be a limitation of the FE method as in-between variations are ignored. Furthermore, only the effect of variable with sufficient variability can be analyses. For example, in a longitudinal study of individual's tourist's perception of the quality of destination attributes, the effect of gender and ethnicity cannot be analyzed. The modeler is expected to make a trade-off between sample variability and omitted variable bias (Allison, 2005). However, as explained before the effects of these time-invariant factors are controlled in the FE model. In circumstances when the in between variation are not relevant, FE model makes use of maximum information, yielding errors terms with smaller variations (Allison, 2005). FE models may additionally include an error component which changes over time but not for each unit, $\tau_t$. $\tau_t$ is treated as a constant in the model.

$$y_{it} = x'_{it}\beta + \mu_i + \tau_t + \varepsilon_{it} \tag{6}$$

Taking the example of the tourist demand model, consider a sample which includes arrivals from 10 sources over a period of 10 years. The aim is the find the income and price elasticities of demand. Each market is a group in the sample. $\mu_{uk}$ will take into account all characteristics of UK other than price and income that will influence arrivals to Australia. UK as a market has certain characteristics which may or may not influence income and relative prices in the country. For example, $\mu_i$ can stand for system of government, democracy, which may or may not be related to income level in UK. However, in the instant that there exists such a relationship, then the inclusion of $\mu_i$ controls for the effect of democracy and allows the estimation of the net effect of income on number of arrivals from UK. The FE technique also assumes that $\mu$ is unique to UK and is not correlated to the characteristics of other countries in the sample. Any correlations between $\mu_{uk}$ and $\mu_{usa}$ are ignored.

If, however, $\mu_i$'s are correlated to one another and to the error terms of other groups, the resulting variance will be high making statistical inference dubious. A better approach in this case will be to use the random effect (RE) technique. In the RE approach, variation across entities is assumed to be random and uncorrelated with independent variables in the model. The RE model is given as

$$y_{it} = x'_{it}\beta + \mu_{it} + \tau_t + \varepsilon_{it} \qquad (7)$$

$\mu_{it}$ is referred to as the between-group error. The advantage of RE is that since variation across the sample are considered it permits the study of time invariant factors such as gender, ethnicity and race in the model. The RE method use variations both within and between individuals, random effects methods typically have less sampling variability than fixed effects methods (Allison, 2005). The problem however, is that all relevant measureable variables need to be included in the model and since data on a few may not be available therefore leading to omitted variable bias in the model.

The choice between FE and FE depends on whether $\mu_i$ is correlated to any of the other explanatory variables of the model (Wooldridge, 2002). When such a correlation exists, the fixed effect technique is superior. Otherwise, the random effect is more parsimonious and gives more efficient estimates (Wooldridge, 2002). A formal test for assessing the correlation between the unobserved heterogeneity and other explanatory variable is the Hausman (1978) specification test. In the tourism literature, FE method has been more frequently applied since the groups under observations are often markets, or destinations which have characteristics which influence the other explanatory variables of the model. The rest of the chapter will focus on the FE modelling method.

*Dynamic Panel Data Models*

By nature, all panel data models are dynamic since they are taking into account the time series dimension of the sample. However, functions which specifically model the effect of lagged dependant variables are referred to as dynamic panel data models. A general dynamic panel data model with FE effect is given as Equation 6 below.

$$y_{it} = \gamma_i y_{i,t-1} + \beta_0 + x'_{ikt}\beta_k + \mu_i + \varepsilon_{it} \qquad (8)$$

It is assumed that:

(a) $\mu_i \sim (N, \sigma_\mu^2)$ and $\varepsilon_{it} \sim (N, \sigma_\varepsilon^2)$ where $\sigma_\mu^2 \geq 0$ and $\sigma_\varepsilon^2 > 0$

(b) The explanatory variables are strictly exogenous, that is they are not correlated with the error terms. i.e. $E(\varepsilon_{it}\, \varepsilon_{js}) = 0$ for $i \neq j$ or $t \neq s$.

(c) The unobserved heterogeneity, if it is present, is random. i.e. $E(\mu_i\, \mu_j) = 0$ for $i \neq j$

(d) The unobserved heterogeneity is uncorrelated within the countries and with the error i.e. $E(\mu_i\, \varepsilon_{js}) = 0$ for $\forall$ i, j, t, s

(e) The explanatory variables are strictly exogenous, that is they are not correlated with the error terms. i.e. $E(x_{it}\, \varepsilon_{js}) = 0$ for $\forall$ i, j, t

(f) The unobserved heterogeneity are correlated with the predetermined variables i.e. $E(x_{it}\, \mu_j) = 0$ for $\forall$ i, j, t

(g) $\gamma_{i0}$ is uncorrelated with the error term i.e. $E(\gamma_{i0}\, \varepsilon_{jt}) = 0$ for $\forall$ i, j, t

(h) $\gamma_{i0}$ can be correlated with the unobserved heterogeneity. i.e. E $(\gamma_{i0}\, \mu_j)$ = unknown for $\forall$ i, j

Dynamic panel data analysis is becoming increasingly popular in the tourism literature modeling. In tourism demand model, this $\boldsymbol{\gamma_i}$ accounts for destination loyalty and repeat visitations. It takes into account to extent to which current visits are dependent on the number of past visits. It takes into account the effect of habit persistence in demand and the extent to which consumer react to *ex-post* information available. $\boldsymbol{\gamma_i}$ is an indication of the efficiency of information diffusion through word-of-mouth.

Generally, the functional form utilized for Equation 8 is that of double logarithm implying that the $\hat{\boldsymbol{\beta}}$'s are the short term elasticities. The long term elasticities the long term elasticities $\hat{\boldsymbol{\beta}}^*$ may be obtained by

$$\hat{\boldsymbol{\beta}}^* = \frac{\hat{\boldsymbol{\beta}}}{\mathbf{1}-\boldsymbol{\gamma}} \tag{9}$$

The most widely used estimation technique for dynamic panel data sets in the tourism literature has been has been the Arrelano Bond (1991). Examples of studie swhihc have applied this technique are Garín-Muños (2006), Khadaroo, and Seetanah (2007) and Eugenio-Martin et al. (2004).

*Estimation Technique*

Estimating $\hat{\boldsymbol{\beta}}$ using standard LSDV method yields biased and inconsistent estimators. $\hat{\boldsymbol{\beta}}_{\mathbf{LSDV}}$ estimator, also referred to as the covariance estimator, is given by:

$$\hat{\boldsymbol{\beta}}_{\mathbf{LSDV}} = \left[ \sum_{\mathbf{i=0}}^{\mathbf{N}} \sum_{\mathbf{t=1}}^{\mathbf{T}} \left( \mathbf{x}_{\mathbf{it}} - \bar{\mathbf{x}}_{\mathbf{it}} \right) \left( \mathbf{x}_{\mathbf{it}} - \bar{\mathbf{x}}_{\mathbf{it}} \right)' \right]^{-1} \left[ \sum_{\mathbf{i=0}}^{\mathbf{N}} \sum_{\mathbf{t=1}}^{\mathbf{T}} \left( \mathbf{x}_{\mathbf{it}} - \bar{\mathbf{x}}_{\mathbf{it}} \right) \left( \mathbf{y}_{\mathbf{it}} - \bar{\mathbf{y}}_{\mathbf{it}} \right)' \right] \qquad (10)$$

where $\bar{\mathbf{y}}_{\mathbf{it}} = \frac{\mathbf{1}}{\mathbf{T}} \sum_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{y}_{\mathbf{it}}$ and $\bar{\mathbf{x}} = \frac{\mathbf{1}}{\mathbf{T}} \sum_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{x}_{\mathbf{it}}$ , and $\hat{\boldsymbol{\beta}}_{\mathbf{LSDV}}$ is the estimated true coefficient of the exogenous variable $\mathbf{x}_{\mathbf{it}}$. $\bar{\mathbf{x}}_{\mathbf{it}}$ and $\bar{\mathbf{y}}_{\mathbf{it}}$ are the mean of $\mathbf{x}_{\mathbf{it}}$ and $\mathbf{y}_{\mathbf{it}}$ respectively. $\hat{\boldsymbol{\beta}}_{\mathbf{LSDV}}$ will be biased and inconsistent unless $\mathbf{T} \to \boldsymbol{\infty}$ (Nickell, 1981; Anderson and Hsiao, 1981; Arellano and Bond, 1991; Kiviet, 1995; Judson and Owen, 1999). This occurs because in Equation 8, $\mathbf{y}_{\mathbf{it-1}}$ will be correlated with the mean of the stochastic error term models $\bar{\boldsymbol{\varepsilon}}_{\mathbf{it}}$ by construction and will be correlated to $\boldsymbol{\varepsilon}_{\mathbf{it-1}}$ which is contained in $\bar{\boldsymbol{\varepsilon}}_{\mathbf{it}}$ (Hsaio, 2003).

Anderson and Hsiao (AH) (1981) and Arellano and Bond (AB) (1991) show that the bias may be reduced by first differencing Equation 8 and using the lagged level values of the $\mathbf{y}_{\mathbf{it}}$ as instruments. Consider Equation 11 below which is similar to Equation 8 but for simplicity, the vector of exogenous variables **x** is left out.

$$\mathbf{y}_{\mathbf{it}} = \boldsymbol{\gamma} \mathbf{y}_{\mathbf{it-1}} + \boldsymbol{\mu}_{\mathbf{i}} + \boldsymbol{\varepsilon}_{\mathbf{it}} \qquad (11)$$

**i = 1, 2....N, t= 1, 2....T**

$\boldsymbol{\mu}_{\mathbf{i}}$ is the fixed effect which is the cause of the bias in the estimation by LSDV. To eliminate $\boldsymbol{\mu}_{\mathbf{i}}$, Anderson and Hsiao (AH) (1981) suggest that first difference transformation be applied to Equation (11). First differencing Equation (11) gives the following:

$$(\mathbf{y}_{\mathbf{it}} - \mathbf{y}_{\mathbf{it-1}}) = \boldsymbol{\gamma}(\mathbf{y}_{\mathbf{it-1}} - \mathbf{y}_{\mathbf{it-2}}) + (\boldsymbol{\varepsilon}_{\mathbf{it}} - \boldsymbol{\varepsilon}_{\mathbf{it-1}}) \qquad (12)$$

**$i = 1, 2....N, t= 1,2....T$**

Equation (12) is a first difference autoregressive process of order one with no exogenous regressors. $\Delta y_{it-1} = (y_{it-1} - y_{it-2})$, is correlated with the error $(\varepsilon_{it} - \varepsilon_{it-1})$. The second lag $y_{it-2}$, and the first difference of this second lag, $\Delta y_{it-2} = (y_{it-2} - y_{it-3})$, are possible instruments, since they are both correlated with $(y_{it-1} - y_{it-2})$ but are uncorrelated with $(\varepsilon_{it} - \varepsilon_{it-1})$, as long as the $\varepsilon_{it}$ themselves are not serially correlated (Anderson and Hsiao,1981). Using the second lag and the first difference of this second lag as instrumental variables, two estimators $\hat{\gamma}_{IV}$ and $\hat{\gamma}^*_{IV}$ can be developed. These are given in (13) and (14).

$$\hat{\gamma}_{IV} = \frac{\sum_{i=1}^{N} \sum_{t=3}^{T} (y_{it} - y_{it-1})(y_{it-2} - y_{it-3})}{\sum_{i=1}^{N} \sum_{t=3}^{T} (y_{it-1} - y_{it-2})(y_{it-2} - y_{it-3})} \tag{13}$$

$$\hat{\gamma}^*_{IV} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} (y_{it} - y_{it-1}) y_{it-2}}{\sum_{i=1}^{N} \sum_{t=2}^{T} (y_{it-1} - y_{it-2}) y_{it-2}} \tag{14}$$

These estimators are consistent when $N \to \infty$ or $T \to \infty$ or both. AB argue that more efficient estimators can be obtained by taking in additional instruments whose validity is based on orthogonality between lagged values of the dependent variable $y_{it}$ and the errors $\varepsilon_{it}$. These results are confirmed by Kiviet (1995) and Judson and Owen (1999). However the bias persists in samples with small **T** (Kiviet, 1995; Judson and Owen, 1999). In fact it increases with the value $\gamma$ and decreases with **T** (Kiviet, 1995). An estimator that relies on lags as instruments under the assumption of white noise errors will lose its consistency if the errors are serially correlated (Kiviet, 1995).

Using simulations to generate data, Judson and Owen (1999) performed an exercise with a panel of varying size. The number of cross sections, **N,** takes the values of 20 or 100 and the number of time periods, **T,** is given the values of 5, 10, 20 and 30. Judson and Owen (1999) show that although the value of the bias falls as **T** increases, it is nevertheless still considerable at **T** = 30 and can be as high as 20 percent of the true value of the parameter.

They prove that estimates resulting from the AB technique have higher variances than the LSDV estimators supporting earlier results of Kiviet (1995).

LSDV estimates are more efficient than any other class of estimates developed for autoregressive panel data models (Kiviet, 1995; Judson and Owen, 1999). The removal of the bias in LSDV estimates opens the possibility of obtaining more robust estimates (Kiviet, 1995). Kiviet (1995) evaluated the bias in the true parameters based on a Monte Carlo study and developed a method to correct for potential bias in the estimated parameters when the true parameters are known. On a practical level however, true parameters are seldom known, in which case Kiviet (1995) suggests that estimates be obtained, using the techniques proposed by AH (1981) and AB (1991). These estimates can then be corrected for the bias by applying the Kiviet (1995) Corrected LSDV (CLSDV) method. This method is only applicable to balanced samples- samples which contain the same number of observations for each cross section. In the tourism literature, authors such as Cortés-Jiménez (2008) in her study of regional tourism in Spain and Italy and Soukiazis and Proença (2008) examining regional tourism in Portugal have employed this technique.

*Unit Root and Cointegration Tests*

Classical statistical inferences rely on data being mean reverting. However, economic variables which tend to evolve over time are not always stationary and failure to account for this will result in spurious regression results. To circumvent such problems, unit roots are carried out to ascertain that regression results are valid. However, while testing for unit root is standard in the time series literature, it is quite recent in panel data (Baltagi, 2005). In the tourism literature although not very common, the availability of samples with fairly large time dimensions has resulted in more testing for unit root in the panel data set up, for example, Lee and Chang (2008) in a study of tourism development and economic growth and Seetaram (2010, 2012) in the context of Australian outbound tourism.

In the panel data setup, panel unit roots tests have higher power than unit root tests based on individual time series for each of the cross section, since the later performs poorly when data periods are short (Baltagi, 2005; Banerjee, 1999; Banerjee et al., 2004; Levin, Lin and Chu, 2002; Im, Persaran and Shin, 2003, and Pedroni, 1999). Several tests for unit roots in panel data have been developed. A few examples are Maddala and Wu (1999), Breitung (2000) and Choi (2001). According to Baltagi (2005) however, the two most efficient tests for stationarity in a panel data setting are Levin, Lin and Chu (hereafter LLC, 2002) and Im, Pesaran and Shin (hereafter IPS, 2003). The fundamental difference between these two tests rests on the assumption made regarding the autoregressive process (Baltagi, 2005):

1. LLC assumes that the autoregressive process is common for all cross sections, that is $\rho = \rho_i$ in Equation 15 give below.

2. IPS assumes that the persistence parameter, $\rho_i$ is allowed to vary across the cross sections.

Both tests are based on estimating the following equation:

$$\Delta y_{it} = \rho_i y_{it-1} + \sum_{j=1}^{z_i} \varphi_{ij} \Delta y_{it-j} + \varepsilon_{it} \tag{15}$$

$y_{it}$ is the dependent variable being tested for unit root. $\Delta$ denotes the first difference in the dependent variable $y_{it}$. $z_i$ is the number of lags to be included in the testing. $\varepsilon_{it}$ are the error terms. $\rho_i$ and $\varphi_{ij}$ are parameters.

LLC assumes that the error term $\varepsilon_{it}$ is independent across the units of the sample and have a fixed variance. LLC tests the hypothesis that each of the series in the panel contains a unit root against the hypothesis that all individual series are stationary. This can be written as:

$H_0$ : $\rho = \rho_i = 0$ for all $i$ ~ *All of the individuals have a unit root.*
$H_1$ : $\rho = \rho_i < 0$ for all $i$ ~ *All the individual series are stationary.*

The IPS test assumes that the panel is balanced. It hypothesises that each of the series contains a unit root against the alternative hypothesis, that at least one of the series is stationary. These can be represented as follows:

$H_0$: $\rho_i = 0$ for all $i$. ~ *All the series have unit roots.*

$$H_1: \begin{cases} \rho_i = 0 \text{ for } i = 1,2,\ldots N1 \\ \rho < 0 \text{ for } i = N+1, N+2,\ldots N \end{cases}$$ ~ *Some of the series have unit roots*

are the cross sections in the panel data set. This method involves regressing the individual series for each of the cross sections. The critical value for the hypotheses is obtained by taking the average of the student t-statistics for the $\boldsymbol{\rho_i}$ from the individual regressions. This is given by:

$$\bar{t}_{NT} = \frac{\left(\sum_{i=t}^{N} t_{iT}(\rho_i)\right)}{N} \tag{16}$$

Series containing unit roots are non-stationary processes which have time-varying mean and variance that increases as sample size grows (Baltagi, 2005).

When variables are individually integrated of the same order such as the ones in this study), a linear combination of these variables can still be stationary (Baltagi, 2005; Banerjee, 1999; Banerjee et al., 2004; Pedroni, 2004). If the series are found to be cointegrated then is at least one cointegrating vector which renders the combination of variables stationary. Furthermore, it implies that there is long run relationship among the variables.

The conventional cointegration tests suffer from low power when applied to samples with small time dimension, such as the one used in this chapter (Baltagi, 2005; Banerjee, 1999; Banerjee et al., 2004 Pedroni, 2004). Panel cointegrating techniques have been developed to allow researchers to pool information regarding common long run relationships from across the panel. In addition, such techniques allow the associated short run dynamic and fixed effects to be heterogeneous across the different members of the panel (Baltagi, 2005; Banerjee, 1999; Banerjee et al., 2004; Pedroni, 1999; 2004). This chapter focuses on the Pedroni (1999) test as it is one of the most widely used test for cointegration in panel data. The first step in this test is to estimate a cointegrating relationship with fixed effects and heterogeneous time trends for each of the cross section of the study individually. Then the cointegration tests are performed based on the residuals obtained (Banerjee, 1999; Banerjee et al., 2004; Pedroni, 1999; 2004).

Pedroni (1999) proposes seven tests for cointegration in the panel data framework. Four of these tests are referred to as the '*panel cointegrating statistics*' or the within-dimension based statistics (Pedroni, 1999, pp. 658). In these tests, he assumes that there is a common cointegrating relationship among the variables. For these four tests, the residuals are pooled across the time dimension of the panel. An autoregressive function is regressed and the autoregressive coefficient is given by $\boldsymbol{\upsilon_i}$. This method assumes that $\boldsymbol{\upsilon_i = \upsilon}$**,** that is, it is common for all countries. The alternate hypothesis is that there is a cointegrating relationship for all the cross sections given as (Pedroni, 1999):

$\mathbf{H_0:}$ : $\boldsymbol{\upsilon = \upsilon_i = 1}$ ~ *All of the individuals are not cointegrated*

$\mathbf{H_0:}$ : $\boldsymbol{\upsilon = \upsilon_i < 1}$ ~ *All of the individuals are cointegrated*

By contrast, the remaining three tests are called the '*group mean cointegrating statistics*' or the between-dimension. These tests statistics are based on pooling the residuals of the regression along the cross sections of the panel (Pedroni, 1999). In these tests, estimators average the individually estimated autoregressive coefficient for each cross section (Pedroni (1999). The hypotheses here are given by:

$\mathbf{H_0:}$ : $\boldsymbol{\upsilon_i = 1}$ ~ *All of the individuals are not cointegrated*

$\mathbf{H_0:}$ : $\boldsymbol{\upsilon_i < 1}$ ~ *All of the individuals are cointegrated*

The group mean statistics can be considered as more accurate, as they allow for more heterogeneity among the countries, and produce consistent estimates (Pedroni, 2001). The higher value of the group mean statistics can be considered to be a more accurate representation of the average long run relationship (Pedroni, 2001).

It can be noted however, that in general the unit root cointegration tests increase the probability of determining if data are stationary or not and whether they are cointegrated (Banerjee, 1999; Banerjee et al. 2004). However, one limitation of these tests is that they assume no cross sectional correlation in the sample (Banerjee, 1999; Banerjee et al., 2004). Banerjee et al. (2004) showed that the results of cointegration tests were susceptible to dependence among the cross sections. This means that the power of the tests is reduced in cases where the cross sections are not independent. In spite of this, in panel data sets, the problem of spurious regression results is not expected to be as serious as in pure time series. As demonstrated by Phillips and Moon (1999), noise in time series regression is lessened by pooling cross section and time series observations, implying that the model may be estimated in level form without risking spurious results.

**Applications of the panel data techniques in tourism data analysis**

The empirical literature based on the panel data methods and applied to the tourism sector, can be divided into two broad areas. The first concerns the determinants of tourism demand. This topic of demand modelling is not recent and goes back Rugg (1972). The study of international tourism arrivals and receipts is a vibrant area of research and it cannot be confined to gravity models due to factors such as seasonality, cultural links or time constraints which are particular to the consumption of tourism products. The second area consists mainly of analysis of the link between tourism and economic growth. Since, several destinations for example, Spain, Italy and Portugal have simultaneously experienced economic development and expansion of their tourism sector, especially following the adoption of the Millennium goals for the economic development, many authors have attempted to estimate the real impact of international tourism on economic growth.

In tourism demand modelling, after a short period of application of standard and classic panel data models (OLS with fixed effects or random effects), the literature had been characterized by a high volume of dynamic panel models. Additionally, from the second half of 2000 onwards, several of the works employ cointegration techniques. The first empirical papers, based on the classic panel data methods, were published in the beginning of the 2000s. These studies focused generally on the prices elasticities or on the impact of political risk and violence on international tourism arrivals. For example, Ledesma- Rodriguez and Navarro-Ibanez (2001) used annual data from 1979 to 1997 to study factors affecting arrivals to Tenerife from 13 markets. They found arrivals to be elastic with respect to income and inelastic with respect to prices and transport cost in the long run. Espinet et al. (2003) used panel data with random effects for a hedonic evaluation based on 86,000 prices between 1991 and 1998. The data concerned hotels in the southern Costa Brava region. Their results indicate a real and significant effect from the quality to the price.

Eilat and Einav (2004) used multinomial logit estimations, based on bilateral data during the period 1985-1998, to study the leisure tourism determinants. This was the first study based on a three dimensional panel data (year, destination, origin). They showed that political risk is a very important determinant of the tourist arrivals. The most recent paper based on the traditional panel data method is that of Arita et al. (2011). They analyzed the impact of ADS (Approved Destination Statues, beginning of the 1990s) agreements on the international tourism arrivals in China. They integrated, in the estimation, fixed effects, to compare results before and after ADS and also between destinations inside and outside the ADS.

As the relevant data available is often annual, there are many empirical works which use dynamic panel data methods to test long-run relationships. The most commonly used estimation technique in the literature is that of AB although authors such as Maloney and Montes Ronjas (2005) made used of General Method of Moment (GMM) suggested by Blundell and Bond (1998). They measured the tourism price elasticity in Caribbean countries, with bilateral data (tourists came from United-States, Canada, United-Kingdom, Germany, Netherlands, Italy and Spain) from 1990 to 2002. They estimated a large price elasticity of 4.9.

As mentioned above, the AB technique is the most widely used in the tourism literature. For example, Neumayer (2004) estimated the effect of violence, risks, freedom and human rights violations on annual international tourism arrivals during 1977-2000. This study was based on a sample which contained more than 100 countries. Neumayer employed two methods for this estimation: a traditional data panel model with fixed effects and a dynamic panel data model. He found that in most cases, these explanatory variables had a real and significant effect on tourism. Garin-Munos and Montero-Martin (2007) studied yearly data from 1991 to 2003 to assess factors affecting the number of arrivals to the Balearic Islands. They found a high level of consumer loyalty to tourism in the Balearic Islands and recommended that suppliers of tourism products should raise the quality of their products and should improve their brand image.

Khadaroo and Seetanah (2007) used data on arrivals to Mauritius during the period 1978 to 2003 to assess the relative importance of transport infrastructure as a demand determinant. Transport infrastructures, was approximated by the net investment in land, air, and sea infrastructure at the current market price while non-transport infrastructures were measured by net investment at current market price on communication, energy, wastewater, and defence infrastructure. Transport infrastructure was found to be an important determinant of demand for travellers from Asia, Europe, and the United States while the latter two groups are also influenced by the non-transport infrastructure. They however use a poor proxy for prices in their model. They sought to capture the price effect by using the real value of the Mauritian rupees in U.S. dollars. The depreciation or appreciation of the Mauritian rupee against the U.S. dollar cannot be expected to reflect changes in the cost for holidays for visitors from Africa, Asia, and Europe. It is not surprising that they found prices to be insignificant in determining international arrivals to Mauritius.

Naude and Saayman (2005) analyzed annual data from 1996 to 2000 to estimate tourism arrivals for 43 African countries. Their estimation yielded a negative coefficient for the lagged dependent variable suggesting that tourism in South African is taking a downward trend as currently level of arrivals were negatively related to past levels. Political stability was a key determinant of arrivals. Their model included lagged values of explanatory variables. It is not unlikely that the variables are highly collinear to their lagged values and therefore the model suffers from multicollinearity. This may explain why only a few of their estimated coefficients were statistically significant.

A dynamic data panel model has the fundamental characteristics to establish the long-run relationship between two variables. As the question of difference between short and long-run estimation is important (the results can differ as a short-run estimation can be amplified or cancelled in the long-run), the dynamic model is often used along with another method of estimation which has the objective to estimate the short-run impact. For example, Kuo et al. (2008) and Kuo et al. (2009) tried to estimate impact of infectious diseases (Avion Flu) and severe acute respiratory syndrome (SARS) on international tourist arrivals in selected Asian, African and European destinations, for the sample period of 2001 to 2006 and 2004 to 2006 respectively. The long-run estimation is based on the GMM procedure and the short-run estimation is provided by ARMAX models for each country. They established different results between short and long-run: SARS had a significant impact on tourist arrivals at long and short-run whereas infectious diseases had only an impact at the long-run. Garin-Munos (2006) used annual data from 1992 to 2002 to estimate factors affecting arrivals to Canary Islands from 15 of its markets. Demand was found to be inelastic in the short run but income and price elasticity was greater than one in the long run. Demand was elastic with respect to changes in transport cost in the short run and in the long run.

Another way of comparing short and long-run results is to employ cointegration techniques and dynamic model. Seetaram (2010) and Seetaram (2012) analyzed tourist arrivals and departures in Australia during the period 1991-2008 respectively. After the testing for stationarity and cointegration (based on the suggestion of Pedroni, 1999 and 2004), she employed GMM and CLSDV (provided by Kiviet, 1995), to calculate short-run and long-run elasticities. The justification for using this estimation technique was that the temporal dimension was relatively small. For the tourist arrivals, she found that demand is inelastic in the short-run with respects to its determinants and elastic in the long-run. Concerning the tourism outbound, their results indicated an elasticity of migration on tourism outbound of 0.2% and amplified in the long-run with a value of 0.6%. However, contrary to the inbound model, there were no effects detected from the price index in the outbound model.

Finally, another way is to use directly an estimator for the cointegrated model: Fully Modified Ordinary Least Squares (FMOLS) or Dynamic Ordinary Least Squares (DOLS). Seetanah et al. (2010) employed a gravity model to test price and income elasticities. However, this paper suffers from a problem of methodology as they integrate many exogeneous variables into the regression. For a set of *n* variables, there can be up to *n-1* independent cointegrating vectors (Harris, 1995). If explanatory variables were cointegrated among them, there would more than *n-1* cointegrating vectors. So it is not possible to put explanatory variables that are cointegrated between them simultaneously in the regression (FMOLS or DOLS).

Regarding tourism and economic growth, the lack of theoretical foundations in explaining the mechanism through which tourism causes economic growth has prompted several authors in applying cointegration techniques in order to explain the causality between the two variables. This has resulting in a small but growing number of empirical studies using dynamic panel data techniques. The first studies have analysed the effect or a growing tourism sector on the economic growth of the destination (tourism growth → economic growth). Eugenio-Martin et al. (2004) estimated the effect from growth of tourism per capita to income per capita with a sample including 21 Latin American countries from 1985 to 1998. They used a dynamic panel data model with AB estimator and they categorised the countries into three groups based on their respective income per capita: rich, medium and poor. The results established that tourism led to economic growth for medium and poor countries while the reverse was true for rich countries.

Cortés-Jiménez (2008) analyzed the impact of international and domestic tourism arrivals on economic growth (by using the GDP per capita) for 17 regions of Spain and 20 regions of Italy during the period 1990-2000. She used two estimators for the dynamic panel model:

GMM and CLSDV. She found that tourism impacted positively to economic growth in for costal and Mediterranean regions. Her results are however, surprising for the inland regions in her sample. Here she found that while domestic tourism fostered economic growth, international tourism had the reverse effect.

Soukiazis and Proença (2008) studied the effect of tourism activity on the economic convergence of Portuguese regions between 1993 and 2001. Convergence was approximated by the difference of GDP per capita. Three estimators were applied: GMM, CLSDV and GLS (with random effects). The authors used bed capacity as proxy for tourism activity. According to their results, a tourism capacity expansion of 1% conducts to a supplement of 0.01% on economic growth. Their results indicated also that tourism accelerates economic convergence across the regions. Without tourism, the convergence was estimated to take 11 years whereas when the effect of tourism is taken into account, this reduces to 10 years only.
Sequeira and Maçãs Nunes (2008) used a sample of 94 countries (poor countries or small countries, with a population less than 5 millions) for the period 1980-2002. Three indicators of the tourism activity were employed (tourism arrival in population proportion, tourism receipts as a percentage of exports and tourism receipts as a percentage of GDP). They found that tourism had a positive effect on economic growth but only for the poor countries.

In the tourism literature, the panel data cointegration technique has also been employed in order to test for the causality between tourism growth and economic development. Narayan et al. (2010) used cointegration tests to estimate long-run elasticities of tourism to Pacific Islands (Tonga, Fiji, Solomon, Papua New-Guinea) for the period extending from 1988 to 2004. They then analyzed the causality from tourism exports to GDP with an ECM model (as its structure permits to check at the short-run and at the long-run). The tourism literature is not clear on the direction of the causality between economic growth and tourism. There are schools of thoughts which suggest that economic growth and development can also impact on tourism demand (economic growth → tourism).

A few studies have empirically tested this link by using cointegration methods for data panel model. For example, Lee and Chang (2008) used panel cointegration methods to determine the relationship and the causality between economic growth and tourism development in 55 countries (developed and developing countries) covering the time span of 1990 to 2002. After performing the IPS unit root and Pedroni (1999) cointegration tests, they applied FMOLS estimator to estimate the long-run relationship between tourism and economic developments. They then performed causality tests, the results of which indicated a unidirectional causality for developed countries (tourism → economic growth) whereas in developing countries, there were bidirectional causality relationships.

It is interesting to note that tourism has not only benefited economic growth but also international trade. Santana-Gallego et al. (2011) employed FMOLS, DOLS while Keum (2011) applied multiple dynamic causal patch analysis to investigate the causality between tourism and trade. Cassette et al. (2009) and Petit (2010) developed panel data cointegration models to analyze whether international trade of goods and services had any effect on income inequalities. Their results indicate that international trade in tourism services results in income inequalities disadvantaging low-skilled workers. Whilst tourism activities seem to benefit economic growth, the low-skilled workers seem to be losing from globalization of the tourism services.

**Limitations of panel data analysis**

The first disadvantage of this technique is that it is fairly complex to estimate and data requirements are high. Observing a number of individuals over a period of time usually results in data collection that can be tedious and expensive Baltagi (2005). From a statistical perspective, panel survey designs have some inherent disadvantages as noted in Kitamura (2000, pp 127).

1. Respondents may find it cumbersome to regularly participate in the same survey, which results in increasing non-response.
2. Attrition or dropout rate from the sample can be high.
3. Overtime the accuracy of data collection may decline. This is known as 'panel fatigue'.
4. The response of individuals may be influence by their responses from previous participations.

These disadvantages can however be addressed although solutions do come at a cost. Solution proposed, include 'refreshing' the survey design and adding fresh participants at later stages. For more in-depth analysis of attrition in panel data see Alderman et al. (2001), Fitzgerald, et al. (1998) and Uhrig et al. (2008).

## Conclusion and Future Directions

The use of panel data modelling techniques is becoming more common in the analysis of tourism data. In less than a decade, the literature has progressed from the development of standard static fixed effect and random effect models in the early 2000's to the more sophisticated dynamic panel data cointegration models. So far, the exploitation of this technique has been restricted to tourism demand studies and others which have explored the relationship between the expansion of tourism sectors and economic growth or international trade. It can be extended to other topics of research within the tourism context.

One interesting area is the investigation of the supply side of the tourism industry. For example, a scrutiny of regional differences in the productivity of this industry will benefit from panel data modelling techniques, as models developed may be used to control for the heterogeneity of each region. Other types of studies where this method can prove to be useful are examinations of labour markets and their outcomes across different districts within a destination or in international comparisons of destinations. Examples of problems that can be addressed are analysis of wage differentials and gender bias or human capital formation and return to education within the tourism industry.

While panel data techniques are powerful, and generally yield reliable estimates, they nevertheless suffer from some weaknesses. Estimations can be complex and data requirements are fairly large especially for dynamic panel models. The main limitations however, is that the panel data survey design is inevitably subject to problems related to attrition. The later drawback however, does not seem to have affected tourism research as the existing studies are based on secondary data or 'macro' data where the occurrence is either easier to deal with or less frequent. It is expected however that future research may be based on survey data when researchers seek to find answers to the behaviour of tourists using micro level data.

Regarding unit roots and cointegration tests, it is noted that the tests applied can be restrictive in that they assume that there is no cross sectional correlations in the sample. In the instances that this assumption is overly binding and inappropriate, researchers may circumvent it by using alternative tests such as those of Westerlund (2007) who has designed tests which perform well in smaller samples and have higher power. Additionally, Westerlund (2006) and Westerlund and Edgerton (2006) have developed cointegration tests for panel data sets with structural breaks. These tests will be useful for researchers who are attempting to examine whether the effect of shocks on the tourism data are permanent or transitory. The exploitation of these tests will be made easier when software such as STATA which is the dominant one for estimating panel data models catch up with theoretical development in this field and incorporate them in the newer version of the programme. Finally it is expected that in the future, the literature will see more use of panel-ECM and that application of panel data techniques for analysing tourism data will extend to non-linear models including binary models.

## REFERENCES

Alderman, H., Behrman, J., Kohler, H.P., Maluccio, J.A. and Watkins, S. (2001). “Attrition in Longitudinal Household Survey Data: Some Tests from Three Developing Countries,” Demographic Research,5(4):79-124

Allison. P. (2005) “Fixed Effects Regression Methods for Longitudinal Data Using SAS” SAS Press, USA

Anderson, T.W. and C. Hsiao (1981), “Estimation of Dynamic Models with Error Components”, *Journal of the American Statistical Association,* 76, 589-606.

Arellano, M. and S. Bond (1991), “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,” *Review of Economic Studies*, 58, 277-297.

Arita, S., C. Edmonds, S. La Croix and J. Mak (2011), “Impact of Approved Destination Status on Chinese travel abroad: an econometric analysis”, *Tourism Economics*, 17(5): 983-996

Baltagi, B. (2005), *Econometric Analysis of Panel Data*, 5$^{th}$ Edition, Wiley and Sons, UK.

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. in J. R. Nesselroade & P. B. Baltes (Eds.), Longitudinal research in the study of behavior and development, pp. 1–39. New York: Academic Press.

Banerjee, A. (1999), "Panel Data Unit Roots and Cointegration: An Overview," *Oxford Bulletin of Economics and Statistics*, 61, 607-629.

Banerjee, A., M. Marcellino and C. Osbat (2004), "Some Cautions on the Use of Panel Methods for Integrated Series of Macroeconomic Data," *Econometrics Journal*, 7, 322-340.

Blundell R.W. and S.R. Bond (1998), “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models”, *Journal of Econometrics*, 87: 115-143.

Breitung, J. (2000), “The Local Power of Some Unit Root Tests for Panel Data,” in *Advances in Econometrics, Vol. 15: Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, ed. by B. Baltagi, JAI Press, Amsterdam, 161-178.

Cassette A., N. Fleury and S.Petit (2009), “Income Inequality and International Trade: Short and Long-Run Evidence and The Specific Case of Tourism Services”, paper presented to the 2nd Conference of the International Association for Tourism Economics, Chiang Mai, Thailand, December 2009

Choi, I. (2001), “Unit Root Tests for Panel Data,” *Journal of International Money and Finance*, 20, 249-272.

Cortés-Jiménez, I. (2008), “Which Type of Tourism Matters to the Regional Economic Growth? The Cases of Spain and Italy”, *International Journal of Tourism Research*, 10: 127-139.

Cunado J., L-A. Gil-Alana and F. Perez De Gracia (2008), “Persistence in International Monthly Arrivals in the Canary Islands”, *Tourism Economics*, 14(1): 123-129.

Eilat , Y. and L. Einav (2004), “Determinants of International Tourism: A Three Dimensional Panel Data Analysis”, *Applied Economics*, 36(12): 1315–1327.

Espinet J.M., M. Saez, C. Coenders and M. Fluvià (2003), “Effect on Prices of the Attributes of Holiday Hotels: A Hedonic Prices Approach”, *Tourism Economics*, 9(2): 165-177

Eugenio-Martín J.L., N.M. Morales and R. Scarpa (2004), “Tourism and Economic Growth in Latin American Countries: A Panel Data Approach”, FEEM working paper No. 2004.26.

Fitzgerald, J., Gottschalk, P. and Moffitt, R. (1998). “An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics,” The Journal of Human Resources, 33, 251-299.

Frees, Edward W. (2004). Longitudinal and Panel Data: Analysis and Applications in the Social Sciences.West Nyack, NY, USA: Cambridge University Press, 2004.

Garín-Muños, T., (2006). “Inbound International Tourism to Canary Island: A Dynamic Panel Data Model.” *Tourism Management*, 27(1): 281-291.

Garín-Muños, T and L. F. Montero-Martín (2007). “Tourism in the Balearic Island: A Dynamic Model for International Demand Using Panel Data.” *Tourism Management,* 27(5): 1224-1235.

Gil-Alana L-A. (2010), “International Arrivals in the Canary Islands: Persistence, Long Memory, Seasonality and Other Implicit Dynamics”, *Tourism Economics*, 16(2): 287-302.

Green, W. (1999), *Econometric Analysis*, 4th Edition, Prentice Hall, USA.

Harris, R. I. D. (1995), *Using Cointegration Analysis in Econometric Modelling*, Prentice Hall Publishing, Essex England.

Hausman, J. A.(1978). "Specification Tests in Econometrics". *Econometrica* 46 (6): 1251–1271

Hsiao, C. *(2003), Analysis of Panel Data*, 2nd Edition, Cambridge University Press, New York.

Judson, R. and A. Owen (1999), “Estimating Dynamic Panel Models: a Guide for Macroeconomist.” *Economics Letters*, 65, 53-78.

Kao, C. (1999), “Spurious Regression and Residual-Based Tests for Cointegration in Panel Data,” *Journal of Econometrics*, 90, 1-44.

Keum, K. (2011), “International Tourism and Trade Flows: A Causality Analysis Using Panel Data”, *Tourism Economics*, 17(5), 949-962.

Khadaroo, J. and B. Seetanah, (2007). "Transport Infrastructure and Tourism Development." *Annals of Tourism Research*, 34(4): 1021-1032.

Khadaroo, J. and B. Seetanah (2008), "The Role of Transport Infrastructure in International Tourism Development: A Gravity Model Approach," *Tourism Management,* 29, 831-840.

Kitamura, R (2000), "Longitudinal Methods" in Handbook of Transport Modelling, D.A. Hensher and K.J Button (eds) Pergamon, Nethelands pp.13-29.

Kiviet, J.F. (1995). "On Bias, Inconsistency and Efficiency of Various Estimators in Dynamic Panel Data Models." *Journal of Econometrics*, 68(1): 53-78.

Kuo, H-I., C-L. Chang, B-W. Huang, C-C. Chen and M. MacAleer (2009), "Estimating the Impact of Avian Flu on International Tourism Demand Using Panel Data", *Tourism Economics*, 15(3), 501-511.

Kuo, H-I., C-C. Chen, W-C. Tseng, L-F. Ju and B-W. Huang (2008), "Assessing Impacts of SARS and Avian Flu on International Tourism Demand to Asia", *Tourism Management*, 29: 917-928.

Maddala, G. S. and S. Wu (1999), "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test," *Oxford Bulletin of Economics and Statistics*, 61, 631-652.

Malloney W-F. and G-V. Montes Ronjas (2005), "How Elastic Are Sea, Sand And Sun? Dynamic Panel Estimates of the Demand for Tourism", *Applied Economic Letters*, 12: 277-280.

Ledesma-Rodríguez, F. J. and M. Navarro- Ibáñez (2001). "Panel Data and Tourism: A Case Study of Tenerife." *Tourism Economics,* 7(1): 75-88.

Lee, C-C. and C-P. Chang (2008), "Tourism Development and Economic Growth: A Closer Look at Panels", *Tourism Management*, 29: 180-192.

Narayan, P.K., S. Narayan, A. Prasad and B.C. Prasad (2010), "Tourism and Economic Growth: A Panel Data Analysis for Pacific Island Countries", *Tourism Economics*, 16(1): 169-185.

Naudé W.A and A. Saayman (2005). "Determinants of Tourist Arrivals in Africa: A Panel data Regression Analysis." *Tourism Economics*, 11(3): 365-391.

Neumayer E. (2003), "The Impact of Political Violence on Tourism–Dynamic Econometric Estimation in a Cross-National Panel", *Journal of Conflict Resolution*, 48(2): 259-281.

Nickell, S., (1981). "Biases in Dynamic Models with Fixed Effects." *Econometrica,* 49(6): 1417–1426.

Pedroni, P., (1999). "Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors." *Oxford Bulletin of Economics and Statistics*, 61(November), 653–670.

Pedroni, P., (2004). "Panel Cointegration; Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis." *Econometric Theory*, 20(3): 597–625.

Petit S. (2010), *An Analysis of International Tourism: Production Fragmented, Two-Way Trade, Redistributive Effects,* PhD. Thesis, University of Lille 1, France.

Phillips, P. C. B. and H. R. Moon (2000), "Nonstationary Panel Data Analysis: An Overview of Some Recent Developments," *Econometric Reviews*, 19, 263-286.

Rugg, D. (1972), "The Choice of Journey Destination: A Theoretical and Empirical Analysis", *Review of Economics and Statistics*, 55 (1): 64-72.

Santana-Gallego M., F. Ledesma-Rodríguez and J.V. Pérez-Rodríguez (2011), "Tourism and Trade in OECD Countries. A Dynamic Heterogeneous Panel Data Analysis", *Empirical Economics*, 41(2): 533-554.

Seetanah B. (2010), "Using the Panel Cointegration Approach to Analyze the Determinants of Tourism Demand in South Africa", *Tourism Economics*, 16(3): 715-729.

Seetaram, N. (2010a). "Use of Dynamic Panel Cointegration Approach to Model International Arrivals to Australia", *Journal of Travel Research,* 49(4), pp 414-422.

Seetaram, N. (2010b). "Computing Airfare Elasticities or Opening Pandora's Box" *Research in Transportation Economics, Special Issue on Tourism and Transport*, 26, pp. 27-36.

Seetaram, N. (2012). “Estimating Demand Elasticities for Australia’s International Outbound Tourism”, *Tourism Economics*, (*in press*)

Sequeira, T.N. and P.M. Nunes (2008), “Does Tourism Influence Economic Growth? A Dynamic Panel Data Approach”, *Applied Economics*, 40: 2431-2441.

Soukiazis, E. and S. Proença (2008), “Tourism as an Alternative Source of Regional Growth in Portugal: A Panel Data Analysis at NUTS II and III Levels”, *Portuguese Economic Journal*, 7: 43-61.

Uhrig, C. N. (2008). “The Nature and Causes of Attrition in the British Household Panel Study,” ISER Working Paper 2008-05.

Westerlund, J. (2006a). Testing for Panel Cointegration with Multiple Structural Breaks, *Oxford Bulletin of Economics & Statistics, 68(1, pp. 101-132.*

Westerlund, J. and Edgerton, D. (2006b). A Simple Test for Cointegration in Dependent Panels with Structural Breaks, *Oxford Bulletin of Economics and Statistics, 70 (5), pp. 665-704.*

Westerlund, J. (2007). Testing for Error Correction in Panel Data, *Oxford Bulletin of Economics and Statistics, 69( 6), pp. 709-748,*