Industrial Management & Data Systems

Selection and industrial applications of panel data based demand forecasting models

Shuyun Ren Tsan-Ming Choi

Article information:

To cite this document:
Shuyun Ren Tsan-Ming Choi , (2016),"Selection and industrial applications of panel data based demand forecasting models", Industrial Management & Data Systems, Vol. 116 Iss 6 pp. 1131 - 1159
Permanent link to this document:
http://dx.doi.org/10.1108/IMDS-08-2015-0345

Selection and industrial applications of panel data based demand forecasting models

Shuyun Ren
The Hong Kong Polytechnic University, Kowloon, Hong Kong, and
Tsan-Ming Choi
Business Division, Institute of Textiles and Clothing,
The Hong Kong Polytechnic University, Kowloon, Hong Kong

Received 20 August 2015
Revised 24 December 2015
Accepted 1 February 2016

Abstract

Purpose – Panel data-based demand forecasting models have been widely adopted in various industrial settings over the past few decades. Despite being a highly versatile and intuitive method, in the literature, there is a lack of comprehensive review examining the strengths, the weaknesses, and the industrial applications of panel data-based demand forecasting models. The purpose of this paper is to fill this gap by reviewing and exploring the features of various main stream panel data-based demand forecasting models. A novel process, in the form of a flowchart, which helps practitioners to select the right panel data models for real world industrial applications, is developed. Future research directions are proposed and discussed.

Design/methodology/approach – It is a review paper. A systematically searched and carefully selected number of panel data-based forecasting models are examined analytically. Their features are also explored and revealed.

Findings – This paper is the first one which reviews the analytical panel data models specifically for demand forecasting applications. A novel model selection process is developed to assist decision makers to select the right panel data models for their specific demand forecasting tasks. The strengths, weaknesses, and industrial applications of different panel data-based demand forecasting models are found. Future research agenda is proposed.

Research limitations/implications – This review covers most commonly used and important panel data-based models for demand forecasting. However, some hybrid models, which combine the panel data-based models with other models, are not covered.

Practical implications – The reviewed panel data-based demand forecasting models are applicable in the real world. The proposed model selection flowchart is implementable in practice and it helps practitioners to select the right panel data-based models for the respective industrial applications.

Originality/value – This paper is the first one which reviews the analytical panel data models specifically for demand forecasting applications. It is original.

Keywords Data systems, Demand forecasting, Model selection, Panel data forecasting, Technical review, Use of information

Paper type Literature review

Nomenclature

i	the cross-section dimension, $i = 1, 2, \ldots, N$
t	the time-series dimension, $i = 1, 2, \ldots, T$
N	N individuals
T	T time periods
y_{it}	the demand of individual i at time period t
β	parameter matrix $(1 \times K)$
X_{it}	the itth observation on K exogenous variables
γ	coefficient
u_{it}	an error term usually be modeled as random variable with a zero mean and a fixed variance
U_t	the disturbance vector form of u_{it}

Industrial Management & Data Systems
Vol. 116 No. 6, 2016
pp. 1131-1159

0263-5577
DOI 10.1108/IMDS-08-2015-0345

This paper is a part of the first author's (Shuyun Ren) PhD dissertation.

α	constant	u	the vector of individual effects
α_i	unobservable individual-specific effect	ϕ_t	the remainder disturbances which are independent of u
λ_t	unobservable time effect	δ	spatial autoregressive coefficient
W	a $N \times N$ spatial weight matrix whose diagonal element are zero	ρ	serial autoregressive coefficient

1. Introduction

Forecasting is an integral part of industrial operations and production management. Demand forecasts are important for understanding market situation and the competition, production planning, including promotions, pricing, advertising, and distribution (Frees and Miller, 2004). However, forecasting the future demand is a truly challenging task. Various methods including statistical methods, intelligent methods, and hybrid methods have been developed to conduct forecasting. In recent years, with the emphasis on "big data" and the data driven knowledge-based operation management, panel data-based forecasting models have been widely adopted in various industrial settings. Panel data, also called time-series and cross-section data or pool data (Hsiao, 2003), follows a given sample of individuals over time. It involves two dimensions: a cross-sectional dimension N[1], and a time-series dimension T, and thus it provides two-dimensional observations on each individual in the sample. The panel data method is timely in the big data era, although the collection of panel data is more costly than the one-dimensional ones[2]. Panel data models have some advantages over the time-series econometric models. They usually give a larger number of data points and incorporate much richer information from both time-series and cross-sectional dimensional data. Panel data models consider variables observed over time and across different units, and can identify effects that simply are not detectable through the purely cross-section or time-serial analysis of data. Hence, panel data methods improve the efficiency of econometric estimates (Hsiao, 2003). In a recent study, Ren *et al.* (2015) suggest that panel data-based forecasting models outperform both time-series methods and artificial intelligent methods in fashion sales forecasting. Panel data approach also reduces the problem of multi-collinearity and provides a higher degree of freedom in the model estimation (Song and Li, 2008). Therefore, it is especially suitable for the forecasting problem when: the time series for all variables are shorter; and cross-sectional information on these variables is available.

Over the past decades, panel data models and forecasting analysis have been used in many research areas. Baltagi (2008b) give a pioneering survey of forecasting with panel data and find that panel data estimators perform well in forecast performance, though the performance of various panel data estimators and their corresponding forecasting performance may vary from one empirical example to another. Different from Baltagi (2008a), the current paper aims at providing guidance for panel data forecasting procedures and further discussing on the strengths, the weaknesses, the application performance, etc., of the panel data-based demand forecasting models. This paper contributes to the literature and advances knowledge in three ways: first, to the best of our knowledge, this paper is the first one which reviews the analytical panel data models specifically for demand forecasting applications. Second, we provide a novel model selection flowchart to let decision makers choose the right panel data method for their specific demand forecasting tasks with respect to the proper data tests. Third, we reveal the strengths, the weaknesses, and the industrial applications of different panel data-based demand forecasting models and discuss future research directions.

As a remark, for the literature searching process: first, we systematically search the major research portals on panel data-related studies published in English journals with the keywords such as panel data, cross-section data, panel data forecast, etc. Then, from the results, we further identify some related models and searched further. Thus, it is a top-down approach in literature searching. In addition, with respect to the objectives of this paper, we mainly focus on the original works on panel data models which report the latest forecasting applications. To help readers better understand the industrial applications, the corresponding details are also extracted and discussed. Figures and tables are added to enhance the presentation of results.

The rest of this paper is organized in the following: Sections 2 and 3 presents the general analytical models of panel data together with estimation methods for each regression model and the testing methods for panel stationary, individual effects, and dependence relationship, respectively. Section 4 suggests a novel workable process for panel data model establishment, and summarizes the industrial applications of panel data models as found from the literature. Section 5 concludes this paper with some important remarks and discussions of future research. To enhance presentation, a list of abbreviation and notation is provided, and Figures A1 and A2 are placed in the Appendix.

2. Analytical models

Panel data models have some advantages over the time-series econometric models in conducting demand forecasting (refer to Table I).There are two main streams of the panel data-based forecasting modeling approaches, namely, the "discrete choice modeling (DCM)" and the "regression choice modeling." Although the main focus of this study is on the regression choice modeling-based forecasting and its applications, we shall also concisely review the application of DCM, based on panel data. The econometric discrete choice analysis is an essential component of studying individual choice behaviors, which allows researchers to analyze and predict how people's choices are influenced by their personal characteristics and by the alternatives available to them. The three most common panel data-based DCMs are logit, nested logit, and probit (Honoréand Kyriazidou, 2000). They are used to describe and predict discrete choices of decision makers or to classify a discrete outcome according to a host of repressors. The panel data-based DCMs are widely used for the analysis of individual choice behaviors (Kim *et al.*, 2005) and can be applied to choice problems in many fields such as economics, tourism (Naudé and Saayman, 2005; Eilat and Einav, 2004), environmental management, urban planning (Arellano and Carrasco, 2003), and transportation. Typical examples of the use of the panel data-based DCM are the travel demand forecasting problem in the transportation industry. The panel data-based DCMs provide the opportunity to construct a dynamic model of travel behaviors which

	Panel data	One-dimensional data
Sample data	Two dimensions (N, T)	Only one dimension T
Forecasting accuracy	More accurate prediction	Less accurate prediction
Learning individual's behavior	By observing the behavior of itself together with others	By observing the behavior of itself
Conducting behavior models	More complicated behavioral models	One-dimension models
Collinearity	Can reduce collinearity	Unavoidable

Table I. Advantages of panel data-based models compared to time-series-based models can improve the prediction ability over time. In the literature, Gopimatj (1995) adopts the panel latent class choice model in capturing heterogeneity in travel demand choice processes with a casual structural formulation. Observe that the logit model commonly assumes that paths are independent of each other and have independence from the irrelevant alternatives property. In reality, however, many paths overlap with each other and are thus not independent.

Another main stream of the panel data-based demand forecasting focusses on the regression analysis. In this section, we introduce several kinds of respective panel data regression models that are commonly used in forecasting problems following the common classification of panel data model in Figure A1. From the stage dependency view, there are static panel data model and dynamic panel data model. Static model is more structural than behavioral while dynamic model is a representation of the behavior of the system's static components. Considering the impact of the individual-specific effects, panel data model can be classified as fixed-effects model and random-effects model. In the fixed-effects model, the effects of omitted individual-specific variables are treated as fixed constants over time; while in the random-effects model, the individual-specific effects are treated as random variables. Besides, based on the dependence relationship in error term, panel data models can be categorized as spatial correlation model and serial correlation model. Spatial correlation model enable decision makers to identify and control for correlations across cross-section units, such as state/region correlation in energy demand forecasting and land-use forecasting. Serial correlation panel data models deal with the correlation existing among error terms from different time periods, which cannot be well described by a constant or an independently distributed error term. After briefly introducing the general notation of panel data models, in the rest of this section, we will give a detailed introduction for panel data models of each category and the estimation methods from an analytical perspective.

2.1 Static models

Following[3], the common panel data regression model is presented as:

$$y_{it} = \alpha + \beta' X_{it} + u_{it}, \tag{1}$$

where $i=1, 2, ..., N$ denotes N individuals. $t=1, 2, ..., T$ denotes T time periods. The i subscript therefore denotes the cross-section dimension, whereas t denotes the time-series dimension, α is a constant, $\beta'(1\times K)$ is fixed but contains unknown parameters and X_{it} is the itth observation on K exogenous variables, u_{it} is a random disturbance term (i.e. noise).

Next, the "fixed-effects" and "random-effects" models are proposed in the literature to account of the individual differences in the panel data.

2.1.1 Fixed-effects model. Considering the following one-way error component model (Balestra and Nerlove, 1966) that is the most widely used specification in economics literature:

$$y_{it} = \beta' X_{it} + \alpha_i + u_{it}, \tag{2}$$

We assume that there are no time-specific effects, and only individual-specific effects are present in this model. Under the fixed-effects case, the individual-specific effects α_i are assumed to be fixed parameters which require estimation. The error term u_{it} denotes the effects which are peculiar to both the individual units and time periods, and it is usually modeled as an i.i.d. random variable with a zero mean and a fixed variance. X_{it} is assumed to be independent of the u_{it} for all i and t. Note that this kind of fixed-effects error component model has been studied in Wallace and Hussain (1969) and Swamy and Arora (1972). The advantage of fixed-effects inference is that there is no need to make an assumption that the effects are independent of α_i, while the disadvantage is that it introduces the issue of incidental parameters (Hsiao, 2003). According to Hsiao (2003), ordinary-least-squares (OLS) estimator is the best linear unbiased estimator (BLUE). The OLS estimation of α_i and β are obtained by minimizing:

$$S = \sum_{i=1}^{N} U_i' U_i = \sum_{i=1}^{N} (y_i - e\alpha_i - X_i\beta)'(y_i - e\alpha_i - X_i\beta). \tag{3}$$

Taking partial derivatives of S with respect to α_i and setting them equal to 0, we have:

$$\hat{\alpha}_i = \bar{y}_i - \beta'\overline{X}_i,\ i = 1, \ldots, N, \tag{4}$$

where:

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it},\ \overline{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}.$$

Then, we can get the estimation of β as follows:

$$\hat{\beta}_{CV} = \left[\sum_{i=1}^{N}\sum_{i=1}^{T}\left(X_{it} - \hat{X}_i\right)\left(X_{it} - \hat{X}_i\right)'\right]^{-1}\left[\sum_{i=1}^{N}\sum_{i=1}^{T}\left(X_{it} - \hat{X}_i\right)\left(y_{it} - \hat{y}_i\right)\right]. \tag{5}$$

Observe the OLS estimator is a consistent (see footnote 3) estimator (Amemiya, 1985) under the fixed-effects assumption when T tends to infinity. A necessary condition for an unbiased and consistent parameter estimation under OLS is that there is no correlation between the error term and any of the explanatory variables.

2.1.2 Random-effects model. Unlike the fixed-effect model in which the effects of omitted individual-specific variables α_i are treated as fixed constants over time, in the random effect (RE) model, the individual-specific effects are treated as random variables. The advantage of random-effects inference is that the number of parameters is fixed which implies that an efficient estimation method can be derived. The disadvantage is that the decision maker has to make specific assumptions about the pattern of correlation (or no correlation) between the effects and the included explanatory variables (Hsiao, 2003). The linear regression, which is called two-way component model in Baltagi (2008a), can be written the following:

$$y_{it} = \beta' X_{it} + \alpha_i + \lambda_t + u_{it}, \tag{6}$$

where λ_t denotes the unobservable time effect and u_{it} is the reminder stochastic disturbance term. Note that under the fixed effect assumption, α_i and λ_t are assumed to be fixed parameters to be estimated and the reminder disturbance noise is stochastic and is modeled as an i.i.d. random variable with a zero mean and a fixed variance.

While for the random case, $\alpha_i \sim i.i.d(0, \sigma_\alpha^2)$, $\lambda_t \sim i.i.d(0, \sigma_\lambda^2)$. X_{it} is independent of α_i, λ_t, and u_{it}, and they are independent of each other. Thus, $E\alpha_i = E\lambda_t = Eu_{it} = 0$, $E\alpha_i\lambda_t = E\alpha_i u_{it} = E\lambda_t u_{it} = 0$, $E\alpha_i\alpha_j = \sigma_\alpha^2$ if $i = j$ or 0 otherwise. $E\lambda_t\lambda_s = \sigma_\lambda^2$ if $t = s$ or 0 otherwise. $Eu_{it}u_{js} = \sigma_u^2$ if $i = j$, $t = s$ or 0 otherwise. The variance of y_{it} is $\sigma_y^2 = \sigma_\alpha^2 + \sigma_\lambda^2 + \sigma_u^2$.

The OLS estimator is unbiased and consistent under the assumption that individual effects are fixed constants, while it is not the BLUE for the random-effects model. It is a consistent and unbiased estimator under the random-effects assumption, even though it is not efficient when T is fixed (Mundlak, 1978). Thus, for the random-effect case, Baltagi (2008a) has proven that the generalized-least-squares (GLS) estimator is the BLUE. Therefore, we can obtain the estimation of β and u_{it} by using the GLS estimation:

$$\begin{aligned}\hat{\beta}_{GLS} &= \left[\frac{1}{T}\sum_{i=1}^{N} X_i' Q X_i + \psi \sum_{i=1}^{N} \left(\overline{X}_i - \overline{X}\right)\left(\overline{X}_i - \overline{X}\right)'\right]^{-1} \\ &\quad \times \left[\frac{1}{T}\sum_{i=1}^{N} X_i' Q y_i + \psi \sum_{i=1}^{N} \left(\overline{X}_i - \overline{X}\right)\left(\overline{y}_i - \overline{y}\right)\right], \\ &= \Delta\hat{\beta}_b + (\mathrm{I}_K - \Delta)\hat{\beta}_{CV} \end{aligned} \tag{7}$$

$$\hat{u}_{GLS} = \overline{y} - \hat{\beta}_{GLS}'\overline{X}, \tag{8}$$

where:

$$\Delta = \psi T\left[\sum_{i=1}^{N} X_i' Q X_i + \psi T \sum_{i=1}^{N} \left(\overline{X}_i - \overline{X}\right)\left(\overline{X}_i - \overline{X}\right)'\right]^{-1} \times \left[\sum_{i=1}^{N} \left(\overline{X}_i - \overline{X}\right)\left(\overline{X}_i - \overline{X}\right)'\right],$$

$$\hat{\beta}_b = \left[\sum_{i=1}^{N} \left(\overline{X}_i - \overline{X}\right)\left(\overline{X}_i - \overline{X}\right)'\right]^{-1}\left[\sum_{i=1}^{N} \left(\overline{X}_i - \overline{X}\right)\left(\overline{y}_i - \overline{y}\right)\right], \text{ and } \psi = \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}.$$

Note that Breuschand Pagan (1980) compared the within estimator with the GLS estimator for the random-effects one-way component model using finite sample and found that feasible GLS is more efficient than covariance estimator (CV) but has the lowest degree of freedom. When T is fixed and N goes to infinity, the maximum-likelihood estimator (MLE) is consistent (Anderson and Hsiao, 1982). In a static model with the strict exogeneity assumption, the presence of individual-specific constants does not affect the consistency of the CV or the MLE estimator of slope coefficients. The CV estimator is consistent for the static model no matter whether the effects are fixed or random.

2.1.3 Comparison of fixed-effects and random-effects. The fixed-effects panel data model and random-effects panel data model represent fundamentally different assumptions of the pooled data, although they employ similar sets of formulas, and sometimes yield similar estimates for the various parameters. Take the fashion products sales data as an example: the random-effects panel data model means that the individual-specific effects from other products or time period fluctuates over units following a distribution. If the effect in a panel data model is modeled as being random, we will learn the features of individual behavior from the features of observed individual behavior, rather than about these particular units themselves. We are interested to test and estimate the variance of these random-effects across different products. While under the fixed-effects assumption, the effects from other product or time period is the average effect of each fashion product, expressed by the regression coefficient. Hsiao (2003), Baltagi (2008a), and Lee and Yu (2010) give a detailed discussion on the choice between random-effects and fixed-effects models. The comparison between models of these two kinds of effects is illustrated in Table II. In practice, the selection of the appropriate model is important to ensure that the various statistics are estimated correctly (Borenstein *et al.*, 2010). Hausman (1978) has proven that using a fixed-effects specification produces significantly different results from a random-effects specification when estimating a wage equation using a sample of 629 high school graduates followed over a six years period. As a remark, Hsiao (2003) finds that whether to treat the effects as fixed or random makes no difference when T is large. For the choice between fixed-effects and random-effects models, Hsiao (1996) gives numerous examples in which the purpose of analysis will determine how to choose these two formulations.

2.2 Dynamic models

The panel data model has been widely used to estimate the parameters of dynamic econometric models. Note that we cannot estimate dynamic models from observations at a single point in time, and it is rare for single cross-section surveys to provide sufficient information about earlier time periods for dynamic relationship to be established. Dynamic panel data models which are able to describe the dynamic relationship between explained variables and explanatory variables are widely used to deal with demand forecasting problems in various research areas, such as energy consumption, tourism demand, water demand, etc. In the following subsection, we briefly introduce some commonly used dynamic models and introduce some consistent and efficient estimation (see footnote 2) methods for each kind of model.

2.2.1 The common regression model. Dynamic models (Baltagi, 2008a), containing lagged dependent variables, are used to estimate behavioral relationships that are dynamic in nature. The common regression model can be written as:

$$y_{it} = \gamma y_{i,t-1} + \beta' X_{it} + \alpha_i + \lambda_t + u_{it} \ (i = 1, ..., N \ \ t = 1, ..., T), \tag{9}$$

where $Eu_{it} = 0$, and $Eu_{it}u_{js} = \sigma_u^2$ if $i = j$ and $t = s$, and $Eu_{it}u_{js} = 0$ otherwise. α_i and λ_t can be fixed or random.

	Fixed-effects models	Random-effects models
Model assumption	The effects α_i and λ_t are the same for different time period and different individual, respectively	The effects α_i and λ_t are vary from time to time and individual to individual, respectively
Effects estimation	Estimate the common effects for all time periods and all individuals	Estimate the mean of the true effects distribution for all time periods and individuals

Table II. Comparisons between fixed-effects and random-effects models

The common regression model without exogenous variables can be expressed as:

$$y_{it} = \gamma y_{i,t-1} + \alpha_i + \lambda_t + u_{it}. \tag{10}$$

Observe that being "dynamic" in the estimation process may lead to the consistency of estimator changing. Thus, the estimation for dynamic models is different from static models' due to the presence of lagged dependent variables.

2.2.2 Estimation for dynamic models. If the model contains exogenous variables in addition to the lagged dependent variable, the situation becomes different. To be specific, not only may the covariance estimator and the MLE be inconsistent, but the interpretation of the model is dependent on the assumption about the initial condition (Anderson and Hsiao, 1982). Anderson and Hsiao (1981) study the problem of estimating a dynamic model with error components model when either the number of time point T or the number of cross-sectional unit N tends to infinity. They examine several models arising from different assumptions about the initial conditions. Their study shows that the MLE is consistent when T tends to infinity no matter what assumptions on the initial conditions are. When T is fixed and N tends to infinity, the consistency of the MLE depends on the assumptions about the initial condition. Anderson and Hsiao (1982) explore the sensitivity of MLE estimators by alternating the assumptions about initial conditions[4] and asymptotic plans. They argue that the advantage of these estimators is their consistency (irrespective of the initial condition and whether T or N or both were tending to infinity). The GMM estimator, developed by Hansen (1982), provides a convenient framework for dynamic models. Bond (2002) reviews the use of GMM estimators in the model which contains endogenous or predetermined explanatory variables, with a large number of cross-section unit observations for a small number of time periods ($N \to \infty$ for a fixed T) and suggests that GMM is useful and efficient for the estimation of this kind of panel data model. The one-step GMM estimation method suggested by Arellano and Bond (1991) has been applied to an unbalanced panel dataset consisting of country data for an employment analysis (Liu, 2004). Note that compared with the OLS estimator, the one-step GMM estimator gives more intuitive results in terms of sign and magnitude. Hsiao (2003) shows that the CV estimator (or the least-squares dummy variable estimator) is always consistent when $T \to \infty$. While it is always inconsistent when T is fixed (finite) no matter whether the individual effects are treated as fixed or random. Nerlove (1971) supports this conclusion by Monte Carlo simulation studies.

For the fixed-effects model. If lagged dependent variables appear as explanatory variables, the strict exogeneity property of the regressors no longer holds. The MLE or the CV estimator under the fixed-effects formulation is no longer consistent in the typical situation in which T is fixed and N tends to infinity (Hsiao, 2003). Although the conventional MLE and CV estimators are inconsistent when T is fixed and N tends to infinity, there exists a transformed likelihood approach that does not involve the incidental parameter and is consistent and efficient under a proper formulation of initial conditions. Hsiao *et al.* (2002) suggest a transformed MLE and a computationally simpler minimum distance estimator (MDE) for fixed-effects formulation and conduct Monte Carlo studies to evaluate the finite sample properties of the MLE, MDE, instrumental variable (IV) estimator and linear generalized method of moments (GMM) estimator. They show that the IV and GMM estimators do not need the formulation of initial conditions. Furthermore, the likelihood approach appears to dominate the GMM approach both in terms of the bias and root mean square error of the estimators and the size and power of the test statistics.

For the random-effects model. When the specific effects are treated as random, they can be considered to be either correlated or uncorrelated with the explanatory variables. If individual-specific effects are correlated with the explanatory variables (the lagged dependent variables), the OLS estimator for dynamic models is biased and inconsistent (Hsiao, 2003). For RE dynamic models, there are various ways to estimate the unknown parameters such as the MLE, the GLS (suggested by Blundell and Smith, 1991), the IV, and the GMM methods. Prior studies such as Nerlove (1971), Sargan and Bhargava (1983), and Nerlove and Balestra (1996) have discussed the ML estimation of the dynamic random-effects model. The MLE, the IV, and the GMM estimators are proven to be consistent (Hsiao, 2003), although the OLS estimator is no longer inconsistent for dynamic error component models with random-effects. With a random-effects formulation, the interpretation of a model always depends on the assumption of initial observation. The consistency property of some estimators also depends on this assumption and on the way in which the number of time-series observations (T) and the number of cross-sectional units (N) tend to infinity, for example the MLE, the CV, the IV estimators and the GLS estimator (Anderson and Hsiao, 1981) have been applied to study the problems of estimating a dynamic model with the error component under assumptions.

Dynamic panel models containing lagged dependent variables allow us to better understand the dynamics of adjustment. We summarize the statistical properties of some common estimators for dynamic models in Table III. Although there are many kinds of theoretical estimators for dynamic panel data models, the estimation performance is different from situation to situation in practice. The OLS methods that are explored in some details and several different methods for estimating parameters in the presence of lagged endogenous variables are proposed in Balestra and Nerlove (1966). Nerlove (1967) conducts Monte Carlo studies and suggests that the OLS method overestimates when N or T or both tend to infinity. Then, Houthakker *et al.* (1974) present a variance component technique developed by Balestra and Nerlove (1966) for estimating dynamic models and suggest that the variance component technique provides satisfactory results, while both OLS and IV cannot help achieve good estimation for the demand of gasoline and residential electricity case. After that, Babel *et al.* (2008) use the OLS estimator to estimate a time-dynamic stochastic model by utilizing a panel data approach in German mortality forecasting and get satisfying estimation results. Garín-Muñoz and Montero-Martín (2007) use GMM-DIFF estimators proposed by Arellano and Bond (1991) to estimate a panel data model which includes lag dependent explanatory variables and get a satisfactorily good performance model.

Estimator	CV	MLE	GMM	GLS	IV	MDE	Transformed MLE
Consistent for fixed-effects						Yes	Yes
Consistent for random-effects			Yes		Yes		
Dependent on initial conditions and the way T and N tend to infinity	Yes	Yes		Yes			

Table III. Statistical properties of some common estimators for dynamic model

2.3 Spatial correlation model

Spatial correlation models include the critical ideas such as distance-decay and spatial interaction Getis (2007), Jang and Shin (2014), and Anselin (1988). The spatial econometrics studies focus on exploring the dependence among observations across space and use the spatial weights matrix to describe the spatial arrangement of the geographical units in the sample (Baltagi, 2008a). Spatial panel data models with spatial error autocorrelation, including a spatially lagged dependent variable, have received great attention in the regional science literature (Elhorst, 2003). Note that the general panel data models allow us to control the heterogeneity across individual units (Baltagi, 2008a), while spatial panel data models can control for both heterogeneity and spatial correlation (Baltagi, 2008a; Baltagi *et al.*, 2003a, b). If the spatial dependence between observations is specified, the model may incorporate a spatial autoregressive process into the error term (which is known as the spatial error model), or contain a spatially autoregressive dependent variable (which yields the spatial lag model). From this perspective, the traditional panel data model only captures "average" or representative behaviors in cross-section diminution. It results in average effects across spatial units that overlook the differences in behaviors among individual spatial units (Quah, 1996). Let W denotes a $(N \times N)$ spatial weight matrix describing the spatial arrangement of the spatial units, w_{ij} denotes the (i, j)th element of W, where i and $j = 1, \ldots N$. The traditional spatial error autocorrelation model can be expressed as follows (Baltagi, 2008a):

$$Y_{it} = \beta' X_{it} + u_{it}, \tag{11}$$

The disturbance vector form is given by:

$$U_t = u + \phi_t, \tag{12}$$

with $\phi_t = \delta W \phi_t + v_t$, where $U_t = (u_{1t}, \ldots, u_{Nt})'$, $u = (u_1, \ldots, u_N)'$, $\phi_t = (\phi_{it}, \ldots, \phi_{Nt})$ and $v_t = (v_{1t}, \ldots, v_{Nt})$. $E(v_{it}) = 0$, $E(v_{it} v_{it}') = \sigma_v^2 I_N$, δ is called the spatial autocorrelation coefficient. The spatially lagged dependent variable model can be specified as:

$$Y_{it} = \delta' W Y_{i-1,t} + \beta' X_{it} + u_{it}, \tag{13}$$

where $E(u_{it}) = 0$, $E(u_{it} u_{it}') = \sigma^2 I_N$.

The spatial econometric literature (LeSage and Pace, 2010) has shown that the OLS estimator (of the response parameters) is unbiased for the spatial error autocorrelation model. For the case when the specification contains a spatially lagged dependent variable, the OLS estimator of the response parameters not only loses the property of being unbiased but is also inconsistent (Elhorst, 2003). To overcome this problem, prior studies Anselin (1988) and Anselin and Hudak (1992) use maximum-likelihood techniques to conduct the estimation. Following that, the GMM estimator is proven to be robust to spatial dependence among the error terms in spatial cross-section models (Anselin, 1999). Yu *et al.* (2008) establish asymptotic properties of the maximum-likelihood (ML) and quasi-maximum-likelihood (QML) estimators for a spatial dynamic panel model with fixed effects when both the number of individuals N and the number of time periods T are large. Then, Yu *et al.* (2012) extend previous study and examine the performance of QML, 2SLS, and GMM estimations for the unstable cases where there are unit roots generated by temporal and spatial correlations. They suggest that, QML estimation's consistency requires the condition that T tends to infinity, while the GMM is applicable even when T is small.

2.4 Serial correlation model

In an error component model, if error terms from different (usually adjacent) time periods (or cross-section observations) are correlated, we say that the error term is serially correlated. Under the serial correlation assumption, u_{it} is correlated with u_{is} for $t \neq s$ in Equation (2), no matter how far t is from s. An unobserved shock in period t will affect the behavioral relationship for the following period s. Serial correlation occurs in time-series studies when the errors associated with a given time period carry over into future time periods. This may be a restrictive assumption for economic relationship, such as investment and production demand forecasting. In serial correlation models, the error term of individual units are serially correlated due to the possible omission of relevant variables, while the existence of these variables is not well described by an error term that is either constant or independently distributed over time periods (Hsiao, 2003). There are different types of serial correlation. With the first-order serial correlation, errors in one time period are correlated directly with errors in the ensuing time period. Note that errors might also be lagged, e.g., if data are collected quarterly, the errors in summer of one year might be correlated with the errors of summer in the next year. With the positive serial correlation, errors in one time period are positively correlated with errors in the next time period.

If the serial correlation is present, the error term ϕ_{it} is expressed as (Baltagi, 2008a):

$$\phi_{it} = \rho\phi_{i,t-1} + u_{it}, \tag{14}$$

where $|\rho| < 1$ and $E(u_{it}) = 0$, $E(u_{it}u_{it}') = \sigma^2$. If the one-way error component model follows an AR (2) process, the error term ϕ_{it} is written as:

$$\phi_{it} = \rho_1\phi_{i,t-1} + \rho_2\phi_{i,t-2} + u_{it}, \tag{15}$$

where $|\rho_2| < 1$, $|\rho_1| < (1-\rho_2)$ and $E(u_{it}) = 0$, $E(u_{it}u_{it}') = \sigma^2$. When $\rho_2 = 0$, this model follows AR(1) process. Arellano and Bond (1991) and Baltagi and Li (1992) consider this serially correlated structure in the error components model. If the one-way error component model follows an MA(1) process, the error term ϕ_{it} is written as:

$$\phi_{it} = u_{it} + \lambda u_{i,t-1}, \tag{16}$$

where $|\lambda| < 1$, and $E(u_{it}) = 0$, $E(u_{it}u_{it}') = \sigma^2$.

This can be extend to the MA(q) case and the autoregressive moving average ARMA (p, q) case on the ϕ_{it}. Drukker (2003) gives a detailed illustration for how to test serial correlation. Serial correlation panel data models have the ability to capture more additional features of the data that may be of interest to an analyst than the common panel data models. It is important to note that the serial correlation will not affect the unbiasedness or consistency of OLS estimators, although it does affect their efficiency (Baltagi, 2008a). The first-differenced GMM estimator for the AR(1) panel data model is developed by Douglas *et al.* (1988) and Arellano and Bond (1991). Besides, the GLS estimator is also adopted in estimating serial correlation panel data models in Frees and Miller (2004). Frees and Miller (2004) use a serial correlation panel data model (called the longitudinal data model) to predict the sales of state lottery tickets. Using the mean absolute error criteria and the mean absolute percentage error criteria, the best forecasts are given by the error component model with AR(1) disturbances followed by the fixed-effects model with AR(1) disturbances. Baltagi *et al.* (2007) consider a spatial panel regression model with serial correlation over time for each spatial unit and spatial dependence across these units at a particular point in time and find that ignoring these correlations, such as spatial at a point in time or serial correlation for a spatial unit over time, may result in misleading inference. Serial correlation that exists among the data that are collected repeatedly across time occurs in time-series studies when the errors associated with a given time period carry over into future time periods.

In this section, we have introduced the commonly used panel data models for demand forecasting and summarized the corresponding estimation methods for each category. We have uncovered that the OLS estimator is unbiased and consistent for static model with both fixed-effects and random-effects. However, the situation becomes different if the model contains exogenous variables in addition to the lagged dependent variables. The OLS estimator is no longer efficient for the dynamic case, while MLE and GMM estimators are suggested to be useful for both fixed-effects and random-effects dynamic models estimation in the literature. We further summarize the estimators for different panel data models in Table IV.

3. Tests

In the above sections, we have introduced several kinds of panel data regression models that are commonly used in demand forecasting problems and the corresponding estimation methods. However, how to decide the right regression model for different industry settings and how to decide the individual effects and cross-section or time-series dependence deserve further investigations. Ignoring any of these will result in inefficient estimates and misleading inference. Thus, a series of pre-tests are important for the model establishment. In this section, we review the testing methods that can help us choose a suitable model for a given application case.

3.1 Panel stationary tests

For panel data applications it is important to know whether an observed panel series is stationary or non-stationary. Over the past decades, there have been several approaches to test for a unit root in panel data (Levin *et al.*, 2002). Prior studies, such as Quah (1992, 1994), Choi (2001), Im *et al.* (2003), Levin *et al.* (2002), and Maddala and Wu (1999), assume that the individual time series in the panel are cross-sectionally independently. For example, Quah (1994) proposes the asymptotically normal tests for a unit root. Levin *et al.* (2002) devise an adjusted *t*-test (LLC) for a unit root for various panel data models. Im *et al.* (2003) propose two tests for examining the unit roots in heterogeneous panels. Baltagi and Kao (2001) give

Panel data model		Estimator	Consistent
Static model	Fixed-effects	OLS, within estimator	OLS
	Random-effects	OLS, GLS, MLE	OLS, GLS
Dynamic model	Fixed-effects	GMM, IV, MDE, MLE, transformed MLE, OLS	GMM, IV, MDE, transformed MLE,
	Random-effects	MLE, IV, GMM, GLS	MLE, IV, GMM
Serial correlation model		OLS, GMM, GLS	OLS
Spatial correlation model		MLE, GMM, QML	GMM

Table IV. Efficient estimators for different panel data regression models a detailed review for this kind of studies. However, in the context of cross-section regression (including cross-country/region), the cross-section dependence should be taken into consideration since there might be common influences to all panel members. Thus, a number of panel unit root test that account for the cross-sectional correlation have been proposed in the literature (Chang, 2002; Phillips and Sul, 2003; Bai and Ng, 2004; Breitung and Das, 2005, 2008; Choi and Chue, 2007; Pesaran 2007). In particular, Gengenbach *et al.* (2009) compare the panel unit root testing methods with a common factor structure and discuss their use in econometric modeling. Bai and Ng (2009) also study whether the difference in finite sample properties can be traced to how the pooled autoregressive coefficient is estimated.

3.2 The individual-specific effects test

Testing for the correlation of unobservable individual effects in panel data regressions is a common practice (Arellano, 1993). Considering a common panel data model as described in Equation (2), the individual-specific effects (α_i) among cross-section individuals are unobservable and may be correlated with X_{it}. Generally, these effects are treated as fixed effects (or REs $=0$). For fixed effects models, the effects are specific to individual cross-sectional units but stay constant over time; or specific to each time period but are the same for all cross-sectional units, while random-effects models treat the effects as random variables. When deciding between these two effects in a panel data model, a Hausman pretest (Hausman, 1978), with the assumption that the REs are uncorrelated with the explanatory variables, is a common approach in many applications (Guggenberger, 2010). For most economics applications since the 1980s, Hausman pretest is also commonly used to make the choice between the REs and FEs estimators (Baltagi *et al.*, 2003a, b). Hausman (1978) has proposed an asymptotic χ^2-test based on the quadratic form obtained from the difference between a consistent estimator under the alternative hypothesis and an efficient estimator under the null hypothesis (Holly, 1982). The null hypothesis is that the efficient estimator is a consistent and efficient estimator of the true parameters. If it is, there should be no systematic difference between the coefficients of the efficient estimator and a comparison estimator that is known to be consistent for the true parameters. If the two models display a systematic difference in the estimated coefficients, then we may doubt the assumption of efficient estimator. This null hypothesis frequently does not withstand empirical tests (Hausman and Taylor, 1981). Note that Hsiao *et al.* (2002) suggest that the Hausman type specification test can also be used to test for fixed or random individual effects in dynamic panel data models, by conducting a Monte Carlo study. Hausman and Taylor (1981) propose a model that introduces an IV estimator using both between- groups and within-groups variation to correct for the correlation of selected repressors with the individual effect. Recently, to examine the equality of both the whole sets of coefficients and that of individual variables that cannot be addressed on the basis of the standard Hausman test, Frondel and Vance (2010) suggest a test variant based on the contrast of between-groups and fixed effects. Besides, the Lagrange multiplier (LM) test developed by Breusch and Pagan (1979, 1980) also can help test the individual effects for panel data models. Extensive Monte Carlo simulation studies on testing in this error component model are performed by Baltagi *et al.* (1992).

3.3 Spatial correlation test

As discussed above, the standard panel data model assumes that there is no spatial correlation. However, there are cases in which spatial correlation does exist.

For instance, for trade flows across a panel of countries, there are spatial effects affecting this trade, depending on the distances among these countries. As a result, the right panel data model for analysis should include spatial correlation. In fact, for the spatial dependence models, some estimation methods such as the GMM methods, have been reviewed above. In addition, there are two kinds of spatial correlation models, namely, the spatial error model and the spatial lag model. In this subsection, the testing methods for these two spatial correlation models will be reviewed. First, Anselin (1988, 2001), Anselin and Bera (1998), and Kelejian and Robinson (1998) develop the LM-based tests for a spatial autocorrelation analysis in cross-sectional spatial data which are observed for a given time point. Then, Baltagi *et al.* (2003a, b) extend the LM test to the spatial error component model and derive a conditional LM test to help test the random region effects in the panel as well as the spatial correlation across these regions. Recently, Jang and Shin (2014) suggest the joint LM and likelihood ratio (LR) tests, the marginal LM and LR tests, and the conditional LM and LR tests for examining both the spatial correlation and time effects. Their limiting null distributions are also derived by conducting a Monte Carlo experiment. Following Baltagi *et al.* (2003b), Baltagi and Liu (2008) derive a joint LM test which simultaneously tests for the absence of spatial lag dependence and random individual effects in a panel data regression model.

3.4 Serial correlation test

Testing for serial correlation has been a standard practice in many areas, such as in applied econometric analysis. The logic is: if the residuals are serially correlated, the least squares estimator may be inefficient and inconsistent if the regressors contain lagged dependent variables (Li and Hsiao, 1998). For time-series data, a lot of prior studies (e.g. Breusch, 1978; Breusch and Pagan, 1980; Godfrey, 1978a, b; Banerjee *et al.*, 1998) have explored the serial correlation testing problem over the past decades. For panel data studies, Gardner (1960) is the first one who extends the error component model to take into account the serial correlation in the remainder disturbance term and test for serial correlation, under the assumption that there are no random effects (using the LM test derived in Godfrey, 1978a, b; Breusch, 1978; Breusch and Pagan, 1980). Then, Bhargava *et al.* (1982) modify the Durbin-Watson statistics (Bhargava *et al.*, 1982) to test the serial correlation when the individual effects are assumed to be fixed. Baltagi and Li (1991) derive a simple LM test which jointly tests the presence of random individual effects and serial correlation. Baltagi and Li (1995) also address this kind of jointly testing problems. By generalizing the testing methods for time-series data, Li and Hsiao (1998) propose two methods to test the zero first-order serial correlation, and the higher-order serial correlations in a residual-based dynamic panel data model. Some Monte Carlo experiments also are conducted to examine the finite sample performances of the proposed tests. After that, a new test for serial correlation in random-effects or fixed-effects one-way models derived by Wooldridge (2002) is proposed, because it can be applied under general conditions and is easy to implement. Baltagi *et al.* (2007) generalize the previous studies by deriving test statistics for the spatial panel data model with serial correlation. In particular, they derive joint and conditional LM and LR tests and look into their small sample properties using Monte Carlo experiments. Extending the time-series test in Baltagi *et al.* (2007), Westerlund (2007) further propose a serial correlation test for panel data analysis based on structural properties that do not impose any common factor restriction. Their simulation results suggest that the proposed test has good small sample properties and possesses a high power compared to other popular residual-based panel co-integration tests.

4. Model selection process and industrial applications

After reviewing and exploring the features of various panel data models and the estimation methods, the typical demand forecasting papers in different industrial sectors based on the panel data approach are analyzed and summarized (as shown in Table V). It is found that panel data methods (both static and dynamic panel data models) are popularly used in tourism demand forecasting. At first, for research on tourism demand forecasting, the price of tourism in destination, the transport (travel) cost, and the tourists' income or the economic level in the origin country are considered as the key factors that mainly affect the tourism demand (Song *et al.*, 2000; Ledesma-Rodríguez *et al.*, 2001; Naudé and Saayman, 2005). Following that, dynamic panel data models with lagged dependent variables are adopted to describe the changing features of consumer preferences. These dynamic models consider the parameter for the lagged dependent variable as a measure of habit formation and interdependent preferences (Garín-Muñoz and Montero-Martín, 2007). It means that if people are satisfied with a destination they maybe more likely to come back and tell others about their favorable experiences; this kind of behavior may affect tourism demand of destination (Garín-Muñoz and Montero-Martín, 2007). Besides, the technology that surrounds tourism activities also cannot be ignored since information sharing, communication,

Model	Effects	Correlation	Application	References
Static model	Fixed-effects		Water demand	Arbués *et al.* (2000), Höglund (1999)
			Tourism demand	Song *et al.* (2000), Ledesma-Rodríguez *et al.* (2001), Naudé and Saayman (2005)
		Spatial	Cigarettes demand	Baltagi and Li (2004)
			Liquor forecast	Baltagi and Li (2006)
	Random-effects		Water demand	Arbués *et al.* (2000), Höglund (1999)
		Spatial	Liquor forecast	Baltagi and Li (2006)
			Land-use	Chakir and Le Gallo (2013)
Dynamic model	Fixed-effects		Water demand	Nauges and Thomas (2000) Arbués *et al.* (2004), Polebitski and Palmer (2009)
			Electricity and natural-gas consumption	Balestra and Nerlove (1966), Baltagi *et al.* (2002)
			Gasoline demand	Baltagi and Griffin (1997)
			Energy demand	Olatubi and Zhang (2003), Liu (2004), Lee and Lee (2010), Miguel Garcia-Cerrutti (2000)
			Tourism demand	Ledesma-Rodríguez *et al.* (2001), Roget and Rodríguez González (2006), Sakai *et al.* (2000), Song and Witt (2000), Garín-Muñoz and Montero-Martín (2007), Falk (2010), Ramos and Rodrigues (2013)
			Sales forecasting of consumer products	Telser (1962), Frees and Miller (2004), Ren *et al.* (2015)
			Others	Babel *et al.* (2008), Li *et al.* (2013)
		Spatial correlation	Emissions forecast	Auffhammer and Steinhauser (2007), Auffhammer and Carson (2008), Burnett *et al.* (2013)
	Random-effects		Tourism demand	Proença and Soukiazis (2005)

Table V. A summary of panel data model applications booking, and purchasing of travel products may have a significance influence on the demand of tourism in some ways (Ramos and Rodrigues, 2013). Thus, dynamic panel data models are suitable to model the relationship between tourism demand and related impact factors. For energy demand forecasting on products such as petroleum, natural gas, and electricity, many studies in the literature have been conducted based on the pure time-series analysis. But these forecasting models failed to consider energy consumption within regions or states. Undoubtedly, country-level studies are critically important for understanding important trends in global energy consumption (Olatubi and Zhang, 2003). Panel data at the county level would accurately illustrate the variations in price, income, and weather that occur within the states (Miguel Garcia-Cerrutti, 2000). Notice that dynamic panel data models which include an adjustment path in the long-run demand are used to model the energy consumption problems in the following studies such as Balestra and Nerlove (1966), Baltagi *et al.* (2002), Olatubi and Zhang (2003), Liu (2004), Lee and Lee (2010), and L. Miguel Garcia-Cerrutti (2000).

Panel data models have not been used in water demand forecasting as widely as tourism demand and energy demand due to the lack of demographic and water consumption data used to construct the panel data model (Polebitski and Palmer, 2009). Considering the household level may be more appropriate when studying pricing effects. In the literature, Höglund (1999) conducts five static and one dynamic panel-data models to investigate pricing effects on water demand in Sweden. He suggests that the OLS and between effects models that control for omitted variables changing over time but are constant between subjects would yield the most reliable results. Nauges and Thomas (2000) create a panel data and estimate four regression models: an OLS, a fixed effects model, a random-effects model using GLS, and an IVs approach. The authors reveal that the IVs model is better than the fixed effects model in their study. Besides, dynamic panel data models are also examined for the water demand analysis in Martinez-Espiñeira (2002), Arbués *et al.* (2004), and Polebitski and Palmer (2009). Moreover, Polebitski and Palmer (2009) examine the performance of fixed effects and random-effects models for conducting the dynamic estimation for Census Tracts using over 100 census tracts and 12 years of demographic, weather, economic, and bimonthly water consumption data. The results indicate that both fixed and random-effects methods work well for modeling water consumption heterogeneity across space. Specifically, the fixed effects model provided low RMSE across all census tracts within the model domain, but is not easily used to forecast demand outside the model domain.

Besides, panel data models have been used to conduct demand forecasting in other industrial areas, e.g. cigarettes demand, liquor demand, land-use, demand of consumer products, and emissions forecast. We can see that each reviewed model has been successfully employed in studying different kinds of demand forecasting problems in the industry. However, it is important to note that different panel data models can be used for the same application and there is hence a need to have further analysis on which model is preferred.

In fact, as we know, the panel data-based models provide a versatile method for dealing with demand forecasting problems in many industrial areas. However, there are many different kinds of analytical panel data models and estimators available. How to select the right model and testing method is a critical issue. Based on the analysis conducted in the above sections, we propose a novel flowchart-based process as shown in Figure A2 for decision makers to identify the right panel data forecasting model and estimator for the analysis. Details of the flowchart are shown in "the novel panel data model selection process." This particular process is important in helping decision makers identify the right panel data-based models for conducting demand forecasting with respect to the data formats and industrial requirements.

Establishing the right panel data model for industrial applications – details:

- Step 1. Input the historical data.
- Step 2. Establish the panel data.
- Step 3. Test for panel stationarity by using the unit root test. If the testing results reject the null hypothesis that the common unit root of panel data is non-stationary, go to Step 4; otherwise, conduct the differential evolution then go back to Step 2, or do the co-integration test and establish the co-integration panel data model. For the details on the testing mechanism and estimation of co-integration panel data model, refer to Banerjee (1999) and Kao and Chiang (2001). For applications, please refer to Lee (2005), Costantini and Martini (2010), Li *et al.* (2013), and Ramos and Rodrigues (2013).
- Step 4. Select a right panel data regression model, static or dynamic model, for the special industrial settings. If the future demand is related to the historical demand, the dynamic model that contains lagged dependent variables will be used to estimate behavioral relationships. The related literature or evidence can be easily found for each industrial setting. After model selection, individual-specific effects and dependent relationship should be tested in the following step.
- Step 5. Test for individual-specific effects: The hypothesis of Hausman/LM test assumes that there are random-effects among cross-sections. Thus, random-effects will be selected if the testing result accepts the hypothesis; otherwise, fixed-effects will be considered.
- Step 6. Correlation test – test for spatial correlation and serial correlation: if the LM spatial testing result rejects the null hypothesis with no spatial correlation, there will be an obvious spatial dependence. If the LM serial testing result rejects the null hypothesis with no spatial correlation, there will be an obvious serial dependence. This step follows the details proposed by Baltagi (2008a).
- Step 7. Considering the testing results above, estimate a panel data-based regression model by using the estimation methods concluded in Section 2.
- Step 8. Output the panel data forecasting model.
- Step 9. The end.

5. Further analysis, insights, and conclusion

Panel data-based demand forecasting models have been proven to be versatile and useful in demand forecasting. In this paper, we have presented an exhaustive literature survey on different panel data-based analytical demand forecasting models, the respective estimation and testing methods, and industrial applications. We have derived a novel workable process for modeling a panel data-based demand forecasting model. As the concluding remarks, we highlight the strengths of panel data models, the data sufficiency requirements, and their applications as follows.

5.1 Strengths of panel data models

First, the panel data models give us a large number of data points in higher dimension than the pure time-series data models. Thus, the panel data models usually contain higher degrees of freedom. Undoubtedly, the forecasting effectiveness can be improved by using the panel data models compared to both the pure time series and the pure cross-sectional data models. Over the past decades, panel data models have been widely applied in demand forecasting problems, especial for those cases in which cross-section data are available while time-series historical data are lacking. Take fashion demand forecasting in (see footnote 4) as an example, since there is an insufficient amount of historical data due to the short life cycle of products, panel data-based demand forecasting model performs very well. Results from real data even show that the panel data models outperform both the traditional statistical and intelligent methods.

Second, panel data forecasting models are able to generate more accurate forecasting performance than either time-series data or cross-section data alone. Since some individual's behaviors (conditional on certain variables) are similar, panel data models provide the possibility of learning an individual's behavior by observing the behavior of others, in addition to the information on that particular individual entity's behavior. They hence generate more accurate description of an individual's behavior than time-series data or cross-section data. The related influential factors which affect demand, such as price promotion, can be modeled in the panel data demand forecasting models so as to obtain a more accurate forecasting result.

Third, panel data-based demand forecasting methods have been well developed and there are many choices available for decision makers to choose from. For example, the dynamic panel data models can accurately describe the dynamic relationship between explained variables and explanatory variables as well as the panel features among individuals. The spatial panel data models enable decision makers to well control the spatial dependence between observations across cross-section. Following our proposed novel process as depicted in Figure A2, decision makers can easily identify the right panel data-based analytical models for conducting demand forecasting.

5.2 Weakness – data sufficiency challenges

Despite being a versatile and intuitive method, panel data methods face some challenges in industrial applications.

First, the panel data that contain both time-series data and cross-section data should be available for employing panel data models. For example, for tourism demand forecasting, we need the panel data of demand which include time periods data and also cross-country data. For predicting the sales of fashion products, we need to have the panel data which include sales of some other correlated items over the given time horizon. Unfortunately, panel datasets which include such a lot of details may be unavailable which directly hinders the application of panel data-based demand forecasting methods.

Second, the number of time-series historical datasets is usually limited. Typical panel data for industrial practice usually cover a short span of time for each individual, such as energy demand forecasting, tourism forecasting, water demand forecasting, etc. Thus, with the limited amount of data, the performance of panel data-based methods will be affected.

5.3 Applications

By reviewing the analytical models of panel data, and industrial applications, some findings are discussed below.

First, panel data-based demand forecasting methods are very useful and highly implementable. They not only outperform the simple time series-based models (Baltagi and Griffin, 1997), but also the other more advanced models (Ren *et al.*, 2015).

Second, panel data-based demand forecasting models can be combined with other existing models to yield great forecasting performance. For example, Li *et al.* (2013) propose a dynamic hybrid panel data particle filter (PDPF) method for the electricity price forecasting problem. They compare their proposed PDPF method with other well-established forecasting methods such as neuralnetworks (NN), fuzzy neural network and support vector machines, and find that PDPF outperforms all of them. Similar results are found in fashion sales forecasting (Ren *et al.*, 2015).

Third, dynamic panel data models are very widely used in demand forecasting problems. Compared to static model, dynamic models contain lagged dependent variables that are used to estimate behavioral relationships. When the current demand is mainly determined by the value of previous demand, a dynamic panel data model is highly suggested to be applied. The applications of dynamic panel data models can be found in electricity demand, natural-gas consumption estimation, and tourism demand forecasting, etc.

Fourth, the spatial correlation panel data models are proven to be more accurate than the ones which ignore the spatial correlation (for the sake of simplicity) in many industrial applications (Auffhammer and Carson, 2008; Burnett *et al.*, 2013; Baltagi and Li, 2006; Chakir and Le Gallo, 2013). As a remark, spatial panel data models can enable decision makers to control the unknown heterogeneity and cross-sectional dependence that are present whenever neighbor correlation exists across cross-sectional units (e.g. for location-related problems, demands over the states, and regions are correlated). Thus, spatial correlation panel data models can be used in many domains such as agricultural economics (Druska and Horrace, 2004), transportation research (Frazier and Kockelman, 2005), public economics (Egger *et al.*, 2005), regional product demand (Baltagi and Li, 2006), growth convergence of countries and regions (Baltagi *et al.*, 2007), regional markets (Keller and Shiue, 2007), labor economics (Foote, 2007), and public economics (Franzese and Hays, 2007). In addition, county/state-level data over time and household-level survey data from villages observed over time can also be used to form the spatial correlation model for studying, e.g., female labor participation rates, or the effects of education on wages (Baltagi *et al.*, 2003a, b).

5.4 Future research directions

Although an extensive literature review on panel data-based demand forecasting methods has been carried out in this paper, there is a need for further research in a number of directions.

First, this paper only provides a concise guidance on how to establish a panel data-based demand forecasting model under different industrial settings. Some further analyses on how the proposed method performs in different industrial settings should be conducted. For example, whether the proposed framework is especially useful for demand forecasting in fashion retailers or in general department stores can be computationally verified by using empirical data. This provides a lot of opportunities for future studies.

Second, even though our proposed framework is novel, there could be an alternative approach to establish the right panel data models-based demand forecasting scheme. Thus, more research can be conducted on it. For instance, one can build another framework by resequencing the flows, and compare the performances of it with the one proposed in this paper with respect to different industrials settings and data features. New insights can be generated.

Third, the panel data-based models can be used together with other existing forecasting models for conducting sophisticated demand forecasting. For example, the particle filter and the panel data mixed model has recently been examined by Ren *et al.* (2015). There are definitely many more different probable candidates which can work with the panel data models. This area is still widely open and a lot of future research can be conducted along this direction.

Fourth, in a supply chain context, how panel data-based models can work with other operations and information systems (such as ERP systems Hasan *et al.*, 2011) with proper information sharing (Chan and Chan, 2009; is an important issue). Since many panel data-based models are analytically tractable, new and innovative analytical studies (e.g. those on inventory control and supply contracting) can be conducted and novel managerial insights can be generated.

Glossary

CV	covariance
GLS	generalized-least-squares
GMM	linear generalized method of moments
FE	fixed effects
IV	instrumental variable
LM	Lagrange multiplier
LR	likelihood ratio
MDE	minimum distance estimator
ML	maximum-likelihood
MLE	maximum-likelihood estimator
OLS	ordinary-least-squares
QML	quasi-maximum likelihood
RE	random effects

Notes

1. N: the number of cross-sectional units. Taking the sales data of apparel products as an example, N denotes the number of product categories.
2. Estimation: using the demand historical data to estimate the unknown parameters of panel regression model.
3. If an estimator converges in probability to the true value of the parameter, it called a consistent estimator (Amemiya 1985).
4. Initial conditions: the assumptions on α_i, λ_t and u_{it}.

References

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press, Cambridge.

Anderson, T.W. and Hsiao, C. (1981), "Estimation of dynamic models with error components", *Journal of the American Statistical Association*, Vol. 76 No. 375, pp. 598-606.

Anderson, T.W. and Hsiao, C. (1982), "Formulation and estimation of dynamic models using panel data", *Journal of Econometrics*, Vol. 18 No. 1, pp. 47-82.

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Vol. 4, Springer, Dordrecht.

Anselin, L. (1999), "The future of spatial analysis in the social sciences", *Geographic Information Sciences*, Vol. 5 No. 2, pp. 67-76.

Anselin, L. (2001), "Rao's score test in spatial econometrics", *Journal of Statistical Planning and Inference*, Vol. 97 No. 1, pp. 113-139.

Anselin, L. and Bera, A.K. (1998), "Spatial dependence in linear regression models with an introduction to spatial econometrics", *Statistics Textbooks and Monographs*, Vol. 155, pp. 237-290.

Anselin, L. and Hudak, S. (1992), "Spatial econometrics in practice: a review of software options", *Regional Science and Urban Economics*, Vol. 22 No. 3, pp. 509-536.

Arbués, F., Barberán, R. and Villanúa, I. (2000), "Water price impact on residential water demand in the city of Zaragoza. A dynamic panel data approach", 40th European Congress of the European Regional Studies Association (ERSA) in Barcelona, Barcelona, Vol. 24, pp. 30-31.

Arbués, F., Barberán, R. and Villanúa, I. (2004), "Price impact on urban residential water demand: a dynamic panel data approach", *Water Resources Research*, Vol. 40 No. 11. doi: 10.1029/2004WR003092.

Arellano, M. (1993), "On the testing of correlated effects with panel data", *Journal of Econometrics*, Vol. 59 Nos 1-2, pp. 87-97.

Arellano, M. and Bond, S. (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *The Review of Economic Studies*, Vol. 58 No. 1, pp. 277-297.

Arellano, M. and Carrasco, R. (2003), "Binary choice panel data models with predetermined variables", *Journal of Econometrics*, Vol. 115 No. 1, pp. 125-157.

Auffhammer, M. and Carson, R.T. (2008), "Forecasting the path of China's CO_2 emissions using province-level information", *Journal of Environmental Economics and Management*, Vol. 55 No. 3, pp. 229-247.

Auffhammer, M. and Steinhauser, R. (2007), "The future trajectory of us CO_2 emissions: the role of state vs aggregate information", *Journal of Regional Science*, Vol. 47, pp. 47-61.

Babel, B., Bomsdorf, E. and Schmidt, R. (2008), "Forecasting German mortality using panel data procedures", *Journal of Population Economics*, Vol. 21 No. 3, pp. 541-555.

Bai, J. and Ng, S. (2004), "A PANIC attack on unit roots and cointegration", *Econometrica*, Vol. 72 No. 4, pp. 1127-1177.

Bai, J. and Ng, S. (2009), "Boosting diffusion indices", *Journal of Applied Econometrics*, Vol. 24 No. 4, pp. 607-629.

Balestra, P. and Nerlove, M. (1966), "Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas", *Econometrica: Journal of the Econometric Society*, Vol. 34, pp. 585-612.

Baltagi, B. (2008a), *Econometric Analysis of Panel Data*, Vol. 1, John Wiley & Sons, Hoboken, NJ.

Baltagi, B.H. (2008b), "Forecasting with panel data", *Journal of Forecasting*, Vol. 27 No. 2, pp. 153-173.

Baltagi, B.H. and Griffin, J.M. (1997), "Pooled estimators vs their heterogeneous counterparts in the context of dynamic demand for gasoline", *Journal of Econometrics*, Vol. 77, pp. 303-327.

Baltagi, B.H. and Kao, C. (2001), *Nonstationary Panels, Cointegration in Panels and Dynamic Panels: A Survey*, Vol. 15, Emerald Group Publishing Limited, Bingley, Bradford.

Baltagi, B.H. and Li, D. (2004), "Prediction in the panel data model with spatial correlation", in Anselin, L., Florax, R.J.G.M. and Rey, S.J. (Eds), *Advances in Spatial Econometrics*, Springer, Heidelberg, pp. 283-295.

Baltagi, B.H. and Li, D. (2006), "Prediction in the panel data model with spatial correlation: the case of liquor", *Spatial Economic Analysis*, Vol. 1 No. 2, pp. 175-185.

Baltagi, B.H. and Li, Q. (1991), "A joint test for serial correlation and random individual effects", *Statistics & Probability Letters*, Vol. 11 No. 3, pp. 277-280.

Baltagi, B.H. and Li, Q. (1992), "Prediction in the one-way error component model with serial correlation", *Journal of Forecasting*, Vol. 11 No. 6, pp. 561-567.

Baltagi, B.H. and Li, Q. (1995), "Testing AR (1) against MA (1) disturbances in an error component model", *Journal of Econometrics*, Vol. 68 No. 1, pp. 133-151.

Baltagi, B.H. and Liu, L. (2008), "Testing for random effects and spatial lag dependence in panel data models", *Statistics & Probability Letters*, Vol. 78 No. 18, pp. 3304-3306.

Baltagi, B.H., Bresson, G. and Pirotte, A. (2002), "Comparison of forecast performance for homogeneous, heterogeneous and shrinkage estimators: some empirical evidence from US electricity and natural-gas consumption", *Economics Letters*, Vol. 76 No. 3, pp. 375-382.

Baltagi, B.H., Bresson, G. and Pirotte, A. (2003a), "Fixed effects, random effects or Hausman-Taylor? A pretest estimator", *Economics Letters*, Vol. 79 No. 3, pp. 361-369.

Baltagi, B.H., Chang, Y.-J. and Li, Q. (1992), "Monte Carlo results on several new and existing tests for the error component model", *Journal of Econometrics*, Vol. 54 Nos 1-3, pp. 95-120.

Baltagi, B.H., Heun Song, S., Cheol Jung, B. and Koh, W. (2007), "Testing for serial correlation, spatial autocorrelation and random effects using panel data", *Journal of Econometrics*, Vol. 140 No. 1, pp. 5-51.

Baltagi, B.H., Song, S.H. and Koh, W. (2003b), "Testing panel data regression models with spatial error correlation", *Journal of Econometrics*, Vol. 117 No. 1, pp. 123-150.

Banerjee, A. (1999), "Panel data unit roots and cointegration: an overview", *Oxford Bulletin of Economics and Statistics*, Vol. 61 No. S1, pp. 607-629.

Banerjee, A., Dolado, J. and Mestre, R. (1998), "Error-correction mechanism tests for cointegration in a single-equation framework", *Journal of Time Series Analysis*, Vol. 19 No. 3, pp. 267-283.

Bhargava, A., Franzini, L. and Narendranathan, W. (1982), "Serial correlation and the fixed effects model", *The Review of Economic Studies*, Vol. 49 No. 4, pp. 533-549.

Blundell, R. and Smith, R.J. (1991), "Initial conditions and efficient estimation in dynamic panel data", Department of Economics, University College London.

Bond, S.R. (2002), "Dynamic panel data models: a guide to micro data methods and practice", *Portuguese Economic Journal*, Vol. 1 No. 2, pp. 141-162.

Borenstein, M., Hedges, L.V., Higgins, J. and Rothstein, H.R. (2010), "A basic introduction to fixed-effect and random-effects models for meta-analysis", *Research Synthesis Methods*, Vol. 1 No. 2, pp. 97-111.

Breitung, J. and Das, S. (2005), "Panel unit root tests under cross-sectional dependence", *Statistica Neerlandica*, Vol. 59 No. 4, pp. 414-433.

Breitung, J. and Das, S. (2008), "Testing for unit roots in panels with a factor structure", *Econometric Theory*, Vol. 24 No. 1, pp. 88-108.

Breusch, T.S. (1978), "Testing for autocorrelation in dynamic linear models", *Australian Economic Papers*, Vol. 17 No. 31, pp. 334-355.

Breusch, T.S. and Pagan, A.R. (1979), "A simple test for heteroscedasticity and random coefficient variation", *Econometrica*, Vol. 47, pp. 1287-1294.

Breusch, T.S. and Pagan, A.R. (1980), "The Lagrange multiplier test and its applications to model specification in econometrics", *The Review of Economic Studies*, Vol. 47 No. 1, pp. 239-253.

Burnett, J., Bergstrom, J.C. and Dorfman, J.H. (2013), "A spatial panel data approach to estimating US state-level energy emissions", *Energy Economics*, Vol. 40, pp. 396-404.

Chakir, R. and Le Gallo, J. (2013), "Predicting land use allocation in France: a spatial panel data analysis", *Ecological Economics*, Vol. 92, pp. 114-125.

Chan, H.K. and Chan, F.T.S. (2009), "Effect of information sharing in supply chains with flexibility", *International Journal of Production Research*, Vol. 47 No. 1, pp. 213-232.

Chang, Y. (2002), "Nonlinear IV unit root tests in panels with cross-sectional dependency", *Journal of Econometrics*, Vol. 110 No. 2, pp. 261-292.

Choi, I. (2001), "Unit root tests for panel data", *Journal of International Money and Finance*, Vol. 20 No. 2, pp. 249-272.

Choi, I. and Chue, T.K. (2007), "Subsampling hypothesis tests for nonstationary panels with applications to exchange rates and stock prices", *Journal of Applied Econometrics*, Vol. 22 No. 2, pp. 233-264.

Costantini, V. and Martini, C. (2010), "The causality between energy consumption and economic growth: a multi-sectoral analysis using non-stationary cointegrated panel data", *Energy Economics*, Vol. 32 No. 3, pp. 591-603.

Douglas, H.-E., Whitney, N. and Harvey, S.R. (1988), "Estimating vector autoregressions with panel data", *Econometrica*, Vol. 56 No. 6, pp. 1371-1395.

Drukker, D.M. (2003), "Testing for serial correlation in linear panel-data models", *Stata Journal*, Vol. 3 No. 2, pp. 168-177.

Druska, V. and Horrace, W.C. (2004), "Generalized moments estimation for spatial panel data: Indonesian rice farming", *American Journal of Agricultural Economics*, Vol. 86 No. 1, pp. 185-198.

Egger, S., Ilie, A., Fu, Y., Chongsathien, J., Kang, D.-J. and Welland, M.E. (2005), "Dynamic shadow mask technique: a universal tool for nanoscience", *Nano Letters*, Vol. 5 No. 1, pp. 15-20.

Eilat, Y. and Einav, L. (2004), "Determinants of international tourism: a three-dimensional panel data analysis", *Applied Economics*, Vol. 36 No. 12, pp. 1315-1327.

Elhorst, J.P. (2003), "Specification and estimation of spatial panel data models", *International Regional Science Review*, Vol. 26 No. 3, pp. 244-268.

Falk, M. (2010), "A dynamic panel data analysis of snow depth and winter tourism", *Tourism Management*, Vol. 31 No. 6, pp. 912-924.

Foote, C.L. (2007), "Space and time in macroeconomic panel data: young workers and state-level unemployment revisited", working paper series, Federal Reserve Bank of Boston, Boston, MA.

Franzese, R.J. and Hays, J.C. (2007), "Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data", *Political Analysis*, Vol. 15 No. 2, pp. 140-164.

Frazier, C. and Kockelman, K.M. (2005), "Spatial econometric models for panel data: incorporating spatial and temporal data", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1902, pp. 80-90.

Frees, E.W. and Miller, T.W. (2004), "Sales forecasting using longitudinal data models", *International Journal of Forecasting*, Vol. 20 No. 1, pp. 99-114.

Frondel, M. and Vance, C. (2010), "Fixed, random, or something in between? A variant of Hausman's specification test for panel data estimators", *Economics Letters*, Vol. 107 No. 3, pp. 327-329.

Gardner, W. (1960), "Dynamic aspects of water availability to plants", *Soil Science*, Vol. 89 No. 2, pp. 63-73.

Garín-Muñoz, T. and Montero-Martín, L.F. (2007), "Tourism in the Balearic Islands: a dynamic model for international demand using panel data", *Tourism Management*, Vol. 28 No. 5, pp. 1224-1235.

Gengenbach, C., Palm, F.C. and Urbain, J.-P. (2009), "Panel unit root tests in the presence of cross-sectional dependencies: comparison and implications for modelling", *Econometric Reviews*, Vol. 29 No. 2, pp. 111-145.

Getis, A. (2007), "Reflections on spatial autocorrelation", *Regional Science and Urban Economics*, Vol. 37 No. 4, pp. 491-496.

Godfrey, L.G. (1978a), "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables", *Econometrica: Journal of the Econometric Society*, Vol. 46 No. 6, pp. 1293-1301.

Godfrey, L.G. (1978b), "Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables", *Econometrica*, Vol. 46 No. 6, pp. 1303-1310.

Gopimatj, D.A. (1995), "Modeling heterogeneity in discrete choice processes: application to travel demand", *Transportation Research Part A: Policy and Practice*, Vol. 31 No. 1, p. 86.

Guggenberger, P. (2010), "The impact of a Hausman pretest on the size of a hypothesis test: the panel data case", *Journal of Econometrics*, Vol. 156 No. 2, pp. 337-343.

Hansen, L.P. (1982), "Large sample properties of generalized method of moments estimators", *Econometrica*, Vol. 50 No. 4, pp. 1029-1054.

Hasan, M., Trinh, N.T., Chan, F.T.S., Chan, H.K. and Chung, S.H. (2011), "Implementation of ERP of the Australian manufacturing companies", *Industrial Management & Data Systems*, Vol. 111 No. 1, pp. 132-145.

Hausman, J.A. (1978), "Specification tests in econometrics", *Econometrica*, Vol. 46 No. 6, pp. 1251-1271.

Hausman, J.A. and Taylor, W.E. (1981), "Panel data and unobservable individual effects", *Econometrica: Journal of the Econometric Society*, Vol. 49 No. 6, pp. 1377-1398.

Höglund, L. (1999), "Household demand for water in Sweden with implications of a potential tax on water use", *Water Resources Research*, Vol. 35 No. 12, pp. 3853-3863.

Holly, A. (1982), "A remark on Hausman's specification test", *Econometrica: Journal of the Econometric Society*, Vol. 50 No. 3, pp. 749-759.

Honoré, B.E. and Kyriazidou, E. (2000), "Panel data discrete choice models with lagged dependent variables", *Econometrica*, Vol. 68 No. 4, pp. 839-874.

Houthakker, H.S., Verleger, P.K. and Sheehan, D.P. (1974), "Dynamic demand analyses for gasoline and residential electricity", *American Journal of Agricultural Economics*, Vol. 56 No. 2, pp. 412-418.

Hsiao, C. (1996), *Identification of Cointegrated Vector Autoregressive Process*, Mimeo.

Hsiao, C. (2003), *Analysis of Panel Data*, Vol. 34, Cambridge University Press, New York, NY.

Hsiao, C., Hashem Pesaran, M. and Kamil Tahmiscioglu, A. (2002), "Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods", *Journal of Econometrics*, Vol. 109 No. 1, pp. 107-150.

Im, K.S., Pesaran, M.H. and Shin, Y. (2003), "Testing for unit roots in heterogeneous panels", *Journal of Econometrics*, Vol. 115 No. 1, pp. 53-74.

Jang, M.J. and Shin, D.W. (2014), "Tests for random time effects and spatial error correlation in panel regression models", *Statistics*, Vol. 48 No. 1, pp. 101-120.

Kao, C. and Chiang, M.-H. (2001), "On the estimation and inference of a cointegrated regression in panel data", *Advances in Econometrics*, Vol. 15 No. 1, pp. 179-222.

Kelejian, H.H. and Robinson, D.P. (1998), "A suggested test for spatial autocorrelation and/or heteroskedasticity and corresponding Monte Carlo results", *Regional Science and Urban Economics*, Vol. 28 No. 4, pp. 389-417.

Keller, W. and Shiue, C.H. (2007), "The origin of spatial interaction", *Journal of Econometrics*, Vol. 140 No. 1, pp. 304-332.

Kim, W.-J., Lee, J.-D. and Kim, T.-Y. (2005), "Demand forecasting for multigenerational products combining discrete choice and dynamics of diffusion under technological trajectories", *Technological Forecasting and Social Change*, Vol. 72 No. 7, pp. 825-849.

Ledesma-Rodríguez, F.J., Navarro-Ibanez, M. and Pérez-Rodríguez, J.V. (2001), "Panel data and tourism: a case study of Tenerife", *Tourism Economics*, Vol. 7 No. 1, pp. 75-88.

Lee, C.-C. (2005), "Energy consumption and GDP in developing countries: a cointegrated panel analysis", *Energy Economics*, Vol. 27 No. 3, pp. 415-427.

Lee, C.-C. and Lee, J.-D. (2010), "A panel data analysis of the demand for total energy and electricity in OECD countries", *The Energy Journal*, Vol. 31 No. 1, pp. 1-24.

Lee, L.-F. and Yu, J. (2010), "Spatial panels: random components vs fixed effects", North American Winter Meeting of the Econometric Society, Atlanta.

LeSage, J. and Pace, R.K. (2010), *Introduction to Spatial Econometrics*, CRC Press, Boca Raton, FL.

Levin, A., Lin, C.-F. and James Chu, C.-S. (2002), "Unit root tests in panel data: asymptotic and finite-sample properties", *Journal of Econometrics*, Vol. 108 No. 1, pp. 1-24.

Li, Q. and Hsiao, C. (1998), "Testing serial correlation in semiparametric panel data models", *Journal of Econometrics*, Vol. 87 No. 2, pp. 207-237.

Li, X., Yu, C., Ren, S., Chiu, C. and Meng, K. (2013), "Day-ahead electricity price forecasting based on panel cointegration and particle filter", *Electric Power Systems Research*, Vol. 95 No. 2013, pp. 66-76.

Liu, G. (2004), "Estimating energy demand elasticities for OECD countries. A dynamic panel data approach", Discussion Paper No. 373, Statistics Norway, Research Department.

Maddala, G.S. and Wu, S. (1999), "A comparative study of unit root tests with panel data and a new simple test", *Oxford Bulletin of Economics and Statistics*, Vol. 61 No. S1, pp. 631-652.

Martinez-Espiñeira, R. (2002), "Residential water demand in the northwest of Spain", *Environmental and Resource Economics*, Vol. 21 No. 2, pp. 161-187.

Miguel Garcia-Cerrutti, L. (2000), "Estimating elasticities of residential energy demand from panel county data using dynamic random variables models with heteroskedastic and correlated error terms", *Resource and Energy Economics*, Vol. 22 No. 4, pp. 355-366.

Mundlak, Y. (1978), "On the pooling of time series and cross section data", *Econometrica: Journal of the Econometric Society*, Vol. 46 No. 1, pp. 69-85.

Naudé, W.A. and Saayman, A. (2005), "Determinants of tourist arrivals in Africa: a panel data regression analysis", *Tourism Economics*, Vol. 11 No. 3, pp. 365-391.

Nauges, C. and Thomas, A. (2000), "Privately operated water utilities, municipal price negotiation, and estimation of residential water demand: the case of France", *Land Economics*, Vol. 76 No. 1, pp. 68-85.

Nerlove, M. (1967), "Experimental evidence on the estimation of dynamic relations from a time series of cross sections", *Econometric Studies Quarterly*, Vol. 33 No. 1, pp. 42-74.

Nerlove, M. (1971), "Further evidence on the estimation of dynamic economic relations from a time series of cross sections", *Econometrica: Journal of the Econometric Society*, Vol. 39 No. 2, pp. 359-382.

Nerlove, M. and Balestra, P. (1996), "Formulation and estimation of econometric models for panel data", in Mátyás, M. and Sevestre, P. (Eds), *The Econometrics of Panel Data*, Vol. 33, Springer, Heidelberg, pp. 3-22.

Olatubi, W.O. and Zhang, Y. (2003), "A dynamic estimation of total energy demand for the Southern States", *The Review of Regional Studies*, Vol. 33 No. 2, pp. 206-228.

Pesaran, M.H. (2007), "A simple panel unit root test in the presence of cross-section dependence", *Journal of Applied Econometrics*, Vol. 22 No. 2, pp. 265-312.

Phillips, P.C. and Sul, D. (2003), "Dynamic panel estimation and homogeneity testing under cross section dependence", *The Econometrics Journal*, Vol. 6 No. 1, pp. 217-259.

Polebitski, A.S. and Palmer, R.N. (2009), "Seasonal residential water demand forecasting for census tracts", *Journal of Water Resources Planning and Management*, Vol. 136 No. 1, pp. 27-36.

Proença, S.A. and Soukiazis, E. (2005), *Demand for Tourism in Portugal: A Panel Data Approach*, Escola Superior Agrária, Instituto Politécnico de Coimbra, Coimbra.

Quah, D. (1992), "The relative importance of permanent and transitory components: identification and some theoretical bounds", *Econometrica: Journal of the Econometric Society*, Vol. 60 No. 1, pp. 107-118.

Quah, D. (1994), "Exploiting cross-section variation for unit root inference in dynamic data", *Economics Letters*, Vol. 44 No. 1, pp. 9-19.

Quah, D.T. (1996), "Empirics for economic growth and convergence", *European Economic Review*, Vol. 40 No. 6, pp. 1353-1375.

Ramos, C.M. and Rodrigues, P.M. (2013), "The importance of ICT for tourism demand: a dynamic panel data analysis", in Matias, Á., Nijkamp, P. and Sarmento, M. (Eds), *Quantitative Methods in Tourism Economics*, Springer, pp. 97-111.

Ren, S., Choi, T.-M. and Liu, N. (2015), "Fashion sales forecasting with a panel data-based particle-filter model", *IEEE Transactions on Systems, Man, and Cybernetics – Systems*, Vol. 45 No. 3, pp. 411-421.

Roget, F.M. and Rodríguez González, X.A. (2006), "Rural tourism demand in Galicia, Spain", *Tourism Economics*, Vol. 12 No. 1, pp. 21-31.

Sakai, M., Brown, J. and Mak, J. (2000), "Population aging and Japanese international travel in the 21st century", *Journal of Travel Research*, Vol. 38 No. 3, pp. 212-220.

Sargan, J.D. and Bhargava, A. (1983), "Testing residuals from least squares regression for being generated by the Gaussian random walk", *Econometrica: Journal of the Econometric Society*, Vol. 51 No. 1, pp. 153-174.

Song, H. and Li, G. (2008), "Tourism demand modelling and forecasting – a review of recent research", *Tourism Management*, Vol. 29 No. 2, pp. 203-220.

Song, H. and Witt, S.F. (2000), *Tourism Demand Modelling and Forecasting: Modern Econometric Approaches*, Routledge, New York, NY.

Song, H., Romilly, P. and Liu, X. (2000), "An empirical study of outbound tourism demand in the UK", *Applied Economics*, Vol. 32 No. 5, pp. 611-624.

Swamy, P. and Arora, S.S. (1972), "The exact finite sample properties of the estimators of coefficients in the error components regression models", *Econometrica: Journal of the Econometric Society*, Vol. 40 No. 2, pp. 261-275.

Telser, L.G. (1962), "The demand for branded goods as estimated from consumer panel data", *The Review of Economics and Statistics*, Vol. 44 No. 3, pp. 300-324.

Wallace, T.D. and Hussain, A. (1969), "The use of error components models in combining cross section with time series data", *Econometrica: Journal of the Econometric Society*, Vol. 37 No. 1, pp. 55-72.

Westerlund, J. (2007), "Testing for error correction in panel data", *Oxford Bulletin of Economics and Statistics*, Vol. 69 No. 6, pp. 709-748.

Wooldridge, J.M. (2002), "Inverse probability weighted M-estimators for sample selection, attrition, and stratification", *Portuguese Economic Journal*, Vol. 1 No. 1, pp. 117-139.

Yu, J., de Jong, R. and Lee, L.-F. (2008), "Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both N and T are large", *Journal of Econometrics*, Vol. 146 No. 1, pp. 118-134.

Yu, J., de Jong, R. and Lee, L.-F. (2012), "Estimation for spatial dynamic panel data with fixed effects: the case of spatial cointegration", *Journal of Econometrics*, Vol. 167 No. 1, pp. 16-37.

Further reading

Alvarez, J. and Arellano, M. (2003), "The time series and cross-section asymptotics of dynamic panel data estimators", *Econometrica*, Vol. 71 No. 4, pp. 1121-1159.

Arbués, F., Garcıa-Valiñas, M.Á. and Martınez-Espiñeira, R. (2003), "Estimation of residential water demand: a state-of-the-art review", *The Journal of Socio-Economics*, Vol. 32 No. 1, pp. 81-102.

Holtz-Eakin, D., Newey, W. and Rosen, H.S. (1988), "Estimating vector autoregressions with panel data", *Econometrica: Journal of the Econometric Society*, Vol. 56 No. 6, pp. 1371-1395.

Hsiao, C. (2007), "Panel data analysis – advantages and challenges", *Test*, Vol. 16 No. 1, pp. 1-22.

Kao, C. (1999), "Spurious regression and residual-based tests for cointegration in panel data", *Journal of Econometrics*, Vol. 90 No. 1, pp. 1-44.

Maddison, D. (2006), "Environmental Kuznets curves: a spatial econometric approach", *Journal of Environmental Economics and Management*, Vol. 51 No. 2, pp. 218-230.

Taylor, W.E. (1980), "Small sample consideration in estimation from panel data", *Journal of Econometrics*, Vol. 10 No. 2, pp. 103-107.

Appendix

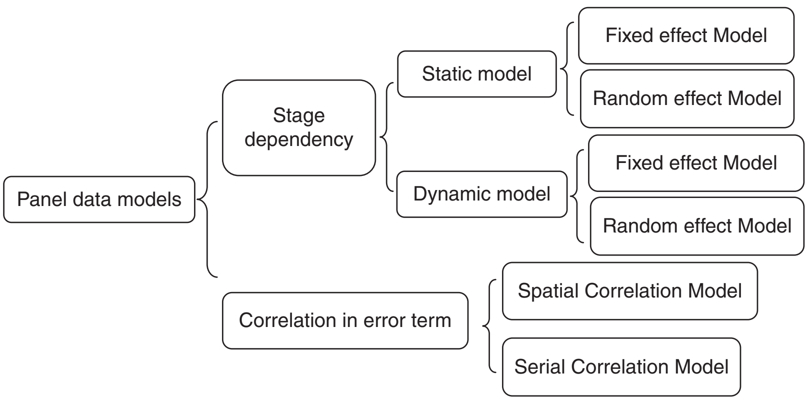

Figure A1. The common classification of panel data models

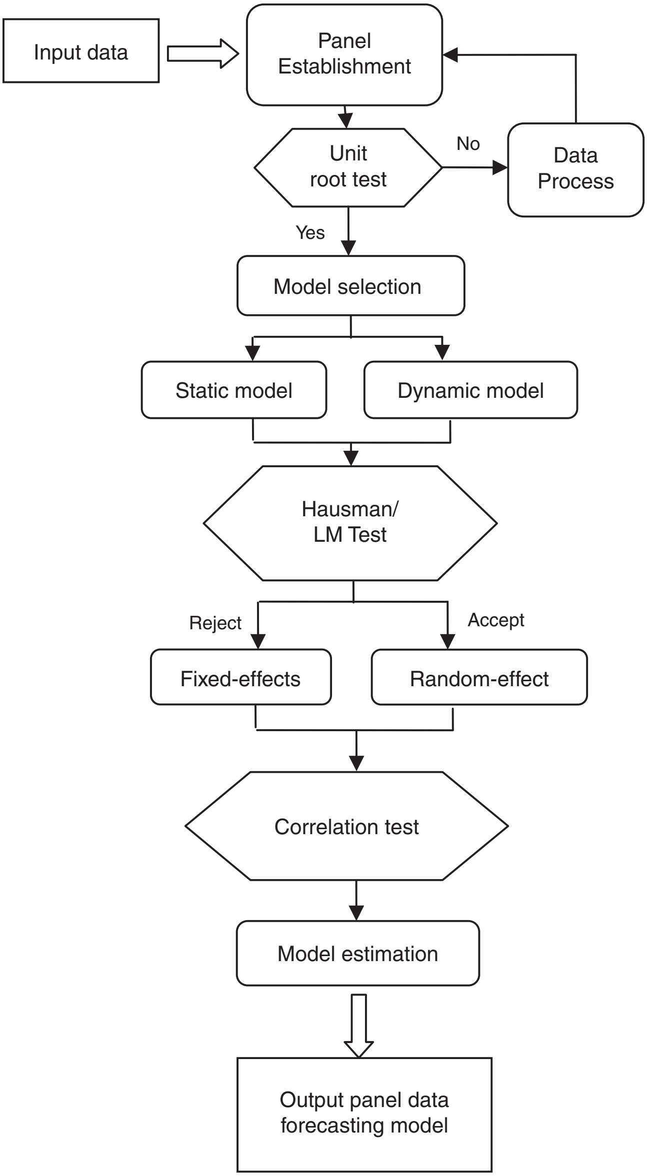

Figure A2.
A novel flow chart developed in this paper for establishing the right panel data model for industrial applications

About the authors

Shuyun Ren is currently a PhD Student at The Hong Kong Polytechnic University. Her current research areas are in forecasting analysis. She has published various papers in academic journals such as *Annals of Operations Research*, *IEEE Transactions on Systems, Man, and Cybernetics – Systems*, *Electric Power Systems Research*, and *Mathematical Problems in Engineering*. She is the Vice Chair of IEEE Hong Kong Section's Student Branch at The Hong Kong Polytechnic University. She is also a Student Member of IEEE and SMC.

Tsan-Ming Choi received his PhD from The Chinese University of Hong Kong. He is currently a Professor of Fashion Business at The Hong Kong Polytechnic University. His recent research interests include demand forecasting driven supply chain operations, and fashion supply chain optimization. He has published extensively in leading academic journals such as *Industrial Management and Data Systems*, *Annals of Operations Research*, *Automatica*, *Decision Support Systems*, *European Journal of Operational Research*, *IEEE Transactions on Automatic Control*, *IEEE Transactions on Cybernetics*, *IEEE Transactions on Engineering Management*, *IEEE Transactions on Industrial Informatics*, *Naval Research Logistics*, *Omega*, *Production and Operations Management*, and many other leading IEEE transactions and journals in information systems/operations management. He authored/edited 16 research handbooks and he is currently a Associate/Senior Editor of journals such as *Production and Operations Management, Decision Support Systems*, *IEEE Transactions of Systems, Man, and Cybernetics – Systems*, *Information Sciences*, etc. He is a Member of IEEE and INFORMS. Tsan-Ming Choi is the corresponding author and can be contacted at: jason.choi@polyu.edu.hk